



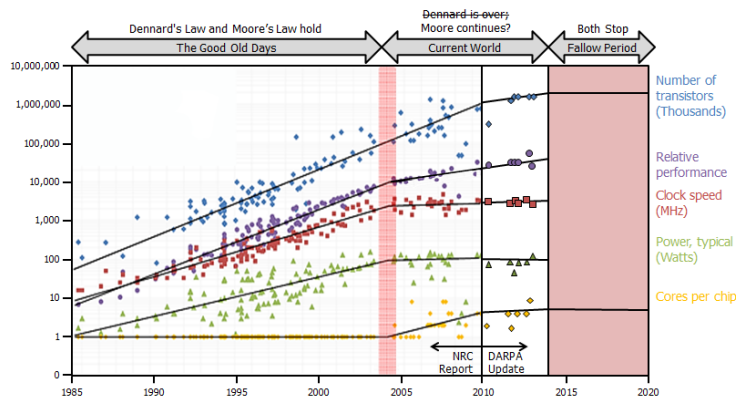
New Approaches for Streaming Data Analytics



ALTERA[®]

Industry Trends: Processing Trends

DARPA The Fallow Period – Where do advancements come from now?

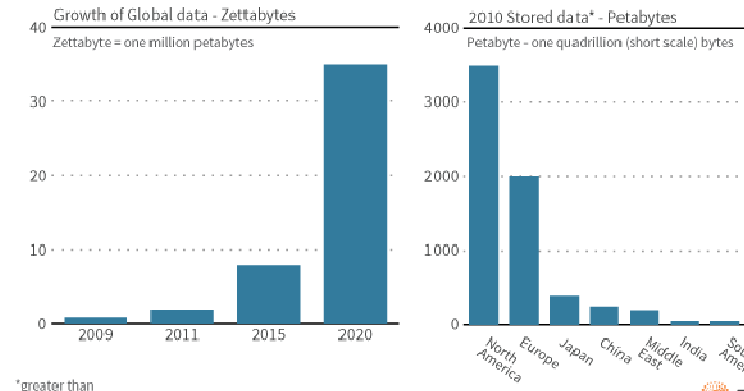


DISTRIBUTION STATEMENT: Distribution authorized to U.S. Government Agencies and their contractors for the Altera Meeting on 7 January 2015. Other requests for this document shall be referred to Microsystems Technology Office, Defense Advanced Research Projects Agency.

4

Big data growth

Big data market is estimated to grow 45% annually to reach \$25 billion by 2015



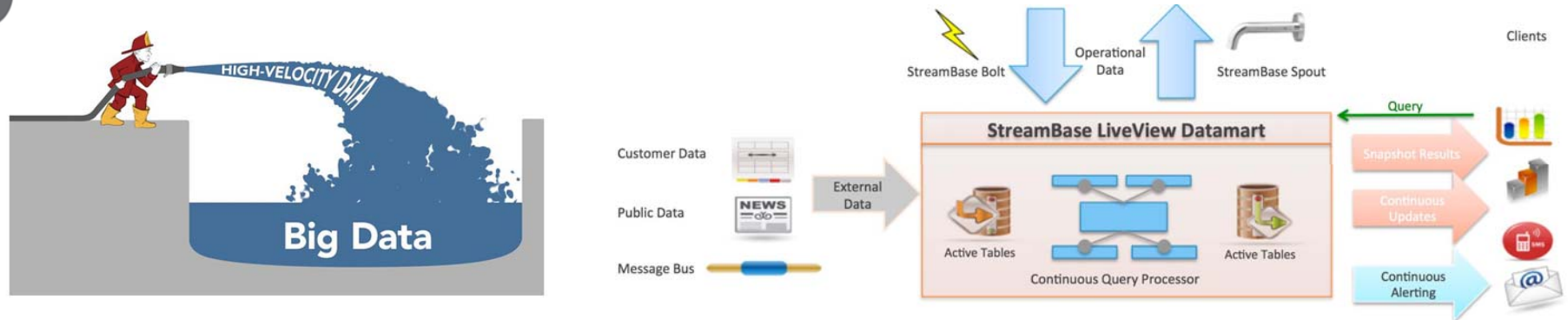
*greater than
Sources: Nasscom - CRISIL GR&A analysis

Reuters graphics/Catherine Trevelyan 05/10/12

REUTERS

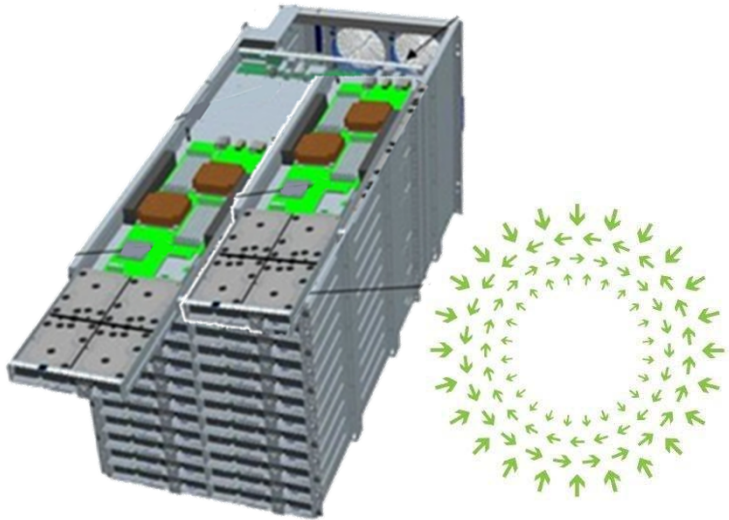
- Decreasing Processor Speeds
- Exponential Data Growth
- Means large distributed processing nodes

Industry Trends: High Velocity Data Driving Need For Real Time Processing



- ◀ Decisions are increasingly made in real time
 - Car Insurance Quotes
 - Fraud Detection
 - Risk Analysis
- ◀ Batch processing depending on frequency
 - Limits the relevancy of data – “More historical”
 - Limits the frequency and timing of decisions
 - Consumes power

