



# CUPTI

DA-05679-001 \_v6.0 | February 2014

## User's Guide



## WHAT'S NEW

CUPTI contains a number of changes and new features as part of the CUDA Toolkit 6.0 release.

- ▶ Two new CUPTI activity kinds have been introduced to enable two new types of source-correlated data collection. The `Instruction Execution` kind collects SASS-level instruction execution counts, divergence data, and predication data. The `Shared Access` kind collects source correlated data indication inefficient shared memory accesses.
- ▶ CUPTI now provides support for CUDA applications using Unified Memory. A new activity record reports Unified Memory activity such as transfers to and from a GPU and the number of Unified Memory related page faults.
- ▶ CUPTI now recognized and reports the special MPS context that is used by CUDA applications running on a system with MPS enabled.
- ▶ The `CUpti_ActivityContext` activity record has been updated in a backwards compatible manner. See [Context Activity Record](#) for more information.
- ▶ To ensure that metric values are calculated as accurately as possible, a new metric API is introduced. Function `cuptiMetricGetRequiredEventGroupSets` can be used to get the groups of events that should be collected at the same time.
- ▶ Execution overheads introduced by CUPTI have been dramatically decreased.
- ▶ The new activity buffer API introduced in CUDA Toolkit 5.5 is now required. The legacy `cuptiActivityEnqueueBuffer` and `cuptiActivityDequeueBuffer` functions have been removed.

# TABLE OF CONTENTS

<b>Chapter 1. Usage</b>	<b>1</b>
1.1. CUPTI Compatibility and Requirements	1
1.2. CUPTI Initialization	1
1.3. CUPTI Activity API	1
1.3.1. Context Activity Record	3
1.3.2. Legacy Activity Records	3
1.4. CUPTI Callback API	3
1.4.1. Driver and Runtime API Callbacks	4
1.4.2. Resource Callbacks	5
1.4.3. Synchronization Callbacks	5
1.4.4. NVIDIA Tools Extension Callbacks	6
1.5. CUPTI Event API	7
1.5.1. Collecting Kernel Execution Events	9
1.5.2. Sampling Events	10
1.6. CUPTI Metric API	10
1.7. Samples	26
<b>Chapter 2. Modules</b>	<b>27</b>
2.1. CUPTI Version	27
cuptiGetVersion	27
CUPTI_API_VERSION	28
2.2. CUPTI Result Codes	28
CUptiResult	28
cuptiGetResultString	30
2.3. CUPTI Activity API	30
CUpti_Activity	31
CUpti_ActivityAPI	31
CUpti_ActivityBranch	31
CUpti_ActivityCdpKernel	31
CUpti_ActivityContext	31
CUpti_ActivityDevice	31
CUpti_ActivityDeviceAttribute	31
CUpti_ActivityEnvironment	31
CUpti_ActivityEvent	31
CUpti_ActivityEventInstance	31
CUpti_ActivityFunction	31
CUpti_ActivityGlobalAccess	31
CUpti_ActivityInstructionExecution	31
CUpti_ActivityKernel	32
CUpti_ActivityKernel2	32
CUpti_ActivityMarker	32

CUpti_ActivityMarkerData.....	32
CUpti_ActivityMemcpy.....	32
CUpti_ActivityMemcpy2.....	32
CUpti_ActivityMemset.....	32
CUpti_ActivityMetric.....	32
CUpti_ActivityMetricInstance.....	32
CUpti_ActivityModule.....	33
CUpti_ActivityName.....	33
CUpti_ActivityObjectKindId.....	33
CUpti_ActivityOverhead.....	33
CUpti_ActivityPreemption.....	33
CUpti_ActivitySharedAccess.....	33
CUpti_ActivitySourceLocator.....	33
CUpti_ActivityUnifiedMemoryCounter.....	33
CUpti_ActivityUnifiedMemoryCounterConfig.....	33
CUpti_ActivityAttribute.....	33
CUpti_ActivityComputeApiKind.....	34
CUpti_ActivityEnvironmentKind.....	34
CUpti_ActivityFlag.....	35
CUpti_ActivityKind.....	36
CUpti_ActivityMemcpyKind.....	40
CUpti_ActivityMemoryKind.....	40
CUpti_ActivityObjectKind.....	41
CUpti_ActivityOverheadKind.....	41
CUpti_ActivityPreemptionKind.....	42
CUpti_ActivityUnifiedMemoryCounterKind.....	42
CUpti_ActivityUnifiedMemoryCounterScope.....	42
CUpti_EnvironmentClocksThrottleReason.....	43
CUpti_BuffersCallbackCompleteFunc.....	44
CUpti_BuffersCallbackRequestFunc.....	44
cuptiActivityConfigureUnifiedMemoryCounter.....	44
cuptiActivityDisable.....	45
cuptiActivityDisableContext.....	45
cuptiActivityEnable.....	46
cuptiActivityEnableContext.....	46
cuptiActivityFlush.....	47
cuptiActivityFlushAll.....	48
cuptiActivityGetAttribute.....	49
cuptiActivityGetNextRecord.....	49
cuptiActivityGetNumDroppedRecords.....	50
cuptiActivityRegisterCallbacks.....	51
cuptiActivitySetAttribute.....	52
cuptiGetContextId.....	53

cuptiGetDeviceld.....	53
cuptiGetStreamId.....	54
cuptiGetTimestamp.....	55
CUPTI_CORRELATION_ID_UNKNOWN.....	55
CUPTI_GRID_ID_UNKNOWN.....	55
CUPTI_SOURCE_LOCATOR_ID_UNKNOWN.....	55
CUPTI_TIMESTAMP_UNKNOWN.....	55
2.4. CUPTI Callback API.....	55
CUpti_CallbackData.....	56
CUpti_ModuleResourceData.....	56
CUpti_NvtxData.....	56
CUpti_ResourceData.....	56
CUpti_SynchronizeData.....	56
CUpti_ApiCallbackSite.....	56
CUpti_CallbackDomain.....	56
CUpti_CallbackIdResource.....	57
CUpti_CallbackIdSync.....	57
CUpti_CallbackFunc.....	58
CUpti_CallbackId.....	58
CUpti_DomainTable.....	58
CUpti_SubscriberHandle.....	58
cuptiEnableAllDomains.....	59
cuptiEnableCallback.....	59
cuptiEnableDomain.....	60
cuptiGetCallbackName.....	61
cuptiGetCallbackState.....	62
cuptiSubscribe.....	63
cuptiSupportedDomains.....	64
cuptiUnsubscribe.....	64
2.5. CUPTI Event API.....	65
CUpti_EventGroupSet.....	65
CUpti_EventGroupSets.....	65
CUpti_DeviceAttribute.....	65
CUpti_DeviceAttributeDeviceClass.....	66
CUpti_EventAttribute.....	66
CUpti_EventCategory.....	66
CUpti_EventCollectionMethod.....	67
CUpti_EventCollectionMode.....	67
CUpti_EventDomainAttribute.....	68
CUpti_EventGroupAttribute.....	68
CUpti_ReadEventFlags.....	69
CUpti_EventDomainID.....	69
CUpti_EventGroup.....	69

CUpti_EventID.....	69
cuptiDeviceEnumEventDomains.....	70
cuptiDeviceGetAttribute.....	71
cuptiDeviceGetEventDomainAttribute.....	71
cuptiDeviceGetNumEventDomains.....	73
cuptiDeviceGetTimestamp.....	73
cuptiDisableKernelReplayMode.....	74
cuptiEnableKernelReplayMode.....	74
cuptiEnumEventDomains.....	75
cuptiEventDomainEnumEvents.....	76
cuptiEventDomainGetAttribute.....	76
cuptiEventDomainGetNumEvents.....	78
cuptiEventGetAttribute.....	78
cuptiEventGetIdFromName.....	79
cuptiEventGroupAddEvent.....	80
cuptiEventGroupCreate.....	81
cuptiEventGroupDestroy.....	82
cuptiEventGroupDisable.....	83
cuptiEventGroupEnable.....	83
cuptiEventGroupGetAttribute.....	84
cuptiEventGroupReadAllEvents.....	85
cuptiEventGroupReadEvent.....	87
cuptiEventGroupRemoveAllEvents.....	88
cuptiEventGroupRemoveEvent.....	89
cuptiEventGroupResetAllEvents.....	90
cuptiEventGroupSetAttribute.....	90
cuptiEventGroupSetDisable.....	91
cuptiEventGroupSetEnable.....	92
cuptiEventGroupSetsCreate.....	93
cuptiEventGroupSetsDestroy.....	94
cuptiGetNumEventDomains.....	94
cuptiSetEventCollectionMode.....	95
CUPTI_EVENT_OVERFLOW.....	96
2.6. CUPTI Metric API.....	96
CUpti_MetricValue.....	96
CUpti_MetricAttribute.....	96
CUpti_MetricCategory.....	96
CUpti_MetricEvaluationMode.....	97
CUpti_MetricPropertyDeviceClass.....	97
CUpti_MetricPropertyID.....	98
CUpti_MetricValueKind.....	98
CUpti_MetricValueUtilizationLevel.....	99
CUpti_MetricID.....	99

cuptiDeviceEnumMetrics.....	99
cuptiDeviceGetNumMetrics.....	100
cuptiEnumMetrics.....	101
cuptiGetNumMetrics.....	101
cuptiMetricCreateEventGroupSets.....	102
cuptiMetricEnumEvents.....	103
cuptiMetricEnumProperties.....	103
cuptiMetricGetAttribute.....	104
cuptiMetricGetIdFromName.....	105
cuptiMetricGetNumEvents.....	106
cuptiMetricGetNumProperties.....	106
cuptiMetricGetRequiredEventGroupSets.....	107
cuptiMetricGetValue.....	108
cuptiMetricGetValue2.....	109
<b>Chapter 3. Data Structures.....</b>	<b>112</b>
CUpti_Activity.....	114
kind.....	114
CUpti_ActivityAPI.....	114
cbid.....	114
correlationId.....	115
end.....	115
kind.....	115
processId.....	115
returnValue.....	115
start.....	115
threadId.....	115
CUpti_ActivityBranch.....	115
correlationId.....	115
diverged.....	116
executed.....	116
kind.....	116
pcOffset.....	116
sourceLocatorId.....	116
threadsExecuted.....	116
CUpti_ActivityCdpKernel.....	116
blockX.....	116
blockY.....	116
blockZ.....	116
completed.....	117
contextId.....	117
correlationId.....	117
deviceId.....	117
dynamicSharedMemory.....	117

end.....	117
executed.....	117
gridId.....	117
gridX.....	117
gridY.....	118
gridZ.....	118
kind.....	118
localMemoryPerThread.....	118
localMemoryTotal.....	118
name.....	118
parentBlockX.....	118
parentBlockY.....	118
parentBlockZ.....	118
parentGridId.....	118
queued.....	119
registersPerThread.....	119
requested.....	119
sharedMemoryConfig.....	119
start.....	119
staticSharedMemory.....	119
streamId.....	119
submitted.....	119
CUpti_ActivityContext.....	119
computeApiKind.....	120
contextId.....	120
deviceId.....	120
kind.....	120
nullStreamId.....	120
CUpti_ActivityDevice.....	120
computeCapabilityMajor.....	120
computeCapabilityMinor.....	120
constantMemorySize.....	120
coreClockRate.....	121
flags.....	121
globalMemoryBandwidth.....	121
globalMemorySize.....	121
id.....	121
kind.....	121
l2CacheSize.....	121
maxBlockDimX.....	121
maxBlockDimY.....	121
maxBlockDimZ.....	121
maxBlocksPerMultiprocessor.....	122

maxGridDimX.....	122
maxGridDimY.....	122
maxGridDimZ.....	122
maxIPC.....	122
maxRegistersPerBlock.....	122
maxSharedMemoryPerBlock.....	122
maxThreadsPerBlock.....	122
maxWarpsPerMultiprocessor.....	122
name.....	122
numMemcpyEngines.....	123
numMultiprocessors.....	123
numThreadsPerWarp.....	123
CUpti_ActivityDeviceAttribute.....	123
attribute.....	123
deviceId.....	123
flags.....	123
kind.....	124
value.....	124
CUpti_ActivityEnvironment.....	124
clocksThrottleReasons.....	124
cooling.....	124
deviceId.....	124
environmentKind.....	124
fanSpeed.....	124
gpuTemperature.....	125
kind.....	125
memoryClock.....	125
pcieLinkGen.....	125
pcieLinkWidth.....	125
power.....	125
power.....	125
powerLimit.....	125
smClock.....	125
speed.....	125
temperature.....	126
timestamp.....	126
CUpti_ActivityEvent.....	126
correlationId.....	126
domain.....	126
id.....	126
kind.....	126
value.....	126
CUpti_ActivityEventInstance.....	127

correlationId.....	127
domain.....	127
id.....	127
instance.....	127
kind.....	127
pad.....	127
value.....	127
CUpti_ActivityFunction.....	128
contextId.....	128
functionIndex.....	128
id.....	128
kind.....	128
moduleId.....	128
name.....	128
CUpti_ActivityGlobalAccess.....	128
correlationId.....	128
executed.....	129
flags.....	129
kind.....	129
l2_transactions.....	129
pcOffset.....	129
sourceLocatorId.....	129
threadsExecuted.....	129
CUpti_ActivityInstructionExecution.....	129
correlationId.....	129
executed.....	130
flags.....	130
functionId.....	130
kind.....	130
notPredOffThreadsExecuted.....	130
pad.....	130
pcOffset.....	130
sourceLocatorId.....	130
threadsExecuted.....	130
CUpti_ActivityKernel.....	131
blockX.....	131
blockY.....	131
blockZ.....	131
cacheConfigExecuted.....	131
cacheConfigRequested.....	131
contextId.....	131
correlationId.....	131
deviceId.....	131

dynamicSharedMemory.....	132
end.....	132
gridX.....	132
gridY.....	132
gridZ.....	132
kind.....	132
localMemoryPerThread.....	132
localMemoryTotal.....	132
name.....	132
pad.....	132
registersPerThread.....	133
reserved0.....	133
runtimeCorrelationId.....	133
start.....	133
staticSharedMemory.....	133
streamId.....	133
CUpti_ActivityKernel2.....	133
blockX.....	133
blockY.....	133
blockZ.....	134
completed.....	134
contextId.....	134
correlationId.....	134
deviceId.....	134
dynamicSharedMemory.....	134
end.....	134
executed.....	134
gridId.....	134
gridX.....	134
gridY.....	135
gridZ.....	135
kind.....	135
localMemoryPerThread.....	135
localMemoryTotal.....	135
name.....	135
registersPerThread.....	135
requested.....	135
reserved0.....	135
sharedMemoryConfig.....	135
start.....	136
staticSharedMemory.....	136
streamId.....	136
CUpti_ActivityMarker.....	136

flags.....	136
id.....	136
kind.....	136
name.....	136
objectId.....	137
objectKind.....	137
timestamp.....	137
CUpti_ActivityMarkerData.....	137
category.....	137
color.....	137
flags.....	137
id.....	137
kind.....	137
payload.....	138
payloadKind.....	138
CUpti_ActivityMemcpy.....	138
bytes.....	138
contextId.....	138
copyKind.....	138
correlationId.....	138
deviceId.....	138
dstKind.....	138
end.....	139
flags.....	139
kind.....	139
reserved0.....	139
runtimeCorrelationId.....	139
srcKind.....	139
start.....	139
streamId.....	140
CUpti_ActivityMemcpy2.....	140
bytes.....	140
contextId.....	140
copyKind.....	140
correlationId.....	140
deviceId.....	140
dstContextId.....	140
dstDeviceId.....	140
dstKind.....	141
end.....	141
flags.....	141
kind.....	141
pad.....	141

reserved0.....	141
srcContextId.....	141
srcDeviceId.....	141
srcKind.....	142
start.....	142
streamId.....	142
CUpti_ActivityMemset.....	142
bytes.....	142
contextId.....	142
correlationId.....	142
deviceId.....	142
end.....	143
kind.....	143
reserved0.....	143
runtimeCorrelationId.....	143
start.....	143
streamId.....	143
value.....	143
CUpti_ActivityMetric.....	143
correlationId.....	144
flags.....	144
id.....	144
kind.....	144
pad.....	144
value.....	144
CUpti_ActivityMetricInstance.....	144
correlationId.....	144
flags.....	145
id.....	145
instance.....	145
kind.....	145
pad.....	145
value.....	145
CUpti_ActivityModule.....	145
contextId.....	145
cubin.....	145
cubinSize.....	146
id.....	146
kind.....	146
pad.....	146
CUpti_ActivityName.....	146
kind.....	146
name.....	146

objectId.....	146
objectKind.....	146
CUpti_ActivityObjectKindId.....	147
dcs.....	147
pt.....	147
CUpti_ActivityOverhead.....	147
end.....	147
kind.....	147
objectId.....	147
objectKind.....	148
overheadKind.....	148
start.....	148
CUpti_ActivityPreemption.....	148
blockX.....	148
blockY.....	148
blockZ.....	148
gridId.....	148
kind.....	148
pad.....	149
preemptionKind.....	149
timestamp.....	149
CUpti_ActivitySharedAccess.....	149
correlationId.....	149
executed.....	149
flags.....	149
functionId.....	149
kind.....	149
pad.....	150
pcOffset.....	150
sharedTransactions.....	150
sourceLocatorId.....	150
theoreticalSharedTransactions.....	150
threadsExecuted.....	150
CUpti_ActivitySourceLocator.....	150
fileName.....	150
id.....	150
kind.....	151
lineNumber.....	151
CUpti_ActivityUnifiedMemoryCounter.....	151
counterKind.....	151
deviceId.....	151
kind.....	151
pad.....	151

processId.....	151
scope.....	152
timestamp.....	152
value.....	152
CUpti_ActivityUnifiedMemoryCounterConfig.....	152
deviceId.....	152
enable.....	152
kind.....	152
scope.....	153
CUpti_CallbackData.....	153
callbackSite.....	153
context.....	153
contextUid.....	153
correlationData.....	153
correlationId.....	153
functionName.....	154
functionParams.....	154
functionReturnValue.....	154
symbolName.....	154
CUpti_EventGroupSet.....	154
eventGroups.....	154
numEventGroups.....	154
CUpti_EventGroupSets.....	155
numSets.....	155
sets.....	155
CUpti_MetricValue.....	155
CUpti_ModuleResourceData.....	155
cubinSize.....	155
moduleId.....	155
pCubin.....	156
CUpti_NvtxData.....	156
functionName.....	156
functionParams.....	156
CUpti_ResourceData.....	156
context.....	156
resourceDescriptor.....	157
stream.....	157
CUpti_SynchronizeData.....	157
context.....	157
stream.....	157
<b>Chapter 4. Data Fields.....</b>	<b>158</b>

# LIST OF TABLES

Table 1 Capability 1.x Metrics ..... 12

Table 2 Capability 2.x Metrics ..... 12

Table 3 Capability 3.x Metrics ..... 19

# Chapter 1.

## USAGE

The *CUDA Profiling Tools Interface* (CUPTI) enables the creation of profiling and tracing tools that target CUDA applications. CUPTI provides four APIs: *the Activity API*, the *Callback API*, the *Event API*, and the *Metric API*. Using these APIs, you can develop profiling tools that give insight into the CPU and GPU behavior of CUDA applications. CUPTI is delivered as a dynamic library on all platforms supported by CUDA.

### 1.1. CUPTI Compatibility and Requirements

New versions of the CUDA driver are backwards compatible with older versions of CUPTI. For example, a developer using a profiling tool based on CUPTI 4.1 can update to a more recently released CUDA driver. However, new versions of CUPTI are not backwards compatible with older versions of the CUDA driver. For example, a developer using a profiling tool based on CUPTI 4.1 must have a version of the CUDA driver released with CUDA Toolkit 4.1 (or later) installed as well. CUPTI calls will fail with `CUPTI_ERROR_NOT_INITIALIZED` if the CUDA driver version is not compatible with the CUPTI version.

### 1.2. CUPTI Initialization

CUPTI initialization occurs lazily the first time you invoke any CUPTI function. For the Activity, Event, Metric, and Callback APIs there are no requirements on when this initialization must occur (i.e. you can invoke the first CUPTI function at any point). See the CUPTI Activity API section for more information on CUPTI initialization requirements for the activity API.

### 1.3. CUPTI Activity API

The CUPTI Activity API allows you to asynchronously collect a trace of an application's CPU and GPU CUDA activity. The following terminology is used by the activity API.

## Activity Record

CPU and GPU activity is reported in C data structures called activity records. There is a different C structure type for each activity kind (e.g. `CUpti_ActivityMemcpy`). Records are generically referred to using the `CUpti_Activity` type. This type contains only a kind field that indicates the kind of the activity record. Using this kind, the object can be cast from the generic `CUpti_Activity` type to the specific type representing the activity. See the `printActivity` function in the [activity\\_trace\\_async](#) sample for an example.

## Activity Buffer

An activity buffer is used to transfer one or more activity records from CUPTI to the client. CUPTI fills activity buffers with activity records as the corresponding activities occur on the CPU and GPU. The CUPTI client is responsible for providing empty activity buffers as necessary to ensure that no records are dropped.

An *asynchronous* buffering API is implemented by `cuptiActivityRegisterCallbacks`, `cuptiActivityFlush`, and `cuptiActivityFlushAll`.

It is not required that the activity API be initialized before CUDA, but if the activity API is not initialized before CUDA some activity records may not be collected. You can force initialization of the activity API by enabling one or more activity kinds using `cuptiActivityEnable` or `cuptiActivityEnableContext`, as shown in the `initTrace` function of the [activity\\_trace\\_async](#) sample. Some activity kinds cannot be directly enabled, see the API documentation for `CUpti_ActivityKind` for details. Functions `cuptiActivityEnable` and `cuptiActivityEnableContext` will return `CUPTI_ERROR_NOT_COMPATIBLE` if the requested activity kind cannot be enabled.

The activity buffer API uses callbacks to request and return buffers of activity records. To use the asynchronous buffering API you must first register two callbacks using `cuptiActivityRegisterCallbacks`. One of these callbacks will be invoked whenever CUPTI needs an empty activity buffer. The other callback is used to deliver a buffer containing one or more activity records to the client. To minimize profiling overhead the client should return as quickly as possible from these callbacks. Functions `cuptiActivityFlush` and `cuptiActivityFlushAll` can be used to force CUPTI to deliver any activity buffers that contain completed activity records. Functions `cuptiActivityGetAttribute` and `cuptiActivitySetAttribute` can be used to read and write attributes that control how the buffering API behaves. See the API documentation for more information.

The [activity\\_trace\\_async](#) sample shows how to use the activity buffer API to collect a trace of CPU and GPU activity for a simple application.

### 1.3.1. Context Activity Record

In 6.0 the context activity record, `CUpti_ActivityContext`, was changed in a manner that introduced a new field into the structure. This new field was introduced in a way that preserves backward compatibility with any persisted versions of this structure.

The 32-bit `computeApiKind` field was replaced with two 16 bit fields, `computeApiKind` and `defaultStreamId`. Because all valid `computeApiKind` values fit within 16 bits, and because all supported CUDA platforms are little-endian, persisted context record data read with the new structure will have the correct value for `computeApiKind` and have a value of zero for `defaultStreamId`. The CUPTI client is responsible for versioning the persisted context data to recognize when the `defaultStreamId` field is valid.

### 1.3.2. Legacy Activity Records

In CUPTI 5.5 the `CUpti_ActivityKernel2` structure replaced `CUpti_ActivityKernel` as the activity record used for the `CUPTI_ACTIVITY_KIND_KERNEL` and `CUPTI_ACTIVITY_KIND_CONCURRENT_KERNEL` activity kinds. The `CUpti_ActivityKernel` definition is retained in CUPTI to enable newer versions of CUPTI to work with persisted activity record data.

The CUPTI client is responsible for versioning the persisted activity record data to recognize when the persisted data is stored using `CUpti_ActivityKernel` or `CUpti_ActivityKernel2`.

## 1.4. CUPTI Callback API

The CUPTI Callback API allows you to register a callback into your own code. Your callback will be invoked when the application being profiled calls a CUDA runtime or driver function, or when certain events occur in the CUDA driver. The following terminology is used by the callback API.

#### Callback Domain

Callbacks are grouped into domains to make it easier to associate your callback functions with groups of related CUDA functions or events. There are currently four callback domains, as defined by `CUpti_CallbackDomain`: a domain for CUDA runtime functions, a domain for CUDA driver functions, a domain for CUDA resource tracking, and a domain for CUDA synchronization notification.

#### Callback ID

Each callback is given a unique ID within the corresponding callback domain so that you can identify it within your callback function. The CUDA driver API IDs are defined in `cupti_driver_cbid.h` and the CUDA runtime API IDs are

defined in `cupti_runtime_cbid.h`. Both of these headers are included for you when you include `cupti.h`. The CUDA resource callback IDs are defined by `CUpti_CallbackIdResource` and the CUDA synchronization callback IDs are defined by `CUpti_CallbackIdSync`.

### Callback Function

Your callback function must be of type `CUpti_CallbackFunc`. This function type has two arguments that specify the callback domain and ID so that you know why the callback is occurring. The type also has a `cbdata` argument that is used to pass data specific to the callback.

### Subscriber

A subscriber is used to associate each of your callback functions with one or more CUDA API functions. There can be at most one subscriber initialized with `cuptiSubscribe()` at any time. Before initializing a new subscriber, the existing subscriber must be finalized with `cuptiUnsubscribe()`.

Each callback domain is described in detail below. Unless explicitly stated, it is not supported to call any CUDA runtime or driver API from within a callback function. Doing so may cause the application to hang.

## 1.4.1. Driver and Runtime API Callbacks

Using the callback API with the `CUPTI_CB_DOMAIN_DRIVER_API` or `CUPTI_CB_DOMAIN_RUNTIME_API` domains, you can associate a callback function with one or more CUDA API functions. When those CUDA functions are invoked in the application, your callback function is invoked as well. For these domains, the `cbdata` argument to your callback function will be of the type `CUpti_CallbackData`.

It is legal to call `cudaThreadSynchronize()`, `cudaDeviceSynchronize()`, `cudaStreamSynchronize()`, `cuCtxSynchronize()`, and `cuStreamSynchronize()` from within a driver or runtime API callback function.

The following code shows a typical sequence used to associate a callback function with one or more CUDA API functions. To simplify the presentation error checking code has been removed.

```
CUpti_SubscriberHandle subscriber;
MyDataStruct *my_data = ...;
...
cuptiSubscribe(&subscriber,
               (CUpti_CallbackFunc)my_callback , my_data);
cuptiEnableDomain(1, subscriber,
                  CUPTI_CB_DOMAIN_RUNTIME_API);
```

First, `cuptiSubscribe` is used to initialize a subscriber with the `my_callback` callback function. Next, `cuptiEnableDomain` is used to associate that callback with all the CUDA runtime API functions. Using this code sequence will cause `my_callback` to be called twice each time any of the CUDA runtime API functions are invoked, once on entry to the CUDA function and once just before exit from the CUDA function. CUPTI callback API functions `cuptiEnableCallback` and `cuptiEnableAllDomains` can

also be used to associate CUDA API functions with a callback (see reference below for more information).

The following code shows a typical callback function.

```
void CUPTI_API
my_callback(void *userdata, CUpti_CallbackDomain domain,
            CUpti_CallbackId cbid, const void *cbdata)
{
    const CUpti_CallbackData *cbInfo = (CUpti_CallbackData *)cbdata;
    MyDataStruct *my_data = (MyDataStruct *)userdata;

    if ((domain == CUPTI_CB_DOMAIN_RUNTIME_API) &&
        (cbid == CUPTI_RUNTIME_TRACE_CBID_cudaMemcpy_v3020)) {
        if (cbInfo->callbackSite == CUPTI_API_ENTER) {
            cudaMemcpy_v3020_params *funcParams =
                (cudaMemcpy_v3020_params *) (cbInfo->
                    functionParams);

            size_t count = funcParams->count;
            enum cudaMemcpyKind kind = funcParams->kind;
            ...
        }
    }
    ...
}
```

In your callback function, you use the `CUpti_CallbackDomain` and `CUpti_CallbackID` parameters to determine which CUDA API function invocation is causing this callback. In the example above, we are checking for the CUDA runtime `cudaMemcpy` function. The `cbdata` parameter holds a structure of useful information that can be used within the callback. In this case we use the `callbackSite` member of the structure to detect that the callback is occurring on entry to `cudaMemcpy`, and we use the `functionParams` member to access the parameters that were passed to `cudaMemcpy`. To access the parameters we first cast `functionParams` to a structure type corresponding to the `cudaMemcpy` function. These parameter structures are contained in `generated_cuda_runtime_api_meta.h`, `generated_cuda_meta.h`, and a number of other files. When possible these files are included for you by `cuprti.h`.

The **callback\_event** and **callback\_timestamp** samples described on the [samples page](#) both show how to use the callback API for the driver and runtime API domains.

## 1.4.2. Resource Callbacks

Using the callback API with the `CUPTI_CB_DOMAIN_RESOURCE` domain, you can associate a callback function with some CUDA resource creation and destruction events. For example, when a CUDA context is created, your callback function will be invoked with a callback ID equal to `CUPTI_CBID_RESOURCE_CONTEXT_CREATED`. For this domain, the `cbdata` argument to your callback function will be of the type `CUpti_ResourceData`.

## 1.4.3. Synchronization Callbacks

Using the callback API with the `CUPTI_CB_DOMAIN_SYNCHRONIZE` domain, you can associate a callback function with CUDA context and stream synchronizations. For

example, when a CUDA context is synchronized, your callback function will be invoked with a callback ID equal to `CUPTI_CBID_SYNCHRONIZE_CONTEXT_SYNCHRONIZED`. For this domain, the `cbdata` argument to your callback function will be of the type `CUpti_SynchronizeData`.

## 1.4.4. NVIDIA Tools Extension Callbacks

Using the callback API with the `CUPTI_CB_DOMAIN_NVTX` domain, you can associate a callback function with NVIDIA Tools Extension (NVTX) API functions. When an NVTX function is invoked in the application, your callback function is invoked as well. For these domains, the `cbdata` argument to your callback function will be of the type `CUpti_NvtxData`.

The NVTX library has its own convention for discovering the profiling library that will provide the implementation of the NVTX callbacks. To receive callbacks you must set the NVTX environment variables appropriately so that when the application calls an NVTX function, your profiling library receive the callbacks. The following code sequence shows a typical initialization sequence to enable NVTX callbacks and activity records.

```
/* Set env so CUPTI-based profiling library loads on first nvtx call. */
char *inj32_path = "/path/to/32-bit/version/of/cupti/based/profiling/library";
char *inj64_path = "/path/to/64-bit/version/of/cupti/based/profiling/library";
setenv("NVTX_INJECTION32_PATH", inj32_path, 1);
setenv("NVTX_INJECTION64_PATH", inj64_path, 1);
```

The following code shows a typical sequence used to associate a callback function with one or more NVTX functions. To simplify the presentation error checking code has been removed.

```
CUpti_SubscriberHandle subscriber;
MyDataStruct *my_data = ...;
...
cuptiSubscribe(&subscriber,
               (CUpti_CallbackFunc)my_callback , my_data);
cuptiEnableDomain(1, subscriber,
                  CUPTI_CB_DOMAIN_NVTX);
```

First, `cuptiSubscribe` is used to initialize a subscriber with the `my_callback` callback function. Next, `cuptiEnableDomain` is used to associate that callback with all the NVTX functions. Using this code sequence will cause `my_callback` to be called once each time any of the NVTX functions are invoked. CUPTI callback API functions `cuptiEnableCallback` and `cuptiEnableAllDomains` can also be used to associate NVTX API functions with a callback (see reference below for more information).

