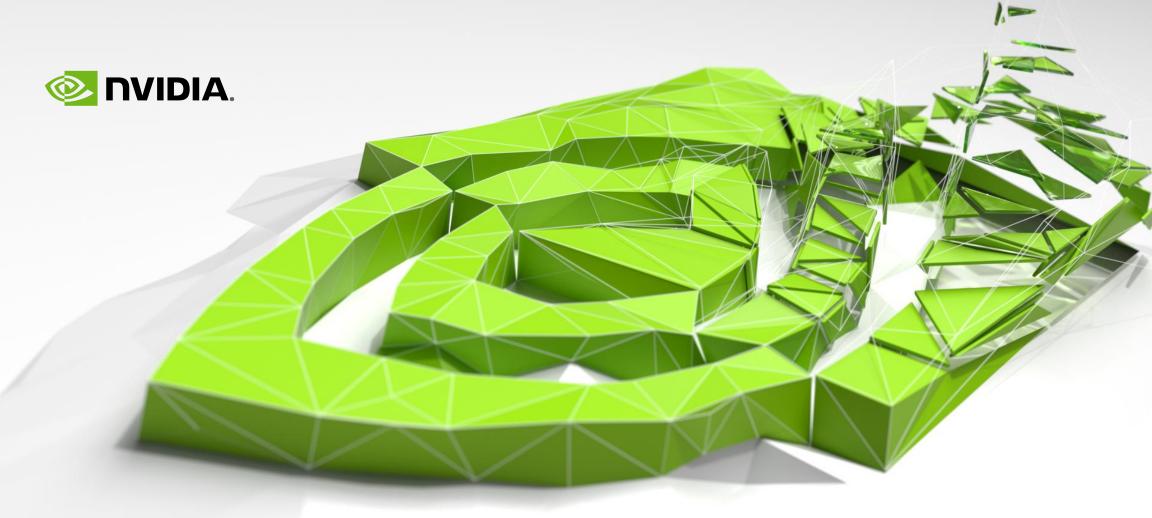
# Driver/Toolkit/Samples



### Software

- GPU Driver
- CUDA toolkit
  - Includes all the software necessary for developers to write applications
    - Compiler (nvcc), libraries, profiler, debugger, documentation
- CUDA Samples
  - Samples illustrating GPU functionality and performance
  - Examples illustrating important programming constructs and techniques.
- <u>www.nvidia.com/getcuda</u> -- all above software is free



# Example GPU hardware and driver

- nvidia-smi
  - -h for help
  - q for long query of all GPUs
  - PCle Bus ID
  - Driver Version
  - ECC state
  - Power State/fans/temps/clockspeed
- runit.nvidia-smi



### nvidia-smi

```
ubuntu@ip-10-97-186-7: ~/qpu-edu-workshops
                                                           ubuntu@ip-10-97-186-7:~/gpu-edu-workshops$ nvidia-smi
Wed Sep 23 21:56:15 2015
 NVIDIA-SMI 346.46 Driver Version: 346.46
 GPU Name Persistence-M| Bus-Id Disp.A | Volatile Uncorr. ECC |
 Fan Temp Perf Pwr:Usage/Cap| Memory-Usage | GPU-Util Compute M.
   0 GRID K520 On | 0000:00:03.0 Off |
                                                              N/A
 N/A 32C P8 17W / 125W | 10MiB / 4095MiB | 0%
                                                          Default
                                                        GPU Memory
 Processes:
          PID Type Process name
                                                        Usage
  No running processes found
ubuntu@ip-10-97-186-7:~/qpu-edu-workshops$
```

### **CUDA** Toolkit

- Compiler (nvcc)
- Libraries
  - BLAS, FFT, cuSPARSE, cuSOLVER, cuDNN, RNG, NPP
- Debuggers
  - cuda-gdb, cuda-memcheck
- Profilers
  - NVVP (GUI), nvprof (command-line)