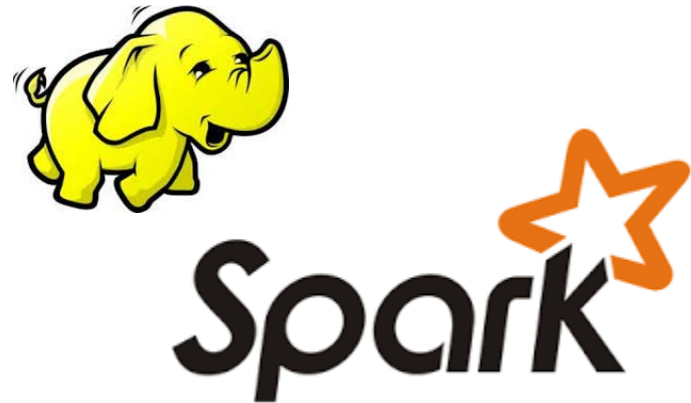
Current solution



OPEN
Compute Project

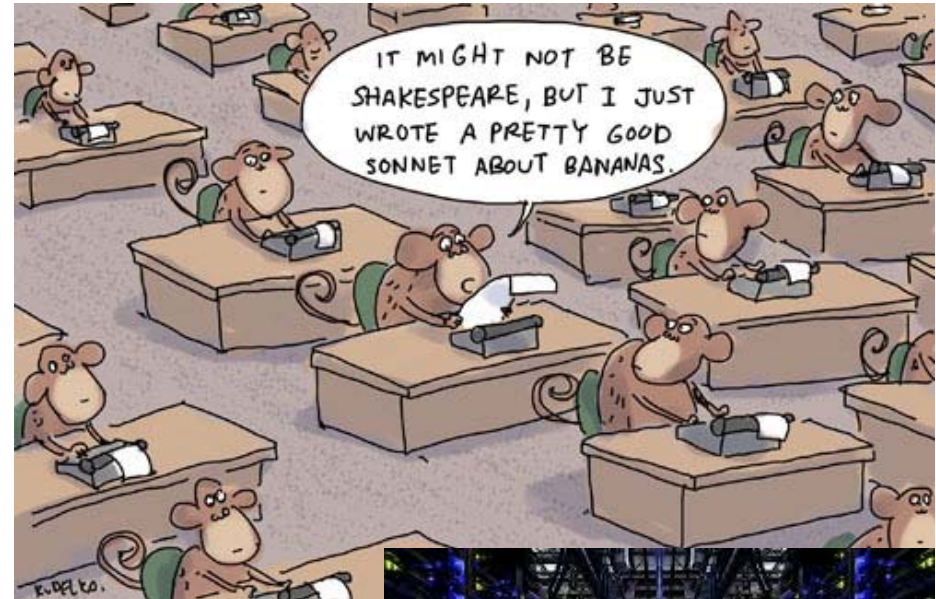
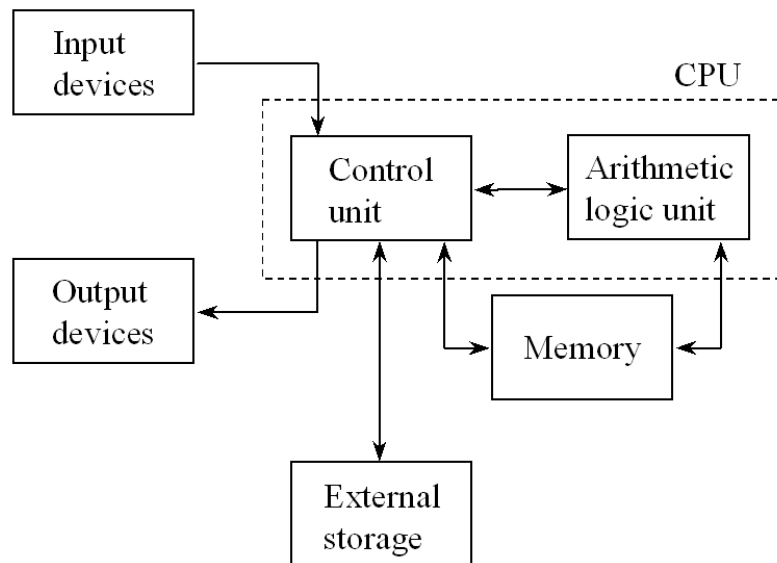
Large homogenous networks
of
cheap hardware