The following code shows a typical callback function.

```
void CUPTI_API
my_callback(void *userdata, CUpti_CallbackDomain domain,
            CUpti_CallbackId cbid, const void *cbdata)
{
    const CUpti_NvtxData *nvtxInfo = (CUpti_NvtxData *)cbdata;
    MyDataStruct *my_data = (MyDataStruct *)userdata;

    if ((domain == CUPTI_CB_DOMAIN_NVTX) &&
        (cbid == NVTX_CBID_CORE_NameOsThreadA)) {
        nvtxNameOsThreadA_params *params = (nvtxNameOsThreadA_params *)nvtxInfo->
            functionParams;
        ...
    }
    ...
}
```

In your callback function, you use the `CUpti_CallbackDomain` and `CUpti_CallbackID` parameters to determine which NVTX API function invocation is causing this callback. In the example above, we are checking for the `nvtxNameOsThreadA` function. The `cbdata` parameter holds a structure of useful information that can be used within the callback. In this case, we use the `functionParams` member to access the parameters that were passed to `nvtxNameOsThreadA`. To access the parameters we first cast `functionParams` to a structure type corresponding to the `nvtxNameOsThreadA` function. These parameter structures are contained in `generated_nvtx_meta.h`.

## 1.5. CUPTI Event API

The CUPTI Event API allows you to query, configure, start, stop, and read the event counters on a CUDA-enabled device. The following terminology is used by the event API.

### Event

An event is a countable activity, action, or occurrence on a device.

### Event ID

Each event is assigned a unique identifier. A named event will represent the same activity, action, or occurrence on all device types. But the named event may have different IDs on different device families. Use `cuptiEventGetIdFromName` to get the ID for a named event on a particular device.

### Event Category

Each event is placed in one of the categories defined by `CUpti_EventCategory`. The category indicates the general type of activity, action, or occurrence measured by the event.

### Event Domain

A device exposes one or more event domains. Each event domain represents a group of related events available on that device. A device may have multiple instances of a domain, indicating that the device can simultaneously record multiple instances of each event within that domain.

## Event Group

An event group is a collection of events that are managed together. The number and type of events that can be added to an event group are subject to device-specific limits. At any given time, a device may be configured to count events from a limited number of event groups. All events in an event group must belong to the same event domain.

## Event Group Set

An event group set is a collection of event groups that can be enabled at the same time. Event group sets are created by `cuptiEventGroupSetsCreate` and `cuptiMetricCreateEventGroupSets`.

You can determine the events available on a device using the `cuptiDeviceEnumEventDomains` and `cuptiEventDomainEnumEvents` functions. The **cupti\_query** sample described on the [samples page](#) shows how to use these functions. You can also enumerate all the CUPTI events available on any device using the `cuptiEnumEventDomains` function.

Configuring and reading event counts requires the following steps. First, select your event collection mode. If you want to count events that occur during the execution of a kernel, use `cuptiSetEventCollectionMode` to set mode `CUPTI_EVENT_COLLECTION_MODE_KERNEL`. If you want to continuously sample the event counts, use mode `CUPTI_EVENT_COLLECTION_MODE_CONTINUOUS`. Next determine the names of the events that you want to count, and then use the `cuptiEventGroupCreate`, `cuptiEventGetIdFromName`, and `cuptiEventGroupAddEvent` functions to create and initialize an event group with those events. If you are unable to add all the events to a single event group then you will need to create multiple event groups. Alternatively, you can use the `cuptiEventGroupSetsCreate` function to automatically create the event group(s) required for a set of events.

To begin counting a set of events, enable the event group or groups that contain those events by using the `cuptiEventGroupEnable` function. If your events are contained in multiple event groups you may be unable to enable all of the event groups at the same time, due to device limitations. In this case, you can gather the events across multiple executions of the application or you can enable kernel replay. If you enable kernel replay using `cuptiEnableKernelReplayMode` you will be able to enable any number of event groups and all the contained events will be collected.

Use the `cuptiEventGroupReadEvent` and/or `cuptiEventGroupReadAllEvents` functions to read the event values. When you are done collecting events, use the `cuptiEventGroupDisable` function to stop counting of the events contained in an event group. The **callback\_event** sample described on the [samples page](#) shows how to use these functions to create, enable, and disable event groups, and how to read event counts.

## 1.5.1. Collecting Kernel Execution Events

A common use of the event API is to count a set of events during the execution of a kernel (as demonstrated by the **callback\_event** sample). The following code shows a typical callback used for this purpose. Assume that the callback was enabled only for a kernel launch using the CUDA runtime (i.e. by `cuptiEnableCallback(1, subscriber, CUPTI_CB_DOMAIN_RUNTIME_API, CUPTI_RUNTIME_TRACE_CBID_cudaLaunch_v3020)`). To simplify the presentation error checking code has been removed.

```
static void CUPTIAPI
getEventValueCallback(void *userdata,
                      CUpti_CallbackDomain domain,
                      CUpti_CallbackId cbid,
                      const void *cbdata)
{
    const CUpti_CallbackData *cbData =
        (CUpti_CallbackData *)cbdata;

    if (cbData->callbackSite == CUPTI_API_ENTER) {
        cudaThreadSynchronize();
        cuptiSetEventCollectionMode(cbInfo->context,
                                    CUPTI_EVENT_COLLECTION_MODE_KERNEL);
        cuptiEventGroupEnable(eventGroup);
    }

    if (cbData->callbackSite == CUPTI_API_EXIT) {
        cudaThreadSynchronize();
        cuptiEventGroupReadEvent(eventGroup,
                                  CUPTI_EVENT_READ_FLAG_NONE,
                                  eventId,
                                  &bytesRead, &eventVal);

        cuptiEventGroupDisable(eventGroup);
    }
}
```

Two synchronization points are used to ensure that events are counted only for the execution of the kernel. If the application contains other threads that launch kernels, then additional thread-level synchronization must also be introduced to ensure that those threads do not launch kernels while the callback is collecting events. When the `cudaLaunch` API is entered (that is, before the kernel is actually launched on the device), `cudaThreadSynchronize` is used to wait until the GPU is idle. The event collection mode is set to `CUPTI_EVENT_COLLECTION_MODE_KERNEL` so that the event counters are automatically started and stopped just before and after the kernel executes. Then event collection is enabled with `cuptiEventGroupEnable`.

When the `cudaLaunch` API is exited (that is, after the kernel is queued for execution on the GPU) another `cudaThreadSynchronize` is used to cause the CPU thread to wait for the kernel to finish execution. Finally, the event counts are read with `cuptiEventGroupReadEvent`.

## 1.5.2. Sampling Events

The event API can also be used to sample event values while a kernel or kernels are executing (as demonstrated by the **event\_sampling** sample). The sample shows one possible way to perform the sampling. The event collection mode is set to `CUPTI_EVENT_COLLECTION_MODE_CONTINUOUS` so that the event counters run continuously. Two threads are used in **event\_sampling**: one thread schedules the kernels and memcpys that perform the computation, while another thread wakes periodically to sample an event counter. In this sample there is no correlation of the event samples with what is happening on the GPU. To get some coarse correlation, you can use `cuptiDeviceGetTimestamp` to collect the GPU timestamp at the time of the sample and also at other interesting points in your application.

## 1.6. CUPTI Metric API

The CUPTI Metric API allows you to collect application metrics calculated from one or more event values. The following terminology is used by the metric API.

### Metric

An characteristic of an application that is calculated from one or more event values.

### Metric ID

Each metric is assigned a unique identifier. A named metric will represent the same characteristic on all device types. But the named metric may have different IDs on different device families. Use `cuptiMetricGetIdFromName` to get the ID for a named metric on a particular device.

### Metric Category

Each metric is placed in one of the categories defined by `CUpti_MetricCategory`. The category indicates the general type of the characteristic measured by the metric.

### Metric Property

Each metric is calculated from input values. These input values can be events or properties of the device or system. The available properties are defined by `CUpti_MetricPropertyID`.

### Metric Value

Each metric has a value that represents one of the kinds defined by `CUpti_MetricValueKind`. For each value kind, there is a corresponding member of the `CUpti_MetricValue` union that is used to hold the metric's value.

The tables included in this section list the metrics available for each device, as determined by the device's compute capability. You can also determine the metrics available on a device using the `cuptiDeviceEnumMetrics` function. The **cupti\_query** sample described on the [samples page](#) shows how to use this function. You can also enumerate all the CUPTI metrics available on any device using the `cuptiEnumMetrics` function.

CUPTI provides two functions for calculating a metric value. `cuptiMetricGetValue2` can be used to calculate a metric value when the device is not available. All required event values and metric properties must be provided by the caller. `cuptiMetricGetValue` can be used to calculate a metric value when the device is available (as a `CUdevice` object). All required event values must be provided by the caller but CUPTI will determine the appropriate property values from the `CUdevice` object.

Configuring and calculating metric values requires the following steps. First, determine the name of the metric that you want to collect, and then use the `cuptiMetricGetIdFromName` to get the metric ID. Use `cuptiMetricEnumEvents` to get the events required to calculate the metric and follow instructions in the CUPTI Event API section to create the event groups for those events. When creating event groups in this manner it is important to use the result of `cuptiMetricGetRequiredEventGroupSets` to properly group together events that must be collected in the same pass to ensure proper metric calculation.

Alternatively, you can use the `cuptiMetricCreateEventGroupSets` function to automatically create the event group(s) required for metric's events. When using this function events will be grouped as required to most accurately calculate the metric, as a result it is not necessary to use `cuptiMetricGetRequiredEventGroupSets`.

If you are using `cuptiMetricGetValue2` the you must also collect the required metric property values using `cuptiMetricEnumProperties`.

Collect event counts as described in the CUPTI Event API section, and then use either `cuptiMetricGetValue` or `cuptiMetricGetValue2` to calculate the metric value from the collected event and property values. The **callback\_metric** sample described on the [samples page](#) shows how to use the functions to calculate event values and calculate a metric using `cuptiMetricGetValue`. Note that, as shown in the example, you should collect event counts from all domain instances and normalize the counts to get the most accurate metric values. It is necessary to normalize the event counts because the number of event counter instances varies by device and by the event being counted.

For example, a device might have 8 multiprocessors but only have event counters for 4 of the multiprocessors, and might have 3 memory units and only have events counters for one memory unit. When calculating a metric that requires a multiprocessor event and a memory unit event, the 4 multiprocessor counters should be summed and multiplied by 2 to normalize the event count across the entire device. Similarly, the one memory unit counter should be multiplied by 3 to normalize the event count across the entire device. The normalized values can then be passed to `cuptiMetricGetValue` or `cuptiMetricGetValue2` to calculate the metric value.

As described, the normalization assumes the kernel executes a sufficient number of blocks to completely load the device. If the kernel has only a small number of blocks, normalizing across the entire device may skew the result.

### Metric Reference - Compute Capability 1.x

Devices with compute capability less than 2.0 implement the metrics shown in the following table. A scope value of single-context indicates that the metric can only be accurately collected when a single context (CUDA or graphic) is executing on the GPU. A scope value of multi-context indicates that the metric can be accurately collected when multiple contexts are executing on the GPU.

**Table 1 Capability 1.x Metrics**

Metric Name	Description	Scope
branch_efficiency	Ratio of non-divergent branches to total branches expressed as percentage	Single-context
gld_efficiency	Ratio of requested global memory load transactions to actual global memory load transactions expressed as percentage	Single-context
gst_efficiency	Ratio of requested global memory store transactions to actual global memory store transactions expressed as percentage	Single-context
gld_requested_throughput	Requested global memory load throughput	Single-context
gst_requested_throughput	Requested global memory store throughput	Single-context

### Metric Reference - Compute Capability 2.x

Devices with compute capability between 2.0, inclusive, and 3.0 implement the metrics shown in the following table. A scope value of single-context indicates that the metric can only be accurately collected when a single context (CUDA or graphic) is executing on the GPU. A scope value of multi-context indicates that the metric can be accurately collected when multiple contexts are executing on the GPU.

**Table 2 Capability 2.x Metrics**

Metric Name	Description	Scope
sm_efficiency	The percentage of time at least one warp is active on a multiprocessor averaged over all multiprocessors on the GPU	Single-context
sm_efficiency_instance	The percentage of time at least one warp is active on a specific multiprocessor	Single-context
achieved_occupancy	Ratio of the average active warps per active cycle to the maximum number of warps supported on a multiprocessor	Multi-context

Metric Name	Description	Scope
issue_slot_utilization	Percentage of issue slots that issued at least one instruction, averaged across all cycles	Multi-context
inst_executed	The number of instructions executed	Multi-context
inst_issued	The number of instructions issued	Multi-context
issue_slots	The number of issue slots used	Multi-context
executed_ipc	Instructions executed per cycle	Multi-context
issued_ipc	Instructions issued per cycle	Multi-context
ipc_instance	Instructions executed per cycle for a single multiprocessor	Multi-context
inst_per_warp	Average number of instructions executed by each warp	Multi-context
cf_issued	Number of issued control-flow instructions	Multi-context
cf_executed	Number of executed control-flow instructions	Multi-context
ldst_issued	Number of issued load and store instructions	Multi-context
ldst_executed	Number of executed load and store instructions	Multi-context
branch_efficiency	Ratio of non-divergent branches to total branches expressed as percentage	Multi-context
warp_execution_efficiency	Ratio of the average active threads per warp to the maximum number of threads per warp supported on a multiprocessor expressed as percentage	Multi-context
inst_replay_overhead	Average number of replays for each instruction executed	Multi-context
shared_replay_overhead	Average number of replays due to shared memory conflicts for each instruction executed	Single-context
global_cache_replay_overhead	Average number of replays due to global memory cache misses for each instruction executed	Single-context
local_replay_overhead	Average number of replays due to local memory accesses for each instruction executed	Single-context

Metric Name	Description	Scope
gld_efficiency	Ratio of requested global memory load throughput to required global memory load throughput expressed as percentage	Single-context
gst_efficiency	Ratio of requested global memory store throughput to required global memory store throughput expressed as percentage	Single-context
gld_transactions	Number of global memory load transactions	Single-context
gst_transactions	Number of global memory store transactions	Single-context
gld_transactions_per_request	Average number of global memory load transactions performed for each global memory load	Single-context
gst_transactions_per_request	Average number of global memory store transactions performed for each global memory store	Single-context
gld_throughput	Global memory load throughput	Single-context
gst_throughput	Global memory store throughput	Single-context
gld_requested_throughput	Requested global memory load throughput	Multi-context
gst_requested_throughput	Requested global memory store throughput	Multi-context
local_load_transactions	Number of local memory load transactions	Single-context
local_store_transactions	Number of local memory store transactions	Single-context
local_load_transactions_per_request	Average number of local memory load transactions performed for each local memory load	Single-context
local_store_transactions_per_request	Average number of local memory store transactions performed for each local memory store	Single-context
local_load_throughput	Local memory load throughput	Single-context
local_store_throughput	Local memory store throughput	Single-context
shared_load_transactions	Number of shared memory load transactions	Single-context
shared_store_transactions	Number of shared memory store transactions	Single-context
shared_load_transactions_per_request	Average number of shared memory load transactions performed for each shared memory load	Single-context

Metric Name	Description	Scope
shared_store_transactions_per_request	Average number of shared memory store transactions performed for each shared memory store	Single-context
shared_load_throughput	Shared memory load throughput	Single-context
shared_store_throughput	Shared memory store throughput	Single-context
shared_efficiency	Ratio of requested shared memory throughput to required shared memory throughput expressed as percentage	Single-context
dram_read_transactions	Device memory read transactions	Single-context
dram_write_transactions	Device memory write transactions	Single-context
dram_read_throughput	Device memory read throughput	Single-context
dram_write_throughput	Device memory write throughput	Single-context
sysmem_read_transactions	System memory read transactions	Single-context
sysmem_write_transactions	System memory write transactions	Single-context
sysmem_read_throughput	System memory read throughput	Single-context
sysmem_write_throughput	System memory write throughput	Single-context
l1_cache_global_hit_rate	Hit rate in L1 cache for global loads	Single-context
l1_cache_local_hit_rate	Hit rate in L1 cache for local loads and stores	Single-context
tex_cache_hit_rate	Texture cache hit rate	Single-context
tex_cache_transactions	Texture cache read transactions	Single-context
tex_cache_throughput	Texture cache throughput	Single-context
l2_read_transactions	Memory read transactions seen at L2 cache for all read requests	Single-context
l2_write_transactions	Memory write transactions seen at L2 cache for all write requests	Single-context
l2_read_throughput	Memory read throughput seen at L2 cache for all read requests	Single-context
l2_write_throughput	Memory write throughput seen at L2 cache for all write requests	Single-context
l2_l1_read_hit_rate	Hit rate at L2 cache for all read requests from L1 cache	Single-context

Metric Name	Description	Scope
l2_l1_read_throughput	Memory read throughput seen at L2 cache for read requests from L1 cache	Single-context
l2_texture_read_hit_rate	Hit rate at L2 cache for all read requests from texture cache	Single-context
l2_texture_read_throughput	Memory read throughput seen at L2 cache for read requests from the texture cache	Single-context
local_memory_overhead	Ratio of local memory traffic to total memory traffic between the L1 and L2 caches expressed as percentage	Single-context
l1_shared_utilization	The utilization level of the L1/shared memory relative to peak utilization on a scale of 0 to 10	Single-context
l2_utilization	The utilization level of the L2 cache relative to the peak utilization on a scale of 0 to 10	Single-context
tex_utilization	The utilization level of the texture cache relative to the peak utilization on a scale of 0 to 10	Single-context
dram_utilization	The utilization level of the device memory relative to the peak utilization on a scale of 0 to 10	Single-context
sysmem_utilization	The utilization level of the system memory relative to the peak utilization on a scale of 0 to 10	Single-context
ldst_fu_utilization	The utilization level of the multiprocessor function units that execute global, local and shared memory instructions on a scale of 0 to 10	Multi-context
cf_fu_utilization	The utilization level of the multiprocessor function units that execute control-flow instructions on a scale of 0 to 10	Multi-context
tex_fu_utilization	The utilization level of the multiprocessor function units that execute texture instructions on a scale of 0 to 10	Multi-context
alu_fu_utilization	The utilization level of the multiprocessor function units that execute integer and	Multi-context

Metric Name	Description	Scope
	floating-point arithmetic instructions on a scale of 0 to 10	
inst_fp_32	Number of single-precision floating-point instructions executed by non-predicated threads (arithmetic, compare, etc.)	Multi-context
inst_fp_64	Number of double-precision floating-point instructions executed by non-predicated threads (arithmetic, compare, etc.)	Multi-context
inst_integer	Number of integer instructions executed by non-predicated threads	Multi-context
inst_bit_convert	Number of bit-conversion instructions executed by non-predicated threads	Multi-context
inst_control	Number of control-flow instructions executed by non-predicated threads (jump, branch, etc.)	Multi-context
inst_compute_ld_st	Number of compute load/store instructions executed by non-predicated threads	Multi-context
inst_inter_thread_communication	Number of inter-thread communication instructions executed by non-predicated threads	Multi-context
inst_misc	Number of miscellaneous instructions executed by non-predicated threads	Multi-context
flops_sp	Single-precision floating point operations executed	Multi-context
flops_sp_add	Single-precision floating point add operations executed	Multi-context
flops_sp_mul	Single-precision floating point multiply operations executed	Multi-context
flops_sp_fma	Single-precision floating point multiply-accumulate operations executed	Multi-context
flops_dp	Double-precision floating point operations executed	Multi-context
flops_dp_add	Double-precision floating point add operations executed	Multi-context

Metric Name	Description	Scope
flops_dp_mul	Double-precision floating point multiply operations executed	Multi-context
flops_dp_fma	Double-precision floating point multiply-accumulate operations executed	Multi-context
flops_sp_special	Single-precision floating point special operations executed	Multi-context
stall_inst_fetch	Percentage of stalls occurring because the next assembly instruction has not yet been fetched	Multi-context
stall_exec_dependency	Percentage of stalls occurring because an input required by the instruction is not yet available	Multi-context
stall_data_request	Percentage of stalls occurring because a memory operation cannot be performed due to the required resources not being available or fully utilized, or because too many requests of a given type are outstanding	Multi-context
stall_sync	Percentage of stalls occurring because the warp is blocked at a __syncthreads() call	Multi-context
stall_texture	Percentage of stalls occurring because the texture sub-system is fully utilized or has too many outstanding requests	Multi-context
stall_other	Percentage of stalls occurring due to miscellaneous reasons	Multi-context

### Metric Reference - Compute Capability 3.x

Devices with compute capability between 3.0, inclusive, and 4.0 implement the metrics shown in the following table. A scope value of single-context indicates that the metric can only be accurately collected when a single context (CUDA or graphic) is executing on the GPU. A scope value of multi-context indicates that the metric can be accurately collected when multiple contexts are executing on the GPU.

Table 3 Capability 3.x Metrics

Metric Name	Description	Scope
sm_efficiency	The percentage of time at least one warp is active on a multiprocessor averaged over all multiprocessors on the GPU	Single-context
sm_efficiency_instance	The percentage of time at least one warp is active on a specific multiprocessor	Single-context
achieved_occupancy	Ratio of the average active warps per active cycle to the maximum number of warps supported on a multiprocessor	Multi-context
issue_slot_utilization	Percentage of issue slots that issued at least one instruction, averaged across all cycles	Multi-context
inst_executed	The number of instructions executed	Multi-context
inst_issued	The number of instructions issued	Multi-context
issue_slots	The number of issue slots used	Multi-context
ipc	Instructions executed per cycle	Multi-context
issued_ipc	Instructions issued per cycle	Multi-context
ipc_instance	Instructions executed per cycle for a single multiprocessor	Multi-context
inst_per_warp	Average number of instructions executed by each warp	Multi-context
cf_issued	Number of issued control-flow instructions	Multi-context
cf_executed	Number of executed control-flow instructions	Multi-context
ldst_issued	Number of issued load and store instructions	Multi-context
ldst_executed	Number of executed load and store instructions	Multi-context
branch_efficiency	Ratio of non-divergent branches to total branches expressed as percentage. This is available for compute capability 3.0.	Multi-context
warp_execution_efficiency	Ratio of the average active threads per warp to the maximum number of threads per warp supported on a multiprocessor expressed as percentage	Multi-context

Metric Name	Description	Scope
warp_nonpred_execution_efficiency	Ratio of the average active threads per warp executing non-predicated instructions to the maximum number of threads per warp supported on a multiprocessor expressed as percentage	Multi-context
inst_replay_overhead	Average number of replays for each instruction executed	Multi-context
shared_replay_overhead	Average number of replays due to shared memory conflicts for each instruction executed	Multi-context
global_replay_overhead	Average number of replays due to global memory cache misses	Multi-context
global_cache_replay_overhead	Average number of replays due to global memory cache misses for each instruction executed	Multi-context
local_replay_overhead	Average number of replays due to local memory accesses for each instruction executed	Multi-context
gld_efficiency	Ratio of requested global memory load throughput to required global memory load throughput expressed as percentage	Single-context
gst_efficiency	Ratio of requested global memory store throughput to required global memory store throughput expressed as percentage	Single-context
gld_transactions	Number of global memory load transactions expressed as percentage	Single-context
gst_transactions	Number of global memory store transactions	Single-context
gld_transactions_per_request	Average number of global memory load transactions performed for each global memory load	Single-context
gst_transactions_per_request	Average number of global memory store transactions performed for each global memory store	Single-context
gld_throughput	Global memory load throughput	Single-context
gst_throughput	Global memory store throughput	Single-context

Metric Name	Description	Scope
gld_requested_throughput	Requested global memory load throughput	Multi-context
gst_requested_throughput	Requested global memory store throughput	Multi-context
nc_l2_read_throughput	Memory read throughput for non coherent global read requests seen at L2 cache	Single-context
nc_l2_read_transactions	Memory read transactions seen at L2 cache for non coherent global read requests	Single-context
nc_cache_global_hit_rate	Hit rate in non coherent cache for global loads	Single-context
nc_gld_throughput	Non coherent global memory load throughput	Single-context
nc_gld_requested_throughput	Requested throughput for global memory loaded via non-coherent cache	Multi-context
nc_gld_efficiency	Ratio of requested non coherent global memory load throughput to required non coherent global memory load throughput expressed as percentage	Single-context
local_load_transactions	Number of local memory load transactions	Single-context
local_store_transactions	Number of local memory store transactions	Single-context
local_load_transactions_per_request	Average number of local memory load transactions performed for each local memory load	Single-context
local_store_transactions_per_request	Average number of local memory store transactions performed for each local memory store	Single-context
local_load_throughput	Local memory load throughput	Single-context
local_store_throughput	Local memory store throughput	Single-context
shared_load_transactions	Number of shared memory load transactions	Single-context
shared_store_transactions	Number of shared memory store transactions	Single-context
shared_load_transactions_per_request	Average number of shared memory load transactions performed for each shared memory load	Single-context
shared_store_transactions_per_request	Average number of shared memory store transactions performed for each shared memory store	Single-context

Metric Name	Description	Scope
shared_load_throughput	Shared memory load throughput	Single-context
shared_store_throughput	Shared memory store throughput	Single-context
shared_efficiency	Ratio of requested shared memory throughput to required shared memory throughput expressed as percentage	Single-context
dram_read_transactions	Device memory read transactions	Single-context
dram_write_transactions	Device memory write transactions	Single-context
dram_read_throughput	Device memory read throughput	Single-context
dram_write_throughput	Device memory write throughput	Single-context
sysmem_read_transactions	System memory read transactions. This is available for compute capability 3.0 and 3.5.	Single-context
sysmem_write_transactions	System memory write transactions. This is available for compute capability 3.0 and 3.5.	Single-context
sysmem_read_throughput	System memory read throughput. This is available for compute capability 3.0 and 3.5.	Single-context
sysmem_write_throughput	System memory write throughput. This is available for compute capability 3.0 and 3.5.	Single-context
l1_cache_global_hit_rate	Hit rate in L1 cache for global loads	Single-context
l1_cache_local_hit_rate	Hit rate in L1 cache for local loads and stores	Single-context
tex_cache_hit_rate	Texture cache hit rate	Single-context
tex_cache_transactions	Texture cache read transactions	Single-context
tex_cache_throughput	Texture cache throughput	Single-context
l2_read_transactions	Memory read transactions seen at L2 cache for all read requests	Single-context
l2_write_transactions	Memory write transactions seen at L2 cache for all write requests	Single-context
l2_read_throughput	Memory read throughput seen at L2 cache for all read requests	Single-context
l2_write_throughput	Memory write throughput seen at L2 cache for all write requests	Single-context

Metric Name	Description	Scope
l2_l1_read_hit_rate	Hit rate at L2 cache for all read requests from L1 cache. This is available for compute capability 3.0 and 3.5.	Single-context
l2_l1_read_throughput	Memory read throughput seen at L2 cache for read requests from L1 cache. This is available for compute capability 3.0 and 3.5.	Single-context
l2_texture_read_hit_rate	Hit rate at L2 cache for all read requests from texture cache	Single-context
l2_texture_read_throughput	Memory read throughput seen at L2 cache for read requests from the texture cache	Single-context
l2_atomic_throughput	Memory read throughput seen at L2 cache for atomic and reduction requests	Single-context
local_memory_overhead	Ratio of local memory traffic to total memory traffic between the L1 and L2 caches expressed as percentage. This is available for compute capability 3.0 and 3.5.	Single-context
l1_shared_utilization	The utilization level of the L1/shared memory relative to peak utilization on a scale of 0 to 10. This is available for compute capability 3.0 and 3.5.	Single-context
l2_utilization	The utilization level of the L2 cache relative to the peak utilization on a scale of 0 to 10	Single-context
tex_utilization	The utilization level of the texture cache relative to the peak utilization on a scale of 0 to 10	Single-context
dram_utilization	The utilization level of the device memory relative to the peak utilization on a scale of 0 to 10	Single-context
system_utilization	The utilization level of the system memory relative to the peak utilization on a scale of 0 to 10. This is available for compute capability 3.0 and 3.5.	Single-context
ldst_fu_utilization	The utilization level of the multiprocessor function units that execute global, local and shared memory instructions on a scale of 0 to 10	Multi-context

Metric Name	Description	Scope
cf_fu_utilization	The utilization level of the multiprocessor function units that execute control-flow instructions on a scale of 0 to 10	Multi-context
tex_fu_utilization	The utilization level of the multiprocessor function units that execute texture instructions on a scale of 0 to 10	Multi-context
alu_fu_utilization	The utilization level of the multiprocessor function units that execute integer and floating-point arithmetic instructions on a scale of 0 to 10	Multi-context
inst_fp_32	Number of single-precision floating-point instructions executed by non-predicated threads (arithmetic, compare, etc.)	Multi-context
inst_fp_64	Number of double-precision floating-point instructions executed by non-predicated threads (arithmetic, compare, etc.)	Multi-context
inst_integer	Number of integer instructions executed by non-predicated threads	Multi-context
inst_bit_convert	Number of bit-conversion instructions executed by non-predicated threads	Multi-context
inst_control	Number of control-flow instructions executed by non-predicated threads (jump, branch, etc.)	Multi-context
inst_compute_ld_st	Number of compute load/store instructions executed by non-predicated threads	Multi-context
inst_inter_thread_communication	Number of inter-thread communication instructions executed by non-predicated threads	Multi-context
inst_misc	Number of miscellaneous instructions executed by non-predicated threads	Multi-context
flops_sp	Single-precision floating point operations executed	Multi-context
flops_sp_add	Single-precision floating point add operations executed	Multi-context

Metric Name	Description	Scope
flops_sp_mul	Single-precision floating point multiply operations executed	Multi-context
flops_sp_fma	Single-precision floating point multiply-accumulate operations executed	Multi-context
flops_dp	Double-precision floating point operations executed	Multi-context
flops_dp_add	Double-precision floating point add operations executed	Multi-context
flops_dp_mul	Double-precision floating point multiply operations executed	Multi-context
flops_dp_fma	Double-precision floating point multiply-accumulate operations executed	Multi-context
flops_sp_special	Single-precision floating point special operations executed	Multi-context
stall_inst_fetch	Percentage of stalls occurring because the next assembly instruction has not yet been fetched	Multi-context
stall_exec_dependency	Percentage of stalls occurring because an input required by the instruction is not yet available	Multi-context
stall_data_request	Percentage of stalls occurring because a memory operation cannot be performed due to the required resources not being available or fully utilized, or because too many requests of a given type are outstanding	Multi-context
stall_sync	Percentage of stalls occurring because the warp is blocked at a __syncthreads() call	Multi-context
stall_texture	Percentage of stalls occurring because the texture sub-system is fully utilized or has too many outstanding requests	Multi-context
stall_other	Percentage of stalls occurring due to miscellaneous reasons	Multi-context
atomic_replay_overhead	Average number of replays due to atomic and reduction bank conflicts for each instruction executed	Multi-context

Metric Name	Description	Scope
atomic_transactions	Global memory atomic and reduction transactions	Multi-context
atomic_transactions_per_request	Average number of global memory atomic and reduction transactions performed for each atomic and reduction instruction	Multi-context

## 1.7. Samples

The CUPTI installation includes several samples that demonstrate the use of the CUPTI APIs. The samples are:

### **activity\_trace\_async**

This sample shows how to collect a trace of CPU and GPU activity using the new asynchronous activity buffer APIs.