### bash runit.bandwidth

```
ubuntu@ip-10-97-186-7: ~/qpu-edu-workshops
                                                                       - - X
ubuntu@ip-10-97-186-7:~/qpu-edu-workshops$ bash runit.bandwidth
[CUDA Bandwidth Test] - Starting...
Running on...
Device 0: GRID K520
Ouick Mode
Host to Device Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)
                                Bandwidth (MB/s)
                                9722.8
  33554432
Device to Host Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)
                                Bandwidth (MB/s)
  33554432
                                8895.3
Device to Device Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)
                                Bandwidth (MB/s)
  33554432
                                119157.8
Result = PASS
NOTE: The CUDA Samples are not meant for performance measurements. Results may v
ary when GPU Boost is enabled.
ubuntu@ip-10-97-186-7:~/qpu-edu-workshops$
```



# bash runit.query

```
- - X
ubuntu@ip-10-97-186-7: ~/gpu-edu-workshops
Device 0: "GRID K520"
                                                7.0 / 7.0
 CUDA Driver Version / Runtime Version
 CUDA Capability Major/Minor version number:
                                                 3.0
 Total amount of global memory:
                                                 4096 MBytes (4294770688 bytes)
  (8) Multiprocessors, (192) CUDA Cores/MP:
                                                1536 CUDA Cores
 GPU Max Clock rate:
                                                797 MHz (0.80 GHz)
 Memory Clock rate:
                                                 2500 Mhz
 Memory Bus Width:
                                                256-bit
 L2 Cache Size:
                                                524288 bytes
 Maximum Texture Dimension Size (x,y,z)
                                                1D=(65536), 2D=(65536, 65536),
3D=(4096, 4096, 4096)
 Maximum Layered 1D Texture Size, (num) layers 1D=(16384), 2048 layers
 Maximum Layered 2D Texture Size, (num) layers 2D=(16384, 16384), 2048 layers
 Total amount of constant memory:
                                                65536 bytes
 Total amount of shared memory per block:
                                                 49152 bytes
 Total number of registers available per block: 65536
 Warp size:
                                                 32
 Maximum number of threads per multiprocessor:
                                                2048
 Maximum number of threads per block:
                                                 1024
 Max dimension size of a thread block (x,v,z): (1024, 1024, 64)
 Max dimension size of a grid size
                                      (x,y,z): (2147483647, 65535, 65535)
 Maximum memory pitch:
                                                2147483647 bytes
 Texture alignment:
                                                 512 bytes
 Concurrent copy and kernel execution:
                                                 Yes with 2 copy engine(s)
```