Running smarter software

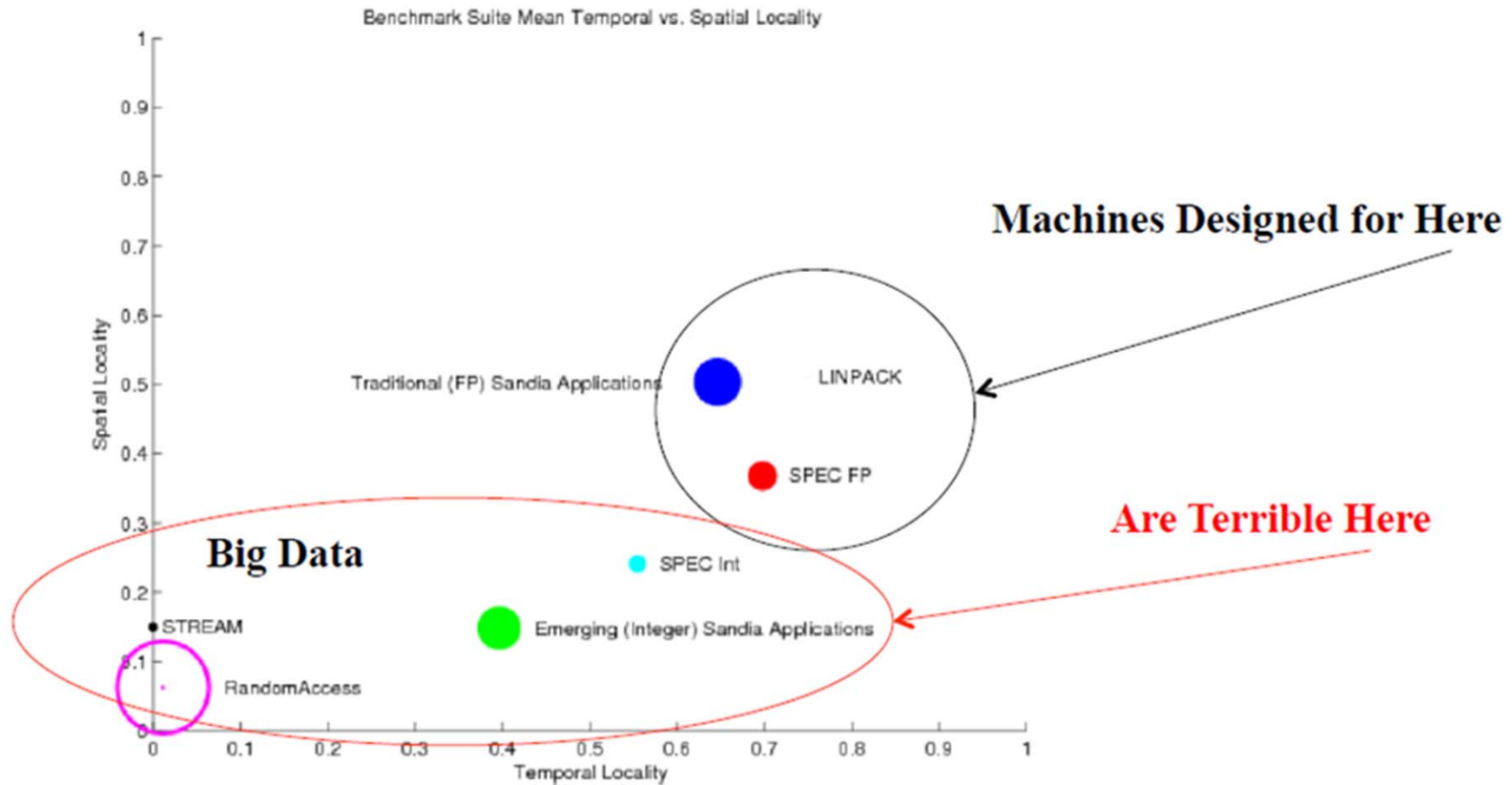


Need to Re-think our Approach to Computing

Von Neumann computer architecture



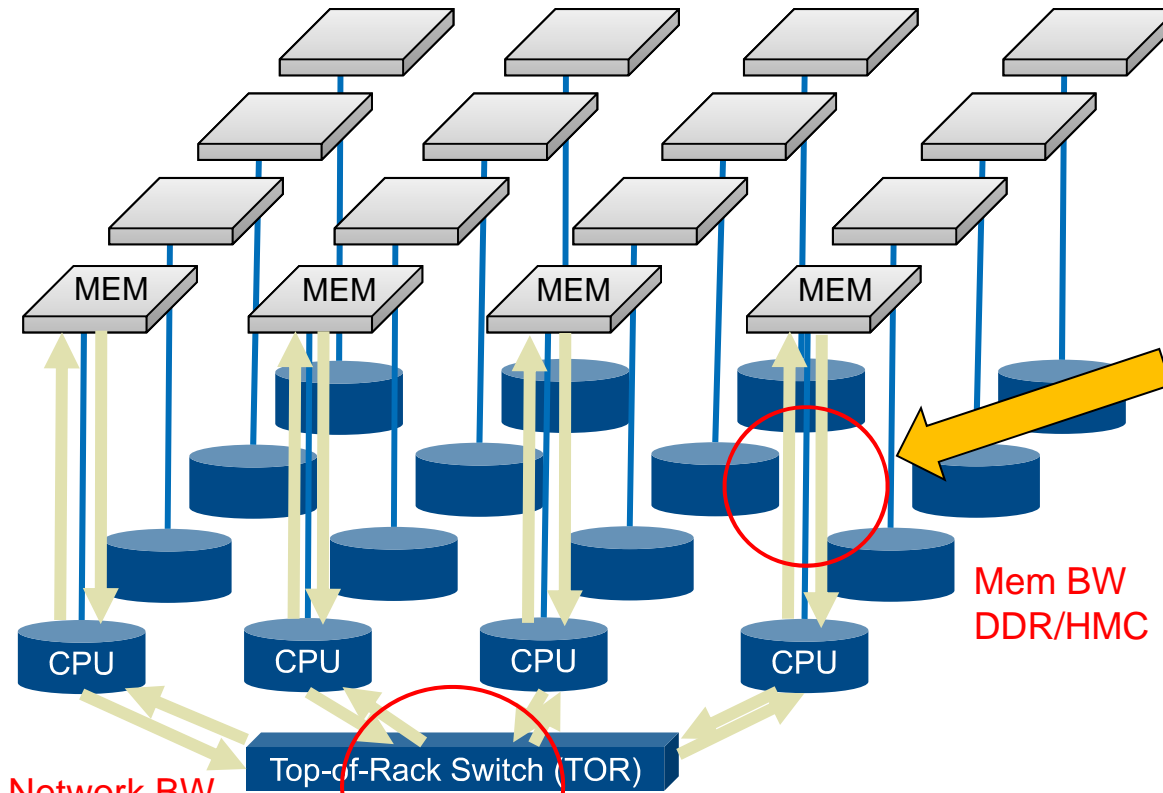
Technical Issues: Application Space is Changing



Murphy, Kogge. On The Memory Access Patterns of Supercomputer Applications: Benchmark Selection and Its Implications, IEEE TC, 7/07

Technical Issues: Mapping Large Datasets on Traditional Networks is Difficult

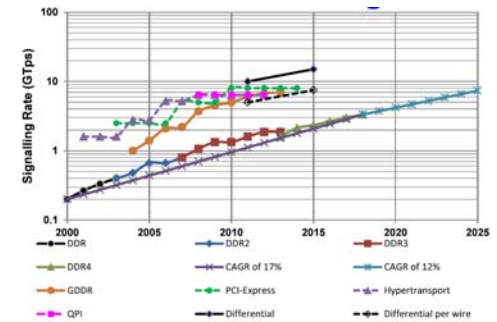
Traditional Network



Network BW
10GbE

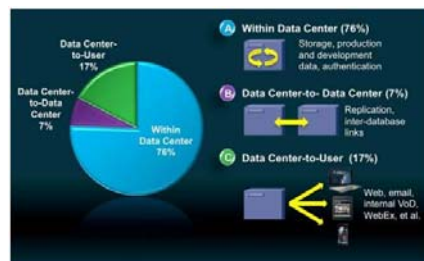
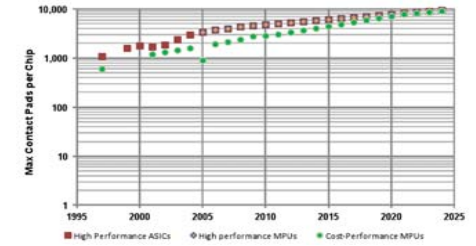
Mem BW
DDR/HMC