### **callback\_event**

This sample shows how to use both the callback and event APIs to record the events that occur during the execution of a simple kernel. The sample shows the required ordering for synchronization, and for event group enabling, disabling and reading.

### **callback\_metric**

This sample shows how to use both the callback and metric APIs to record the metric's events during the execution of a simple kernel, and then use those events to calculate the metric value.

### **callback\_timestamp**

This sample shows how to use the callback API to record a trace of API start and stop times.

### **cupti\_query**

This sample shows how to query CUDA-enabled devices for their event domains, events, and metrics.

### **event\_sampling**

This sample shows how to use the event API to sample events using a separate host thread.

# Chapter 2.

## MODULES

Here is a list of all modules:

- ▶ CUPTI Version
- ▶ CUPTI Result Codes
- ▶ CUPTI Activity API
- ▶ CUPTI Callback API
- ▶ CUPTI Event API
- ▶ CUPTI Metric API

### 2.1. CUPTI Version

Function and macro to determine the CUPTI version.

#### **CuptiResult cuptiGetVersion (uint32\_t \*version)**

Get the CUPTI API version.

##### **Parameters**

###### **version**

Returns the version

##### **Returns**

- ▶ CUPTI\_SUCCESS  
on success
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if `version` is NULL

**Description**

Return the API version in `*version`.

**See also:**

[CUPTI\\_API\\_VERSION](#)

**#define CUPTI\_API\_VERSION 5**

The API version for this implementation of CUPTI.

The API version for this implementation of CUPTI. This define along with [cuptiGetVersion](#) can be used to dynamically detect if the version of CUPTI compiled against matches the version of the loaded CUPTI library.

v1 : CUDAToolsSDK 4.0 v2 : CUDAToolsSDK 4.1 v3 : CUDA Toolkit 5.0 v4 : CUDA Toolkit 5.5 v5 : CUDA Toolkit 6.0

## 2.2. CUPTI Result Codes

Error and result codes returned by CUPTI functions.

**enum CUptiResult**

CUPTI result codes.

Error and result codes returned by CUPTI functions.

**Values**

**CUPTI\_SUCCESS = 0**

No error.

**CUPTI\_ERROR\_INVALID\_PARAMETER = 1**

One or more of the parameters is invalid.

**CUPTI\_ERROR\_INVALID\_DEVICE = 2**

The device does not correspond to a valid CUDA device.

**CUPTI\_ERROR\_INVALID\_CONTEXT = 3**

The context is NULL or not valid.

**CUPTI\_ERROR\_INVALID\_EVENT\_DOMAIN\_ID = 4**

The event domain id is invalid.

**CUPTI\_ERROR\_INVALID\_EVENT\_ID = 5**

The event id is invalid.

**CUPTI\_ERROR\_INVALID\_EVENT\_NAME = 6**

The event name is invalid.

**CUPTI\_ERROR\_INVALID\_OPERATION = 7**

The current operation cannot be performed due to dependency on other factors.

**CUPTI\_ERROR\_OUT\_OF\_MEMORY = 8**

Unable to allocate enough memory to perform the requested operation.

**CUPTI\_ERROR\_HARDWARE = 9**

An error occurred on the performance monitoring hardware.

**CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT = 10**

The output buffer size is not sufficient to return all requested data.

**CUPTI\_ERROR\_API\_NOT\_IMPLEMENTED = 11**

API is not implemented.

**CUPTI\_ERROR\_MAX\_LIMIT\_REACHED = 12**

The maximum limit is reached.

**CUPTI\_ERROR\_NOT\_READY = 13**

The object is not yet ready to perform the requested operation.

**CUPTI\_ERROR\_NOT\_COMPATIBLE = 14**

The current operation is not compatible with the current state of the object

**CUPTI\_ERROR\_NOT\_INITIALIZED = 15**

CUPTI is unable to initialize its connection to the CUDA driver.

**CUPTI\_ERROR\_INVALID\_METRIC\_ID = 16**

The metric id is invalid.

**CUPTI\_ERROR\_INVALID\_METRIC\_NAME = 17**

The metric name is invalid.

**CUPTI\_ERROR\_QUEUE\_EMPTY = 18**

The queue is empty.

**CUPTI\_ERROR\_INVALID\_HANDLE = 19**

Invalid handle (internal?).

**CUPTI\_ERROR\_INVALID\_STREAM = 20**

Invalid stream.

**CUPTI\_ERROR\_INVALID\_KIND = 21**

Invalid kind.

**CUPTI\_ERROR\_INVALID\_EVENT\_VALUE = 22**

Invalid event value.

**CUPTI\_ERROR\_DISABLED = 23**

CUPTI is disabled due to conflicts with other enabled profilers

**CUPTI\_ERROR\_INVALID\_MODULE = 24**

Invalid module.

**CUPTI\_ERROR\_INVALID\_METRIC\_VALUE = 25**

Invalid metric value.

**CUPTI\_ERROR\_HARDWARE\_BUSY = 26**

The performance monitoring hardware is in use by other client.

**CUPTI\_ERROR\_NOT\_SUPPORTED = 27**

The attempted operation is not supported on the current system or device.

**CUPTI\_ERROR\_UNKNOWN = 999**

An unknown internal error has occurred.

**CUPTI\_ERROR\_FORCE\_INT = 0x7fffffff**

## CUptiResult cuptiGetResultString (CUptiResult result, const char \*\*str)

Get the descriptive string for a CUptiResult.

### Parameters

#### result

The result to get the string for

#### str

Returns the string

### Returns

- ▶ CUPTI\_SUCCESS  
on success
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if `str` is NULL or `result` is not a valid CUptiResult

### Description

Return the descriptive string for a CUptiResult in `*str`.



**Thread-safety:** this function is thread safe.

## 2.3. CUPTI Activity API

Functions, types, and enums that implement the CUPTI Activity API.

## **struct CUpti\_Activity**

The base activity record.

## **struct CUpti\_ActivityAPI**

The activity record for a driver or runtime API invocation.

## **struct CUpti\_ActivityBranch**

The activity record for source level result branch.

## **struct CUpti\_ActivityCdpKernel**

The activity record for CDP (CUDA Dynamic Parallelism) kernel.

## **struct CUpti\_ActivityContext**

The activity record for a context.

## **struct CUpti\_ActivityDevice**

The activity record for a device.

## **struct CUpti\_ActivityDeviceAttribute**

The activity record for a device attribute.

## **struct CUpti\_ActivityEnvironment**

The activity record for CUPTI environmental data.

## **struct CUpti\_ActivityEvent**

The activity record for a CUPTI event.

## **struct CUpti\_ActivityEventInstance**

The activity record for a CUPTI event with instance information.

## **struct CUpti\_ActivityFunction**

The activity record for global/device functions.

## **struct CUpti\_ActivityGlobalAccess**

The activity record for source-level global access.

## **struct CUpti\_ActivityInstructionExecution**

The activity record for source-level sass/source line-by-line correlation.

## struct CUpti\_ActivityKernel

The activity record for kernel. (deprecated).

## struct CUpti\_ActivityKernel2

The activity record for a kernel (CUDA 5.5 onwards).

## struct CUpti\_ActivityMarker

The activity record providing a marker which is an instantaneous point in time.

## struct CUpti\_ActivityMarkerData

The activity record providing detailed information for a marker.

## struct CUpti\_ActivityMemcpy

The activity record for memory copies.

## struct CUpti\_ActivityMemcpy2

The activity record for peer-to-peer memory copies.

## struct CUpti\_ActivityMemset

The activity record for memset.

## struct CUpti\_ActivityMetric

The activity record for a CUPTI metric.

## struct CUpti\_ActivityMetricInstance

The activity record for a CUPTI metric with instance information. This activity record represents a CUPTI metric value for a specific metric domain instance (CUPTI\_ACTIVITY\_KIND\_METRIC\_INSTANCE). This activity record kind is not produced by the activity API but is included for completeness and ease-of-use. Profile frameworks built on top of CUPTI that collect metric data may choose to use this type to store the collected metric data. This activity record should be used when metric domain instance information needs to be associated with the metric.

## struct CUpti\_ActivityModule

The activity record for a CUDA module.

## struct CUpti\_ActivityName

The activity record providing a name.

## union CUpti\_ActivityObjectKindId

Identifiers for object kinds as specified by CUpti\_ActivityObjectKind.

## struct CUpti\_ActivityOverhead

The activity record for CUPTI and driver overheads.

## struct CUpti\_ActivityPreemption

The activity record for a preemption of a CDP kernel.

## struct CUpti\_ActivitySharedAccess

The activity record for source-level shared access.

## struct CUpti\_ActivitySourceLocator

The activity record for source locator.

## struct CUpti\_ActivityUnifiedMemoryCounter

The activity record for Unified Memory counters.

## struct CUpti\_ActivityUnifiedMemoryCounterConfig

Unified Memory counters configuration structure.

## enum CUpti\_ActivityAttribute

Activity attributes.

These attributes are used to control the behavior of the activity API.

### Values

#### CUPTI\_ACTIVITY\_ATTR\_DEVICE\_BUFFER\_SIZE = 0

The device memory reserved for storing profiling data for non-CDP operations for each stream. The value is a `size_t`. Larger buffers require less flush operations but consume more device memory. Small buffers might increase the risk of missing timestamps for concurrent kernel records in the asynchronous buffer handling mode if too many kernels are launched/replayed between context synchronizations. This value only applies to new allocations. Set this value before initializing CUDA or

before creating a stream to ensure it is considered for the following allocations. Note: The actual amount of device memory per stream reserved by CUPTI might be larger.

**CUPTI\_ACTIVITY\_ATTR\_DEVICE\_BUFFER\_SIZE\_CDP = 1**

The device memory reserved for storing profiling data for CDP operations for each stream. The value is a `size_t`. Larger buffers require less flush operations but consume more device memory. This value only applies to new allocations. Set this value before initializing CUDA or before creating a stream to ensure it is considered for the following allocations. Note: The actual amount of device memory per stream reserved by CUPTI might be larger.

**CUPTI\_ACTIVITY\_ATTR\_DEVICE\_BUFFER\_POOL\_LIMIT = 2**

The maximum number of device memory buffers stored for reuse by CUPTI. The value is a `size_t`. Buffers can be reused by streams of the same context. Increasing this value reduces the profiling overhead when the application creates and destroys many streams. Setting this value will not modify the number of memory buffers currently stored. Set this value before initializing CUDA to ensure the limit is not exceeded.

## enum CUpti\_ActivityComputeApiKind

The kind of a compute API.

### Values

**CUPTI\_ACTIVITY\_COMPUTE\_API\_UNKNOWN = 0**

The compute API is not known.

**CUPTI\_ACTIVITY\_COMPUTE\_API\_CUDA = 1**

The compute APIs are for CUDA.

**CUPTI\_ACTIVITY\_COMPUTE\_API\_CUDA\_MPS = 2**

The compute APIs are for CUDA running in MPS (Multi-Process Service) environment.

**CUPTI\_ACTIVITY\_COMPUTE\_API\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityEnvironmentKind

The kind of environment data. Used to indicate what type of data is being reported by an environment activity record.

### Values

**CUPTI\_ACTIVITY\_ENVIRONMENT\_UNKNOWN = 0**

Unknown data.

**CUPTI\_ACTIVITY\_ENVIRONMENT\_SPEED = 1**

The environment data is related to speed.

**CUPTI\_ACTIVITY\_ENVIRONMENT\_TEMPERATURE = 2**

The environment data is related to temperature.

**CUPTI\_ACTIVITY\_ENVIRONMENT\_POWER = 3**

The environment data is related to power.

**CUPTI\_ACTIVITY\_ENVIRONMENT\_COOLING = 4**

The environment data is related to cooling.

**CUPTI\_ACTIVITY\_ENVIRONMENT\_COUNT**

**CUPTI\_ACTIVITY\_ENVIRONMENT\_KIND\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityFlag

Flags associated with activity records.

Activity record flags. Flags can be combined by bitwise OR to associated multiple flags with an activity record. Each flag is specific to a certain activity kind, as noted below.

### Values

**CUPTI\_ACTIVITY\_FLAG\_NONE = 0**

Indicates the activity record has no flags.

**CUPTI\_ACTIVITY\_FLAG\_DEVICE\_CONCURRENT\_KERNELS = 1<<0**

Indicates the activity represents a device that supports concurrent kernel execution.

Valid for CUPTI\_ACTIVITY\_KIND\_DEVICE.

**CUPTI\_ACTIVITY\_FLAG\_DEVICE\_ATTRIBUTE\_CUDEVICE = 1<<0**

Indicates if the activity represents a CUdevice\_attribute

value or a CUpti\_DeviceAttribute value. Valid for

CUPTI\_ACTIVITY\_KIND\_DEVICE\_ATTRIBUTE.

**CUPTI\_ACTIVITY\_FLAG\_MEMCPY\_ASYNC = 1<<0**

Indicates the activity represents an asynchronous memcpy operation. Valid for

CUPTI\_ACTIVITY\_KIND\_MEMCPY.

**CUPTI\_ACTIVITY\_FLAG\_MARKER\_INSTANTANEOUS = 1<<0**

Indicates the activity represents an instantaneous marker. Valid for

CUPTI\_ACTIVITY\_KIND\_MARKER.

**CUPTI\_ACTIVITY\_FLAG\_MARKER\_START = 1<<1**

Indicates the activity represents a region start marker. Valid for

CUPTI\_ACTIVITY\_KIND\_MARKER.

**CUPTI\_ACTIVITY\_FLAG\_MARKER\_END = 1<<2**

Indicates the activity represents a region end marker. Valid for

CUPTI\_ACTIVITY\_KIND\_MARKER.

**CUPTI\_ACTIVITY\_FLAG\_MARKER\_COLOR\_NONE = 1<<0**

Indicates the activity represents a marker that does not specify a color. Valid for

CUPTI\_ACTIVITY\_KIND\_MARKER\_DATA.

**CUPTI\_ACTIVITY\_FLAG\_MARKER\_COLOR\_ARGB = 1<<1**

Indicates the activity represents a marker that specifies a color in alpha-red-green-

blue format. Valid for CUPTI\_ACTIVITY\_KIND\_MARKER\_DATA.

**CUPTI\_ACTIVITY\_FLAG\_GLOBAL\_ACCESS\_KIND\_SIZE\_MASK = 0xFF<<0**

The number of bytes requested by each thread Valid for [CUpti\\_ActivityGlobalAccess](#).

**CUPTI\_ACTIVITY\_FLAG\_GLOBAL\_ACCESS\_KIND\_LOAD = 1<<8**

If bit in this flag is set, the access was load, else it is a store access. Valid for [CUpti\\_ActivityGlobalAccess](#).

**CUPTI\_ACTIVITY\_FLAG\_GLOBAL\_ACCESS\_KIND\_CACHED = 1<<9**

If this bit in flag is set, the load access was cached else it is uncached. Valid for [CUpti\\_ActivityGlobalAccess](#).

**CUPTI\_ACTIVITY\_FLAG\_METRIC\_OVERFLOWED = 1<<0**

If this bit in flag is set, the metric value overflowed. Valid for [CUpti\\_ActivityMetric](#) and [CUpti\\_ActivityMetricInstance](#).

**CUPTI\_ACTIVITY\_FLAG\_METRIC\_VALUE\_INVALID = 1<<1**

If this bit in flag is set, the metric value couldn't be calculated. This occurs when a value(s) required to calculate the metric is missing. Valid for [CUpti\\_ActivityMetric](#) and [CUpti\\_ActivityMetricInstance](#).

**CUPTI\_ACTIVITY\_FLAG\_INSTRUCTION\_VALUE\_INVALID = 1<<0**

If this bit in flag is set, the source level metric value couldn't be calculated. This occurs when a value(s) required to calculate the source level metric cannot be evaluated. Valid for [CUpti\\_ActivityInstructionExecution](#).

**CUPTI\_ACTIVITY\_FLAG\_FLUSH\_FORCED = 1<<0**

When calling `cuptiActivityFlush` or `cuptiActivityFlushAll`, this flag can be set to force CUPTI to flush all records in the buffer, whether finished or not

**CUPTI\_ACTIVITY\_FLAG\_SHARED\_ACCESS\_KIND\_SIZE\_MASK = 0xFF<<0**

The number of bytes requested by each thread Valid for [CUpti\\_ActivitySharedAccess](#).

**CUPTI\_ACTIVITY\_FLAG\_SHARED\_ACCESS\_KIND\_LOAD = 1<<8**

If bit in this flag is set, the access was load, else it is a store access. Valid for [CUpti\\_ActivitySharedAccess](#).

**CUPTI\_ACTIVITY\_FLAG\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityKind

The kinds of activity records.

Each activity record kind represents information about a GPU or an activity occurring on a CPU or GPU. Each kind is associated with a activity record structure that holds the information associated with the kind.

**See also:**

[CUpti\\_Activity](#)

[CUpti\\_ActivityAPI](#)

[CUpti\\_ActivityContext](#)

[CUpti\\_ActivityDevice](#)

[CUpti\\_ActivityDeviceAttribute](#)

[CUpti\\_ActivityEvent](#)

[CUpti\\_ActivityEventInstance](#)

CUpti\_ActivityKernel  
 CUpti\_ActivityKernel2  
 CUpti\_ActivityCdpKernel  
 CUpti\_ActivityPreemption  
 CUpti\_ActivityMemcpy  
 CUpti\_ActivityMemcpy2  
 CUpti\_ActivityMemset  
 CUpti\_ActivityMetric  
 CUpti\_ActivityMetricInstance  
 CUpti\_ActivityName  
 CUpti\_ActivityMarker  
 CUpti\_ActivityMarkerData  
 CUpti\_ActivitySourceLocator  
 CUpti\_ActivityGlobalAccess  
 CUpti\_ActivityBranch  
 CUpti\_ActivityOverhead  
 CUpti\_ActivityEnvironment  
 CUpti\_ActivityInstructionExecution  
 CUpti\_ActivityUnifiedMemoryCounter  
 CUpti\_ActivityFunction  
 CUpti\_ActivityModule  
 CUpti\_ActivityDeviceAttribute  
 CUpti\_ActivitySharedAccess

## Values

**CUPTI\_ACTIVITY\_KIND\_INVALID = 0**

The activity record is invalid.

**CUPTI\_ACTIVITY\_KIND\_MEMCPY = 1**

A host->host, host->device, or device->device memory copy. The corresponding activity record structure is [CUpti\\_ActivityMemcpy](#).

**CUPTI\_ACTIVITY\_KIND\_MEMSET = 2**

A memory set executing on the GPU. The corresponding activity record structure is [CUpti\\_ActivityMemset](#).

**CUPTI\_ACTIVITY\_KIND\_KERNEL = 3**

A kernel executing on the GPU. The corresponding activity record structure is [CUpti\\_ActivityKernel2](#).

**CUPTI\_ACTIVITY\_KIND\_DRIVER = 4**

A CUDA driver API function execution. The corresponding activity record structure is [CUpti\\_ActivityAPI](#).

**CUPTI\_ACTIVITY\_KIND\_RUNTIME = 5**

A CUDA runtime API function execution. The corresponding activity record structure is [CUpti\\_ActivityAPI](#).

**CUPTI\_ACTIVITY\_KIND\_EVENT = 6**

An event value. The corresponding activity record structure is [CUpti\\_ActivityEvent](#).

**CUPTI\_ACTIVITY\_KIND\_METRIC = 7**

A metric value. The corresponding activity record structure is [CUpti\\_ActivityMetric](#).

**CUPTI\_ACTIVITY\_KIND\_DEVICE = 8**

Information about a device. The corresponding activity record structure is [CUpti\\_ActivityDevice](#).

**CUPTI\_ACTIVITY\_KIND\_CONTEXT = 9**

Information about a context. The corresponding activity record structure is [CUpti\\_ActivityContext](#).

**CUPTI\_ACTIVITY\_KIND\_CONCURRENT\_KERNEL = 10**

A (potentially concurrent) kernel executing on the GPU. The corresponding activity record structure is [CUpti\\_ActivityKernel2](#).

**CUPTI\_ACTIVITY\_KIND\_NAME = 11**

Thread, device, context, etc. name. The corresponding activity record structure is [CUpti\\_ActivityName](#).

**CUPTI\_ACTIVITY\_KIND\_MARKER = 12**

Instantaneous, start, or end marker.

**CUPTI\_ACTIVITY\_KIND\_MARKER\_DATA = 13**

Extended, optional, data about a marker.

**CUPTI\_ACTIVITY\_KIND\_SOURCE\_LOCATOR = 14**

Source information about source level result. The corresponding activity record structure is [CUpti\\_ActivitySourceLocator](#).

**CUPTI\_ACTIVITY\_KIND\_GLOBAL\_ACCESS = 15**

Results for source-level global access. The corresponding activity record structure is [CUpti\\_ActivityGlobalAccess](#).

**CUPTI\_ACTIVITY\_KIND\_BRANCH = 16**

Results for source-level branch. The corresponding activity record structure is [CUpti\\_ActivityBranch](#).

**CUPTI\_ACTIVITY\_KIND\_OVERHEAD = 17**

Overhead activity records. The corresponding activity record structure is [CUpti\\_ActivityOverhead](#).

**CUPTI\_ACTIVITY\_KIND\_CDP\_KERNEL = 18**

A CDP (CUDA Dynamic Parallel) kernel executing on the GPU. The corresponding activity record structure is [CUpti\\_ActivityCdpKernel](#). This activity can not be directly enabled or disabled. It is enabled and disabled through concurrent kernel activity [CUPTI\\_ACTIVITY\\_KIND\\_CONCURRENT\\_KERNEL](#)

#### **CUPTI\_ACTIVITY\_KIND\_PREEMPTION = 19**

Preemption activity record indicating a preemption of a CDP (CUDA Dynamic Parallel) kernel executing on the GPU. The corresponding activity record structure is [CUpti\\_ActivityPreemption](#).

#### **CUPTI\_ACTIVITY\_KIND\_ENVIRONMENT = 20**

Environment activity records indicating power, clock, thermal, etc. levels of the GPU. The corresponding activity record structure is [CUpti\\_ActivityEnvironment](#).

#### **CUPTI\_ACTIVITY\_KIND\_EVENT\_INSTANCE = 21**

An event value associated with a specific event domain instance. The corresponding activity record structure is [CUpti\\_ActivityEventInstance](#).

#### **CUPTI\_ACTIVITY\_KIND\_MEMCPY2 = 22**

A peer to peer memory copy. The corresponding activity record structure is [CUpti\\_ActivityMemcpy2](#).

#### **CUPTI\_ACTIVITY\_KIND\_METRIC\_INSTANCE = 23**

A metric value associated with a specific metric domain instance. The corresponding activity record structure is [CUpti\\_ActivityMetricInstance](#).

#### **CUPTI\_ACTIVITY\_KIND\_INSTRUCTION\_EXECUTION = 24**

SASS/Source line-by-line correlation record. The corresponding activity record structure is [CUpti\\_ActivityInstructionExecution](#).

#### **CUPTI\_ACTIVITY\_KIND\_UNIFIED\_MEMORY\_COUNTER = 25**

Unified Memory counter record. The corresponding activity record structure is [CUpti\\_ActivityUnifiedMemoryCounter](#).

#### **CUPTI\_ACTIVITY\_KIND\_FUNCTION = 26**

Device global/function record. The corresponding activity record structure is [CUpti\\_ActivityFunction](#).

#### **CUPTI\_ACTIVITY\_KIND\_MODULE = 27**

CUDA Module record. The corresponding activity record structure is [CUpti\\_ActivityModule](#).

#### **CUPTI\_ACTIVITY\_KIND\_DEVICE\_ATTRIBUTE = 28**

A device attribute value. The corresponding activity record structure is [CUpti\\_ActivityDeviceAttribute](#).

#### **CUPTI\_ACTIVITY\_KIND\_SHARED\_ACCESS = 29**

Results for source-level shared access. The corresponding activity record structure is [CUpti\\_ActivitySharedAccess](#).

#### **CUPTI\_ACTIVITY\_KIND\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityMemcpyKind

The kind of a memory copy, indicating the source and destination targets of the copy.

Each kind represents the source and destination targets of a memory copy. Targets are host, device, and array.

### Values

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_UNKNOWN = 0**

The memory copy kind is not known.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_HTOD = 1**

A host to device memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_DTOH = 2**

A device to host memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_HTOA = 3**

A host to device array memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_ATOH = 4**

A device array to host memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_ATOA = 5**

A device array to device array memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_ATOD = 6**

A device array to device memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_DTOA = 7**

A device to device array memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_DTOD = 8**

A device to device memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_HTOH = 9**

A host to host memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_PTOP = 10**

A peer to peer memory copy.

**CUPTI\_ACTIVITY\_MEMCPY\_KIND\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityMemoryKind

The kinds of memory accessed by a memory copy.

Each kind represents the type of the source or destination memory accessed by a memory copy.

### Values

**CUPTI\_ACTIVITY\_MEMORY\_KIND\_UNKNOWN = 0**

The source or destination memory kind is unknown.

**CUPTI\_ACTIVITY\_MEMORY\_KIND\_PAGEABLE = 1**

The source or destination memory is pageable.

**CUPTI\_ACTIVITY\_MEMORY\_KIND\_PINNED = 2**

The source or destination memory is pinned.

**CUPTI\_ACTIVITY\_MEMORY\_KIND\_DEVICE = 3**

The source or destination memory is on the device.

**CUPTI\_ACTIVITY\_MEMORY\_KIND\_ARRAY = 4**

The source or destination memory is an array.

**CUPTI\_ACTIVITY\_MEMORY\_KIND\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityObjectKind

The kinds of activity objects.

See also:

[CUpti\\_ActivityObjectKindId](#)

### Values

**CUPTI\_ACTIVITY\_OBJECT\_UNKNOWN = 0**

The object kind is not known.

**CUPTI\_ACTIVITY\_OBJECT\_PROCESS = 1**

A process.

**CUPTI\_ACTIVITY\_OBJECT\_THREAD = 2**

A thread.

**CUPTI\_ACTIVITY\_OBJECT\_DEVICE = 3**

A device.

**CUPTI\_ACTIVITY\_OBJECT\_CONTEXT = 4**

A context.

**CUPTI\_ACTIVITY\_OBJECT\_STREAM = 5**

A stream.

**CUPTI\_ACTIVITY\_OBJECT\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityOverheadKind

The kinds of activity overhead.

### Values

**CUPTI\_ACTIVITY\_OVERHEAD\_UNKNOWN = 0**

The overhead kind is not known.

**CUPTI\_ACTIVITY\_OVERHEAD\_DRIVER\_COMPILER = 1**

Compiler(JIT) overhead.

**CUPTI\_ACTIVITY\_OVERHEAD\_CUPTI\_BUFFER\_FLUSH = 1<<16**

Activity buffer flush overhead.

**CUPTI\_ACTIVITY\_OVERHEAD\_CUPTI\_INSTRUMENTATION = 2<<16**

CUPTI instrumentation overhead.

**CUPTI\_ACTIVITY\_OVERHEAD\_CUPTI\_RESOURCE = 3<<16**

CUPTI resource creation and destruction overhead.

**CUPTI\_ACTIVITY\_OVERHEAD\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityPreemptionKind

The kind of a preemption activity.

### Values

**CUPTI\_ACTIVITY\_PREEMPTION\_KIND\_UNKNOWN = 0**

The preemption kind is not known.

**CUPTI\_ACTIVITY\_PREEMPTION\_KIND\_SAVE = 1**

Preemption to save CDP block.

**CUPTI\_ACTIVITY\_PREEMPTION\_KIND\_RESTORE = 2**

Preemption to restore CDP block.

**CUPTI\_ACTIVITY\_PREEMPTION\_KIND\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityUnifiedMemoryCounterKind

Kind of the Unified Memory counter.

Many activities are associated with Unified Memory mechanism; among them are transfer from host to device, device to host, page fault at host side.

### Values

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_KIND\_UNKNOWN = 0**

The unified memory counter kind is not known.

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_KIND\_BYTES\_TRANSFER\_HTOD = 1**

Number of bytes transferred from host to device

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_KIND\_BYTES\_TRANSFER\_DTOH = 2**

Number of bytes transferred from device to host

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_KIND\_CPU\_PAGE\_FAULT\_COUNT = 3**

Number of CPU page faults

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_KIND\_COUNT**

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_KIND\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ActivityUnifiedMemoryCounterScope

Scope of the unified memory counter.

**Values**

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_SCOPE\_UNKNOWN = 0**

The unified memory counter scope is not known.

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_SCOPE\_PROCESS\_SINGLE\_DEVICE = 1**

Collect unified memory counter for single process on one device

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_SCOPE\_PROCESS\_ALL\_DEVICES = 2**

Collect unified memory counter for single process across all devices

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_SCOPE\_COUNT**

**CUPTI\_ACTIVITY\_UNIFIED\_MEMORY\_COUNTER\_SCOPE\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_EnvironmentClocksThrottleReason

Reasons for clock throttling.

The possible reasons that a clock can be throttled. There can be more than one reason that a clock is being throttled so these types can be combined by bitwise OR. These are used in the clocksThrottleReason field in the Environment Activity Record.

**Values**

**CUPTI\_CLOCKS\_THROTTLE\_REASON\_GPU\_IDLE = 0x00000001**

Nothing is running on the GPU and the clocks are dropping to idle state.

**CUPTI\_CLOCKS\_THROTTLE\_REASON\_USER\_DEFINED\_CLOCKS = 0x00000002**

The GPU clocks are limited by a user specified limit.

**CUPTI\_CLOCKS\_THROTTLE\_REASON\_SW\_POWER\_CAP = 0x00000004**

A software power scaling algorithm is reducing the clocks below requested clocks.

**CUPTI\_CLOCKS\_THROTTLE\_REASON\_HW\_SLOWDOWN = 0x00000008**

Hardware slowdown to reduce the clock by a factor of two or more is engaged. This is an indicator of one of the following: 1) Temperature is too high, 2) External power brake assertion is being triggered (e.g. by the system power supply), 3) Change in power state.

**CUPTI\_CLOCKS\_THROTTLE\_REASON\_UNKNOWN = 0x80000000**

Some unspecified factor is reducing the clocks.

**CUPTI\_CLOCKS\_THROTTLE\_REASON\_UNSUPPORTED = 0x40000000**

Throttle reason is not supported for this GPU.

**CUPTI\_CLOCKS\_THROTTLE\_REASON\_NONE = 0x00000000**

No clock throttling.

**CUPTI\_CLOCKS\_THROTTLE\_REASON\_FORCE\_INT = 0x7fffffff**

```
typedef (*CUpti_BuffersCallbackCompleteFunc)
(CUcontext context, uint32_t streamId, uint8_t* buffer,
size_t size, size_t validSize)
```

Function type for callback used by CUPTI to return a buffer of activity records.