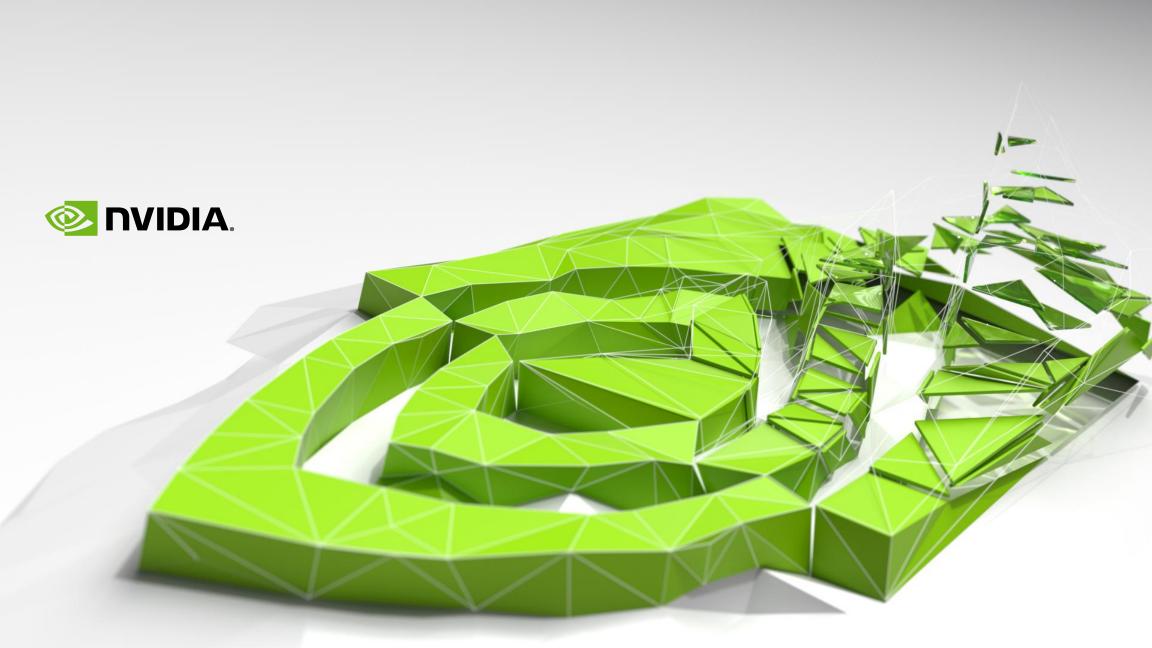


### bash runit.matmul

```
ubuntu@ip-10-97-186-7: ~/gpu-edu-workshops
                                                                       - - X
[Matrix Multiply Using CUDA] - Starting...
GPU Device 0: "GRID K520" with compute capability 3.0
MatrixA(320,320), MatrixB(640,320)
Computing result using CUDA Kernel...
done
Performance = 192.73 GFlop/s, Time= 0.680 msec, Size= 131072000 Ops, WorkgroupSiz
e= 1024 threads/block
Checking computed result for correctness: Result = PASS
NOTE: The CUDA Samples are not meant for performance measurements. Results may v
ary when GPU Boost is enabled.
[Matrix Multiply CUBLAS] - Starting...
GPU Device 0: "GRID K520" with compute capability 3.0
MatrixA(320,640), MatrixB(320,640), MatrixC(320,640)
Computing result using CUBLAS...done.
Performance = 880.97 GFlop/s, Time= 0.149 msec, Size= 131072000 Ops
Computing result using host CFU...done.
Comparing CUBLAS Matrix Multiply with CPU results: PASS
NOTE: The CUDA Samples are not meant for performance measurements. Results may v
ary when GPU Boost is enabled.
ubuntu@ip-10-97-186-7:~/gpu-edu-workshops$
```

# Recap

- Driver
  - nvidia-smi to query GPU hardware and state
- CUDA Toolkit
  - Development tools for GPU software development
- CUDA Samples
  - Sample code and diagnostic tests





# CUDA Samples (bundled with Toolkit)

- ~/NVIDIA\_CUDA-7.0\_Samples/bin/x86\_64/linux/release
- Sample programs to illustrate GPU programming constructs and algorithms

Useful diagnostic tests to query GPU and its performance

cd gpu-edu-workshops

# Content slide: 36 pt black, trebuchet font, up to 2 lines maximum, capitalize only first word

Subtitle: 24 pt, one line maximum

Body/bullet text no longer has a bullet icon

Use 20 pt font

No sub-bullets allowed

No more than five bullets; one idea per bullet

Example of highlighted text

# Same as previous slide

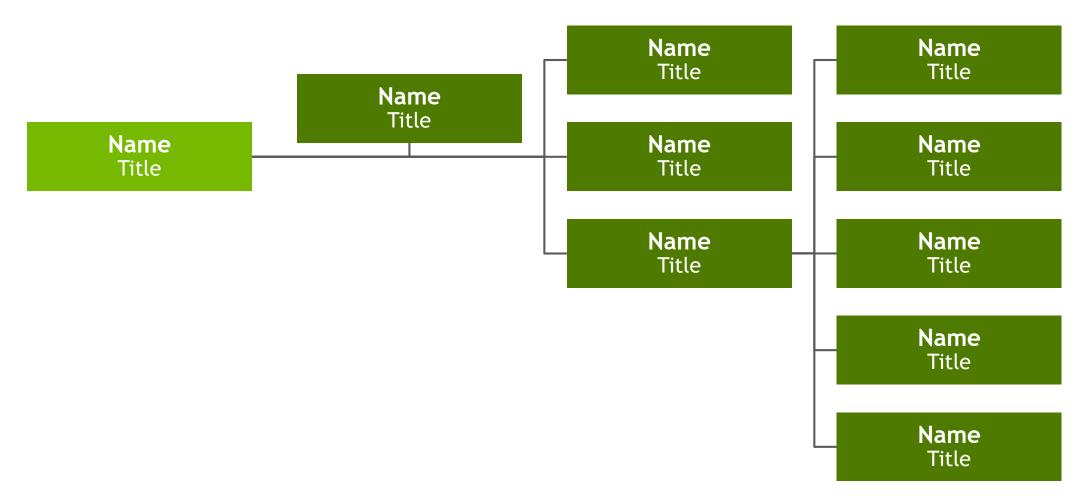
But includes logo and page number in lower right corner