IO signaling speed



Number of Memory Interfaces

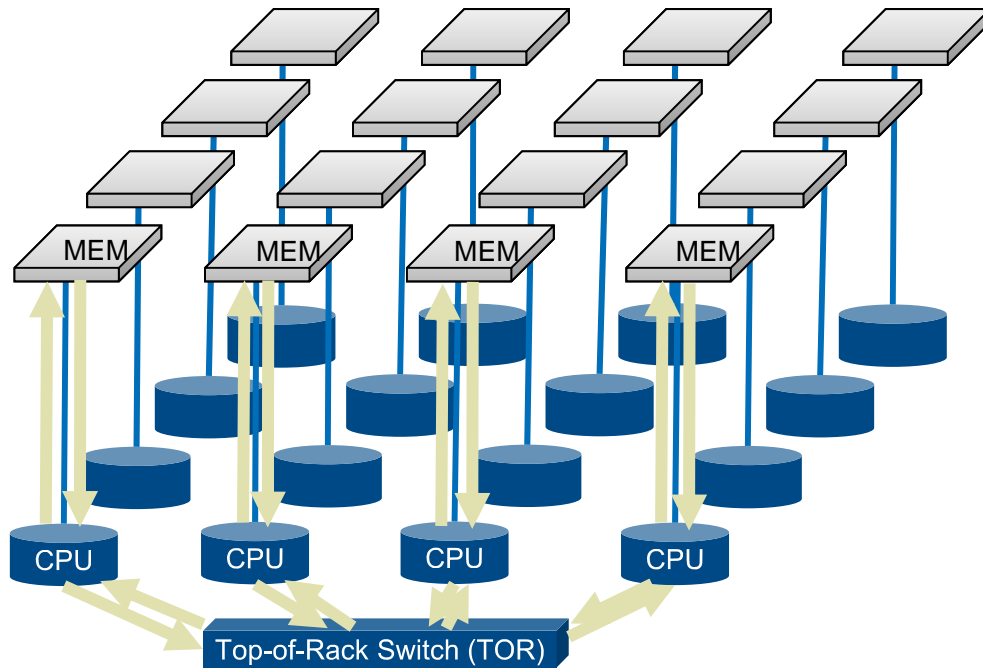
But Growth In Chip I/O is at Best Slow



Exponential
East/West Growth

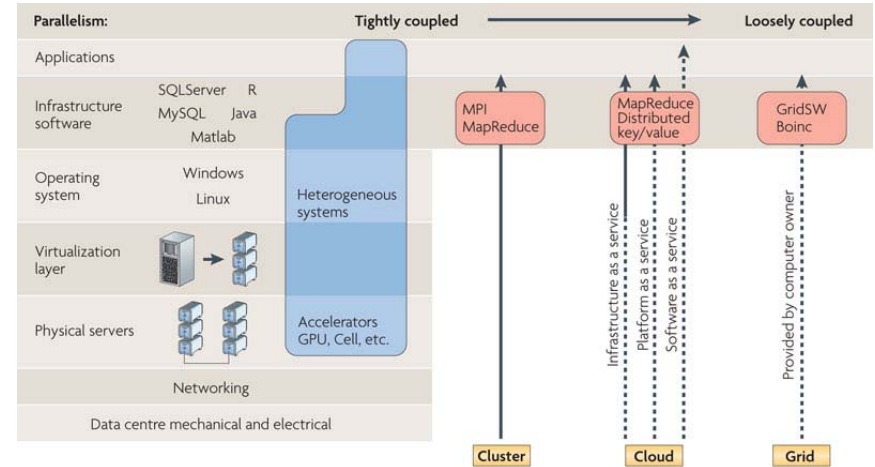
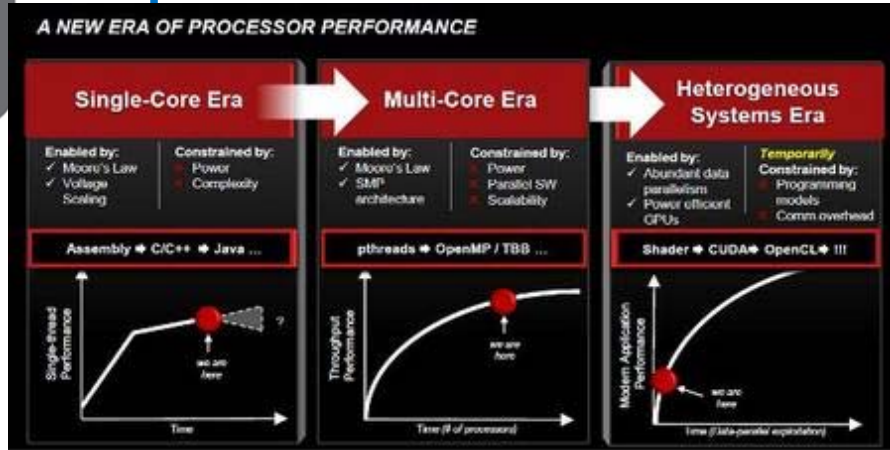


Technical Issues: Cloud Computing- a Sea of Processors/ Islands of Memory



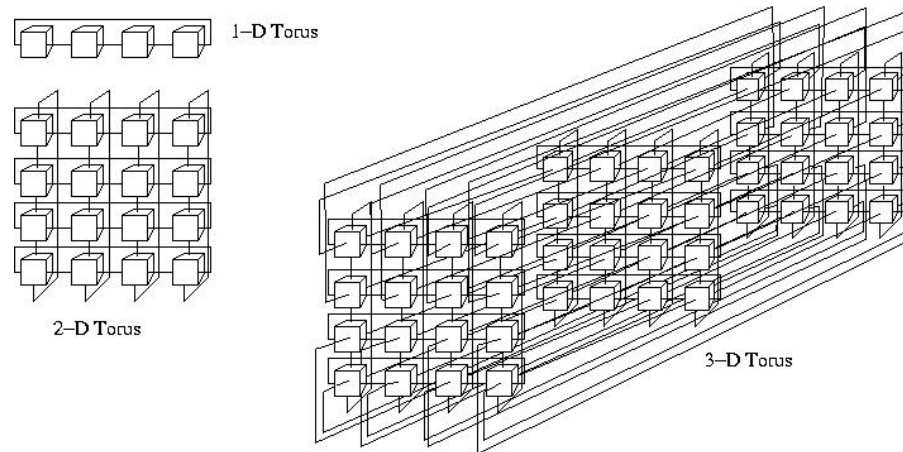
- Virtualization creates seas of processors
 - Hyper threading allows distributing loads
 - Processors are dynamically allocated
- Memory is static
 - No coherency between memory elements
 - Can not be dynamically allocated