This callback function returns to the CUPTI client a buffer containing activity records. The buffer contains `validSize` bytes of activity records which should be read using `cuptiActivityGetNextRecord`. The number of dropped records can be read using `cuptiActivityGetNumDroppedRecords`. After this call CUPTI relinquished ownership of the buffer and will not use it anymore. The client may return the buffer to CUPTI using the `CUpti_BuffersCallbackRequestFunc` callback.

```
typedef (*CUpti_BuffersCallbackRequestFunc) (uint8_t*
*buffer, size_t* size, size_t* maxNumRecords)
```

Function type for callback used by CUPTI to request an empty buffer for storing activity records.

This callback function signals the CUPTI client that an activity buffer is needed by CUPTI. The activity buffer is used by CUPTI to store activity records. The callback function can decline the request by setting `*buffer` to NULL. In this case CUPTI may drop activity records.

## CUptiResult

```
cuptiActivityConfigureUnifiedMemoryCounter
(CUpti_ActivityUnifiedMemoryCounterConfig *config,
uint32_t count)
```

Set Unified Memory Counter configuration.

### Parameters

#### **config**

A pointer to `CUpti_ActivityUnifiedMemoryCounterConfig` structures containing Unified Memory counter configuration.

#### **count**

Number of Unified Memory counter configuration structures

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED

- ▶ `CUPTI_ERROR_INVALID_PARAMETER`  
if `config` is `NULL` or any parameter in the `config` structures is not a valid value
- ▶ `CUPTI_ERROR_NOT_SUPPORTED`  
Indicates that the system/device does not support the unified memory counters

## CUptiResult cuptiActivityDisable (CUpti\_ActivityKind kind)

Disable collection of a specific kind of activity record.

### Parameters

#### kind

The kind of activity record to stop collecting

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_INVALID_KIND`  
if the activity kind is not supported

### Description

Disable collection of a specific kind of activity record. Multiple kinds can be disabled by calling this function multiple times. By default all activity kinds are disabled for collection.

## CUptiResult cuptiActivityDisableContext (CUcontext context, CUpti\_ActivityKind kind)

Disable collection of a specific kind of activity record for a context.

### Parameters

#### context

The context for which activity is to be disabled

#### kind

The kind of activity record to stop collecting

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`

- ▶ `CUPTI_ERROR_INVALID_KIND`  
if the activity kind is not supported

### Description

Disable collection of a specific kind of activity record for a context. This setting done by this API will supersede the global settings for activity records. Multiple kinds can be enabled by calling this function multiple times.

## CUptiResult cuptiActivityEnable (CUpti\_ActivityKind kind)

Enable collection of a specific kind of activity record.

### Parameters

#### kind

The kind of activity record to collect

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_NOT_COMPATIBLE`  
if the activity kind cannot be enabled
- ▶ `CUPTI_ERROR_INVALID_KIND`  
if the activity kind is not supported

### Description

Enable collection of a specific kind of activity record. Multiple kinds can be enabled by calling this function multiple times. By default all activity kinds are disabled for collection.

## CUptiResult cuptiActivityEnableContext (CUcontext context, CUpti\_ActivityKind kind)

Enable collection of a specific kind of activity record for a context.

### Parameters

#### context

The context for which activity is to be enabled

**kind**

The kind of activity record to collect

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_NOT\_COMPATIBLE  
if the activity kind cannot be enabled
- ▶ CUPTI\_ERROR\_INVALID\_KIND  
if the activity kind is not supported

**Description**

Enable collection of a specific kind of activity record for a context. This setting done by this API will supersede the global settings for activity records enabled by [cuptiActivityEnable](#). Multiple kinds can be enabled by calling this function multiple times.

## CUptiResult cuptiActivityFlush (CUcontext context, uint32\_t streamId, uint32\_t flag)

Wait for all activity records are delivered via the completion callback.

**Parameters****context**

A valid CUcontext or NULL.

**streamId**

The stream ID.

**flag**

The flag can be set to indicate a forced flush. See CUpti\_ActivityFlag

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_CUPTI\_ERROR\_INVALID\_OPERATION  
if not preceeded by a successful call to cuptiActivityRegisterCallbacks
- ▶ CUPTI\_ERROR\_UNKNOWN  
an internal error occurred

## Description

This function does not return until all activity records associated with the specified context/stream are returned to the CUPTI client using the callback registered in `cuptiActivityRegisterCallbacks`. To ensure that all activity records are complete, the requested stream(s), if any, are synchronized.

If `context` is `NULL`, the global activity records (i.e. those not associated with a particular stream) are flushed (in this case no streams are synchronized). If `context` is a valid `CUcontext` and `streamId` is 0, the buffers of all streams of this context are flushed. Otherwise, the buffers of the specified stream in this context is flushed.

Before calling this function, the buffer handling callback api must be activated by calling `cuptiActivityRegisterCallbacks`.

**\*\*DEPRECATED\*\*** This method is deprecated and will be removed in a future release. `CONTEXT` and `STREAMID` will be ignored. Use `cuptiActivityFlushAll` to flush all data.

## CUptiResult cuptiActivityFlushAll (uint32\_t flag)

Wait for all activity records are delivered via the completion callback.

### Parameters

#### flag

The flag can be set to indicate a forced flush. See `CUpti_ActivityFlag`

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_INVALID_OPERATION`  
if not preceeded by a successful call to `cuptiActivityRegisterCallbacks`
- ▶ `CUPTI_ERROR_UNKNOWN`  
an internal error occurred

## Description

This function does not return until all activity records associated with all contexts/streams (and the global buffers not associated with any stream) are returned to the CUPTI client using the callback registered in `cuptiActivityRegisterCallbacks`. To ensure that all activity records are complete, the requested stream(s), if any, are synchronized.

Before calling this function, the buffer handling callback api must be activated by calling `cuptiActivityRegisterCallbacks`.

## CUptiResult cuptiActivityGetAttribute (CUpti\_ActivityAttribute attr, size\_t \*valueSize, void \*value)

Read an activity API attribute.

### Parameters

#### attr

The attribute to read

#### valueSize

Size of buffer pointed by the value, and returns the number of bytes written to value

#### value

Returns the value of the attribute

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if valueSize or value is NULL, or if attr is not an activity attribute
- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT
  - Indicates that the value buffer is too small to hold the attribute value.

### Description

Read an activity API attribute and return it in \*value.

## CUptiResult cuptiActivityGetNextRecord (uint8\_t \*buffer, size\_t validBufferSizeBytes, CUpti\_Activity \*\*record)

Iterate over the activity records in a buffer.

### Parameters

#### buffer

The buffer containing activity records

#### validBufferSizeBytes

The number of valid bytes in the buffer.

**record**

Inputs the previous record returned by `cuptiActivityGetNextRecord` and returns the next activity record from the buffer. If input value is `NULL`, returns the first activity record in the buffer. Records of kind `CUPTI_ACTIVITY_KIND_CONCURRENT_KERNEL` may contain invalid (0) timestamps, indicating that no timing information could be collected for lack of device memory.

**Returns**

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_MAX_LIMIT_REACHED`  
if no more records in the buffer
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`  
if `buffer` is `NULL`.

**Description**

This is a helper function to iterate over the activity records in a buffer. A buffer of activity records is typically obtained by using the `cuptiActivityDequeueBuffer()` function or by receiving a `CUpti_BuffersCallbackCompleteFunc` callback.

An example of typical usage:

```
↑ CUpti_Activity *record = NULL;
CUptiResult status = CUPTI_SUCCESS;
do {
    status = cuptiActivityGetNextRecord(buffer, validSize, &record);
    if(status == CUPTI_SUCCESS) {
        // Use record here...
    }
    else if (status == CUPTI_ERROR_MAX_LIMIT_REACHED)
        break;
    else {
        goto Error;
    }
} while (1);
```

## CUptiResult cuptiActivityGetNumDroppedRecords (CUcontext context, uint32\_t streamId, size\_t \*dropped)

Get the number of activity records that were dropped of insufficient buffer space.

**Parameters****context**

The context, or `NULL` to get dropped count from global queue

**streamId**

The stream ID

**dropped**

The number of records that were dropped since the last call to this function.

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if dropped is NULL

**Description**

Get the number of records that were dropped because of insufficient buffer space. The dropped count includes records that could not be recorded because CUPTI did not have activity buffer space available for the record (because the CUpti\_BuffersCallbackRequestFunc callback did not return an empty buffer of sufficient size) and also CDP records that could not be record because the device-size buffer was full (size is controlled by the CUPTI\_ACTIVITY\_ATTR\_DEVICE\_BUFFER\_SIZE\_CDP attribute). The dropped count maintained for the queue is reset to zero when this function is called.

## CUptiResult cuptiActivityRegisterCallbacks (CUpti\_BuffersCallbackRequestFunc funcBufferRequested, CUpti\_BuffersCallbackCompleteFunc funcBufferCompleted)

Registers callback functions with CUPTI for activity buffer handling.

**Parameters****funcBufferRequested**

callback which is invoked when an empty buffer is requested by CUPTI

**funcBufferCompleted**

callback which is invoked when a buffer containing activity records is available from CUPTI

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if either `funcBufferRequested` or `funcBufferCompleted` is NULL

### Description

This function registers two callback functions to be used in asynchronous buffer handling. If registered, activity record buffers are handled using asynchronous requested/completed callbacks from CUPTI.

Registering these callbacks prevents the client from using CUPTI's blocking enqueue/dequeue functions.

## CuptiResult cuptiActivitySetAttribute (Cupti\_ActivityAttribute attr, size\_t \*valueSize, void \*value)

Write an activity API attribute.

### Parameters

#### **attr**

The attribute to write

#### **valueSize**

The size, in bytes, of the value

#### **value**

The attribute value to write

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `valueSize` or `value` is NULL, or if `attr` is not an activity attribute
- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT
  - Indicates that the `value` buffer is too small to hold the attribute value.

### Description

Write an activity API attribute.

## CUptiResult cuptiGetContextId (CUcontext context, uint32\_t \*contextId)

Get the ID of a context.

### Parameters

#### **context**

The context

#### **contextId**

Returns a process-unique ID for the context

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_CONTEXT
  - The context is NULL or not valid.
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if contextId is NULL

### Description

Get the ID of a context.

## CUptiResult cuptiGetDeviceId (CUcontext context, uint32\_t \*deviceId)

Get the ID of a device.

### Parameters

#### **context**

The context, or NULL to indicate the current context.

#### **deviceId**

Returns the ID of the device that is current for the calling thread.

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_DEVICE

- if unable to get device ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
- if `deviceId` is NULL

### Description

If `context` is NULL, returns the ID of the device that contains the currently active context. If `context` is non-NULL, returns the ID of the device which contains that context. Operates in a similar manner to `cudaGetDevice()` or `cuCtxGetDevice()` but may be called from within callback functions.

## CuptiResult cuptiGetStreamId (CUcontext context, CUstream stream, uint32\_t \*streamId)

Get the ID of a stream.

### Parameters

#### context

If non-NULL then the stream is checked to ensure that it belongs to this context. Typically this parameter should be null.

#### stream

The stream

#### streamId

Returns a context-unique ID for the stream

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_STREAM
  - if unable to get stream ID, or if `context` is non-NULL and `stream` does not belong to the context
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `streamId` is NULL

### Description

Get the ID of a stream. The stream ID is unique within a context (i.e. all streams within a context will have unique stream IDs).

See also:

`cuptiActivityEnqueueBuffer`

`cuptiActivityDequeueBuffer`

## **CUptiResult cuptiGetTimestamp (uint64\_t \*timestamp)**

Get the CUPTI timestamp.

### **Parameters**

#### **timestamp**

Returns the CUPTI timestamp

### **Returns**

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`

if `timestamp` is `NULL`

### **Description**

Returns a timestamp normalized to correspond with the start and end timestamps reported in the CUPTI activity records. The timestamp is reported in nanoseconds.

## **#define CUPTI\_CORRELATION\_ID\_UNKNOWN 0**

An invalid/unknown correlation ID. A correlation ID of this value indicates that there is no correlation for the activity record.

## **#define CUPTI\_GRID\_ID\_UNKNOWN 0LL**

An invalid/unknown grid ID.

## **#define CUPTI\_SOURCE\_LOCATOR\_ID\_UNKNOWN 0**

The source-locator ID that indicates an unknown source location. There is not an actual `CUpti_ActivitySourceLocator` object corresponding to this value.

## **#define CUPTI\_TIMESTAMP\_UNKNOWN 0LL**

An invalid/unknown timestamp for a start, end, queued, submitted, or completed time.

## **2.4. CUPTI Callback API**

Functions, types, and enums that implement the CUPTI Callback API.

## struct CUpti\_CallbackData

Data passed into a runtime or driver API callback function.

## struct CUpti\_ModuleResourceData

Module data passed into a resource callback function.

## struct CUpti\_NvtxData

Data passed into a NVTX callback function.

## struct CUpti\_ResourceData

Data passed into a resource callback function.

## struct CUpti\_SynchronizeData

Data passed into a synchronize callback function.

## enum CUpti\_ApiCallbackSite

Specifies the point in an API call that a callback is issued.

Specifies the point in an API call that a callback is issued. This value is communicated to the callback function via `CUpti_CallbackData::callbackSite`.

### Values

**CUPTI\_API\_ENTER = 0**

The callback is at the entry of the API call.

**CUPTI\_API\_EXIT = 1**

The callback is at the exit of the API call.

**CUPTI\_API\_CBSITE\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_CallbackDomain

Callback domains.

Callback domains. Each domain represents callback points for a group of related API functions or CUDA driver activity.

### Values

**CUPTI\_CB\_DOMAIN\_INVALID = 0**

Invalid domain.

**CUPTI\_CB\_DOMAIN\_DRIVER\_API = 1**

Domain containing callback points for all driver API functions.

**CUPTI\_CB\_DOMAIN\_RUNTIME\_API = 2**

Domain containing callback points for all runtime API functions.

**CUPTI\_CB\_DOMAIN\_RESOURCE = 3**

Domain containing callback points for CUDA resource tracking.

**CUPTI\_CB\_DOMAIN\_SYNCHRONIZE = 4**

Domain containing callback points for CUDA synchronization.

**CUPTI\_CB\_DOMAIN\_NVTX = 5**

Domain containing callback points for NVTX API functions.

**CUPTI\_CB\_DOMAIN\_SIZE = 6**

**CUPTI\_CB\_DOMAIN\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_CallbackIdResource

Callback IDs for resource domain.

Callback IDs for resource domain, CUPTI\_CB\_DOMAIN\_RESOURCE. This value is communicated to the callback function via the `cbid` parameter.

### Values

**CUPTI\_CBID\_RESOURCE\_INVALID = 0**

Invalid resource callback ID.

**CUPTI\_CBID\_RESOURCE\_CONTEXT\_CREATED = 1**

A new context has been created.

**CUPTI\_CBID\_RESOURCE\_CONTEXT\_DESTROY\_STARTING = 2**

A context is about to be destroyed.

**CUPTI\_CBID\_RESOURCE\_STREAM\_CREATED = 3**

A new stream has been created.

**CUPTI\_CBID\_RESOURCE\_STREAM\_DESTROY\_STARTING = 4**

A stream is about to be destroyed.

**CUPTI\_CBID\_RESOURCE\_CU\_INIT\_FINISHED = 5**

The driver has finished initializing.

**CUPTI\_CBID\_RESOURCE\_MODULE\_LOADED = 6**

A module has been loaded.

**CUPTI\_CBID\_RESOURCE\_MODULE\_UNLOAD\_STARTING = 7**

A module is about to be unloaded.

**CUPTI\_CBID\_RESOURCE\_MODULE\_PROFILED = 8**

The current module which is being profiled.

**CUPTI\_CBID\_RESOURCE\_SIZE**

**CUPTI\_CBID\_RESOURCE\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_CallbackIdSync

Callback IDs for synchronization domain.

Callback IDs for synchronization domain, CUPTI\_CB\_DOMAIN\_SYNCHRONIZE. This value is communicated to the callback function via the `cbid` parameter.

## Values

**CUPTI\_CBID\_SYNCHRONIZE\_INVALID = 0**

Invalid synchronize callback ID.

**CUPTI\_CBID\_SYNCHRONIZE\_STREAM\_SYNCHRONIZED = 1**

Stream synchronization has completed for the stream.

**CUPTI\_CBID\_SYNCHRONIZE\_CONTEXT\_SYNCHRONIZED = 2**

Context synchronization has completed for the context.

**CUPTI\_CBID\_SYNCHRONIZE\_SIZE**

**CUPTI\_CBID\_SYNCHRONIZE\_FORCE\_INT = 0x7fffffff**

**typedef (\*CUpti\_CallbackFunc) (void\* userdata,  
CUpti\_CallbackDomain domain, CUpti\_CallbackId cbid,  
const void\* cbdata)**

Function type for a callback.

Function type for a callback. The type of the data passed to the callback in `cbdata` depends on the `domain`. If `domain` is `CUPTI_CB_DOMAIN_DRIVER_API` or `CUPTI_CB_DOMAIN_RUNTIME_API` the type of `cbdata` will be [CUpti\\_CallbackData](#). If `domain` is `CUPTI_CB_DOMAIN_RESOURCE` the type of `cbdata` will be [CUpti\\_ResourceData](#). If `domain` is `CUPTI_CB_DOMAIN_SYNCHRONIZE` the type of `cbdata` will be [CUpti\\_SynchronizeData](#). If `domain` is `CUPTI_CB_DOMAIN_NVTX` the type of `cbdata` will be [CUpti\\_NvtxData](#).

**typedef uint32\_t CUpti\_CallbackId**

An ID for a driver API, runtime API, resource or synchronization callback.

An ID for a driver API, runtime API, resource or synchronization callback. Within a driver API callback this should be interpreted as a `CUpti_driver_api_trace_cbid` value (these values are defined in `cupti_driver_cbid.h`). Within a runtime API callback this should be interpreted as a `CUpti_runtime_api_trace_cbid` value (these values are defined in `cupti_runtime_cbid.h`). Within a resource API callback this should be interpreted as a [CUpti\\_CallbackIdResource](#) value. Within a synchronize API callback this should be interpreted as a [CUpti\\_CallbackIdSync](#) value.

**typedef CUpti\_DomainTable**

Pointer to an array of callback domains.

**typedef struct CUpti\_Subscriber\_st  
\*CUpti\_SubscriberHandle**

A callback subscriber.

## CUptiResult cuptiEnableAllDomains (uint32\_t enable, CUpti\_SubscriberHandle subscriber)

Enable or disable all callbacks in all domains.

### Parameters

#### enable

New enable state for all callbacks in all domain. Zero disables all callbacks, non-zero enables all callbacks.

#### subscriber

- Handle to callback subscription

### Returns

- ▶ CUPTI\_SUCCESS  
on success
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED  
if unable to initialize CUPTI
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if `subscriber` is invalid

### Description

Enable or disable all callbacks in all domains.



**Thread-safety:** a subscriber must serialize access to `cuptiGetCallbackState`, `cuptiEnableCallback`, `cuptiEnableDomain`, and `cuptiEnableAllDomains`. For example, if `cuptiGetCallbackState(sub, d, *)` and `cuptiEnableAllDomains(sub)` are called concurrently, the results are undefined.

## CUptiResult cuptiEnableCallback (uint32\_t enable, CUpti\_SubscriberHandle subscriber, CUpti\_CallbackDomain domain, CUpti\_CallbackId cbid)

Enable or disabled callbacks for a specific domain and callback ID.

### Parameters

#### enable

New enable state for the callback. Zero disables the callback, non-zero enables the callback.

**subscriber**

- Handle to callback subscription

**domain**

The domain of the callback

**cbid**

The ID of the callback

**Returns**

- ▶ CUPTI\_SUCCESS  
on success
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED  
if unable to initialize CUPTI
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if subscriber, domain or cbid is invalid.

**Description**

Enable or disabled callbacks for a subscriber for a specific domain and callback ID.



**Thread-safety:** a subscriber must serialize access to `cuptiGetCallbackState`, `cuptiEnableCallback`, `cuptiEnableDomain`, and `cuptiEnableAllDomains`. For example, if `cuptiGetCallbackState(sub, d, c)` and `cuptiEnableCallback(sub, d, c)` are called concurrently, the results are undefined.

## CuptiResult cuptiEnableDomain (uint32\_t enable, CUpti\_SubscriberHandle subscriber, CUpti\_CallbackDomain domain)

Enable or disabled all callbacks for a specific domain.

**Parameters****enable**

New enable state for all callbacks in the domain. Zero disables all callbacks, non-zero enables all callbacks.

**subscriber**

- Handle to callback subscription

**domain**

The domain of the callback

**Returns**

- ▶ `CUPTI_SUCCESS`  
on success
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`  
if unable to initialize CUPTI
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`  
if `subscriber` or `domain` is invalid

**Description**

Enable or disabled all callbacks for a specific domain.



**Thread-safety:** a subscriber must serialize access to `cuprtiGetCallbackState`, `cuprtiEnableCallback`, `cuprtiEnableDomain`, and `cuprtiEnableAllDomains`. For example, if `cuprtiGetCallbackEnabled(sub, d, *)` and `cuprtiEnableDomain(sub, d)` are called concurrently, the results are undefined.

## CuptiResult cuprtiGetCallbackName (Cupti\_CallbackDomain domain, uint32\_t cbid, const char \*\*name)

Get the name of a callback for a specific domain and callback ID.

**Parameters****domain**

The domain of the callback

**cbid**

The ID of the callback

**name**

Returns pointer to the name string on success, NULL otherwise

**Returns**

- ▶ `CUPTI_SUCCESS`  
on success
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`  
if `name` is NULL, or if `domain` or `cbid` is invalid.

## Description

Returns a pointer to the name `c_string` in `**name`.



**Names** are available only for the DRIVER and RUNTIME domains.

## CUptiResult cuptiGetCallbackState (uint32\_t \*enable, CUpti\_SubscriberHandle subscriber, CUpti\_CallbackDomain domain, CUpti\_CallbackId cbid)

Get the current enabled/disabled state of a callback for a specific domain and function ID.

### Parameters

#### **enable**

Returns non-zero if callback enabled, zero if not enabled

#### **subscriber**

Handle to the initialize subscriber

#### **domain**

The domain of the callback

#### **cbid**

The ID of the callback

### Returns

- ▶ CUPTI\_SUCCESS  
on success
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED  
if unable to initialize CUPTI
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if `enable` is NULL, or if `subscriber`, `domain` or `cbid` is invalid.

## Description

Returns non-zero in `*enable` if the callback for a domain and callback ID is enabled, and zero if not enabled.



**Thread-safety:** a subscriber must serialize access to `cuptiGetCallbackState`, `cuptiEnableCallback`, `cuptiEnableDomain`, and `cuptiEnableAllDomains`. For example, if `cuptiGetCallbackState(sub, d, c)` and `cuptiEnableCallback(sub, d, c)` are called concurrently, the results are undefined.

## CUptiResult cuptiSubscribe (CUpti\_SubscriberHandle \*subscriber, CUpti\_CallbackFunc callback, void \*userdata)

Initialize a callback subscriber with a callback function and user data.

### Parameters

#### subscriber

Returns handle to initialize subscriber

#### callback

The callback function

#### userdata

A pointer to user data. This data will be passed to the callback function via the `userdata` parameter.

### Returns

- ▶ CUPTI\_SUCCESS  
on success
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED  
if unable to initialize CUPTI
- ▶ CUPTI\_ERROR\_MAX\_LIMIT\_REACHED  
if there is already a CUPTI subscriber
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if `subscriber` is NULL

### Description

Initializes a callback subscriber with a callback function and (optionally) a pointer to user data. The returned subscriber handle can be used to enable and disable the callback for specific domains and callback IDs.



- ▶ Only a single subscriber can be registered at a time.
- ▶ This function does not enable any callbacks.
- ▶ **Thread-safety:** this function is thread safe.

## CUptiResult cuptiSupportedDomains (size\_t \*domainCount, CUpti\_DomainTable \*domainTable)

Get the available callback domains.

### Parameters

#### domainCount

Returns number of callback domains

#### domainTable

Returns pointer to array of available callback domains

### Returns

- ▶ CUPTI\_SUCCESS  
on success
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED  
if unable to initialize CUPTI
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if domainCount or domainTable are NULL

### Description

Returns in \*domainTable an array of size \*domainCount of all the available callback domains.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiUnsubscribe (CUpti\_SubscriberHandle subscriber)

Unregister a callback subscriber.

### Parameters

#### subscriber

Handle to the initialize subscriber

### Returns

- ▶ CUPTI\_SUCCESS  
on success
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED

- if unable to initialize CUPTI
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`
- if `subscriber` is `NULL` or not initialized

### Description

Removes a callback subscriber so that no future callbacks will be issued to that subscriber.



**Thread-safety:** this function is thread safe.

## 2.5. CUPTI Event API

Functions, types, and enums that implement the CUPTI Event API.

### struct Cupti\_EventGroupSet

A set of event groups.

### struct Cupti\_EventGroupSets

A set of event group sets.

### enum Cupti\_DeviceAttribute

Device attributes.

CUPTI device attributes. These attributes can be read using `cuptiDeviceGetAttribute`.

### Values

**CUPTI\_DEVICE\_ATTR\_MAX\_EVENT\_ID = 1**

Number of event IDs for a device. Value is a `uint32_t`.

**CUPTI\_DEVICE\_ATTR\_MAX\_EVENT\_DOMAIN\_ID = 2**

Number of event domain IDs for a device. Value is a `uint32_t`.

**CUPTI\_DEVICE\_ATTR\_GLOBAL\_MEMORY\_BANDWIDTH = 3**

Get global memory bandwidth in Kbytes/sec. Value is a `uint64_t`.

**CUPTI\_DEVICE\_ATTR\_INSTRUCTION\_PER\_CYCLE = 4**

Get theoretical maximum number of instructions per cycle. Value is a `uint32_t`.

**CUPTI\_DEVICE\_ATTR\_INSTRUCTION\_THROUGHPUT\_SINGLE\_PRECISION = 5**

Get theoretical maximum number of single precision instructions that can be executed per second. Value is a `uint64_t`.

**CUPTI\_DEVICE\_ATTR\_MAX\_FRAME\_BUFFERS = 6**

Get number of frame buffers for device. Value is a `uint64_t`.

**CUPTI\_DEVICE\_ATTR\_PCIE\_LINK\_RATE = 7**

Get PCIe link rate in Mega bits/sec for device. Return 0 if bus-type is non-PCIe. Value is a `uint64_t`.

**CUPTI\_DEVICE\_ATTR\_PCIE\_LINK\_WIDTH = 8**

Get PCIe link width for device. Return 0 if bus-type is non-PCIe. Value is a `uint64_t`.

**CUPTI\_DEVICE\_ATTR\_PCIE\_GEN = 9**

Get PCIe generation for device. Return 0 if bus-type is non-PCIe. Value is a `uint64_t`.

**CUPTI\_DEVICE\_ATTR\_DEVICE\_CLASS = 10**

Get the class for the device. Value is a `CUpti_DeviceAttributeDeviceClass`.

**CUPTI\_DEVICE\_ATTR\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_DeviceAttributeDeviceClass

Device class.

Enumeration of device classes for device attribute

CUPTI\_DEVICE\_ATTR\_DEVICE\_CLASS.

### Values

**CUPTI\_DEVICE\_ATTR\_DEVICE\_CLASS\_TESLA = 0**

**CUPTI\_DEVICE\_ATTR\_DEVICE\_CLASS\_QUADRO = 1**

**CUPTI\_DEVICE\_ATTR\_DEVICE\_CLASS\_GEFORCE = 2**

## enum CUpti\_EventAttribute

Event attributes.

Event attributes. These attributes can be read using `cuprtiEventGetAttribute`.

### Values

**CUPTI\_EVENT\_ATTR\_NAME = 0**

Event name. Value is a null terminated const c-string.

**CUPTI\_EVENT\_ATTR\_SHORT\_DESCRIPTION = 1**

Short description of event. Value is a null terminated const c-string.

**CUPTI\_EVENT\_ATTR\_LONG\_DESCRIPTION = 2**

Long description of event. Value is a null terminated const c-string.

**CUPTI\_EVENT\_ATTR\_CATEGORY = 3**

Category of event. Value is `CUpti_EventCategory`.

**CUPTI\_EVENT\_ATTR\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_EventCategory

An event category.

Each event is assigned to a category that represents the general type of the event. A event's category is accessed using `cuprtiEventGetAttribute` and the `CUPTI_EVENT_ATTR_CATEGORY` attribute.

**Values****CUPTI\_EVENT\_CATEGORY\_INSTRUCTION = 0**

An instruction related event.

**CUPTI\_EVENT\_CATEGORY\_MEMORY = 1**

A memory related event.

**CUPTI\_EVENT\_CATEGORY\_CACHE = 2**

A cache related event.

**CUPTI\_EVENT\_CATEGORY\_PROFILE\_TRIGGER = 3**

A profile-trigger event.

**CUPTI\_EVENT\_CATEGORY\_FORCE\_INT = 0x7fffffff****enum CUpti\_EventCollectionMethod**

The collection method used for an event.

The collection method indicates how an event is collected.

**Values****CUPTI\_EVENT\_COLLECTION\_METHOD\_PM = 0**

Event is collected using a hardware global performance monitor.

**CUPTI\_EVENT\_COLLECTION\_METHOD\_SM = 1**

Event is collected using a hardware SM performance monitor.

**CUPTI\_EVENT\_COLLECTION\_METHOD\_INSTRUMENTED = 2**

Event is collected using software instrumentation.

**CUPTI\_EVENT\_COLLECTION\_METHOD\_FORCE\_INT = 0x7fffffff****enum CUpti\_EventCollectionMode**

Event collection modes.

The event collection mode determines the period over which the events within the enabled event groups will be collected.

**Values****CUPTI\_EVENT\_COLLECTION\_MODE\_CONTINUOUS = 0**Events are collected for the entire duration between the `cuptiEventGroupEnable` and `cuptiEventGroupDisable` calls. This is the default mode. For devices with compute capability less than 2.0, event values are reset when a kernel is launched. For all other devices event values are only reset when the events are read.**CUPTI\_EVENT\_COLLECTION\_MODE\_KERNEL = 1**Events are collected only for the durations of kernel executions that occur between the `cuptiEventGroupEnable` and `cuptiEventGroupDisable` calls. Event collection begins when a kernel execution begins, and stops when kernel execution completes. Event values are reset to zero when each kernel execution begins. If multiple kernel executions occur between the `cuptiEventGroupEnable` and `cuptiEventGroupDisable`

calls then the event values must be read after each kernel launch if those events need to be associated with the specific kernel launch.

**CUPTI\_EVENT\_COLLECTION\_MODE\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_EventDomainAttribute

Event domain attributes.

Event domain attributes. Except where noted, all the attributes can be read using either [cuptiDeviceGetEventDomainAttribute](#) or [cuptiEventDomainGetAttribute](#).

### Values

**CUPTI\_EVENT\_DOMAIN\_ATTR\_NAME = 0**

Event domain name. Value is a null terminated const c-string.

**CUPTI\_EVENT\_DOMAIN\_ATTR\_INSTANCE\_COUNT = 1**

Number of instances of the domain for which event counts will be collected.

The domain may have additional instances that cannot be profiled (see

**CUPTI\_EVENT\_DOMAIN\_ATTR\_TOTAL\_INSTANCE\_COUNT**). Can be read only with [cuptiDeviceGetEventDomainAttribute](#). Value is a `uint32_t`.