### **NVIDIA** confidential

Use this for slides that include "confidential" information/data

# Title only slide

# Org chart example



# Agenda

Action item number one

Action item number two

Action item number three

Action item number four

Action item number five

# Agenda

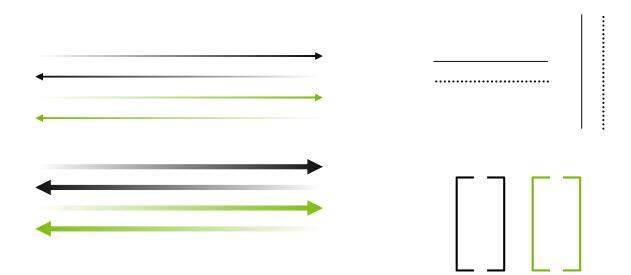
9:00 am - 9:30 am	Welcome	Presenter Name
9:30 am - 10:30 am	Keynote	Presenter Name
10:30 am - 11:30 am	GPU/gaming	Presenter Name
11:30 am - 12:30 pm	Break	All
12:30 pm - 1:30 pm	Automotive	Presenter Name
1:30 pm - 2:30 pm	Enterprise	Presenter Name
2:30 pm - 3:30 pm	Financials	Presenter Name
3:30 pm - 5:00 pm	Q&A	All

