

A silhouette of a person in mid-air, jumping over a gap between two dark, jagged rock formations. The background is a bright, hazy sky with soft clouds, suggesting a sunset or sunrise. The overall mood is one of achievement and overcoming challenges.

# Gaps Between Big Data and Big Compute

**Steve Scott**

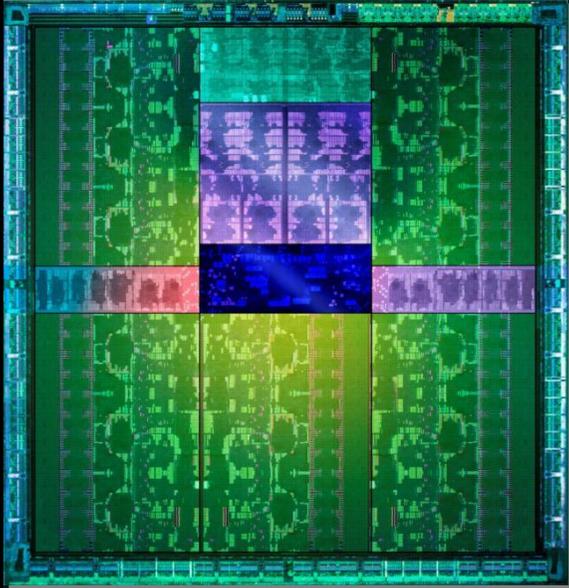
**CTO Tesla**

**SOS17**

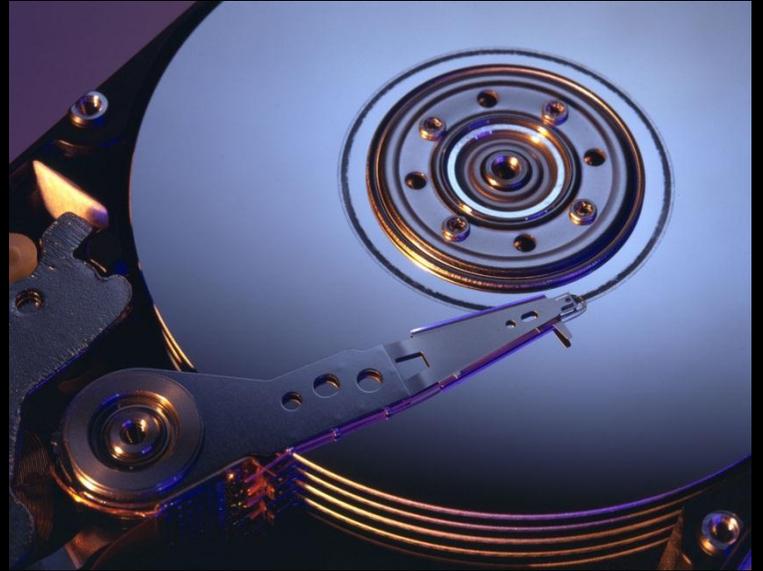
**March 26, 2013**

# Big Data

Data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process the data within a tolerable elapsed time.



& &



# Lots of Common Needs

- Compute power
- Interconnect bandwidth
- Global file system
- Memory
- Visualization
- Scaling
- Resiliency
- etc.



# A Matter of Balance

Network  
Bandwidth

File System  
Capacity & Bandwidth



Compute

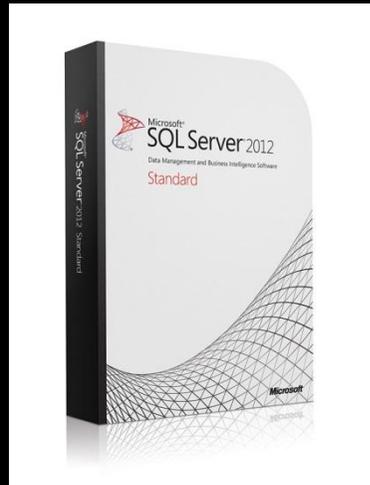
Memory

We have these same trade-offs *within* HPC

# Making Supercomputers Better for Big Data

- More network bandwidth (injection and global)
- More file system bandwidth (possibly distributed)
- Larger memories (likely add layer of flash/NV RAM)
- **And for Big Graph problems:**
  - Global address space with TLB support to map whole machine concurrently
  - Network that excels at single-word references
  - Lots of thread-level parallelism to tolerate global latencies with low concurrency per thread
  - Lightweight synchronization
  - Sophisticated runtime to manage parallelism

# Much Greater Gap is in Software...





 NVIDIA®