**CUPTI\_EVENT\_DOMAIN\_ATTR\_TOTAL\_INSTANCE\_COUNT = 3**

Total number of instances of the domain, including instances that cannot be profiled. Use **CUPTI\_EVENT\_DOMAIN\_ATTR\_INSTANCE\_COUNT** to get the number of instances that can be profiled. Can be read only with [cuptiDeviceGetEventDomainAttribute](#). Value is a `uint32_t`.

**CUPTI\_EVENT\_DOMAIN\_ATTR\_COLLECTION\_METHOD = 4**

Collection method used for events contained in the event domain. Value is a [CUpti\\_EventCollectionMethod](#).

**CUPTI\_EVENT\_DOMAIN\_ATTR\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_EventGroupAttribute

Event group attributes.

Event group attributes. These attributes can be read using [cuptiEventGroupGetAttribute](#). Attributes marked [rw] can also be written using [cuptiEventGroupSetAttribute](#).

### Values

**CUPTI\_EVENT\_GROUP\_ATTR\_EVENT\_DOMAIN\_ID = 0**

The domain to which the event group is bound. This attribute is set when the first event is added to the group. Value is a `CUpti_EventDomainID`.

**CUPTI\_EVENT\_GROUP\_ATTR\_PROFILE\_ALL\_DOMAIN\_INSTANCES = 1**

[rw] Profile all the instances of the domain for this eventgroup. This feature can be used to get load balancing across all instances of a domain. Value is an integer.

**CUPTI\_EVENT\_GROUP\_ATTR\_USER\_DATA = 2**

[rw] Reserved for user data.

**CUPTI\_EVENT\_GROUP\_ATTR\_NUM\_EVENTS = 3**

Number of events in the group. Value is a uint32\_t.

**CUPTI\_EVENT\_GROUP\_ATTR\_EVENTS = 4**

Enumerates events in the group. Value is a pointer to buffer of size `sizeof(CUpti_EventID) * num_of_events` in the eventgroup. `num_of_events` can be queried using `CUPTI_EVENT_GROUP_ATTR_NUM_EVENTS`.

**CUPTI\_EVENT\_GROUP\_ATTR\_INSTANCE\_COUNT = 5**

Number of instances of the domain bound to this event group that will be counted. Value is a uint32\_t.

**CUPTI\_EVENT\_GROUP\_ATTR\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_ReadEventFlags

Flags for `cuptiEventGroupReadEvent` and `cuptiEventGroupReadAllEvents`.

Flags for `cuptiEventGroupReadEvent` and `cuptiEventGroupReadAllEvents`.

### Values

**CUPTI\_EVENT\_READ\_FLAG\_NONE = 0**

No flags.

**CUPTI\_EVENT\_READ\_FLAG\_FORCE\_INT = 0x7fffffff**

## typedef uint32\_t CUpti\_EventDomainID

ID for an event domain.

ID for an event domain. An event domain represents a group of related events. A device may have multiple instances of a domain, indicating that the device can simultaneously record multiple instances of each event within that domain.

## typedef void \*CUpti\_EventGroup

A group of events.

An event group is a collection of events that are managed together. All events in an event group must belong to the same domain.

## typedef uint32\_t CUpti\_EventID

ID for an event.

An event represents a countable activity, action, or occurrence on the device.

## CUptiResult cuptiDeviceEnumEventDomains (CUdevice device, size\_t \*arraySizeBytes, CUpti\_EventDomainID \*domainArray)

Get the event domains for a device.

### Parameters

#### device

The CUDA device

#### arraySizeBytes

The size of domainArray in bytes, and returns the number of bytes written to domainArray

#### domainArray

Returns the IDs of the event domains for the device

### Returns

- ▶ CUPTI\_SUCCESS
  - ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
  - ▶ CUPTI\_ERROR\_INVALID\_DEVICE
  - ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
- if arraySizeBytes or domainArray are NULL

### Description

Returns the event domains IDs in domainArray for a device. The size of the domainArray buffer is given by \*arraySizeBytes. The size of the domainArray buffer must be at least numdomains \* sizeof(CUpti\_EventDomainID) or else all domains will not be returned. The value returned in \*arraySizeBytes contains the number of bytes returned in domainArray.



**Thread-safety:** this function is thread safe.

## CuptiResult cuptiDeviceGetAttribute (CUdevice device, CUpti\_DeviceAttribute attrib, size\_t \*valueSize, void \*value)

Read a device attribute.

### Parameters

#### device

The CUDA device

#### attrib

The attribute to read

#### valueSize

Size of buffer pointed by the value, and returns the number of bytes written to value

#### value

Returns the value of the attribute

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_DEVICE
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if valueSize or value is NULL, or if attrib is not a device attribute
- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT
  - For non-c-string attribute values, indicates that the value buffer is too small to hold the attribute value.

### Description

Read a device attribute and return it in \*value.



**Thread-safety:** this function is thread safe.

## CuptiResult cuptiDeviceGetEventDomainAttribute (CUdevice device, CUpti\_EventDomainID eventDomain,

## CUpti\_EventDomainAttribute attrib, size\_t \*valueSize, void \*value)

Read an event domain attribute.

### Parameters

#### **device**

The CUDA device

#### **eventDomain**

ID of the event domain

#### **attrib**

The event domain attribute to read

#### **valueSize**

The size of the `value` buffer in bytes, and returns the number of bytes written to `value`

#### **value**

Returns the attribute's value

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_DEVICE
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_DOMAIN\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `valueSize` or `value` is NULL, or if `attrib` is not an event domain attribute
- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT
  - For non-c-string attribute values, indicates that the `value` buffer is too small to hold the attribute value.

### Description

Returns an event domain attribute in `*value`. The size of the `value` buffer is given by `*valueSize`. The value returned in `*valueSize` contains the number of bytes returned in `value`.

If the attribute value is a c-string that is longer than `*valueSize`, then only the first `*valueSize` characters will be returned and there will be no terminating null byte.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiDeviceGetNumEventDomains (CUdevice device, uint32\_t \*numDomains)

Get the number of domains for a device.

### Parameters

#### device

The CUDA device

#### numDomains

Returns the number of domains

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_DEVICE
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if numDomains is NULL

### Description

Returns the number of domains in numDomains for a device.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiDeviceGetTimestamp (CUcontext context, uint64\_t \*timestamp)

Read a device timestamp.

### Parameters

#### context

A context on the device from which to get the timestamp

#### timestamp

Returns the device timestamp

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED

- ▶ CUPTI\_ERROR\_INVALID\_CONTEXT
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

is timestamp is NULL

### Description

Returns the device timestamp in \*timestamp. The timestamp is reported in nanoseconds and indicates the time since the device was last reset.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiDisableKernelReplayMode (CUcontext context)

Disable kernel replay mode.

### Parameters

#### context

The context

### Returns

- ▶ CUPTI\_SUCCESS

### Description

Set profiling mode for the context to non-replay (default) mode. Event collection mode will be set to CUPTI\_EVENT\_COLLECTION\_MODE\_CONTINUOUS. All previously enabled event groups and event group sets will be disabled.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiEnableKernelReplayMode (CUcontext context)

Enable kernel replay mode.

### Parameters

#### context

The context

**Returns**

- ▶ CUPTI\_SUCCESS

**Description**

Set profiling mode for the context to replay mode. In this mode, any number of events can be collected in one run of the kernel. The event collection mode will automatically switch to CUPTI\_EVENT\_COLLECTION\_MODE\_KERNEL. In this mode, `cuptiSetEventCollectionMode` will return CUPTI\_ERROR\_INVALID\_OPERATION.



- ▶ **Kernels** might take longer to run if many events are enabled.
- ▶ **Thread-safety:** this function is thread safe.

## CuptiResult cuptiEnumEventDomains (size\_t \*arraySizeBytes, CUpti\_EventDomainID \*domainArray)

Get the event domains available on any device.

**Parameters****arraySizeBytes**

The size of `domainArray` in bytes, and returns the number of bytes written to `domainArray`

**domainArray**

Returns all the event domains

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if `arraySizeBytes` or `domainArray` are NULL

**Description**

Returns all the event domains available on any CUDA-capable device. Event domain IDs are returned in `domainArray`. The size of the `domainArray` buffer is given by `*arraySizeBytes`. The size of the `domainArray` buffer must be at least `numDomains * sizeof(CUpti_EventDomainID)` or all domains will not be returned. The value returned in `*arraySizeBytes` contains the number of bytes returned in `domainArray`.



- ▶ **Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventDomainEnumEvents (CUpti\_EventDomainID eventDomain, size\_t \*arraySizeBytes, CUpti\_EventID \*eventArray)

Get the events in a domain.

### Parameters

#### eventDomain

ID of the event domain

#### arraySizeBytes

The size of `eventArray` in bytes, and returns the number of bytes written to `eventArray`

#### eventArray

Returns the IDs of the events in the domain

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_DOMAIN\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `arraySizeBytes` or `eventArray` are NULL

### Description

Returns the event IDs in `eventArray` for a domain. The size of the `eventArray` buffer is given by `*arraySizeBytes`. The size of the `eventArray` buffer must be at least `numdomainevents * sizeof(CUpti_EventID)` or else all events will not be returned. The value returned in `*arraySizeBytes` contains the number of bytes returned in `eventArray`.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventDomainGetAttribute (CUpti\_EventDomainID eventDomain,

## CUpti\_EventDomainAttribute attrib, size\_t \*valueSize, void \*value)

Read an event domain attribute.

### Parameters

#### eventDomain

ID of the event domain

#### attrib

The event domain attribute to read

#### valueSize

The size of the `value` buffer in bytes, and returns the number of bytes written to `value`

#### value

Returns the attribute's value

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_DOMAIN\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `valueSize` or `value` is NULL, or if `attrib` is not an event domain attribute
- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT
  - For non-c-string attribute values, indicates that the `value` buffer is too small to hold the attribute value.

### Description

Returns an event domain attribute in `*value`. The size of the `value` buffer is given by `*valueSize`. The value returned in `*valueSize` contains the number of bytes returned in `value`.

If the attribute value is a c-string that is longer than `*valueSize`, then only the first `*valueSize` characters will be returned and there will be no terminating null byte.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventDomainGetNumEvents (CUpti\_EventDomainID eventDomain, uint32\_t \*numEvents)

Get number of events in a domain.

### Parameters

#### eventDomain

ID of the event domain

#### numEvents

Returns the number of events in the domain

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_DOMAIN\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `numEvents` is NULL

### Description

Returns the number of events in `numEvents` for a domain.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventGetAttribute (CUpti\_EventID event, CUpti\_EventAttribute attrib, size\_t \*valueSize, void \*value)

Get an event attribute.

### Parameters

#### event

ID of the event

#### attrib

The event attribute to read

**valueSize**

The size of the `value` buffer in bytes, and returns the number of bytes written to `value`

**value**

Returns the attribute's value

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `valueSize` or `value` is NULL, or if `attrib` is not an event attribute

- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT

For non-c-string attribute values, indicates that the `value` buffer is too small to hold the attribute value.

**Description**

Returns an event attribute in `*value`. The size of the `value` buffer is given by `*valueSize`. The value returned in `*valueSize` contains the number of bytes returned in `value`.

If the attribute value is a c-string that is longer than `*valueSize`, then only the first `*valueSize` characters will be returned and there will be no terminating null byte.



**Thread-safety:** this function is thread safe.

## CuptiResult cuptiEventGetIdFromName (CUdevice device, const char \*eventName, CUpti\_EventID \*event)

Find an event by name.

**Parameters****device**

The CUDA device

**eventName**

The name of the event to find

**event**

Returns the ID of the found event or undefined if unable to find the event

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_DEVICE
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_NAME
  - if unable to find an event with name `eventName`. In this case `*event` is undefined
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `eventName` or `event` are NULL

**Description**

Find an event by name and return the event ID in `*event`.



**Thread-safety:** this function is thread safe.

## CuptiResult cuptiEventGroupAddEvent (CUpti\_EventGroup eventGroup, CUpti\_EventID event)

Add an event to an event group.

**Parameters****eventGroup**

The event group

**event**

The event to add to the group

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_ID
- ▶ CUPTI\_ERROR\_OUT\_OF\_MEMORY
- ▶ CUPTI\_ERROR\_INVALID\_OPERATION
  - if `eventGroup` is enabled
- ▶ CUPTI\_ERROR\_NOT\_COMPATIBLE

if `event` belongs to a different event domain than the events already in `eventGroup`, or if a device limitation prevents `event` from being collected at the same time as the events already in `eventGroup`

- ▶ `CUPTI_ERROR_MAX_LIMIT_REACHED`

if `eventGroup` is full

- ▶ `CUPTI_ERROR_INVALID_PARAMETER`

if `eventGroup` is `NULL`

## Description

Add an event to an event group. The event add can fail for a number of reasons:

- ▶ The event group is enabled
- ▶ The event does not belong to the same event domain as the events that are already in the event group
- ▶ Device limitations on the events that can belong to the same group
- ▶ The event group is full



**Thread-safety:** this function is thread safe.

## CuptiResult cuptiEventGroupCreate (CUcontext context, CUpti\_EventGroup \*eventGroup, uint32\_t flags)

Create a new event group for a context.

### Parameters

#### **context**

The context for the event group

#### **eventGroup**

Returns the new event group

#### **flags**

Reserved - must be zero

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_INVALID_CONTEXT`
- ▶ `CUPTI_ERROR_OUT_OF_MEMORY`
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`

if `eventGroup` is `NULL`

### Description

Creates a new event group for `context` and returns the new group in `*eventGroup`.



- ▶ `flags` are reserved for future use and should be set to zero.
- ▶ **Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventGroupDestroy (CUpti\_EventGroup eventGroup)

Destroy an event group.

### Parameters

#### **eventGroup**

The event group to destroy

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_INVALID_OPERATION`  
if the event group is enabled
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`  
if `eventGroup` is `NULL`

### Description

Destroy an `eventGroup` and free its resources. An event group cannot be destroyed if it is enabled.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventGroupDisable (CUpti\_EventGroup eventGroup)

Disable an event group.

### Parameters

#### eventGroup

The event group

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_HARDWARE
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if eventGroup is NULL

### Description

Disable an event group. Disabling an event group stops collection of events contained in the group.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventGroupEnable (CUpti\_EventGroup eventGroup)

Enable an event group.

### Parameters

#### eventGroup

The event group

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_HARDWARE
- ▶ CUPTI\_ERROR\_NOT\_READY

- if `eventGroup` does not contain any events
- ▶ `CUPTI_ERROR_NOT_COMPATIBLE`
- if `eventGroup` cannot be enabled due to other already enabled event groups
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`
- if `eventGroup` is `NULL`
- ▶ `CUPTI_ERROR_HARDWARE_BUSY`
- if another client is profiling and hardware is busy

### Description

Enable an event group. Enabling an event group zeros the value of all the events in the group and then starts collection of those events.



**Thread-safety:** this function is thread safe.

**CUptiResult** `cuptiEventGroupGetAttribute`  
 (`CUpti_EventGroup` `eventGroup`,  
`CUpti_EventGroupAttribute` `attrib`, `size_t` `*valueSize`,  
`void` `*value`)

Read an event group attribute.

### Parameters

#### **eventGroup**

The event group

#### **attrib**

The attribute to read

#### **valueSize**

Size of buffer pointed by the value, and returns the number of bytes written to `value`

#### **value**

Returns the value of the attribute

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`
- if `valueSize` or `value` is `NULL`, or if `attrib` is not an eventgroup attribute
- ▶ `CUPTI_ERROR_PARAMETER_SIZE_NOT_SUFFICIENT`

For non-c-string attribute values, indicates that the `value` buffer is too small to hold the attribute value.

## Description

Read an event group attribute and return it in `*value`.



**Thread-safety:** this function is thread safe but client must guard against simultaneous destruction or modification of `eventGroup` (for example, client must guard against simultaneous calls to `cuptiEventGroupDestroy`, `cuptiEventGroupAddEvent`, etc.), and must guard against simultaneous destruction of the context in which `eventGroup` was created (for example, client must guard against simultaneous calls to `cudaDeviceReset`, `cuCtxDestroy`, etc.).

**CUptiResult cuptiEventGroupReadAllEvents**  
 (CUpti\_EventGroup eventGroup, CUpti\_ReadEventFlags flags, size\_t \*eventValueBufferSizeBytes, uint64\_t \*eventValueBuffer, size\_t \*eventIdArraySizeBytes, CUpti\_EventID \*eventIdArray, size\_t \*numEventIdsRead)

Read the values for all the events in an event group.

## Parameters

### eventGroup

The event group

### flags

Flags controlling the reading mode

### eventValueBufferSizeBytes

The size of `eventValueBuffer` in bytes, and returns the number of bytes written to `eventValueBuffer`

### eventValueBuffer

Returns the event values

### eventIdArraySizeBytes

The size of `eventIdArray` in bytes, and returns the number of bytes written to `eventIdArray`

### eventIdArray

Returns the IDs of the events in the same order as the values return in `eventValueBuffer`.

### numEventIdsRead

Returns the number of event IDs returned in `eventIdArray`

## Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_HARDWARE
- ▶ CUPTI\_ERROR\_INVALID\_OPERATION
  - if eventGroup is disabled
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if eventGroup, eventValueBufferSizeBytes, eventValueBuffer, eventIdArraySizeBytes, eventIdArray or numEventIdsRead is NULL

## Description

Read the values for all the events in an event group. The event values are returned in the eventValueBuffer buffer. eventValueBufferSizeBytes indicates the size of eventValueBuffer. The buffer must be at least (sizeof(uint64) \* number of events in group) if CUPTI\_EVENT\_GROUP\_ATTR\_PROFILE\_ALL\_DOMAIN\_INSTANCES is not set on the group containing the events. The buffer must be at least (sizeof(uint64) \* number of domain instances \* number of events in group) if CUPTI\_EVENT\_GROUP\_ATTR\_PROFILE\_ALL\_DOMAIN\_INSTANCES is set on the group.

The data format returned in eventValueBuffer is:

- ▶ domain instance 0: event0 event1 ... eventN
- ▶ domain instance 1: event0 event1 ... eventN
- ▶ ...
- ▶ domain instance M: event0 event1 ... eventN

The event order in eventValueBuffer is returned in eventIdArray. The size of eventIdArray is specified in eventIdArraySizeBytes. The size should be at least (sizeof(CUpti\_EventID) \* number of events in group).

If any instance of any event counter overflows, the value returned for that event instance will be CUPTI\_EVENT\_OVERFLOW.

The only allowed value for flags is CUPTI\_EVENT\_READ\_FLAG\_NONE.

Reading events from a disabled event group is not allowed. After being read, an event's value is reset to zero.



**Thread-safety:** this function is thread safe but client must guard against simultaneous destruction or modification of eventGroup (for example, client must guard against simultaneous calls to `cuptiEventGroupDestroy`, `cuptiEventGroupAddEvent`, etc.), and must guard against simultaneous destruction of the context in which

`eventGroup` was created (for example, client must guard against simultaneous calls to `cudaDeviceReset`, `cuCtxDestroy`, etc.). If `cuptiEventGroupResetAllEvents` is called simultaneously with this function, then returned event values are undefined.

**CUptiResult cuptiEventGroupReadEvent**  
 (CUpti\_EventGroup eventGroup, CUpti\_ReadEventFlags flags, CUpti\_EventID event, size\_t \*eventValueBufferSizeBytes, uint64\_t \*eventValueBuffer)

Read the value for an event in an event group.

#### Parameters

##### **eventGroup**

The event group

##### **flags**

Flags controlling the reading mode

##### **event**

The event to read

##### **eventValueBufferSizeBytes**

The size of `eventValueBuffer` in bytes, and returns the number of bytes written to `eventValueBuffer`

##### **eventValueBuffer**

Returns the event value(s)

#### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_ID
- ▶ CUPTI\_ERROR\_HARDWARE
- ▶ CUPTI\_ERROR\_INVALID\_OPERATION
  - if `eventGroup` is disabled
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `eventGroup`, `eventValueBufferSizeBytes` or `eventValueBuffer` is NULL

## Description

Read the value for an event in an event group. The event value is returned in the `eventValueBuffer` buffer. `eventValueBufferSizeBytes` indicates the size of the `eventValueBuffer` buffer. The buffer must be at least `sizeof(uint64)` if `CUPTI_EVENT_GROUP_ATTR_PROFILE_ALL_DOMAIN_INSTANCES` is not set on the group containing the event. The buffer must be at least `(sizeof(uint64) * number of domain instances)` if `CUPTI_EVENT_GROUP_ATTR_PROFILE_ALL_DOMAIN_INSTANCES` is set on the group.

If any instance of an event counter overflows, the value returned for that event instance will be `CUPTI_EVENT_OVERFLOW`.

The only allowed value for `flags` is `CUPTI_EVENT_READ_FLAG_NONE`.

Reading an event from a disabled event group is not allowed. After being read, an event's value is reset to zero.



**Thread-safety:** this function is thread safe but client must guard against simultaneous destruction or modification of `eventGroup` (for example, client must guard against simultaneous calls to `cuptiEventGroupDestroy`, `cuptiEventGroupAddEvent`, etc.), and must guard against simultaneous destruction of the context in which `eventGroup` was created (for example, client must guard against simultaneous calls to `cudaDeviceReset`, `cuCtxDestroy`, etc.). If `cuptiEventGroupResetAllEvents` is called simultaneously with this function, then returned event values are undefined.

## CuptiResult cuptiEventGroupRemoveAllEvents (CUpti\_EventGroup eventGroup)

Remove all events from an event group.

### Parameters

#### **eventGroup**

The event group

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_INVALID_OPERATION`
  - if `eventGroup` is enabled
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`

if `eventGroup` is NULL

### Description

Remove all events from an event group. Events cannot be removed if the event group is enabled.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventGroupRemoveEvent (CUpti\_EventGroup eventGroup, CUpti\_EventID event)

Remove an event from an event group.

### Parameters

#### **eventGroup**

The event group

#### **event**

The event to remove from the group

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_ID
- ▶ CUPTI\_ERROR\_INVALID\_OPERATION
  - if `eventGroup` is enabled
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `eventGroup` is NULL

### Description

Remove `event` from the an event group. The event cannot be removed if the event group is enabled.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventGroupResetAllEvents (CUpti\_EventGroup eventGroup)

Zero all the event counts in an event group.

### Parameters

#### **eventGroup**

The event group

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_HARDWARE
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `eventGroup` is NULL

### Description

Zero all the event counts in an event group.



**Thread-safety:** this function is thread safe but client must guard against simultaneous destruction or modification of `eventGroup` (for example, client must guard against simultaneous calls to `cuptiEventGroupDestroy`, `cuptiEventGroupAddEvent`, etc.), and must guard against simultaneous destruction of the context in which `eventGroup` was created (for example, client must guard against simultaneous calls to `cudaDeviceReset`, `cuCtxDestroy`, etc.).

## CUptiResult cuptiEventGroupSetAttribute (CUpti\_EventGroup eventGroup, CUpti\_EventGroupAttribute attrib, size\_t valueSize, void \*value)

Write an event group attribute.

### Parameters

#### **eventGroup**

The event group

#### **attrib**

The attribute to write

**valueSize**

The size, in bytes, of the value

**value**

The attribute value to write

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `valueSize` or `value` is NULL, or if `attrib` is not an event group attribute, or if `attrib` is not a writable attribute

- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT

Indicates that the `value` buffer is too small to hold the attribute value.

**Description**

Write an event group attribute.



**Thread-safety:** this function is thread safe.

## CuptiResult cuptiEventGroupSetDisable (Cupti\_EventGroupSet \*eventGroupSet)

Disable an event group set.

**Parameters****eventGroupSet**

The pointer to the event group set

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_HARDWARE
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `eventGroupSet` is NULL

## Description

Disable a set of event groups. Disabling a set of event groups stops collection of events contained in the groups.



- ▶ **Thread-safety:** this function is thread safe.
- ▶ If this call fails, some of the event groups in the set may be disabled and other event groups may remain enabled.

## CuptiResult cuptiEventGroupSetEnable (Cupti\_EventGroupSet \*eventGroupSet)

Enable an event group set.

### Parameters

#### eventGroupSet

The pointer to the event group set

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_HARDWARE
- ▶ CUPTI\_ERROR\_NOT\_READY
  - if eventGroup does not contain any events
- ▶ CUPTI\_ERROR\_NOT\_COMPATIBLE
  - if eventGroup cannot be enabled due to other already enabled event groups
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if eventGroupSet is NULL
- ▶ CUPTI\_ERROR\_HARDWARE\_BUSY
  - if other client is profiling and hardware is busy

## Description

Enable a set of event groups. Enabling a set of event groups zeros the value of all the events in all the groups and then starts collection of those events.



- ▶ **Thread-safety:** this function is thread safe.

## CUptiResult cuptiEventGroupSetsCreate (CUcontext context, size\_t eventIdArraySizeBytes, CUpti\_EventID \*eventIdArray, CUpti\_EventGroupSets \*\*eventGroupPasses)

For a set of events, get the grouping that indicates the number of passes and the event groups necessary to collect the events.

### Parameters

#### context

The context for event collection

#### eventIdArraySizeBytes

Size of eventIdArray in bytes

#### eventIdArray

Array of event IDs that need to be grouped

#### eventGroupPasses

Returns a [CUpti\\_EventGroupSets](#) object that indicates the number of passes required to collect the events and the events to collect on each pass

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_CONTEXT
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if eventIdArray or eventGroupPasses is NULL

### Description

The number of events that can be collected simultaneously varies by device and by the type of the events. When events can be collected simultaneously, they may need to be grouped into multiple event groups because they are from different event domains. This function takes a set of events and determines how many passes are required to collect all those events, and which events can be collected simultaneously in each pass.

The [CUpti\\_EventGroupSets](#) returned in eventGroupPasses indicates how many passes are required to collect the events with the numSets field. Within each event

group set, the `sets` array indicates the event groups that should be collected on each pass.



**Thread-safety:** this function is thread safe, but client must guard against another thread simultaneously destroying `context`.

## CUptiResult cuptiEventGroupSetsDestroy (CUpti\_EventGroupSets \*eventGroupSets)

Destroy a CUpti\_EventGroupSets object.

### Parameters

#### eventGroupSets

The object to destroy

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_OPERATION
  - if any of the event groups contained in the sets is enabled
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if eventGroupSets is NULL

### Description

Destroy a CUpti\_EventGroupSets object.



**Thread-safety:** this function is thread safe.

## CUptiResult cuptiGetNumEventDomains (uint32\_t \*numDomains)

Get the number of event domains available on any device.

### Parameters

#### numDomains

Returns the number of domains

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `numDomains` is NULL

**Description**

Returns the total number of event domains available on any CUDA-capable device.



**Thread-safety:** this function is thread safe.

## CuptiResult cuptiSetEventCollectionMode (CUcontext context, CUpti\_EventCollectionMode mode)

Set the event collection mode.

**Parameters****context**

The context

**mode**

The event collection mode

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_CONTEXT
- ▶ CUPTI\_ERROR\_INVALID\_OPERATION
  - if called when replay mode is enabled

**Description**

Set the event collection mode for a `context`. The `mode` controls the event collection behavior of all events in event groups created in the `context`. This API is invalid in kernel replay mode.



**Thread-safety:** this function is thread safe.

```
#define CUPTI_EVENT_OVERFLOW
((uint64_t)0xFFFFFFFFFFFFFFFFULL)
```

The overflow value for a CUPTI event.

The CUPTI event value that indicates an overflow.

## 2.6. CUPTI Metric API

Functions, types, and enums that implement the CUPTI Metric API.

```
union CUpti_MetricValue
```

A metric value.

```
enum CUpti_MetricAttribute
```

Metric attributes.

Metric attributes describe properties of a metric. These attributes can be read using [cuprtMetricGetAttribute](#).

### Values

```
CUPTI_METRIC_ATTR_NAME = 0
```

Metric name. Value is a null terminated const c-string.

```
CUPTI_METRIC_ATTR_SHORT_DESCRIPTION = 1
```

Short description of metric. Value is a null terminated const c-string.

```
CUPTI_METRIC_ATTR_LONG_DESCRIPTION = 2
```

Long description of metric. Value is a null terminated const c-string.

```
CUPTI_METRIC_ATTR_CATEGORY = 3
```

Category of the metric. Value is of type CUpti\_MetricCategory.

```
CUPTI_METRIC_ATTR_VALUE_KIND = 4
```

Value type of the metric. Value is of type CUpti\_MetricValueKind.

```
CUPTI_METRIC_ATTR_EVALUATION_MODE = 5
```

Metric evaluation mode. Value is of type CUpti\_MetricEvaluationMode.

```
CUPTI_METRIC_ATTR_FORCE_INT = 0x7fffffff
```

```
enum CUpti_MetricCategory
```

A metric category.

Each metric is assigned to a category that represents the general type of the metric. A metric's category is accessed using [cuprtMetricGetAttribute](#) and the CUPTI\_METRIC\_ATTR\_CATEGORY attribute.

## Values

**CUPTI\_METRIC\_CATEGORY\_MEMORY = 0**

A memory related metric.

**CUPTI\_METRIC\_CATEGORY\_INSTRUCTION = 1**

An instruction related metric.

**CUPTI\_METRIC\_CATEGORY\_MULTIPROCESSOR = 2**

A multiprocessor related metric.

**CUPTI\_METRIC\_CATEGORY\_CACHE = 3**

A cache related metric.

**CUPTI\_METRIC\_CATEGORY\_TEXTURE = 4**

A texture related metric.

**CUPTI\_METRIC\_CATEGORY\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_MetricEvaluationMode

A metric evaluation mode.

A metric can be evaluated per hardware instance to know the load balancing across instances of a domain or the metric can be evaluated in aggregate mode when the events involved in metric evaluation are from different event domains. It might be possible to evaluate some metrics in both modes for convenience. A metric's evaluation mode is accessed using [CUpti\\_MetricEvaluationMode](#) and the `CUPTI_METRIC_ATTR_EVALUATION_MODE` attribute.

## Values

**CUPTI\_METRIC\_EVALUATION\_MODE\_PER\_INSTANCE = 1**

If this bit is set, the metric can be profiled for each instance of the domain. The event values passed to [cuprtiMetricGetValue](#) can contain values for one instance of the domain. And [cuprtiMetricGetValue](#) can be called for each instance.

**CUPTI\_METRIC\_EVALUATION\_MODE\_AGGREGATE = 1<<1**

If this bit is set, the metric can be profiled over all instances. The event values passed to [cuprtiMetricGetValue](#) can be aggregated values of events for all instances of the domain.

**CUPTI\_METRIC\_EVALUATION\_MODE\_FORCE\_INT = 0x7fffffff**

## enum CUpti\_MetricPropertyDeviceClass

Device class.

Enumeration of device classes for metric property

`CUPTI_METRIC_PROPERTY_DEVICE_CLASS`.

## Values

**CUPTI\_METRIC\_PROPERTY\_DEVICE\_CLASS\_TESLA = 0**

**CUPTI\_METRIC\_PROPERTY\_DEVICE\_CLASS\_QUADRO = 1**

**CUPTI\_METRIC\_PROPERTY\_DEVICE\_CLASS\_GEFORCE = 2**

## enum CUpti\_MetricPropertyID

Metric device properties.

Metric device properties describe device properties which are needed for a metric. Some of these properties can be collected using `cuDeviceGetAttribute`.