Specialization is the Answer



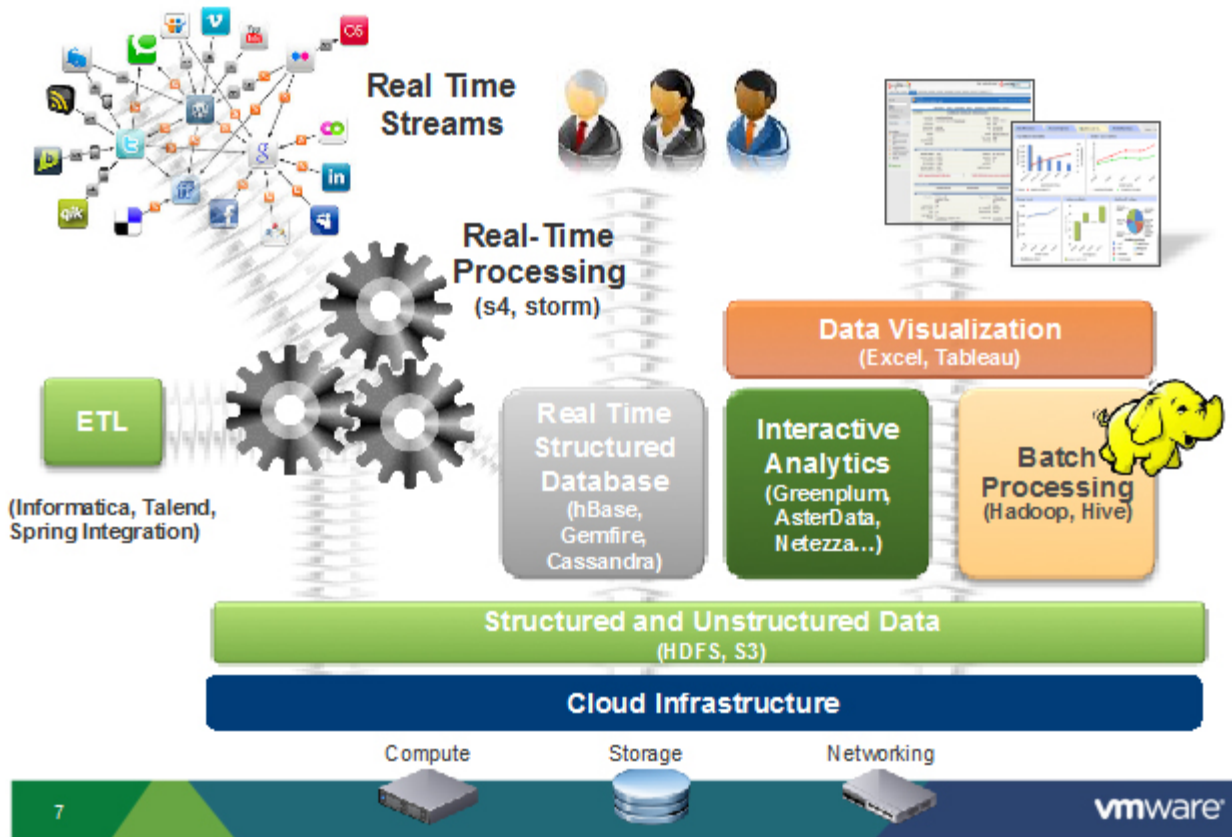
Large Datasets require massive parallelism