# Section divider or transition slide

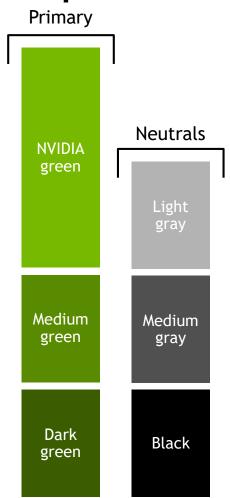
# Graphic styles

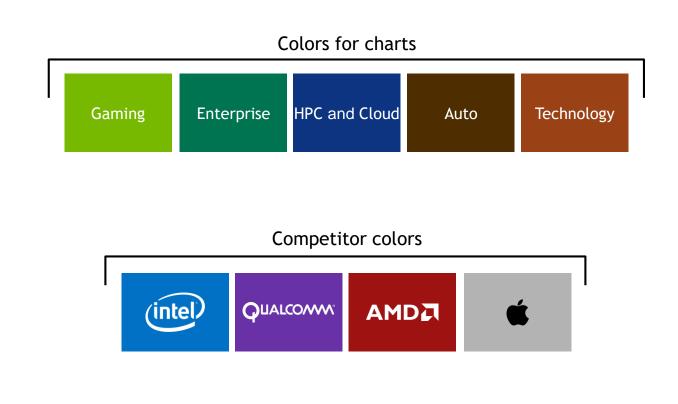
Green Box Style

Gray Box Style



# Color palette



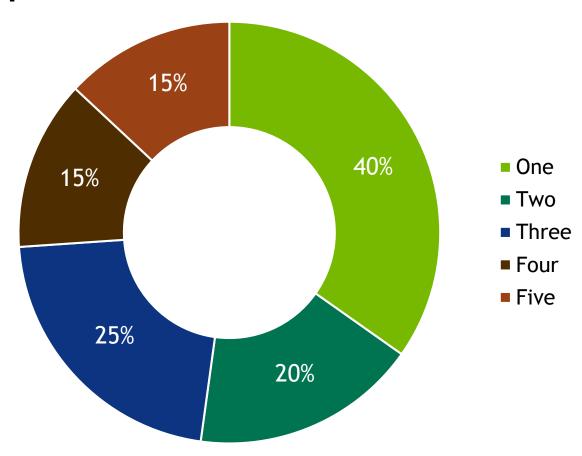


# Table sample

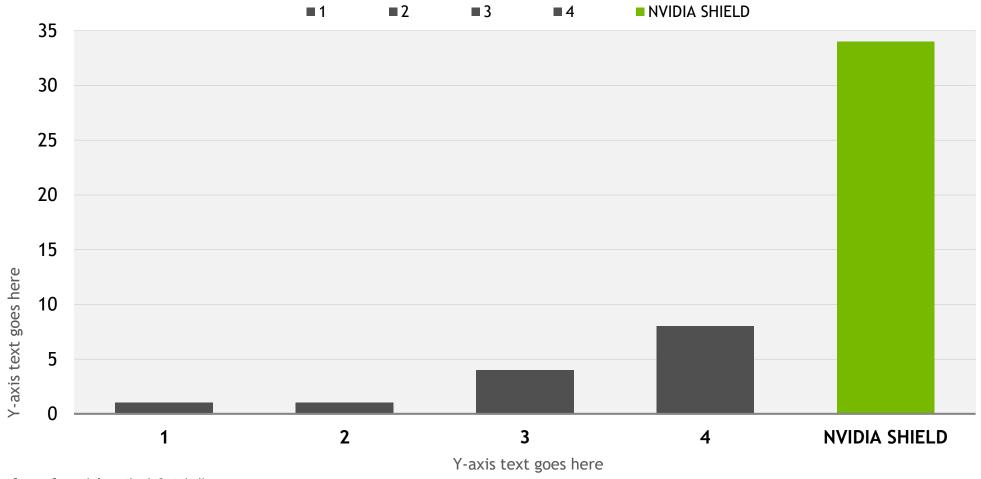
	CATEGORY ONE	CATEGORY TWO	CATEGORY THREE
Trebuchet bold, 18 pt	Trebuchet, 16 pt	Text	Text
Row one	Text	Text	Text
Row two	Text	Text	Text
Row three	Text	Text	Text