### Values

**CUPTI\_METRIC\_PROPERTY\_MULTIPROCESSOR\_COUNT**  
**CUPTI\_METRIC\_PROPERTY\_WARPS\_PER\_MULTIPROCESSOR**  
**CUPTI\_METRIC\_PROPERTY\_KERNEL\_GPU\_TIME**  
**CUPTI\_METRIC\_PROPERTY\_CLOCK\_RATE**  
**CUPTI\_METRIC\_PROPERTY\_FRAME\_BUFFER\_COUNT**  
**CUPTI\_METRIC\_PROPERTY\_GLOBAL\_MEMORY\_BANDWIDTH**  
**CUPTI\_METRIC\_PROPERTY\_PCIE\_LINK\_RATE**  
**CUPTI\_METRIC\_PROPERTY\_PCIE\_LINK\_WIDTH**  
**CUPTI\_METRIC\_PROPERTY\_PCIE\_GEN**  
**CUPTI\_METRIC\_PROPERTY\_DEVICE\_CLASS**

## enum CUpti\_MetricValueKind

Kinds of metric values.

Metric values can be one of several different kinds. Corresponding to each kind is a member of the `CUpti_MetricValue` union. The metric value returned by `cuptiMetricGetValue` should be accessed using the appropriate member of that union based on its value kind.

### Values

**CUPTI\_METRIC\_VALUE\_KIND\_DOUBLE = 0**

The metric value is a 64-bit double.

**CUPTI\_METRIC\_VALUE\_KIND\_UINT64 = 1**

The metric value is a 64-bit unsigned integer.

**CUPTI\_METRIC\_VALUE\_KIND\_PERCENT = 2**

The metric value is a percentage represented by a 64-bit double. For example, 57.5% is represented by the value 57.5.

**CUPTI\_METRIC\_VALUE\_KIND\_THROUGHPUT = 3**

The metric value is a throughput represented by a 64-bit integer. The unit for throughput values is bytes/second.

**CUPTI\_METRIC\_VALUE\_KIND\_INT64 = 4**

The metric value is a 64-bit signed integer.

**CUPTI\_METRIC\_VALUE\_KIND\_UTILIZATION\_LEVEL = 5**

The metric value is a utilization level, as represented by

`CUpti_MetricValueUtilizationLevel`.

`CUPTI_METRIC_VALUE_KIND_FORCE_INT = 0x7fffffff`

## enum `CUpti_MetricValueUtilizationLevel`

Enumeration of utilization levels for metrics values of kind

`CUPTI_METRIC_VALUE_KIND_UTILIZATION_LEVEL`. Utilization values can vary from IDLE (0) to MAX (10) but the enumeration only provides specific names for a few values.

### Values

`CUPTI_METRIC_VALUE_UTILIZATION_IDLE = 0`

`CUPTI_METRIC_VALUE_UTILIZATION_LOW = 2`

`CUPTI_METRIC_VALUE_UTILIZATION_MID = 5`

`CUPTI_METRIC_VALUE_UTILIZATION_HIGH = 8`

`CUPTI_METRIC_VALUE_UTILIZATION_MAX = 10`

`CUPTI_METRIC_VALUE_UTILIZATION_FORCE_INT = 0x7fffffff`

## typedef `uint32_t CUpti_MetricID`

ID for a metric.

A metric provides a measure of some aspect of the device.

## `CUptiResult cuptiDeviceEnumMetrics (CUdevice device, size_t *arraySizeBytes, CUpti_MetricID *metricArray)`

Get the metrics for a device.

### Parameters

#### **device**

The CUDA device

#### **arraySizeBytes**

The size of `metricArray` in bytes, and returns the number of bytes written to `metricArray`

#### **metricArray**

Returns the IDs of the metrics for the device

### Returns

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_INVALID_DEVICE`

- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if `arraySizeBytes` or `metricArray` are NULL

### Description

Returns the metric IDs in `metricArray` for a device. The size of the `metricArray` buffer is given by `*arraySizeBytes`. The size of the `metricArray` buffer must be at least `numMetrics * sizeof(CUpti_MetricID)` or else all metric IDs will not be returned. The value returned in `*arraySizeBytes` contains the number of bytes returned in `metricArray`.

## CUptiResult cuptiDeviceGetNumMetrics (CUdevice device, uint32\_t \*numMetrics)

Get the number of metrics for a device.

### Parameters

#### device

The CUDA device

#### numMetrics

Returns the number of metrics available for the device

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_DEVICE
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER  
if `numMetrics` is NULL

### Description

Returns the number of metrics available for a device.

## CUptiResult cuptiEnumMetrics (size\_t \*arraySizeBytes, CUpti\_MetricID \*metricArray)

Get all the metrics available on any device.

### Parameters

#### arraySizeBytes

The size of `metricArray` in bytes, and returns the number of bytes written to `metricArray`

#### metricArray

Returns the IDs of the metrics

### Returns

- ▶ CUPTI\_SUCCESS
  - ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
- if `arraySizeBytes` or `metricArray` are NULL

### Description

Returns the metric IDs in `metricArray` for all CUDA-capable devices. The size of the `metricArray` buffer is given by `*arraySizeBytes`. The size of the `metricArray` buffer must be at least `numMetrics * sizeof(CUpti_MetricID)` or all metric IDs will not be returned. The value returned in `*arraySizeBytes` contains the number of bytes returned in `metricArray`.

## CUptiResult cuptiGetNumMetrics (uint32\_t \*numMetrics)

Get the total number of metrics available on any device.

### Parameters

#### numMetrics

Returns the number of metrics

### Returns

- ▶ CUPTI\_SUCCESS
  - ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
- if `numMetrics` is NULL

### Description

Returns the total number of metrics available on any CUDA-capable devices.

## CUptiResult cuptiMetricCreateEventGroupSets (CUcontext context, size\_t metricIdArraySizeBytes, CUpti\_MetricID \*metricIdArray, CUpti\_EventGroupSets \*\*eventGroupPasses)

For a set of metrics, get the grouping that indicates the number of passes and the event groups necessary to collect the events required for those metrics.

### Parameters

#### context

The context for event collection

#### metricIdArraySizeBytes

Size of the metricIdArray in bytes

#### metricIdArray

Array of metric IDs

#### eventGroupPasses

Returns a [CUpti\\_EventGroupSets](#) object that indicates the number of passes required to collect the events and the events to collect on each pass

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_CONTEXT
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if metricIdArray or eventGroupPasses is NULL

### Description

For a set of metrics, get the grouping that indicates the number of passes and the event groups necessary to collect the events required for those metrics.

### See also:

[cuptiEventGroupSetsCreate](#) for details on event group set creation.

## CUptiResult cuptiMetricEnumEvents (CUpti\_MetricID metric, size\_t \*eventIdArraySizeBytes, CUpti\_EventID \*eventIdArray)

Get the events required to calculating a metric.

### Parameters

#### **metric**

ID of the metric

#### **eventIdArraySizeBytes**

The size of `eventIdArray` in bytes, and returns the number of bytes written to `eventIdArray`

#### **eventIdArray**

Returns the IDs of the events required to calculate `metric`

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `eventIdArraySizeBytes` or `eventIdArray` are NULL.

### Description

Gets the event IDs in `eventIdArray` required to calculate a `metric`. The size of the `eventIdArray` buffer is given by `*eventIdArraySizeBytes` and must be at least `numEvents * sizeof(CUpti_EventID)` or all events will not be returned. The value returned in `*eventIdArraySizeBytes` contains the number of bytes returned in `eventIdArray`.

## CUptiResult cuptiMetricEnumProperties (CUpti\_MetricID metric, size\_t \*propIdArraySizeBytes, CUpti\_MetricPropertyID \*propIdArray)

Get the properties required to calculating a metric.

### Parameters

#### **metric**

ID of the metric

**propIdArraySizeBytes**

The size of `propIdArray` in bytes, and returns the number of bytes written to `propIdArray`

**propIdArray**

Returns the IDs of the properties required to calculate `metric`

**Returns**

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `propIdArraySizeBytes` or `propIdArray` are NULL.

**Description**

Gets the property IDs in `propIdArray` required to calculate a `metric`. The size of the `propIdArray` buffer is given by `*propIdArraySizeBytes` and must be at least `numProp * sizeof(CUpti_DeviceAttribute)` or all properties will not be returned. The value returned in `*propIdArraySizeBytes` contains the number of bytes returned in `propIdArray`.

## CUptiResult cuptiMetricGetAttribute (CUpti\_MetricID metric, CUpti\_MetricAttribute attrib, size\_t \*valueSize, void \*value)

Get a metric attribute.

**Parameters****metric**

ID of the metric

**attrib**

The metric attribute to read

**valueSize**

The size of the `value` buffer in bytes, and returns the number of bytes written to `value`

**value**

Returns the attribute's value

**Returns**

- ▶ CUPTI\_SUCCESS

- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER
  - if `valueSize` or `value` is NULL, or if `attrib` is not a metric attribute
- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT
  - For non-c-string attribute values, indicates that the `value` buffer is too small to hold the attribute value.

### Description

Returns a metric attribute in `*value`. The size of the `value` buffer is given by `*valueSize`. The value returned in `*valueSize` contains the number of bytes returned in `value`.

If the attribute value is a c-string that is longer than `*valueSize`, then only the first `*valueSize` characters will be returned and there will be no terminating null byte.

## CuptiResult cuptiMetricGetIdFromName (CUdevice device, const char \*metricName, CUpti\_MetricID \*metric)

Find an metric by name.

### Parameters

#### **device**

The CUDA device

#### **metricName**

The name of metric to find

#### **metric**

Returns the ID of the found metric or undefined if unable to find the metric

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_DEVICE
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_NAME
  - if unable to find a metric with name `metricName`. In this case `*metric` is undefined
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `metricName` or `metric` are NULL.

### Description

Find a metric by name and return the metric ID in `*metric`.

## CUptiResult cuptiMetricGetNumEvents (CUpti\_MetricID metric, uint32\_t \*numEvents)

Get number of events required to calculate a metric.

### Parameters

#### **metric**

ID of the metric

#### **numEvents**

Returns the number of events required for the metric

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if `numEvents` is NULL

### Description

Returns the number of events in `numEvents` that are required to calculate a metric.

## CUptiResult cuptiMetricGetNumProperties (CUpti\_MetricID metric, uint32\_t \*numProp)

Get number of properties required to calculate a metric.

### Parameters

#### **metric**

ID of the metric

#### **numProp**

Returns the number of properties required for the metric

### Returns

- ▶ CUPTI\_SUCCESS

- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_ID
- ▶ CUPTI\_ERROR\_INVALID\_PARAMETER

if numProp is NULL

### Description

Returns the number of properties in numProp that are required to calculate a metric.

## CUptiResult cuptiMetricGetRequiredEventGroupSets (CUcontext context, CUpti\_MetricID metric, CUpti\_EventGroupSets \*\*eventGroupSets)

For a metric get the groups of events that must be collected in the same pass.

### Parameters

#### context

The context for event collection

#### metric

The metric ID

#### eventGroupSets

Returns a CUpti\_EventGroupSets object that indicates the events that must be collected in the same pass to ensure the metric is calculated correctly. Returns NULL if no grouping is required for metric

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_ID

### Description

For a metric get the groups of events that must be collected in the same pass to ensure that the metric is calculated correctly. If the events are not collected as specified then the metric value may be inaccurate.

The function returns NULL if a metric does not have any required event group. In this case the events needed for the metric can be grouped in any manner for collection.

**CUptiResult cuptiMetricGetValue (CUdevice device, CUpti\_MetricID metric, size\_t eventIdArraySizeBytes, CUpti\_EventID \*eventIdArray, size\_t eventValueArraySizeBytes, uint64\_t \*eventValueArray, uint64\_t timeDuration, CUpti\_MetricValue \*metricValue)**

Calculate the value for a metric.

### Parameters

#### **device**

The CUDA device that the metric is being calculated for

#### **metric**

The metric ID

#### **eventIdArraySizeBytes**

The size of `eventIdArray` in bytes

#### **eventIdArray**

The event IDs required to calculate `metric`

#### **eventValueArraySizeBytes**

The size of `eventValueArray` in bytes

#### **eventValueArray**

The normalized event values required to calculate `metric`. The values must be order to match the order of events in `eventIdArray`

#### **timeDuration**

The duration over which the events were collected, in ns

#### **metricValue**

Returns the value for the metric

### Returns

- ▶ CUPTI\_SUCCESS
- ▶ CUPTI\_ERROR\_NOT\_INITIALIZED
- ▶ CUPTI\_ERROR\_INVALID\_METRIC\_ID
- ▶ CUPTI\_ERROR\_INVALID\_OPERATION
- ▶ CUPTI\_ERROR\_PARAMETER\_SIZE\_NOT\_SUFFICIENT
  - if the `eventIdArray` does not contain all the events needed for `metric`
- ▶ CUPTI\_ERROR\_INVALID\_EVENT\_VALUE
  - if any of the event values required for the metric is CUPTI\_EVENT\_OVERFLOW

► CUPTI\_ERROR\_INVALID\_METRIC\_VALUE

if the computed metric value cannot be represented in the metric's value type. For example, if the metric value type is unsigned and the computed metric value is negative

► CUPTI\_ERROR\_INVALID\_PARAMETER

if `metricValue`, `eventIdArray` or `eventValueArray` is NULL

## Description

Use the events collected for a metric to calculate the metric value. Metric value evaluation depends on the evaluation mode `CUpti_MetricEvaluationMode` that the metric supports. If a metric has evaluation mode as `CUPTI_METRIC_EVALUATION_MODE_PER_INSTANCE`, then it assumes that the input event value is for one domain instance. If a metric has evaluation mode as `CUPTI_METRIC_EVALUATION_MODE_AGGREGATE`, it assumes that input event values are normalized to represent all domain instances on a device. For the most accurate metric collection, the events required for the metric should be collected for all profiled domain instances. For example, to collect all instances of an event, set the `CUPTI_EVENT_GROUP_ATTR_PROFILE_ALL_DOMAIN_INSTANCES` attribute on the group containing the event to 1. The normalized value for the event is then:  $(\text{sum\_event\_values} * \text{totalInstanceCount}) / \text{instanceCount}$ , where `sum_event_values` is the summation of the event values across all profiled domain instances, `totalInstanceCount` is obtained from querying `CUPTI_EVENT_DOMAIN_ATTR_TOTAL_INSTANCE_COUNT` and `instanceCount` is obtained from querying `CUPTI_EVENT_GROUP_ATTR_INSTANCE_COUNT` (or `CUPTI_EVENT_DOMAIN_ATTR_INSTANCE_COUNT`).

**CUptiResult cuptiMetricGetValue2 (CUpti\_MetricID metric, size\_t eventIdArraySizeBytes, CUpti\_EventID \*eventIdArray, size\_t eventValueArraySizeBytes, uint64\_t \*eventValueArray, size\_t propIdArraySizeBytes, CUpti\_MetricPropertyID \*propIdArray, size\_t propValueArraySizeBytes, uint64\_t \*propValueArray, CUpti\_MetricValue \*metricValue)**

Calculate the value for a metric.

## Parameters

**metric**

The metric ID

**eventIdArraySizeBytes**

The size of `eventIdArray` in bytes

**eventIdArray**

The event IDs required to calculate `metric`

**eventValueArraySizeBytes**

The size of `eventValueArray` in bytes

**eventValueArray**

The normalized event values required to calculate `metric`. The values must be order to match the order of events in `eventIdArray`

**propIdArraySizeBytes**

The size of `propIdArray` in bytes

**propIdArray**

The metric property IDs required to calculate `metric`

**propValueArraySizeBytes**

The size of `propValueArray` in bytes

**propValueArray**

The metric property values required to calculate `metric`. The values must be order to match the order of metric properties in `propIdArray`

**metricValue**

Returns the value for the metric

**Returns**

- ▶ `CUPTI_SUCCESS`
- ▶ `CUPTI_ERROR_NOT_INITIALIZED`
- ▶ `CUPTI_ERROR_INVALID_METRIC_ID`
- ▶ `CUPTI_ERROR_INVALID_OPERATION`
- ▶ `CUPTI_ERROR_PARAMETER_SIZE_NOT_SUFFICIENT`
  - if the `eventIdArray` does not contain all the events needed for `metric`
- ▶ `CUPTI_ERROR_INVALID_EVENT_VALUE`
  - if any of the event values required for the metric is `CUPTI_EVENT_OVERFLOW`
- ▶ `CUPTI_ERROR_NOT_COMPATIBLE`
  - if the computed metric value cannot be represented in the metric's value type. For example, if the metric value type is unsigned and the computed metric value is negative
- ▶ `CUPTI_ERROR_INVALID_PARAMETER`
  - if `metricValue`, `eventIdArray` or `eventValueArray` is `NULL`

## Description

Use the events and properties collected for a metric to calculate the metric value. Metric value evaluation depends on the evaluation mode `CUpti_MetricEvaluationMode` that the metric supports. If a metric has evaluation mode as `CUPTI_METRIC_EVALUATION_MODE_PER_INSTANCE`, then it assumes that the input event value is for one domain instance. If a metric has evaluation mode as `CUPTI_METRIC_EVALUATION_MODE_AGGREGATE`, it assumes that input event values are normalized to represent all domain instances on a device. For the most accurate metric collection, the events required for the metric should be collected for all profiled domain instances. For example, to collect all instances of an event, set the `CUPTI_EVENT_GROUP_ATTR_PROFILE_ALL_DOMAIN_INSTANCES` attribute on the group containing the event to 1. The normalized value for the event is then:  $(\text{sum\_event\_values} * \text{totalInstanceCount}) / \text{instanceCount}$ , where `sum_event_values` is the summation of the event values across all profiled domain instances, `totalInstanceCount` is obtained from querying `CUPTI_EVENT_DOMAIN_ATTR_TOTAL_INSTANCE_COUNT` and `instanceCount` is obtained from querying `CUPTI_EVENT_GROUP_ATTR_INSTANCE_COUNT` (or `CUPTI_EVENT_DOMAIN_ATTR_INSTANCE_COUNT`).

# Chapter 3.

## DATA STRUCTURES

Here are the data structures with brief descriptions:

### **CUpti\_Activity**

The base activity record

### **CUpti\_ActivityAPI**

The activity record for a driver or runtime API invocation

### **CUpti\_ActivityBranch**

The activity record for source level result branch

### **CUpti\_ActivityCdpKernel**

The activity record for CDP (CUDA Dynamic Parallelism) kernel

### **CUpti\_ActivityContext**

The activity record for a context

### **CUpti\_ActivityDevice**

The activity record for a device

### **CUpti\_ActivityDeviceAttribute**

The activity record for a device attribute

### **CUpti\_ActivityEnvironment**

The activity record for CUPTI environmental data

### **CUpti\_ActivityEvent**

The activity record for a CUPTI event

### **CUpti\_ActivityEventInstance**

The activity record for a CUPTI event with instance information

### **CUpti\_ActivityFunction**

The activity record for global/device functions

### **CUpti\_ActivityGlobalAccess**

The activity record for source-level global access

### **CUpti\_ActivityInstructionExecution**

The activity record for source-level sass/source line-by-line correlation

### **CUpti\_ActivityKernel**

The activity record for kernel. (deprecated)

**CUpti\_ActivityKernel2**

The activity record for a kernel (CUDA 5.5 onwards)

**CUpti\_ActivityMarker**

The activity record providing a marker which is an instantaneous point in time

**CUpti\_ActivityMarkerData**

The activity record providing detailed information for a marker

**CUpti\_ActivityMemcpy**

The activity record for memory copies

**CUpti\_ActivityMemcpy2**

The activity record for peer-to-peer memory copies

**CUpti\_ActivityMemset**

The activity record for memset

**CUpti\_ActivityMetric**

The activity record for a CUPTI metric

**CUpti\_ActivityMetricInstance**

The activity record for a CUPTI metric with instance information. This activity record represents a CUPTI metric value for a specific metric domain instance (CUPTI\_ACTIVITY\_KIND\_METRIC\_INSTANCE). This activity record kind is not produced by the activity API but is included for completeness and ease-of-use. Profile frameworks built on top of CUPTI that collect metric data may choose to use this type to store the collected metric data. This activity record should be used when metric domain instance information needs to be associated with the metric

**CUpti\_ActivityModule**

The activity record for a CUDA module

**CUpti\_ActivityName**

The activity record providing a name

**CUpti\_ActivityObjectKindId**

Identifiers for object kinds as specified by CUpti\_ActivityObjectKind

**CUpti\_ActivityOverhead**

The activity record for CUPTI and driver overheads

**CUpti\_ActivityPreemption**

The activity record for a preemption of a CDP kernel

**CUpti\_ActivitySharedAccess**

The activity record for source-level shared access

**CUpti\_ActivitySourceLocator**

The activity record for source locator

**CUpti\_ActivityUnifiedMemoryCounter**

The activity record for Unified Memory counters

**CUpti\_ActivityUnifiedMemoryCounterConfig**

Unified Memory counters configuration structure

**CUpti\_CallbackData**

Data passed into a runtime or driver API callback function

**CUpti\_EventGroupSet**

A set of event groups

**CUpti\_EventGroupSets**

A set of event group sets

**CUpti\_MetricValue**

A metric value

**CUpti\_ModuleResourceData**

Module data passed into a resource callback function

**CUpti\_NvtxData**

Data passed into a NVTX callback function

**CUpti\_ResourceData**

Data passed into a resource callback function

**CUpti\_SynchronizeData**

Data passed into a synchronize callback function

## 3.1. CUpti\_Activity Struct Reference

The base activity record.

The activity API uses a [CUpti\\_Activity](#) as a generic representation for any activity.

The 'kind' field is used to determine the specific activity kind, and from that the [CUpti\\_Activity](#) object can be cast to the specific activity record type appropriate for that kind.

Note that all activity record types are padded and aligned to ensure that each member of the record is naturally aligned.

**See also:**

[CUpti\\_ActivityKind](#)

### CUpti\_ActivityKind CUpti\_Activity::kind

The kind of this activity.

## 3.2. CUpti\_ActivityAPI Struct Reference

The activity record for a driver or runtime API invocation.

This activity record represents an invocation of a driver or runtime API (CUPTI\_ACTIVITY\_KIND\_DRIVER and CUPTI\_ACTIVITY\_KIND\_RUNTIME).

### CUpti\_CallbackId CUpti\_ActivityAPI::cbid

The ID of the driver or runtime function.

## `uint32_t CUpti_ActivityAPI::correlationId`

The correlation ID of the driver or runtime CUDA function. Each function invocation is assigned a unique correlation ID that is identical to the correlation ID in the memcopy, memset, or kernel activity record that is associated with this function.

## `uint64_t CUpti_ActivityAPI::end`

The end timestamp for the function, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the function.

## `CUpti_ActivityKind CUpti_ActivityAPI::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_DRIVER` or `CUPTI_ACTIVITY_KIND_RUNTIME`.

## `uint32_t CUpti_ActivityAPI::processId`

The ID of the process where the driver or runtime CUDA function is executing.

## `uint32_t CUpti_ActivityAPI::returnValue`

The return value for the function. For a CUDA driver function with will be a `CUresult` value, and for a CUDA runtime function this will be a `cudaError_t` value.

## `uint64_t CUpti_ActivityAPI::start`

The start timestamp for the function, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the function.

## `uint32_t CUpti_ActivityAPI::threadId`

The ID of the thread where the driver or runtime CUDA function is executing.

## 3.3. `CUpti_ActivityBranch` Struct Reference

The activity record for source level result branch.

This activity record the locations of the branches in the source (`CUPTI_ACTIVITY_KIND_BRANCH`).

## `uint32_t CUpti_ActivityBranch::correlationId`

The correlation ID of the kernel to which this result is associated.

## `uint32_t CUpti_ActivityBranch::diverged`

Number of times this branch diverged

## `uint32_t CUpti_ActivityBranch::executed`

The number of times this branch was executed

## `CUpti_ActivityKind CUpti_ActivityBranch::kind`

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_BRANCH.

## `uint32_t CUpti_ActivityBranch::pcOffset`

The pc offset for the branch.

## `uint32_t CUpti_ActivityBranch::sourceLocatorId`

The ID for source locator.

## `uint64_t CUpti_ActivityBranch::threadsExecuted`

This increments each time when this instruction is executed by number of threads that executed this instruction

## 3.4. CUpti\_ActivityCdpKernel Struct Reference

The activity record for CDP (CUDA Dynamic Parallelism) kernel.

This activity record represents a CDP kernel execution.

## `int32_t CUpti_ActivityCdpKernel::blockX`

The X-dimension block size for the kernel.

## `int32_t CUpti_ActivityCdpKernel::blockY`

The Y-dimension block size for the kernel.

## `int32_t CUpti_ActivityCdpKernel::blockZ`

The Z-dimension grid size for the kernel.

## `uint64_t CUpti_ActivityCdpKernel::completed`

The timestamp when kernel is marked as completed, in ns. A value of `CUPTI_TIMESTAMP_UNKNOWN` indicates that the completion time is unknown.

## `uint32_t CUpti_ActivityCdpKernel::contextId`

The ID of the context where the kernel is executing.

## `uint32_t CUpti_ActivityCdpKernel::correlationId`

The correlation ID of the kernel. Each kernel execution is assigned a unique correlation ID that is identical to the correlation ID in the driver API activity record that launched the kernel.

## `uint32_t CUpti_ActivityCdpKernel::deviceId`

The ID of the device where the kernel is executing.

## `int32_t`

## `CUpti_ActivityCdpKernel::dynamicSharedMemory`

The dynamic shared memory reserved for the kernel, in bytes.

## `uint64_t CUpti_ActivityCdpKernel::end`

The end timestamp for the kernel execution, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the kernel.

## `uint8_t CUpti_ActivityCdpKernel::executed`

The cache configuration used for the kernel. The value is one of the `CUfunc_cache` enumeration values from `cuda.h`.

## `int64_t CUpti_ActivityCdpKernel::gridId`

The grid ID of the kernel. Each kernel execution is assigned a unique grid ID.

## `int32_t CUpti_ActivityCdpKernel::gridX`

The X-dimension grid size for the kernel.

**int32\_t CUpti\_ActivityCdpKernel::gridY**

The Y-dimension grid size for the kernel.

**int32\_t CUpti\_ActivityCdpKernel::gridZ**

The Z-dimension grid size for the kernel.

**CUpti\_ActivityKind CUpti\_ActivityCdpKernel::kind**

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_CDP\_KERNEL

**uint32\_t**

**CUpti\_ActivityCdpKernel::localMemoryPerThread**

The amount of local memory reserved for each thread, in bytes.

**uint32\_t CUpti\_ActivityCdpKernel::localMemoryTotal**

The total amount of local memory reserved for the kernel, in bytes.

**const char \*CUpti\_ActivityCdpKernel::name**

The name of the kernel. This name is shared across all activity records representing the same kernel, and so should not be modified.

**uint32\_t CUpti\_ActivityCdpKernel::parentBlockX**

The X-dimension of the parent block.

**uint32\_t CUpti\_ActivityCdpKernel::parentBlockY**

The Y-dimension of the parent block.

**uint32\_t CUpti\_ActivityCdpKernel::parentBlockZ**

The Z-dimension of the parent block.

**int64\_t CUpti\_ActivityCdpKernel::parentGridId**

The grid ID of the parent kernel.

## uint64\_t CUpti\_ActivityCdpKernel::queued

The timestamp when kernel is queued up, in ns. A value of CUPTI\_TIMESTAMP\_UNKNOWN indicates that the queued time is unknown.

## uint16\_t CUpti\_ActivityCdpKernel::registersPerThread

The number of registers required for each thread executing the kernel.

## uint8\_t CUpti\_ActivityCdpKernel::requested

The cache configuration requested by the kernel. The value is one of the CUfunc\_cache enumeration values from cuda.h.

## uint8\_t CUpti\_ActivityCdpKernel::sharedMemoryConfig

The shared memory configuration used for the kernel. The value is one of the CUsharedconfig enumeration values from cuda.h.

## uint64\_t CUpti\_ActivityCdpKernel::start

The start timestamp for the kernel execution, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the kernel.

## int32\_t CUpti\_ActivityCdpKernel::staticSharedMemory

The static shared memory allocated for the kernel, in bytes.

## uint32\_t CUpti\_ActivityCdpKernel::streamId

The ID of the stream where the kernel is executing.

## uint64\_t CUpti\_ActivityCdpKernel::submitted

The timestamp when kernel is submitted to the gpu, in ns. A value of CUPTI\_TIMESTAMP\_UNKNOWN indicates that the submission time is unknown.

## 3.5. CUpti\_ActivityContext Struct Reference

The activity record for a context.

This activity record represents information about a context (CUPTI\_ACTIVITY\_KIND\_CONTEXT).

## `uint16_t CUpti_ActivityContext::computeApiKind`

The compute API kind.

**See also:**

`CUpti_ActivityComputeApiKind`

## `uint32_t CUpti_ActivityContext::contextId`

The context ID.

## `uint32_t CUpti_ActivityContext::deviceId`

The device ID.

## `CUpti_ActivityKind CUpti_ActivityContext::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_CONTEXT`.

## `uint16_t CUpti_ActivityContext::nullStreamId`

The ID for the NULL stream in this context

## 3.6. `CUpti_ActivityDevice` Struct Reference

The activity record for a device.

This activity record represents information about a GPU device (`CUPTI_ACTIVITY_KIND_DEVICE`).

## `uint32_t CUpti_ActivityDevice::computeCapabilityMajor`

Compute capability for the device, major number.

## `uint32_t CUpti_ActivityDevice::computeCapabilityMinor`

Compute capability for the device, minor number.

## `uint32_t CUpti_ActivityDevice::constantMemorySize`

The amount of constant memory on the device, in bytes.

## `uint32_t CUpti_ActivityDevice::coreClockRate`

The core clock rate of the device, in kHz.

## `CUpti_ActivityFlag CUpti_ActivityDevice::flags`

The flags associated with the device.

**See also:**

`CUpti_ActivityFlag`

## `uint64_t CUpti_ActivityDevice::globalMemoryBandwidth`

The global memory bandwidth available on the device, in kBytes/sec.

## `uint64_t CUpti_ActivityDevice::globalMemorySize`

The amount of global memory on the device, in bytes.

## `uint32_t CUpti_ActivityDevice::id`

The device ID.

## `CUpti_ActivityKind CUpti_ActivityDevice::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_DEVICE`.

## `uint32_t CUpti_ActivityDevice::l2CacheSize`

The size of the L2 cache on the device, in bytes.

## `uint32_t CUpti_ActivityDevice::maxBlockDimX`

Maximum allowed X dimension for a block.

## `uint32_t CUpti_ActivityDevice::maxBlockDimY`

Maximum allowed Y dimension for a block.

## `uint32_t CUpti_ActivityDevice::maxBlockDimZ`

Maximum allowed Z dimension for a block.

**uint32\_t**  
**CUpti\_ActivityDevice::maxBlocksPerMultiprocessor**

Maximum number of blocks that can be present on a multiprocessor at any given time.

**uint32\_t CUpti\_ActivityDevice::maxGridDimX**

Maximum allowed X dimension for a grid.

**uint32\_t CUpti\_ActivityDevice::maxGridDimY**

Maximum allowed Y dimension for a grid.

**uint32\_t CUpti\_ActivityDevice::maxGridDimZ**

Maximum allowed Z dimension for a grid.