A sea of function specific accelerators



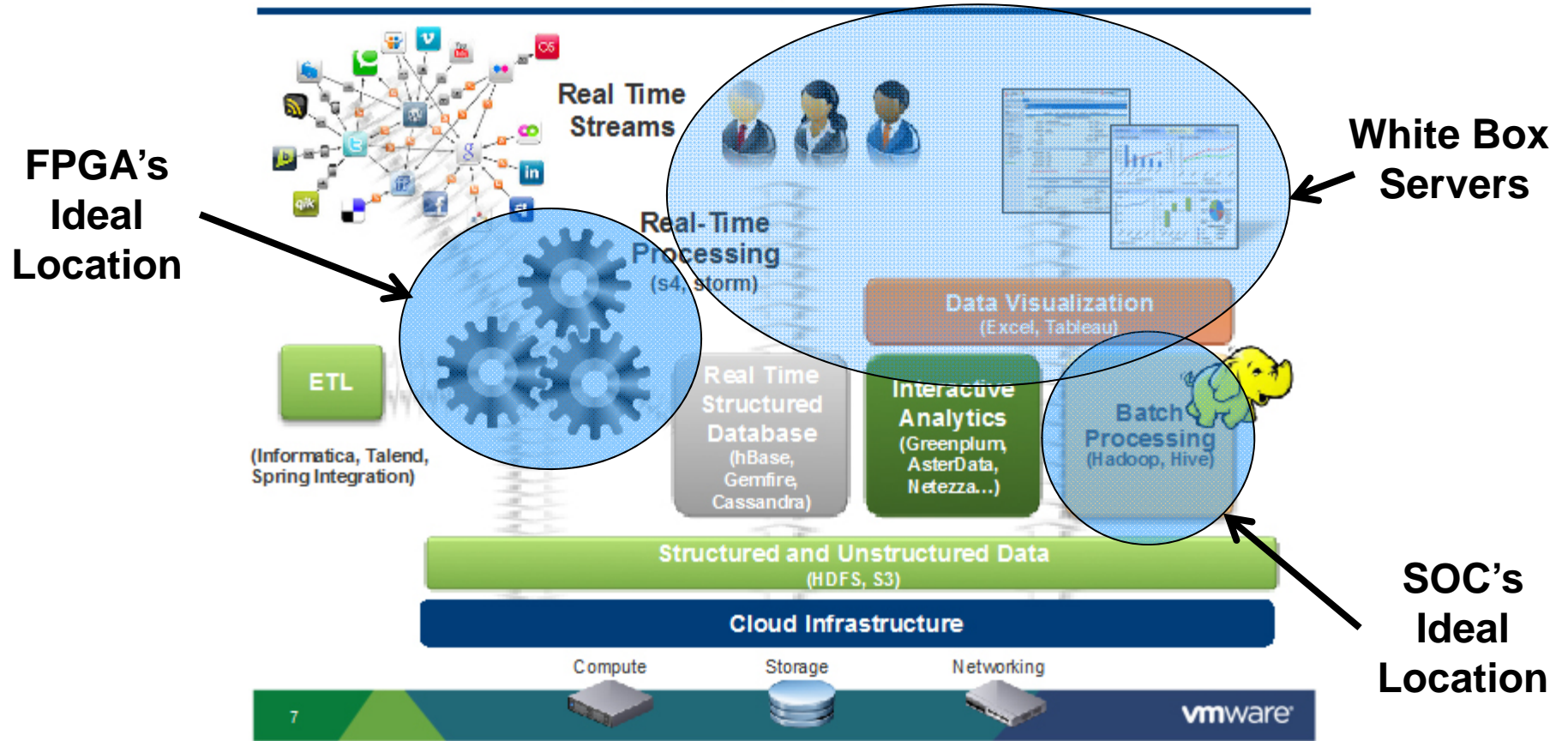
FPGAs in Big Data: Real Time Processing

A Holistic View of a Big Data System



FPGAs in Big Data: Real Time Processing and Data Reduction

A Holistic View of a Big Data System



Current Deployments



Microsoft Open Compute Server

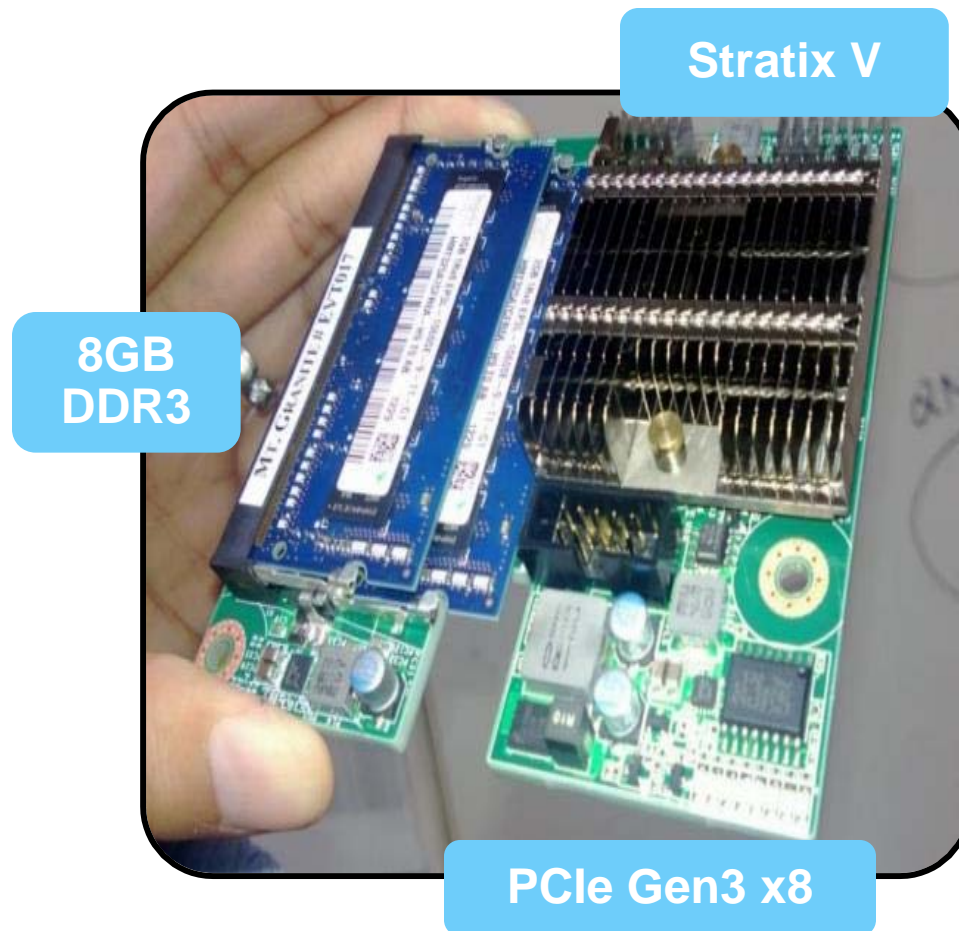


- Two 8-core Xeon 2.1 GHz CPUs
- 64 GB DRAM
- 4 HDDs, 2 SSDs
- 10 Gb Ethernet

Air flow

Catapult FPGA Accelerator Card

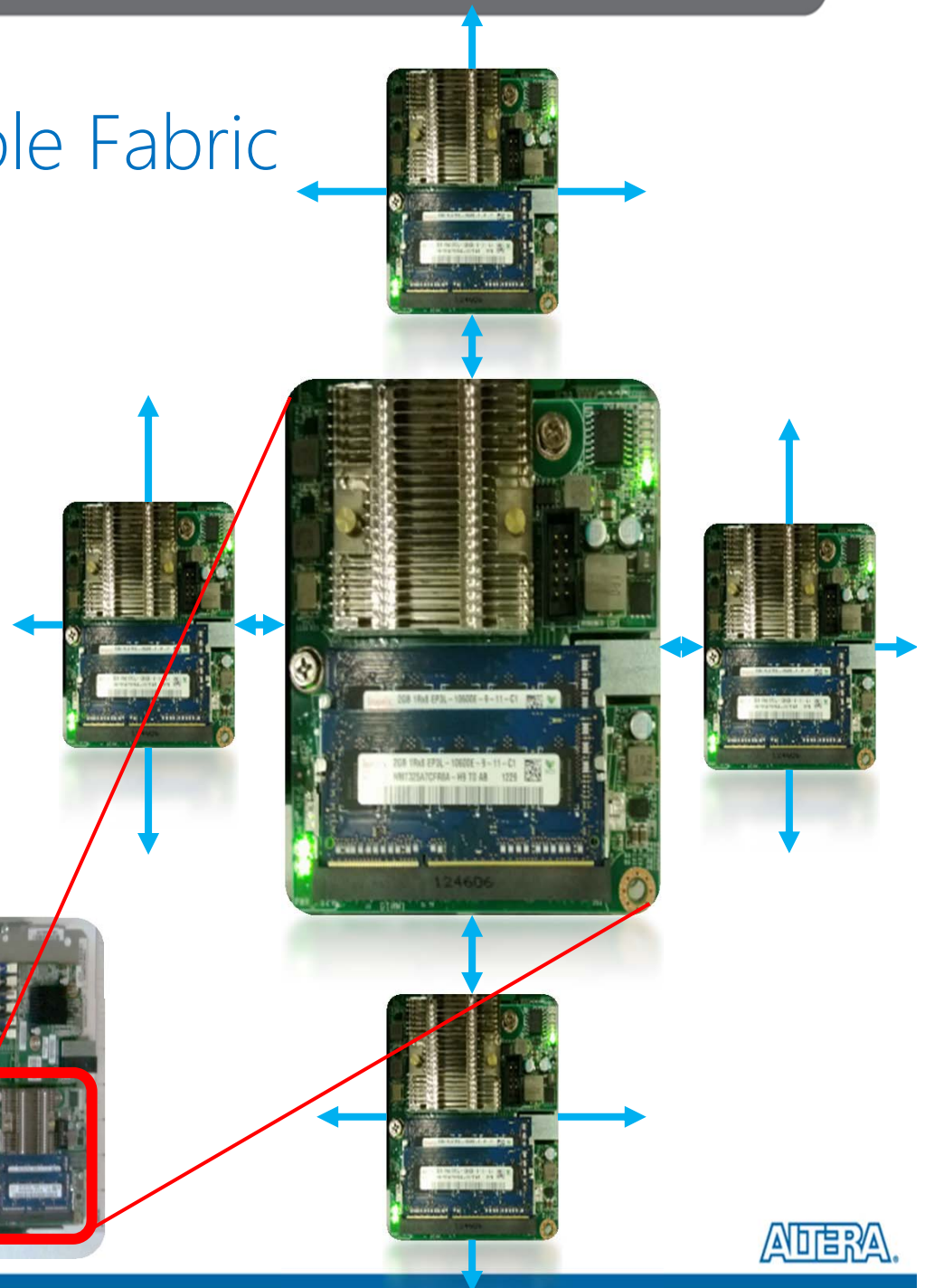
- Altera Stratix V D5
- 172,600 ALMs, 2,014 M20Ks, 1,590 DSPs
- PCIe Gen 3 x8
- 8GB DDR3-1333
- Powered by PCIe slot
- Torus Network