Source: Source information is 8 pt, italic

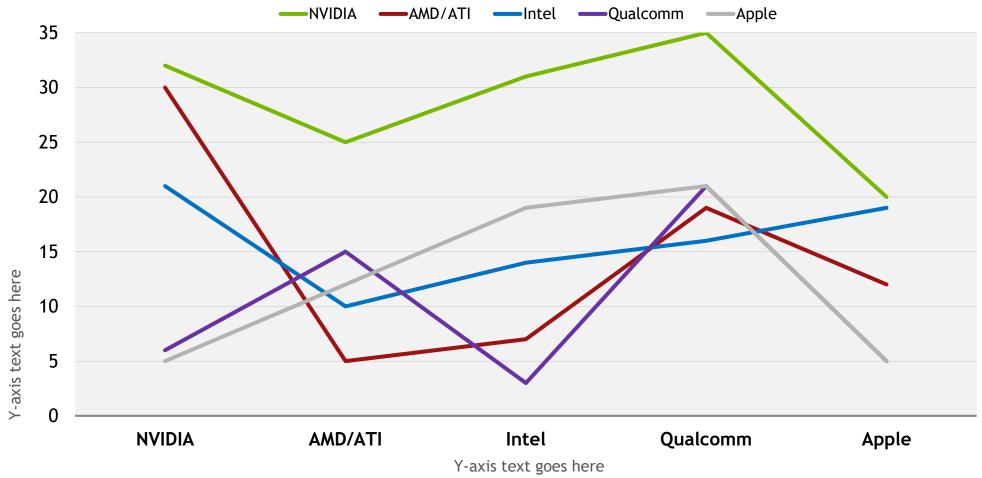
# Pie chart sample

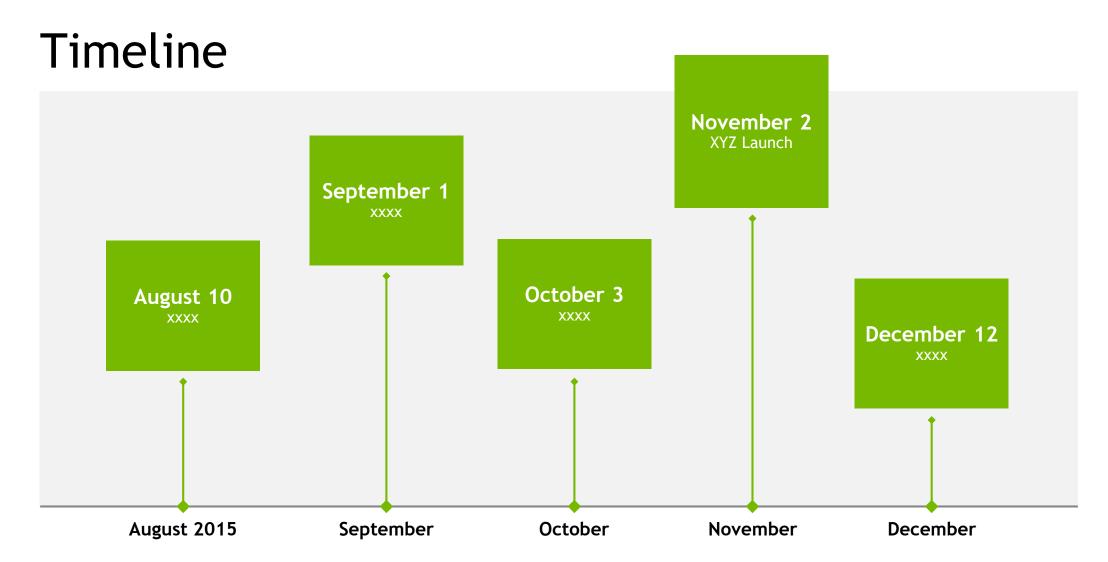


# Bar chart sample



# Line chart sample



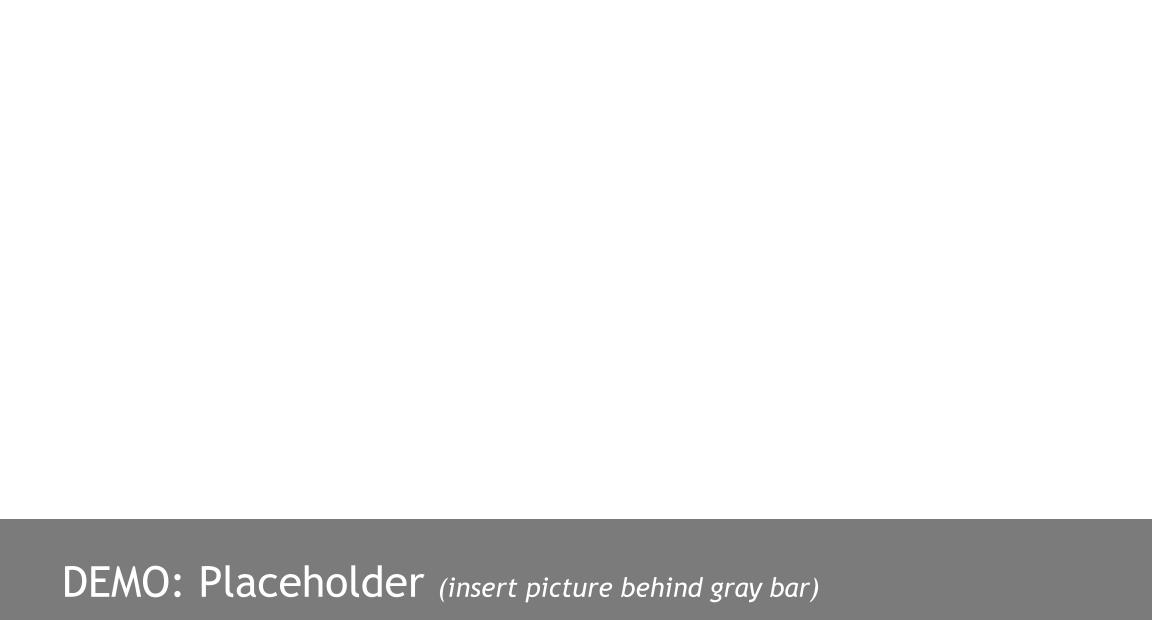


Inserting video: Insert/Video/Video from File. Insert video by browsing your directory and selecting OK.