**uint32\_t CUpti\_ActivityDevice::maxIPC**

The maximum "instructions per cycle" possible on each device multiprocessor.

**uint32\_t CUpti\_ActivityDevice::maxRegistersPerBlock**

Maximum number of registers that can be allocated to a block.

**uint32\_t**  
**CUpti\_ActivityDevice::maxSharedMemoryPerBlock**

Maximum amount of shared memory that can be assigned to a block, in bytes.

**uint32\_t CUpti\_ActivityDevice::maxThreadsPerBlock**

Maximum number of threads allowed in a block.

**uint32\_t**  
**CUpti\_ActivityDevice::maxWarpsPerMultiprocessor**

Maximum number of warps that can be present on a multiprocessor at any given time.

**const char \*CUpti\_ActivityDevice::name**

The device name. This name is shared across all activity records representing instances of the device, and so should not be modified.

## `uint32_t CUpti_ActivityDevice::numMemcpyEngines`

Number of memory copy engines on the device.

## `uint32_t CUpti_ActivityDevice::numMultiprocessors`

Number of multiprocessors on the device.

## `uint32_t CUpti_ActivityDevice::numThreadsPerWarp`

The number of threads per warp on the device.

## 3.7. CUpti\_ActivityDeviceAttribute Struct Reference

The activity record for a device attribute.

This activity record represents information about a GPU device: either a CUpti\_DeviceAttribute or CUdevice\_attribute value (CUPTI\_ACTIVITY\_KIND\_DEVICE\_ATTRIBUTE).

### `CUpti_ActivityDeviceAttribute::@6` `CUpti_ActivityDeviceAttribute::attribute`

The attribute, either a CUpti\_DeviceAttribute or CUdevice\_attribute. Flag CUPTI\_ACTIVITY\_FLAG\_DEVICE\_ATTRIBUTE\_CUDEVICE is used to indicate what kind of attribute this is. If CUPTI\_ACTIVITY\_FLAG\_DEVICE\_ATTRIBUTE\_CUDEVICE is 1 then CUdevice\_attribute field is value, otherwise CUpti\_DeviceAttribute field is valid.

## `uint32_t CUpti_ActivityDeviceAttribute::deviceId`

The ID of the device that this attribute applies to.

## `CUpti_ActivityFlag CUpti_ActivityDeviceAttribute::flags`

The flags associated with the device.

**See also:**

[`CUpti\_ActivityFlag`](#)

## CUpti\_ActivityKind CUpti\_ActivityDeviceAttribute::kind

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_DEVICE\_ATTRIBUTE.

## CUpti\_ActivityDeviceAttribute::@7 CUpti\_ActivityDeviceAttribute::value

The value for the attribute. See CUpti\_DeviceAttribute and CUdevice\_attribute for the type of the value for a given attribute.

## 3.8. CUpti\_ActivityEnvironment Struct Reference

The activity record for CUPTI environmental data.

This activity record provides CUPTI environmental data, include power, clocks, and thermals. This information is sampled at various rates and returned in this activity record. The consumer of the record needs to check the environmentKind field to figure out what kind of environmental record this is.

## CUpti\_EnvironmentClocksThrottleReason CUpti\_ActivityEnvironment::clocksThrottleReasons

The clocks throttle reasons.

## CUpti\_ActivityEnvironment::@8::@12 CUpti\_ActivityEnvironment::cooling

Data returned for CUPTI\_ACTIVITY\_ENVIRONMENT\_COOLING environment kind.

## uint32\_t CUpti\_ActivityEnvironment::deviceId

The ID of the device

## CUpti\_ActivityEnvironmentKind CUpti\_ActivityEnvironment::environmentKind

The kind of data reported in this record.

## uint32\_t CUpti\_ActivityEnvironment::fanSpeed

The fan speed as percentage of maximum.

## `uint32_t CUpti_ActivityEnvironment::gpuTemperature`

The GPU temperature in degrees C.

## `CUpti_ActivityKind CUpti_ActivityEnvironment::kind`

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_ENVIRONMENT.

## `uint32_t CUpti_ActivityEnvironment::memoryClock`

The memory frequency in MHz

## `uint32_t CUpti_ActivityEnvironment::pcieLinkGen`

The PCIe link generation.

## `uint32_t CUpti_ActivityEnvironment::pcieLinkWidth`

The PCIe link width.

## `CUpti_ActivityEnvironment::@8::@11`

## `CUpti_ActivityEnvironment::power`

Data returned for CUPTI\_ACTIVITY\_ENVIRONMENT\_POWER environment kind.

## `uint32_t CUpti_ActivityEnvironment::power`

The power in milliwatts consumed by GPU and associated circuitry.

## `uint32_t CUpti_ActivityEnvironment::powerLimit`

The power in milliwatts that will trigger power management algorithm.

## `uint32_t CUpti_ActivityEnvironment::smClock`

The SM frequency in MHz

## `CUpti_ActivityEnvironment::@8::@9`

## `CUpti_ActivityEnvironment::speed`

Data returned for CUPTI\_ACTIVITY\_ENVIRONMENT\_SPEED environment kind.

## CUpti\_ActivityEnvironment::@8::@10 CUpti\_ActivityEnvironment::temperature

Data returned for CUPTI\_ACTIVITY\_ENVIRONMENT\_TEMPERATURE environment kind.

## uint64\_t CUpti\_ActivityEnvironment::timestamp

The timestamp when this sample was retrieved, in ns. A value of 0 indicates that timestamp information could not be collected for the marker.

## 3.9. CUpti\_ActivityEvent Struct Reference

The activity record for a CUPTI event.

This activity record represents a CUPTI event value (CUPTI\_ACTIVITY\_KIND\_EVENT). This activity record kind is not produced by the activity API but is included for completeness and ease-of-use. Profile frameworks built on top of CUPTI that collect event data may choose to use this type to store the collected event data.

## uint32\_t CUpti\_ActivityEvent::correlationId

The correlation ID of the event. Use of this ID is user-defined, but typically this ID value will equal the correlation ID of the kernel for which the event was gathered.

## CUpti\_EventDomainID CUpti\_ActivityEvent::domain

The event domain ID.

## CUpti\_EventID CUpti\_ActivityEvent::id

The event ID.

## CUpti\_ActivityKind CUpti\_ActivityEvent::kind

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_EVENT.

## uint64\_t CUpti\_ActivityEvent::value

The event value.

## 3.10. CUpti\_ActivityEventInstance Struct Reference

The activity record for a CUPTI event with instance information.

This activity record represents the a CUPTI event value for a specific event domain instance (CUPTI\_ACTIVITY\_KIND\_EVENT\_INSTANCE). This activity record kind is not produced by the activity API but is included for completeness and ease-of-use. Profile frameworks built on top of CUPTI that collect event data may choose to use this type to store the collected event data. This activity record should be used when event domain instance information needs to be associated with the event.

### uint32\_t CUpti\_ActivityEventInstance::correlationId

The correlation ID of the event. Use of this ID is user-defined, but typically this ID value will equal the correlation ID of the kernel for which the event was gathered.

### CUpti\_EventDomainID CUpti\_ActivityEventInstance::domain

The event domain ID.

### CUpti\_EventID CUpti\_ActivityEventInstance::id

The event ID.

### uint32\_t CUpti\_ActivityEventInstance::instance

The event domain instance.

### CUpti\_ActivityKind CUpti\_ActivityEventInstance::kind

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_EVENT\_INSTANCE.

### uint32\_t CUpti\_ActivityEventInstance::pad

Undefined. Reserved for internal use.

### uint64\_t CUpti\_ActivityEventInstance::value

The event value.

### 3.11. CUpti\_ActivityFunction Struct Reference

The activity record for global/device functions.

This activity records function name and corresponding module information. (CUPTI\_ACTIVITY\_KIND\_FUNCTION).

**uint32\_t CUpti\_ActivityFunction::contextId**

The ID of the context where the function is launched.

**uint32\_t CUpti\_ActivityFunction::functionIndex**

The function's unique symbol index in the module.

**uint32\_t CUpti\_ActivityFunction::id**

ID to uniquely identify the record

**CUpti\_ActivityKind CUpti\_ActivityFunction::kind**

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_FUNCTION.

**uint32\_t CUpti\_ActivityFunction::moduleId**

The module ID in which this global/device function is present.

**const char \*CUpti\_ActivityFunction::name**

The name of the function. This name is shared across all activity records representing the same kernel, and so should not be modified.

### 3.12. CUpti\_ActivityGlobalAccess Struct Reference

The activity record for source-level global access.

This activity records the locations of the global accesses in the source (CUPTI\_ACTIVITY\_KIND\_GLOBAL\_ACCESS).

**uint32\_t CUpti\_ActivityGlobalAccess::correlationId**

The correlation ID of the kernel to which this result is associated.

## `uint32_t CUpti_ActivityGlobalAccess::executed`

The number of times this instruction was executed

## `CUpti_ActivityFlag CUpti_ActivityGlobalAccess::flags`

The properties of this global access.

## `CUpti_ActivityKind CUpti_ActivityGlobalAccess::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_GLOBAL_ACCESS`.

## `uint64_t CUpti_ActivityGlobalAccess::l2_transactions`

The total number of 32 bytes transactions to L2 cache generated by this access

## `uint32_t CUpti_ActivityGlobalAccess::pcOffset`

The pc offset for the access.

## `uint32_t CUpti_ActivityGlobalAccess::sourceLocatorId`

The ID for source locator.

## `uint64_t CUpti_ActivityGlobalAccess::threadsExecuted`

This increments each time when this instruction is executed by number of threads that executed this instruction with predicate and condition code evaluating to true.

## 3.13. `CUpti_ActivityInstructionExecution` Struct Reference

The activity record for source-level sass/source line-by-line correlation.

This activity records source level sass/source correlation information. (`CUPTI_ACTIVITY_KIND_INSTRUCTION_EXECUTION`).

### `uint32_t`

## `CUpti_ActivityInstructionExecution::correlationId`

The correlation ID of the kernel to which this result is associated.

## uint32\_t CUpti\_ActivityInstructionExecution::executed

The number of times this instruction was executed.

## CUpti\_ActivityFlag

## CUpti\_ActivityInstructionExecution::flags

The properties of this instruction execution.

## uint32\_t CUpti\_ActivityInstructionExecution::functionId

Correlation ID with global/device function name

## CUpti\_ActivityKind

## CUpti\_ActivityInstructionExecution::kind

The activity record kind, must be

CUPTI\_ACTIVITY\_KIND\_INSTRUCTION\_EXECUTION.

## uint64\_t

## CUpti\_ActivityInstructionExecution::notPredOffThreadsExecuted

This increments each time when this instruction is executed by number of threads that executed this instruction with predicate and condition code evaluating to true.

## uint32\_t CUpti\_ActivityInstructionExecution::pad

Undefined. Reserved for internal use.

## uint32\_t CUpti\_ActivityInstructionExecution::pcOffset

The pc offset for the instruction.

## uint32\_t

## CUpti\_ActivityInstructionExecution::sourceLocatorId

The ID for source locator.

## uint64\_t

## CUpti\_ActivityInstructionExecution::threadsExecuted

This increments each time when this instruction is executed by number of threads that executed this instruction, regardless of predicate or condition code.

## 3.14. CUpti\_ActivityKernel Struct Reference

The activity record for kernel. (deprecated).

This activity record represents a kernel execution (CUPTI\_ACTIVITY\_KIND\_KERNEL and CUPTI\_ACTIVITY\_KIND\_CONCURRENT\_KERNEL) but is no longer generated by CUPTI. Kernel activities are now reported using the [CUpti\\_ActivityKernel2](#) activity record.

### int32\_t CUpti\_ActivityKernel::blockX

The X-dimension block size for the kernel.

### int32\_t CUpti\_ActivityKernel::blockY

The Y-dimension block size for the kernel.

### int32\_t CUpti\_ActivityKernel::blockZ

The Z-dimension grid size for the kernel.

### uint8\_t CUpti\_ActivityKernel::cacheConfigExecuted

The cache configuration used for the kernel. The value is one of the CUfunc\_cache enumeration values from cuda.h.

### uint8\_t CUpti\_ActivityKernel::cacheConfigRequested

The cache configuration requested by the kernel. The value is one of the CUfunc\_cache enumeration values from cuda.h.

### uint32\_t CUpti\_ActivityKernel::contextId

The ID of the context where the kernel is executing.

### uint32\_t CUpti\_ActivityKernel::correlationId

The correlation ID of the kernel. Each kernel execution is assigned a unique correlation ID that is identical to the correlation ID in the driver API activity record that launched the kernel.

### uint32\_t CUpti\_ActivityKernel::deviceId

The ID of the device where the kernel is executing.

## `int32_t CUpti_ActivityKernel::dynamicSharedMemory`

The dynamic shared memory reserved for the kernel, in bytes.

## `uint64_t CUpti_ActivityKernel::end`

The end timestamp for the kernel execution, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the kernel.

## `int32_t CUpti_ActivityKernel::gridX`

The X-dimension grid size for the kernel.

## `int32_t CUpti_ActivityKernel::gridY`

The Y-dimension grid size for the kernel.

## `int32_t CUpti_ActivityKernel::gridZ`

The Z-dimension grid size for the kernel.

## `CUpti_ActivityKind CUpti_ActivityKernel::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_KERNEL` or `CUPTI_ACTIVITY_KIND_CONCURRENT_KERNEL`.

## `uint32_t CUpti_ActivityKernel::localMemoryPerThread`

The amount of local memory reserved for each thread, in bytes.

## `uint32_t CUpti_ActivityKernel::localMemoryTotal`

The total amount of local memory reserved for the kernel, in bytes.

## `const char *CUpti_ActivityKernel::name`

The name of the kernel. This name is shared across all activity records representing the same kernel, and so should not be modified.

## `uint32_t CUpti_ActivityKernel::pad`

Undefined. Reserved for internal use.

## `uint16_t CUpti_ActivityKernel::registersPerThread`

The number of registers required for each thread executing the kernel.

## `void *CUpti_ActivityKernel::reserved0`

Undefined. Reserved for internal use.

## `uint32_t CUpti_ActivityKernel::runtimeCorrelationId`

The runtime correlation ID of the kernel. Each kernel execution is assigned a unique runtime correlation ID that is identical to the correlation ID in the runtime API activity record that launched the kernel.

## `uint64_t CUpti_ActivityKernel::start`

The start timestamp for the kernel execution, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the kernel.

## `int32_t CUpti_ActivityKernel::staticSharedMemory`

The static shared memory allocated for the kernel, in bytes.

## `uint32_t CUpti_ActivityKernel::streamId`

The ID of the stream where the kernel is executing.

## 3.15. CUpti\_ActivityKernel2 Struct Reference

The activity record for a kernel (CUDA 5.5 onwards).

This activity record represents a kernel execution (CUPTI\_ACTIVITY\_KIND\_KERNEL and CUPTI\_ACTIVITY\_KIND\_CONCURRENT\_KERNEL).

## `int32_t CUpti_ActivityKernel2::blockX`

The X-dimension block size for the kernel.

## `int32_t CUpti_ActivityKernel2::blockY`

The Y-dimension block size for the kernel.

## `int32_t CUpti_ActivityKernel2::blockZ`

The Z-dimension grid size for the kernel.

## `uint64_t CUpti_ActivityKernel2::completed`

The completed timestamp for the kernel execution, in ns. It represents the completion of all it's child kernels and the kernel itself. A value of `CUPTI_TIMESTAMP_UNKNOWN` indicates that the completion time is unknown.

## `uint32_t CUpti_ActivityKernel2::contextId`

The ID of the context where the kernel is executing.

## `uint32_t CUpti_ActivityKernel2::correlationId`

The correlation ID of the kernel. Each kernel execution is assigned a unique correlation ID that is identical to the correlation ID in the driver or runtime API activity record that launched the kernel.

## `uint32_t CUpti_ActivityKernel2::deviceId`

The ID of the device where the kernel is executing.

## `int32_t CUpti_ActivityKernel2::dynamicSharedMemory`

The dynamic shared memory reserved for the kernel, in bytes.

## `uint64_t CUpti_ActivityKernel2::end`

The end timestamp for the kernel execution, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the kernel.

## `uint8_t CUpti_ActivityKernel2::executed`

The cache configuration used for the kernel. The value is one of the `CUfunc_cache` enumeration values from `cuda.h`.

## `int64_t CUpti_ActivityKernel2::gridId`

The grid ID of the kernel. Each kernel is assigned a unique grid ID at runtime.

## `int32_t CUpti_ActivityKernel2::gridX`

The X-dimension grid size for the kernel.

## `int32_t CUpti_ActivityKernel2::gridY`

The Y-dimension grid size for the kernel.

## `int32_t CUpti_ActivityKernel2::gridZ`

The Z-dimension grid size for the kernel.

## `CUpti_ActivityKind CUpti_ActivityKernel2::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_KERNEL` or `CUPTI_ACTIVITY_KIND_CONCURRENT_KERNEL`.

## `uint32_t CUpti_ActivityKernel2::localMemoryPerThread`

The amount of local memory reserved for each thread, in bytes.

## `uint32_t CUpti_ActivityKernel2::localMemoryTotal`

The total amount of local memory reserved for the kernel, in bytes.

## `const char *CUpti_ActivityKernel2::name`

The name of the kernel. This name is shared across all activity records representing the same kernel, and so should not be modified.

## `uint16_t CUpti_ActivityKernel2::registersPerThread`

The number of registers required for each thread executing the kernel.

## `uint8_t CUpti_ActivityKernel2::requested`

The cache configuration requested by the kernel. The value is one of the `CUfunc_cache` enumeration values from `cuda.h`.

## `void *CUpti_ActivityKernel2::reserved0`

Undefined. Reserved for internal use.

## `uint8_t CUpti_ActivityKernel2::sharedMemoryConfig`

The shared memory configuration used for the kernel. The value is one of the `CUsharedconfig` enumeration values from `cuda.h`.

## uint64\_t CUpti\_ActivityKernel2::start

The start timestamp for the kernel execution, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the kernel.

## int32\_t CUpti\_ActivityKernel2::staticSharedMemory

The static shared memory allocated for the kernel, in bytes.

## uint32\_t CUpti\_ActivityKernel2::streamId

The ID of the stream where the kernel is executing.

## 3.16. CUpti\_ActivityMarker Struct Reference

The activity record providing a marker which is an instantaneous point in time.

The marker is specified with a descriptive name and unique id (CUPTI\_ACTIVITY\_KIND\_MARKER).

## CUpti\_ActivityFlag CUpti\_ActivityMarker::flags

The flags associated with the marker.

**See also:**

[CUpti\\_ActivityFlag](#)

## uint32\_t CUpti\_ActivityMarker::id

The marker ID.

## CUpti\_ActivityKind CUpti\_ActivityMarker::kind

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_MARKER.

## const char \*CUpti\_ActivityMarker::name

The marker name for an instantaneous or start marker. This will be NULL for an end marker.

## CUpti\_ActivityMarker::objectId

The identifier for the activity object associated with this marker. 'objectKind' indicates which ID is valid for this record.

## CUpti\_ActivityObjectKind

## CUpti\_ActivityMarker::objectKind

The kind of activity object associated with this marker.

## uint64\_t CUpti\_ActivityMarker::timestamp

The timestamp for the marker, in ns. A value of 0 indicates that timestamp information could not be collected for the marker.

## 3.17. CUpti\_ActivityMarkerData Struct Reference

The activity record providing detailed information for a marker.

The marker data contains color, payload, and category.  
(CUPTI\_ACTIVITY\_KIND\_MARKER\_DATA).

## uint32\_t CUpti\_ActivityMarkerData::category

The category for the marker.

## uint32\_t CUpti\_ActivityMarkerData::color

The color for the marker.

## CUpti\_ActivityFlag CUpti\_ActivityMarkerData::flags

The flags associated with the marker.

**See also:**

[CUpti\\_ActivityFlag](#)

## uint32\_t CUpti\_ActivityMarkerData::id

The marker ID.

## CUpti\_ActivityKind CUpti\_ActivityMarkerData::kind

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_MARKER\_DATA.

## `CUpti_ActivityMarkerData::payload`

The payload value.

## `CUpti_MetricValueKind`

## `CUpti_ActivityMarkerData::payloadKind`

Defines the payload format for the value associated with the marker.

## 3.18. `CUpti_ActivityMemcpy` Struct Reference

The activity record for memory copies.

This activity record represents a memory copy (`CUPTI_ACTIVITY_KIND_MEMCPY`).

## `uint64_t CUpti_ActivityMemcpy::bytes`

The number of bytes transferred by the memory copy.

## `uint32_t CUpti_ActivityMemcpy::contextId`

The ID of the context where the memory copy is occurring.

## `uint8_t CUpti_ActivityMemcpy::copyKind`

The kind of the memory copy, stored as a byte to reduce record size.

**See also:**

`CUpti_ActivityMemcpyKind`

## `uint32_t CUpti_ActivityMemcpy::correlationId`

The correlation ID of the memory copy. Each memory copy is assigned a unique correlation ID that is identical to the correlation ID in the driver API activity record that launched the memory copy.

## `uint32_t CUpti_ActivityMemcpy::deviceId`

The ID of the device where the memory copy is occurring.

## `uint8_t CUpti_ActivityMemcpy::dstKind`

The destination memory kind read by the memory copy, stored as a byte to reduce record size.

**See also:**

[CUpti\\_ActivityMemoryKind](#)

## uint64\_t CUpti\_ActivityMemcpy::end

The end timestamp for the memory copy, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the memory copy.

## uint8\_t CUpti\_ActivityMemcpy::flags

The flags associated with the memory copy.

**See also:**

[CUpti\\_ActivityFlag](#)

## CUpti\_ActivityKind CUpti\_ActivityMemcpy::kind

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_MEMCPY.

## void \*CUpti\_ActivityMemcpy::reserved0

Undefined. Reserved for internal use.

## uint32\_t CUpti\_ActivityMemcpy::runtimeCorrelationId

The runtime correlation ID of the memory copy. Each memory copy is assigned a unique runtime correlation ID that is identical to the correlation ID in the runtime API activity record that launched the memory copy.

## uint8\_t CUpti\_ActivityMemcpy::srcKind

The source memory kind read by the memory copy, stored as a byte to reduce record size.

**See also:**

[CUpti\\_ActivityMemoryKind](#)

## uint64\_t CUpti\_ActivityMemcpy::start

The start timestamp for the memory copy, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the memory copy.

## `uint32_t CUpti_ActivityMemcpy::streamId`

The ID of the stream where the memory copy is occurring.

## 3.19. CUpti\_ActivityMemcpy2 Struct Reference

The activity record for peer-to-peer memory copies.

This activity record represents a peer-to-peer memory copy (CUPTI\_ACTIVITY\_KIND\_MEMCPY2).

## `uint64_t CUpti_ActivityMemcpy2::bytes`

The number of bytes transferred by the memory copy.

## `uint32_t CUpti_ActivityMemcpy2::contextId`

The ID of the context where the memory copy is occurring.

## `uint8_t CUpti_ActivityMemcpy2::copyKind`

The kind of the memory copy, stored as a byte to reduce record size.

**See also:**

`CUpti_ActivityMemcpyKind`

## `uint32_t CUpti_ActivityMemcpy2::correlationId`

The correlation ID of the memory copy. Each memory copy is assigned a unique correlation ID that is identical to the correlation ID in the driver and runtime API activity record that launched the memory copy.

## `uint32_t CUpti_ActivityMemcpy2::deviceId`

The ID of the device where the memory copy is occurring.

## `uint32_t CUpti_ActivityMemcpy2::dstContextId`

The ID of the context owning the memory being copied to.

## `uint32_t CUpti_ActivityMemcpy2::dstDeviceId`

The ID of the device where memory is being copied to.

## uint8\_t CUpti\_ActivityMemcpy2::dstKind

The destination memory kind read by the memory copy, stored as a byte to reduce record size.

**See also:**

[CUpti\\_ActivityMemoryKind](#)

## uint64\_t CUpti\_ActivityMemcpy2::end

The end timestamp for the memory copy, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the memory copy.

## uint8\_t CUpti\_ActivityMemcpy2::flags

The flags associated with the memory copy.

**See also:**

[CUpti\\_ActivityFlag](#)

## CUpti\_ActivityKind CUpti\_ActivityMemcpy2::kind

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_MEMCPY2.

## uint32\_t CUpti\_ActivityMemcpy2::pad

Undefined. Reserved for internal use.

## void \*CUpti\_ActivityMemcpy2::reserved0

Undefined. Reserved for internal use.

## uint32\_t CUpti\_ActivityMemcpy2::srcContextId

The ID of the context owning the memory being copied from.

## uint32\_t CUpti\_ActivityMemcpy2::srcDeviceId

The ID of the device where memory is being copied from.

## uint8\_t CUpti\_ActivityMemcpy2::srcKind

The source memory kind read by the memory copy, stored as a byte to reduce record size.

**See also:**

[CUpti\\_ActivityMemoryKind](#)

## uint64\_t CUpti\_ActivityMemcpy2::start

The start timestamp for the memory copy, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the memory copy.

## uint32\_t CUpti\_ActivityMemcpy2::streamId

The ID of the stream where the memory copy is occurring.

## 3.20. CUpti\_ActivityMemset Struct Reference

The activity record for memset.

This activity record represents a memory set operation (CUPTI\_ACTIVITY\_KIND\_MEMSET).

## uint64\_t CUpti\_ActivityMemset::bytes

The number of bytes being set by the memory set.

## uint32\_t CUpti\_ActivityMemset::contextId

The ID of the context where the memory set is occurring.

## uint32\_t CUpti\_ActivityMemset::correlationId

The correlation ID of the memory set. Each memory set is assigned a unique correlation ID that is identical to the correlation ID in the driver API activity record that launched the memory set.

## uint32\_t CUpti\_ActivityMemset::deviceId

The ID of the device where the memory set is occurring.

## `uint64_t CUpti_ActivityMemset::end`

The end timestamp for the memory set, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the memory set.

## `CUpti_ActivityKind CUpti_ActivityMemset::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_MEMSET`.

## `void *CUpti_ActivityMemset::reserved0`

Undefined. Reserved for internal use.

## `uint32_t CUpti_ActivityMemset::runtimeCorrelationId`

The runtime correlation ID of the memory set. Each memory set is assigned a unique runtime correlation ID that is identical to the correlation ID in the runtime API activity record that launched the memory set.

## `uint64_t CUpti_ActivityMemset::start`

The start timestamp for the memory set, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the memory set.

## `uint32_t CUpti_ActivityMemset::streamId`

The ID of the stream where the memory set is occurring.

## `uint32_t CUpti_ActivityMemset::value`

The value being assigned to memory by the memory set.

## 3.21. CUpti\_ActivityMetric Struct Reference

The activity record for a CUPTI metric.

This activity record represents the collection of a CUPTI metric value (`CUPTI_ACTIVITY_KIND_METRIC`). This activity record kind is not produced by the activity API but is included for completeness and ease-of-use. Profile frameworks built on top of CUPTI that collect metric data may choose to use this type to store the collected metric data.

## `uint32_t CUpti_ActivityMetric::correlationId`

The correlation ID of the metric. Use of this ID is user-defined, but typically this ID value will equal the correlation ID of the kernel for which the metric was gathered.

## `uint8_t CUpti_ActivityMetric::flags`

The properties of this metric.

**See also:**

`CUpti_ActivityFlag`

## `CUpti_MetricID CUpti_ActivityMetric::id`

The metric ID.

## `CUpti_ActivityKind CUpti_ActivityMetric::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_METRIC`.

## `uint8_t CUpti_ActivityMetric::pad`

Undefined. Reserved for internal use.

## `CUpti_ActivityMetric::value`

The metric value.

## 3.22. `CUpti_ActivityMetricInstance` Struct Reference

The activity record for a CUPTI metric with instance information. This activity record represents a CUPTI metric value for a specific metric domain instance (`CUPTI_ACTIVITY_KIND_METRIC_INSTANCE`). This activity record kind is not produced by the activity API but is included for completeness and ease-of-use. Profile frameworks built on top of CUPTI that collect metric data may choose to use this type to store the collected metric data. This activity record should be used when metric domain instance information needs to be associated with the metric.

## `uint32_t CUpti_ActivityMetricInstance::correlationId`

The correlation ID of the metric. Use of this ID is user-defined, but typically this ID value will equal the correlation ID of the kernel for which the metric was gathered.

## `uint8_t CUpti_ActivityMetricInstance::flags`

The properties of this metric.

**See also:**

`CUpti_ActivityFlag`

## `CUpti_MetricID CUpti_ActivityMetricInstance::id`

The metric ID.

## `uint32_t CUpti_ActivityMetricInstance::instance`

The metric domain instance.

## `CUpti_ActivityKind CUpti_ActivityMetricInstance::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_METRIC_INSTANCE`.

## `uint8_t CUpti_ActivityMetricInstance::pad`

Undefined. Reserved for internal use.

## `CUpti_ActivityMetricInstance::value`

The metric value.

## 3.23. `CUpti_ActivityModule` Struct Reference

The activity record for a CUDA module.

This activity record represents a CUDA module (`CUPTI_ACTIVITY_KIND_MODULE`). This activity record kind is not produced by the activity API but is included for completeness and ease-of-use. Profile frameworks built on top of CUPTI that collect module data from the module callback may choose to use this type to store the collected module data.

## `uint32_t CUpti_ActivityModule::contextId`

The ID of the context where the module is loaded.

## `const void *CUpti_ActivityModule::cubin`

The pointer to cubin.

## `uint32_t CUpti_ActivityModule::cubinSize`

The cubin size.

## `uint32_t CUpti_ActivityModule::id`

The module ID.

## `CUpti_ActivityKind CUpti_ActivityModule::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_MODULE`.

## `uint32_t CUpti_ActivityModule::pad`

Undefined. Reserved for internal use.

## 3.24. `CUpti_ActivityName` Struct Reference

The activity record providing a name.

This activity record provides a name for a device, context, thread, etc. (`CUPTI_ACTIVITY_KIND_NAME`).

## `CUpti_ActivityKind CUpti_ActivityName::kind`

The activity record kind, must be `CUPTI_ACTIVITY_KIND_NAME`.

## `const char *CUpti_ActivityName::name`

The name.

## `CUpti_ActivityName::objectId`

The identifier for the activity object. 'objectKind' indicates which ID is valid for this record.

## `CUpti_ActivityObjectKind CUpti_ActivityName::objectKind`

The kind of activity object being named.

## 3.25. CUpti\_ActivityObjectKindId Union Reference

Identifiers for object kinds as specified by CUpti\_ActivityObjectKind.

See also:

[CUpti\\_ActivityObjectKind](#)

**CUpti\_ActivityObjectKindId::@1**

**CUpti\_ActivityObjectKindId::dcs**

A device object requires that we identify the device ID. A context object requires that we identify both the device and context ID. A stream object requires that we identify device, context, and stream ID.

**CUpti\_ActivityObjectKindId::@0**

**CUpti\_ActivityObjectKindId::pt**

A process object requires that we identify the process ID. A thread object requires that we identify both the process and thread ID.

## 3.26. CUpti\_ActivityOverhead Struct Reference

The activity record for CUPTI and driver overheads.

This activity record provides CUPTI and driver overhead information (CUPTI\_ACTIVITY\_OVERHEAD).

**uint64\_t CUpti\_ActivityOverhead::end**

The end timestamp for the overhead, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the overhead.

