


# The runtime APIs:

---

- High-level interface to CUDA.
- Harness the power of NVIDIA GPUs. 
- Managing GPU devices, memory allocation, and execution of parallel kernels.
- Example:
  - The GPU's limits and constraints such the maximum number of threads per block.
  - **APIs provide functions** that allow you to query that value.

# The runtime APIs Key points:

- API's functions operated differently. 

```
__host__ __device__ cudaError_t cudaGetDeviceCount ( int* count )
```

Returns the number of compute-capable devices.

- Return an error status:

```
__host__ __device__ cudaError_t cudaGetDeviceCount ( int* count )
```

Returns the number of compute-capable devices.



- Saves time.



# The runtime APIs:

---

Hamdy Sultan

The course is prepared for [udemy.com](https://www.udemy.com)