File type that works best in PowerPoint is: .wmv



### Video file



# Compare or contrast sample

### Column one

Sample text placed here

Best to keep information no more than 5 bullet points long

Should be used to distinguish buckets of information across multiple columns

### Column two

Sample text placed here

Best to keep information no more than 5 bullet points long

Should be used to distinguish buckets of information across multiple columns

### Column three

Sample text placed here

Best to keep information no more than 5 bullet points long

Should be used to distinguish buckets of information across multiple columns

# Box diagram

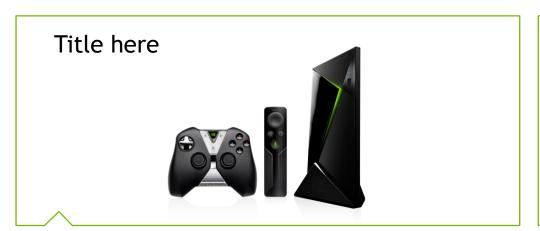
Text only descriptors

### Title here

Supporting text here

# Box diagram

### Title with images









## Product slide



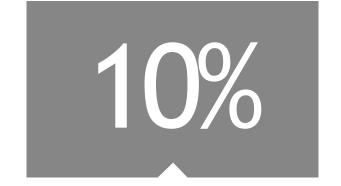
# Stat example

65%

Supporting text goes here
[Use green for POSITIVE]



Text goes here
[Use red for NEGATIVE]



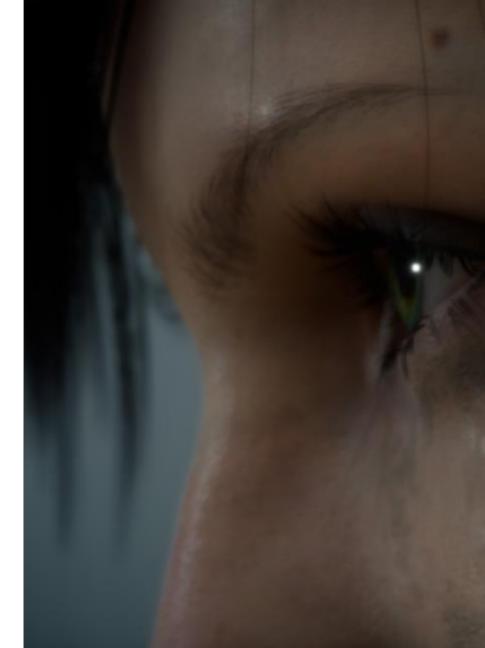
Text goes here
[Use gray for NUETRAL]



# Text slide with photograph Optional subtitle

Text goes here

Text goes here



# Text slide with photographs

Multiple images

Allow a consistent "gutter" space between images

Use a gray outline (R: 90, G: 90, B: 90) at 0.5 pt







"Quoted text example goes here. Trebuchet, 42 pt. Use with shorter (headline type) quotes."

Quote Source, Trebuchet 20 pt

"Quoted text example goes here. Trebuchet, 42 pt. Use with shorter (headline type) quotes."

### "Multi-quote slide sample."

- Source: Either a name or publication text here, OR, a company logo to the right



### "Multi-quote slide sample."

- Source: Either a name or publication text here, OR, a company logo to the right



### "Multi-quote slide sample."

- Source: Either a name or publication text here, OR, a company logo to the right



# Customer success story

### Optional subtitle



### Challenge

Ei eos viris putant, veritus deserunt et qui.

Ea his diam labore prodesset, ei justo atomorum per.

Ei ullum utinam recteque cum. Quem tota dici

### **Solution**

Ei eos viris putant, veritus deserunt et qui.

Ea his diam labore prodesset, ei justo atomorum per.

Ei ullum utinam recteque cum. Quem tota dici

### **Impact**

Ei eos viris putant, veritus deserunt et qui.

Ea his diam labore prodesset, ei justo atomorum per.

Ei ullum utinam recteque cum. Quem tota dici