**CUpti\_ActivityKind CUpti\_ActivityOverhead::kind**

The activity record kind, must be CUPTI\_ACTIVITY\_OVERHEAD.

**CUpti\_ActivityOverhead::objectId**

The identifier for the activity object. 'objectKind' indicates which ID is valid for this record.

**CUpti\_ActivityObjectKind**  
**CUpti\_ActivityOverhead::objectKind**

The kind of activity object that the overhead is associated with.

**CUpti\_ActivityOverheadKind**  
**CUpti\_ActivityOverhead::overheadKind**

The kind of overhead, CUPTI, DRIVER, COMPILER etc.

**uint64\_t CUpti\_ActivityOverhead::start**

The start timestamp for the overhead, in ns. A value of 0 for both the start and end timestamps indicates that timestamp information could not be collected for the overhead.

## 3.27. CUpti\_ActivityPreemption Struct Reference

The activity record for a preemption of a CDP kernel.

This activity record represents a preemption of a CDP kernel.

**uint32\_t CUpti\_ActivityPreemption::blockX**

The X-dimension of the block that is preempted

**uint32\_t CUpti\_ActivityPreemption::blockY**

The Y-dimension of the block that is preempted

**uint32\_t CUpti\_ActivityPreemption::blockZ**

The Z-dimension of the block that is preempted

**int64\_t CUpti\_ActivityPreemption::gridId**

The grid-id of the block that is preempted

**CUpti\_ActivityKind CUpti\_ActivityPreemption::kind**

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_PREEMPTION

**uint32\_t CUpti\_ActivityPreemption::pad**

Undefined. Reserved for internal use.

**CUpti\_ActivityPreemptionKind**

**CUpti\_ActivityPreemption::preemptionKind**

kind of the preemption

**uint64\_t CUpti\_ActivityPreemption::timestamp**

The timestamp of the preemption, in ns. A value of 0 indicates that timestamp information could not be collected for the preemption.

## 3.28. CUpti\_ActivitySharedAccess Struct Reference

The activity record for source-level shared access.

This activity records the locations of the shared accesses in the source (CUPTI\_ACTIVITY\_KIND\_SHARED\_ACCESS).

**uint32\_t CUpti\_ActivitySharedAccess::correlationId**

The correlation ID of the kernel to which this result is associated.

**uint32\_t CUpti\_ActivitySharedAccess::executed**

The number of times this instruction was executed

**CUpti\_ActivityFlag CUpti\_ActivitySharedAccess::flags**

The properties of this shared access.

**uint32\_t CUpti\_ActivitySharedAccess::functionId**

Correlation ID with global/device function name

**CUpti\_ActivityKind CUpti\_ActivitySharedAccess::kind**

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_SHARED\_ACCESS.

**uint32\_t CUpti\_ActivitySharedAccess::pad**

Undefined. Reserved for internal use.

**uint32\_t CUpti\_ActivitySharedAccess::pcOffset**

The pc offset for the access.

**uint64\_t**

**CUpti\_ActivitySharedAccess::sharedTransactions**

The total number of shared memory transactions generated by this access

**uint32\_t CUpti\_ActivitySharedAccess::sourceLocatorId**

The ID for source locator.

**uint64\_t**

**CUpti\_ActivitySharedAccess::theoreticalSharedTransactions**

The minimum number of shared memory transactions possible based on the access pattern.

**uint64\_t CUpti\_ActivitySharedAccess::threadsExecuted**

This increments each time when this instruction is executed by number of threads that executed this instruction with predicate and condition code evaluating to true.

## 3.29. CUpti\_ActivitySourceLocator Struct Reference

The activity record for source locator.

This activity record represents a source locator (CUPTI\_ACTIVITY\_KIND\_SOURCE\_LOCATOR).

**const char \*CUpti\_ActivitySourceLocator::fileName**

The path for the file.

**uint32\_t CUpti\_ActivitySourceLocator::id**

The ID for the source path, will be used in all the source level results.

## **CUpti\_ActivityKind CUpti\_ActivitySourceLocator::kind**

The activity record kind, must be CUPTI\_ACTIVITY\_KIND\_SOURCE\_LOCATOR.

## **uint32\_t CUpti\_ActivitySourceLocator::lineNumber**

The line number in the source .

## **3.30. CUpti\_ActivityUnifiedMemoryCounter Struct Reference**

The activity record for Unified Memory counters.

This activity record represents a Unified Memory counter (CUPTI\_ACTIVITY\_KIND\_UNIFIED\_MEMORY\_COUNTER).

## **CUpti\_ActivityUnifiedMemoryCounterKind CUpti\_ActivityUnifiedMemoryCounter::counterKind**

The Unified Memory counter kind. See /ref CUpti\_ActivityUnifiedMemoryCounterKind

## **uint32\_t CUpti\_ActivityUnifiedMemoryCounter::deviceId**

The ID of the device involved in the memory transfer operation. It is not relevant if the scope of the counter is global (all devices).

## **CUpti\_ActivityKind CUpti\_ActivityUnifiedMemoryCounter::kind**

The activity record kind, must be  
CUPTI\_ACTIVITY\_KIND\_UNIFIED\_MEMORY\_COUNTER

## **uint32\_t CUpti\_ActivityUnifiedMemoryCounter::pad**

Undefined. Reserved for internal use.

## **uint32\_t CUpti\_ActivityUnifiedMemoryCounter::processId**

The ID of the process to which this record belongs to. In case of global scope, processId is undefined.

## CUpti\_ActivityUnifiedMemoryCounterScope

### CUpti\_ActivityUnifiedMemoryCounter::scope

Scope of the Unified Memory counter. See /ref  
CUpti\_ActivityUnifiedMemoryCounterScope

## uint64\_t

### CUpti\_ActivityUnifiedMemoryCounter::timestamp

The timestamp when this sample was retrieved, in ns. A value of 0 indicates that timestamp information could not be collected

## uint64\_t CUpti\_ActivityUnifiedMemoryCounter::value

Value of the counter

## 3.31. CUpti\_ActivityUnifiedMemoryCounterConfig Struct Reference

Unified Memory counters configuration structure.

This structure controls the enable/disable of the various Unified Memory counters consisting of scope, kind and other parameters. See function /ref  
cuptiActivityConfigureUnifiedMemoryCounter

## uint32\_t

### CUpti\_ActivityUnifiedMemoryCounterConfig::deviceId

Device id of the target device. This is relevant only for single device scopes.

## uint32\_t

### CUpti\_ActivityUnifiedMemoryCounterConfig::enable

Control to enable/disable the counter. To enable the counter set it to non-zero value while disable is indicated by zero.

## CUpti\_ActivityUnifiedMemoryCounterKind

### CUpti\_ActivityUnifiedMemoryCounterConfig::kind

Unified Memory counter Counter kind

## CUpti\_ActivityUnifiedMemoryCounterScope

### CUpti\_ActivityUnifiedMemoryCounterConfig::scope

Unified Memory counter Counter scope

## 3.32. CUpti\_CallbackData Struct Reference

Data passed into a runtime or driver API callback function.

Data passed into a runtime or driver API callback function as the `cbdata` argument to [CUpti\\_CallbackFunc](#). The `cbdata` will be this type for domain equal to `CUPTI_CB_DOMAIN_DRIVER_API` or `CUPTI_CB_DOMAIN_RUNTIME_API`. The callback data is valid only within the invocation of the callback function that is passed the data. If you need to retain some data for use outside of the callback, you must make a copy of that data. For example, if you make a shallow copy of [CUpti\\_CallbackData](#) within a callback, you cannot dereference `functionParams` outside of that callback to access the function parameters. `functionName` is an exception: the string pointed to by `functionName` is a global constant and so may be accessed outside of the callback.

### CUpti\_ApiCallbackSite CUpti\_CallbackData::callbackSite

Point in the runtime or driver function from where the callback was issued.

### CUcontext CUpti\_CallbackData::context

Driver context current to the thread, or null if no context is current. This value can change from the entry to exit callback of a runtime API function if the runtime initializes a context.

### uint32\_t CUpti\_CallbackData::contextUid

Unique ID for the CUDA context associated with the thread. The UIDs are assigned sequentially as contexts are created and are unique within a process.

### uint64\_t \*CUpti\_CallbackData::correlationData

Pointer to data shared between the entry and exit callbacks of a given runtime or driver API function invocation. This field can be used to pass 64-bit values from the entry callback to the corresponding exit callback.

### uint32\_t CUpti\_CallbackData::correlationId

The activity record correlation ID for this callback. For a driver domain callback (i.e. domain `CUPTI_CB_DOMAIN_DRIVER_API`) this ID will equal the correlation ID in

the `CUpti_ActivityAPI` record corresponding to the CUDA driver function call. For a runtime domain callback (i.e. `domain CUPTI_CB_DOMAIN_RUNTIME_API`) this ID will equal the correlation ID in the `CUpti_ActivityAPI` record corresponding to the CUDA runtime function call. Within the callback, this ID can be recorded to correlate user data with the activity record. This field is new in 4.1.

## `const char *CUpti_CallbackData::functionName`

Name of the runtime or driver API function which issued the callback. This string is a global constant and so may be accessed outside of the callback.

## `const void *CUpti_CallbackData::functionParams`

Pointer to the arguments passed to the runtime or driver API call. See `generated_cuda_runtime_api_meta.h` and `generated_cuda_meta.h` for structure definitions for the parameters for each runtime and driver API function.

## `void *CUpti_CallbackData::functionReturnValue`

Pointer to the return value of the runtime or driver API call. This field is only valid within the `exit::CUPTI_API_EXIT` callback. For a runtime API `functionReturnValue` points to a `cudaError_t`. For a driver API `functionReturnValue` points to a `CUresult`.

## `const char *CUpti_CallbackData::symbolName`

Name of the symbol operated on by the runtime or driver API function which issued the callback. This entry is valid only for driver and runtime launch callbacks, where it returns the name of the kernel.

## 3.33. CUpti\_EventGroupSet Struct Reference

A set of event groups.

A set of event groups. When returned by `cuptiEventGroupSetsCreate` and `cuptiMetricCreateEventGroupSets` a set indicates that event groups that can be enabled at the same time (i.e. all the events in the set can be collected simultaneously).

## `CUpti_EventGroup *CUpti_EventGroupSet::eventGroups`

An array of `numEventGroups` event groups.

## `uint32_t CUpti_EventGroupSet::numEventGroups`

The number of event groups in the set.

### 3.34. CUpti\_EventGroupSets Struct Reference

A set of event group sets.

A set of event group sets. When returned by `cuptiEventGroupSetsCreate` and `cuptiMetricCreateEventGroupSets` a `CUpti_EventGroupSets` indicates the number of passes required to collect all the events, and the event groups that should be collected during each pass.

`uint32_t CUpti_EventGroupSets::numSets`

Number of event group sets.

`CUpti_EventGroupSet *CUpti_EventGroupSets::sets`

An array of `numSets` event group sets.

### 3.35. CUpti\_MetricValue Union Reference

A metric value.

Metric values can be one of several different kinds. Corresponding to each kind is a member of the `CUpti_MetricValue` union. The metric value returned by `cuptiMetricGetValue` should be accessed using the appropriate member of that union based on its value kind.

### 3.36. CUpti\_ModuleResourceData Struct Reference

Module data passed into a resource callback function.

CUDA module data passed into a resource callback function as the `cbdata` argument to `CUpti_CallbackFunc`. The `cbdata` will be this type for `domain` equal to `CUPTI_CB_DOMAIN_RESOURCE`. The module data is valid only within the invocation of the callback function that is passed the data. If you need to retain some data for use outside of the callback, you must make a copy of that data.

`size_t CUpti_ModuleResourceData::cubinSize`

The size of the cubin.

`uint32_t CUpti_ModuleResourceData::moduleId`

Identifier to associate with the CUDA module.

## `const char *CUpti_ModuleResourceData::pCubin`

Pointer to the associated cubin.

## 3.37. CUpti\_NvtxData Struct Reference

Data passed into a NVTX callback function.

Data passed into a NVTX callback function as the `cbdata` argument to [CUpti\\_CallbackFunc](#). The `cbdata` will be this type for `domain` equal to `CUPTI_CB_DOMAIN_NVTX`. Unless otherwise notes, the callback data is valid only within the invocation of the callback function that is passed the data. If you need to retain some data for use outside of the callback, you must make a copy of that data.

### `const char *CUpti_NvtxData::functionName`

Name of the NVTX API function which issued the callback. This string is a global constant and so may be accessed outside of the callback.

### `const void *CUpti_NvtxData::functionParams`

Pointer to the arguments passed to the NVTX API call. See `generated_nvtx_meta.h` for structure definitions for the parameters for each NVTX API function.

## 3.38. CUpti\_ResourceData Struct Reference

Data passed into a resource callback function.

Data passed into a resource callback function as the `cbdata` argument to [CUpti\\_CallbackFunc](#). The `cbdata` will be this type for `domain` equal to `CUPTI_CB_DOMAIN_RESOURCE`. The callback data is valid only within the invocation of the callback function that is passed the data. If you need to retain some data for use outside of the callback, you must make a copy of that data.

### `CUcontext CUpti_ResourceData::context`

For `CUPTI_CBID_RESOURCE_CONTEXT_CREATED` and `CUPTI_CBID_RESOURCE_CONTEXT_DESTROY_STARTING`, the context being created or destroyed. For `CUPTI_CBID_RESOURCE_STREAM_CREATED` and `CUPTI_CBID_RESOURCE_STREAM_DESTROY_STARTING`, the context containing the stream being created or destroyed.

## `void *CUpti_ResourceData::resourceDescriptor`

Reserved for future use.

## `CUstream CUpti_ResourceData::stream`

For `CUPTI_CBID_RESOURCE_STREAM_CREATED` and `CUPTI_CBID_RESOURCE_STREAM_DESTROY_STARTING`, the stream being created or destroyed.

## 3.39. CUpti\_SynchronizeData Struct Reference

Data passed into a synchronize callback function.

Data passed into a synchronize callback function as the `cbdata` argument to `CUpti_CallbackFunc`. The `cbdata` will be this type for domain equal to `CUPTI_CB_DOMAIN_SYNCHRONIZE`. The callback data is valid only within the invocation of the callback function that is passed the data. If you need to retain some data for use outside of the callback, you must make a copy of that data.

## `CUcontext CUpti_SynchronizeData::context`

The context of the stream being synchronized.

## `CUstream CUpti_SynchronizeData::stream`

The stream being synchronized.

# Chapter 4.

## DATA FIELDS

Here is a list of all documented struct and union fields with links to the struct/union documentation for each field:

### A

#### **attribute**

[CUpti\\_ActivityDeviceAttribute](#)

### B

#### **blockX**

[CUpti\\_ActivityKernel](#)  
[CUpti\\_ActivityKernel2](#)  
[CUpti\\_ActivityPreemption](#)  
[CUpti\\_ActivityCdpKernel](#)

#### **blockY**

[CUpti\\_ActivityPreemption](#)  
[CUpti\\_ActivityKernel](#)  
[CUpti\\_ActivityKernel2](#)  
[CUpti\\_ActivityCdpKernel](#)

#### **blockZ**

[CUpti\\_ActivityCdpKernel](#)  
[CUpti\\_ActivityKernel](#)  
[CUpti\\_ActivityKernel2](#)  
[CUpti\\_ActivityPreemption](#)

#### **bytes**

[CUpti\\_ActivityMemset](#)  
[CUpti\\_ActivityMemcpy](#)  
[CUpti\\_ActivityMemcpy2](#)

**C****cacheConfigExecuted**

CUpti\_ActivityKernel

**cacheConfigRequested**

CUpti\_ActivityKernel

**callbackSite**

CUpti\_CallbackData

**category**

CUpti\_ActivityMarkerData

**cbid**

CUpti\_ActivityAPI

**clocksThrottleReasons**

CUpti\_ActivityEnvironment

**color**

CUpti\_ActivityMarkerData

**completed**

CUpti\_ActivityKernel2

CUpti\_ActivityCdpKernel

**computeApiKind**

CUpti\_ActivityContext

**computeCapabilityMajor**

CUpti\_ActivityDevice

**computeCapabilityMinor**

CUpti\_ActivityDevice

**constantMemorySize**

CUpti\_ActivityDevice

**context**

CUpti\_SynchronizeData

CUpti\_CallbackData

CUpti\_ResourceData

**contextId**

CUpti\_ActivityMemcpy

CUpti\_ActivityMemcpy2

CUpti\_ActivityMemset

CUpti\_ActivityKernel

CUpti\_ActivityKernel2

CUpti\_ActivityCdpKernel

CUpti\_ActivityContext

CUpti\_ActivityFunction

CUpti\_ActivityModule

**contextUid**

CUpti\_CallbackData

**cooling**

CUpti\_ActivityEnvironment

**copyKind**

CUpti\_ActivityMemcpy

CUpti\_ActivityMemcpy2

**coreClockRate**

CUpti\_ActivityDevice

**correlationData**

CUpti\_CallbackData

**correlationId**

CUpti\_ActivityGlobalAccess

CUpti\_ActivityKernel

CUpti\_ActivityEvent

CUpti\_ActivityMetricInstance

CUpti\_ActivityMemcpy

CUpti\_ActivityInstructionExecution

CUpti\_ActivityMetric

CUpti\_ActivityKernel2

CUpti\_CallbackData

CUpti\_ActivityCdpKernel

CUpti\_ActivityBranch

CUpti\_ActivityMemset

CUpti\_ActivityEventInstance

CUpti\_ActivityMemcpy2

CUpti\_ActivitySharedAccess

CUpti\_ActivityAPI

**counterKind**

CUpti\_ActivityUnifiedMemoryCounter

**cubin**

CUpti\_ActivityModule

**cubinSize**

CUpti\_ModuleResourceData

CUpti\_ActivityModule

**D****dcs**

CUpti\_ActivityObjectKindId

**deviceId**

CUpti\_ActivityUnifiedMemoryCounterConfig

CUpti\_ActivityMemcpy2

CUpti\_ActivityCdpKernel

CUpti\_ActivityDeviceAttribute

CUpti\_ActivityMemset

- CUpti\_ActivityContext
- CUpti\_ActivityEnvironment
- CUpti\_ActivityMemcpy
- CUpti\_ActivityKernel
- CUpti\_ActivityUnifiedMemoryCounter
- CUpti\_ActivityKernel2
- diverged**
  - CUpti\_ActivityBranch
- domain**
  - CUpti\_ActivityEventInstance
  - CUpti\_ActivityEvent
- dstContextId**
  - CUpti\_ActivityMemcpy2
- dstDeviceId**
  - CUpti\_ActivityMemcpy2
- dstKind**
  - CUpti\_ActivityMemcpy2
  - CUpti\_ActivityMemcpy
- dynamicSharedMemory**
  - CUpti\_ActivityCdpKernel
  - CUpti\_ActivityKernel
  - CUpti\_ActivityKernel2
- E**
- enable**
  - CUpti\_ActivityUnifiedMemoryCounterConfig
- end**
  - CUpti\_ActivityMemcpy
  - CUpti\_ActivityMemset
  - CUpti\_ActivityAPI
  - CUpti\_ActivityOverhead
  - CUpti\_ActivityKernel
  - CUpti\_ActivityMemcpy2
  - CUpti\_ActivityKernel2
  - CUpti\_ActivityCdpKernel
- environmentKind**
  - CUpti\_ActivityEnvironment
- eventGroups**
  - CUpti\_EventGroupSet
- executed**
  - CUpti\_ActivityCdpKernel
  - CUpti\_ActivityKernel2
  - CUpti\_ActivityInstructionExecution

CUpti\_ActivityBranch  
 CUpti\_ActivityGlobalAccess  
 CUpti\_ActivitySharedAccess

**F****fanSpeed**

CUpti\_ActivityEnvironment

**fileName**

CUpti\_ActivitySourceLocator

**flags**

CUpti\_ActivityMemcpy2  
 CUpti\_ActivityDevice  
 CUpti\_ActivityDeviceAttribute  
 CUpti\_ActivityMetric  
 CUpti\_ActivityMarker  
 CUpti\_ActivityMarkerData  
 CUpti\_ActivityMemcpy  
 CUpti\_ActivitySharedAccess  
 CUpti\_ActivityMetricInstance  
 CUpti\_ActivityInstructionExecution  
 CUpti\_ActivityGlobalAccess

**functionId**

CUpti\_ActivitySharedAccess  
 CUpti\_ActivityInstructionExecution

**functionIndex**

CUpti\_ActivityFunction

**functionName**

CUpti\_NvtxData  
 CUpti\_CallbackData

**functionParams**

CUpti\_CallbackData  
 CUpti\_NvtxData

**functionReturnValue**

CUpti\_CallbackData

**G****globalMemoryBandwidth**

CUpti\_ActivityDevice

**globalMemorySize**

CUpti\_ActivityDevice

**gpuTemperature**

CUpti\_ActivityEnvironment

**gridId**

CUpti\_ActivityKernel2  
 CUpti\_ActivityCdpKernel  
 CUpti\_ActivityPreemption

**gridX**

CUpti\_ActivityKernel2  
 CUpti\_ActivityCdpKernel  
 CUpti\_ActivityKernel

**gridY**

CUpti\_ActivityKernel2  
 CUpti\_ActivityKernel  
 CUpti\_ActivityCdpKernel

**gridZ**

CUpti\_ActivityKernel  
 CUpti\_ActivityKernel2  
 CUpti\_ActivityCdpKernel

**I****id**

CUpti\_ActivityEvent  
 CUpti\_ActivityEventInstance  
 CUpti\_ActivityMetricInstance  
 CUpti\_ActivityModule  
 CUpti\_ActivityMarkerData  
 CUpti\_ActivityFunction  
 CUpti\_ActivityMarker  
 CUpti\_ActivityDevice  
 CUpti\_ActivitySourceLocator  
 CUpti\_ActivityMetric

**instance**

CUpti\_ActivityEventInstance  
 CUpti\_ActivityMetricInstance

**K****kind**

CUpti\_ActivityUnifiedMemoryCounterConfig  
 CUpti\_ActivitySharedAccess  
 CUpti\_ActivityModule  
 CUpti\_ActivityFunction  
 CUpti\_ActivityUnifiedMemoryCounter  
 CUpti\_ActivityInstructionExecution  
 CUpti\_ActivityEnvironment  
 CUpti\_ActivityOverhead

CUpti\_ActivityMarkerData  
 CUpti\_ActivityMarker  
 CUpti\_ActivityName  
 CUpti\_ActivityContext  
 CUpti\_ActivityDeviceAttribute  
 CUpti\_ActivityDevice  
 CUpti\_ActivityBranch  
 CUpti\_ActivityGlobalAccess  
 CUpti\_ActivitySourceLocator  
 CUpti\_ActivityMetricInstance  
 CUpti\_ActivityMetric  
 CUpti\_ActivityEventInstance  
 CUpti\_ActivityEvent  
 CUpti\_ActivityAPI  
 CUpti\_ActivityPreemption  
 CUpti\_ActivityCdpKernel  
 CUpti\_ActivityKernel2  
 CUpti\_ActivityKernel  
 CUpti\_ActivityMemset  
 CUpti\_ActivityMemcpy2  
 CUpti\_ActivityMemcpy  
 CUpti\_Activity

## L

### **l2\_transactions**

CUpti\_ActivityGlobalAccess

### **l2CacheSize**

CUpti\_ActivityDevice

### **lineNumber**

CUpti\_ActivitySourceLocator

### **localMemoryPerThread**

CUpti\_ActivityKernel

CUpti\_ActivityKernel2

CUpti\_ActivityCdpKernel

### **localMemoryTotal**

CUpti\_ActivityCdpKernel

CUpti\_ActivityKernel

CUpti\_ActivityKernel2

## M

### **maxBlockDimX**

CUpti\_ActivityDevice

**maxBlockDimY**  
 CUpti\_ActivityDevice  
**maxBlockDimZ**  
 CUpti\_ActivityDevice  
**maxBlocksPerMultiprocessor**  
 CUpti\_ActivityDevice  
**maxGridDimX**  
 CUpti\_ActivityDevice  
**maxGridDimY**  
 CUpti\_ActivityDevice  
**maxGridDimZ**  
 CUpti\_ActivityDevice  
**maxIPC**  
 CUpti\_ActivityDevice  
**maxRegistersPerBlock**  
 CUpti\_ActivityDevice  
**maxSharedMemoryPerBlock**  
 CUpti\_ActivityDevice  
**maxThreadsPerBlock**  
 CUpti\_ActivityDevice  
**maxWarpsPerMultiprocessor**  
 CUpti\_ActivityDevice  
**memoryClock**  
 CUpti\_ActivityEnvironment  
**moduleId**  
 CUpti\_ModuleResourceData  
 CUpti\_ActivityFunction

## N

**name**  
 CUpti\_ActivityKernel  
 CUpti\_ActivityKernel2  
 CUpti\_ActivityDevice  
 CUpti\_ActivityName  
 CUpti\_ActivityCdpKernel  
 CUpti\_ActivityMarker  
 CUpti\_ActivityFunction  
**notPredOffThreadsExecuted**  
 CUpti\_ActivityInstructionExecution  
**nullStreamId**  
 CUpti\_ActivityContext  
**numEventGroups**  
 CUpti\_EventGroupSet

**numMemcpyEngines**

CUpti\_ActivityDevice

**numMultiprocessors**

CUpti\_ActivityDevice

**numSets**

CUpti\_EventGroupSets

**numThreadsPerWarp**

CUpti\_ActivityDevice

**O****objectId**

CUpti\_ActivityName

CUpti\_ActivityMarker

CUpti\_ActivityOverhead

**objectKind**

CUpti\_ActivityMarker

CUpti\_ActivityName

CUpti\_ActivityOverhead

**overheadKind**

CUpti\_ActivityOverhead

**P****pad**

CUpti\_ActivityMemcpy2

CUpti\_ActivityKernel

CUpti\_ActivityEventInstance

CUpti\_ActivityUnifiedMemoryCounter

CUpti\_ActivityModule

CUpti\_ActivityMetric

CUpti\_ActivitySharedAccess

CUpti\_ActivityPreemption

CUpti\_ActivityMetricInstance

CUpti\_ActivityInstructionExecution

**parentBlockX**

CUpti\_ActivityCdpKernel

**parentBlockY**

CUpti\_ActivityCdpKernel

**parentBlockZ**

CUpti\_ActivityCdpKernel

**parentGridId**

CUpti\_ActivityCdpKernel

**payload**

CUpti\_ActivityMarkerData

**payloadKind**

CUpti\_ActivityMarkerData

**pcieLinkGen**

CUpti\_ActivityEnvironment

**pcieLinkWidth**

CUpti\_ActivityEnvironment

**pcOffset**

CUpti\_ActivityBranch

CUpti\_ActivityInstructionExecution

CUpti\_ActivitySharedAccess

CUpti\_ActivityGlobalAccess

**pCubin**

CUpti\_ModuleResourceData

**power**

CUpti\_ActivityEnvironment

**powerLimit**

CUpti\_ActivityEnvironment

**preemptionKind**

CUpti\_ActivityPreemption

**processId**

CUpti\_ActivityUnifiedMemoryCounter

CUpti\_ActivityAPI

**pt**

CUpti\_ActivityObjectKindId

**Q****queued**

CUpti\_ActivityCdpKernel

**R****registersPerThread**

CUpti\_ActivityKernel

CUpti\_ActivityKernel2

CUpti\_ActivityCdpKernel

**requested**

CUpti\_ActivityKernel2

CUpti\_ActivityCdpKernel

**reserved0**

CUpti\_ActivityMemset

CUpti\_ActivityKernel

CUpti\_ActivityKernel2

CUpti\_ActivityMemcpy

CUpti\_ActivityMemcpy2

**resourceDescriptor**

CUpti\_ResourceData

**returnValue**

CUpti\_ActivityAPI

**runtimeCorrelationId**

CUpti\_ActivityMemset

CUpti\_ActivityMemcpy

CUpti\_ActivityKernel

**S****scope**

CUpti\_ActivityUnifiedMemoryCounterConfig

CUpti\_ActivityUnifiedMemoryCounter

**sets**

CUpti\_EventGroupSets

**sharedMemoryConfig**

CUpti\_ActivityKernel2

CUpti\_ActivityCdpKernel

**sharedTransactions**

CUpti\_ActivitySharedAccess

**smClock**

CUpti\_ActivityEnvironment

**sourceLocatorId**

CUpti\_ActivityGlobalAccess

CUpti\_ActivityBranch

CUpti\_ActivityInstructionExecution

CUpti\_ActivitySharedAccess

**speed**

CUpti\_ActivityEnvironment

**srcContextId**

CUpti\_ActivityMemcpy2

**srcDeviceId**

CUpti\_ActivityMemcpy2

**srcKind**

CUpti\_ActivityMemcpy2

CUpti\_ActivityMemcpy

**start**

CUpti\_ActivityMemcpy

CUpti\_ActivityMemcpy2

CUpti\_ActivityMemset

CUpti\_ActivityKernel

CUpti\_ActivityKernel2

CUpti\_ActivityCdpKernel

CUpti\_ActivityAPI  
 CUpti\_ActivityOverhead  
**staticSharedMemory**  
 CUpti\_ActivityCdpKernel  
 CUpti\_ActivityKernel2  
 CUpti\_ActivityKernel  
**stream**  
 CUpti\_SynchronizeData  
 CUpti\_ResourceData  
**streamId**  
 CUpti\_ActivityMemcpy  
 CUpti\_ActivityMemcpy2  
 CUpti\_ActivityKernel  
 CUpti\_ActivityMemset  
 CUpti\_ActivityKernel2  
 CUpti\_ActivityCdpKernel  
**submitted**  
 CUpti\_ActivityCdpKernel  
**symbolName**  
 CUpti\_CallbackData

**T**  
**temperature**  
 CUpti\_ActivityEnvironment  
**theoreticalSharedTransactions**  
 CUpti\_ActivitySharedAccess  
**threadId**  
 CUpti\_ActivityAPI  
**threadsExecuted**  
 CUpti\_ActivityGlobalAccess  
 CUpti\_ActivityBranch  
 CUpti\_ActivitySharedAccess  
 CUpti\_ActivityInstructionExecution  
**timestamp**  
 CUpti\_ActivityMarker  
 CUpti\_ActivityUnifiedMemoryCounter  
 CUpti\_ActivityEnvironment  
 CUpti\_ActivityPreemption

**V**  
**value**  
 CUpti\_ActivityMemset  
 CUpti\_ActivityUnifiedMemoryCounter

CUpti\_ActivityDeviceAttribute  
CUpti\_ActivityMetricInstance  
CUpti\_ActivityMetric  
CUpti\_ActivityEventInstance  
CUpti\_ActivityEvent

## **Notice**

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE.

Information furnished is believed to be accurate and reliable. However, NVIDIA Corporation assumes no responsibility for the consequences of use of such information or for any infringement of patents or other rights of third parties that may result from its use. No license is granted by implication of otherwise under any patent rights of NVIDIA Corporation. Specifications mentioned in this publication are subject to change without notice. This publication supersedes and replaces all other information previously supplied. NVIDIA Corporation products are not authorized as critical components in life support devices or systems without express written approval of NVIDIA Corporation.

## **Trademarks**

NVIDIA and the NVIDIA logo are trademarks or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

## **Copyright**

© 2007-2014 NVIDIA Corporation. All rights reserved.