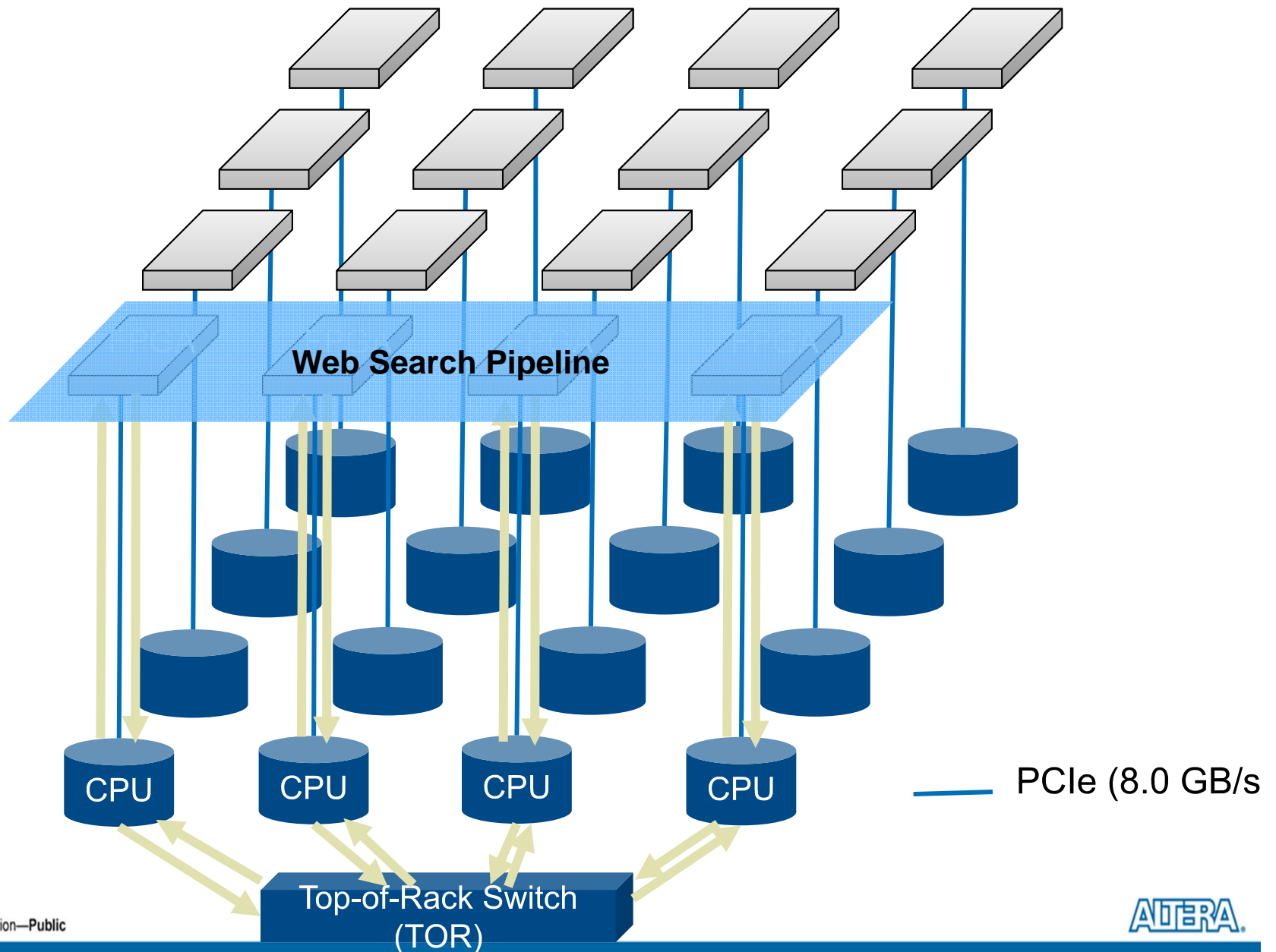
Scalable Reconfigurable Fabric

- 1 FPGA board per Server
- 48 Servers per ½ Rack
- 6x8 Torus Network among FPGAs
 - 20 Gb over SAS SFF-8088 cables

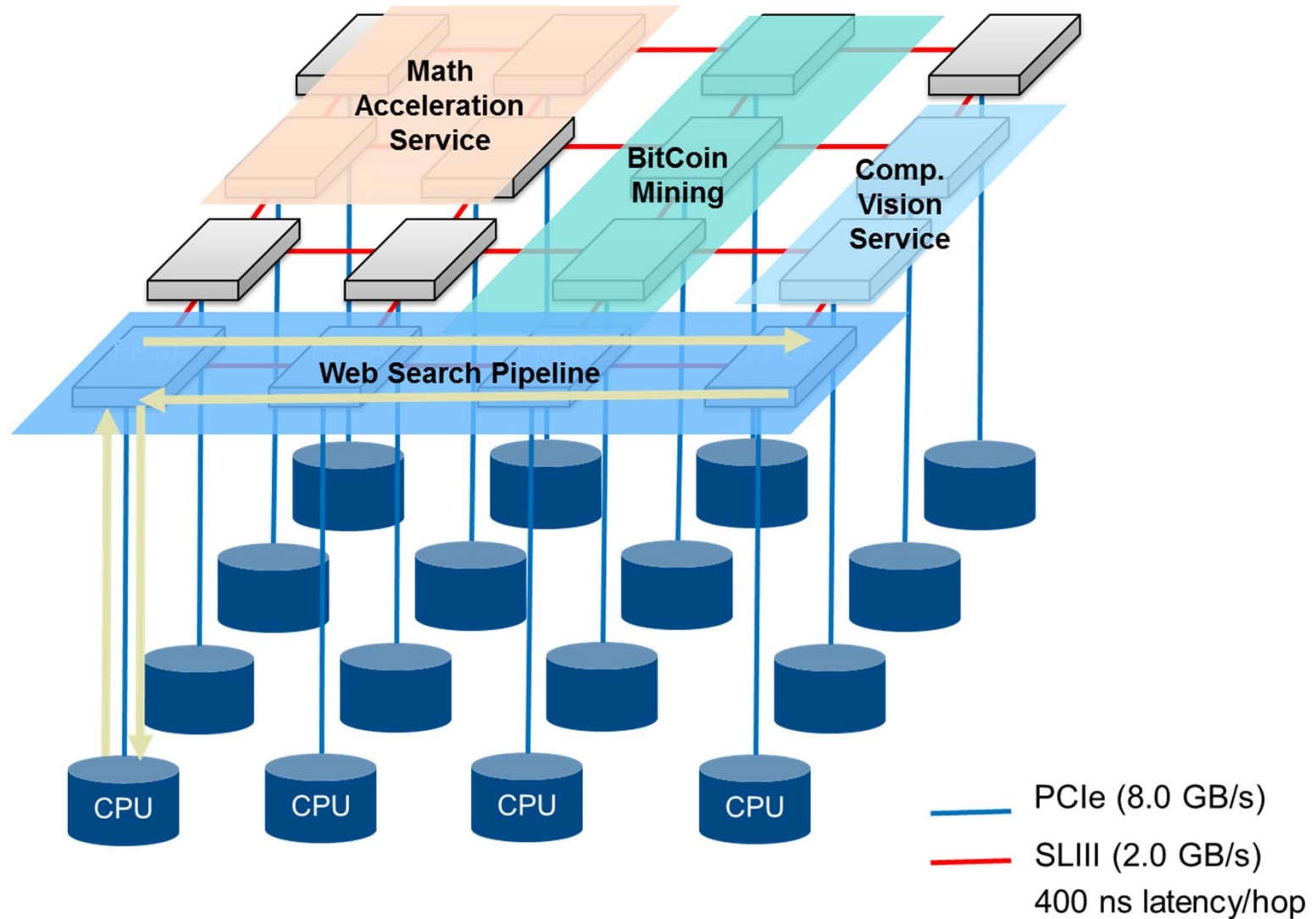
Data Center Server (1U, ½ width)



An Elastic Reconfigurable Fabric

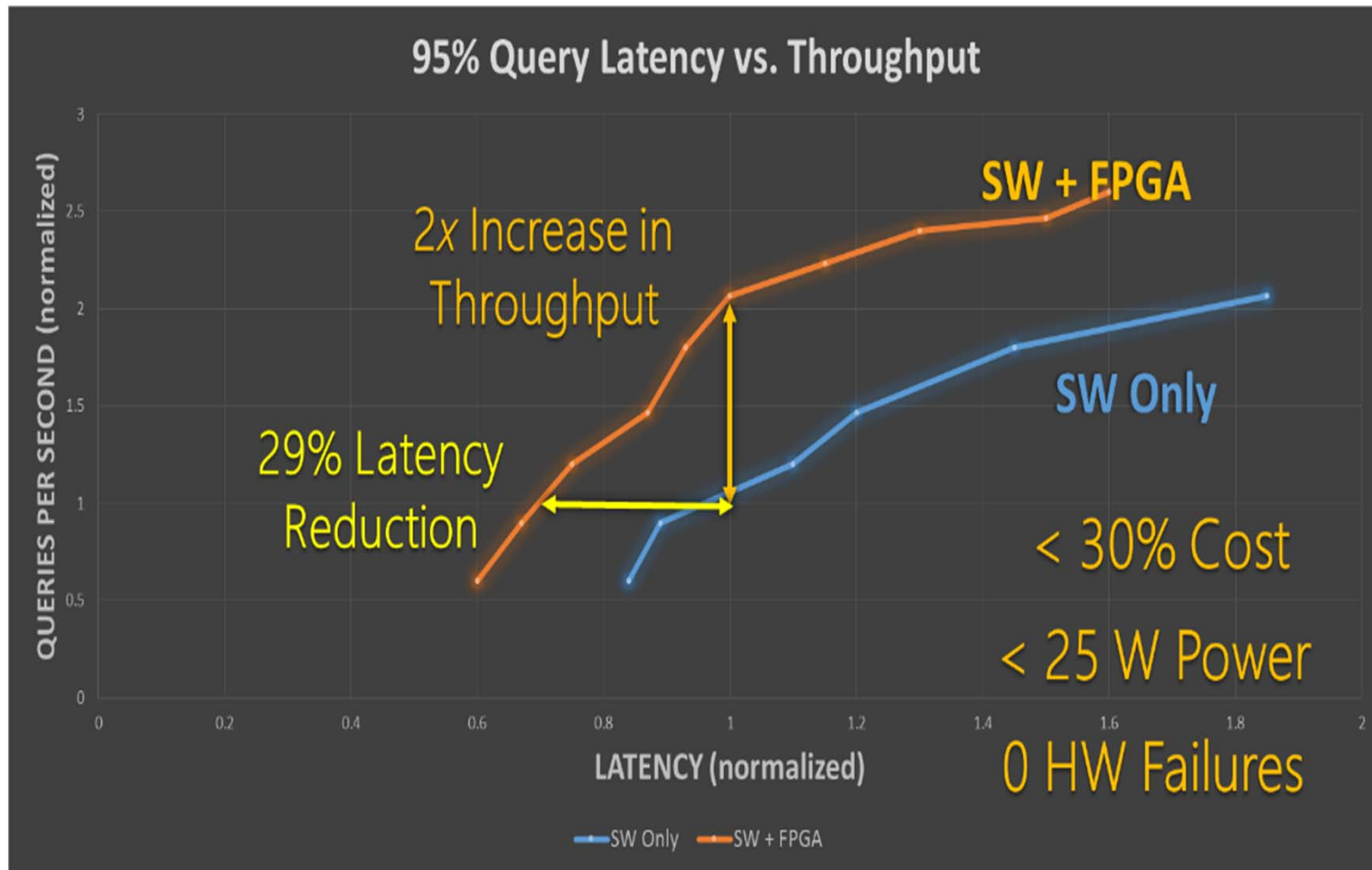


An Elastic Reconfigurable Fabric



Accelerating Large-Scale Services – Bing Search

1,632 Servers with FPGAs Running Bing Page Ranking Service (~30,000 lines of C++)



Summary

- ◀ Computing is changing
 - Higher data volume and velocity
 - Need for real time data analytics
 - End of Moore's Law
- ◀ Traditional Von Neumann approaches won't work
- ◀ Need to look at heterogeneous architectures