Worlds Colliding:
Why Big Data Changes How To Think About Enterprise Storage

Intersect360
RESEARCH

Addison Snell
addison@intersect360.com

Intersect360
RESEARCH

Overview, Mostly in Order

- Okay, who invited the HPC guy?
- The hazards of “forecasting” Big Data
- What is Big Data, if not Hadoop?
- What’s the real opportunity?
- (Vendor) hype vs. (end user) reality
- The satisfaction gap
- Worlds colliding = bad. Trends colliding = worse.
- The big finish: I don’t care how you solve the problem. Just don’t call it HPC!

Intersect360 RESEARCH

Technical vs. Enterprise Computing

Technical Computing

- Top-line missions:
  - Find the oil
  - Design the minivan
  - Cure the disease
- Driven by price/performance
- Fast adoption of new technologies, algorithms, and approaches

Enterprise Computing

- Keeps business running
  - Communicate/collaborate
  - Market and sell the product
  - Accounting, HR, finance, …
- Driven by RAS: reliability, availability, serviceability
- Slow adoption of new technologies, algorithms, and approaches

New Survey Data

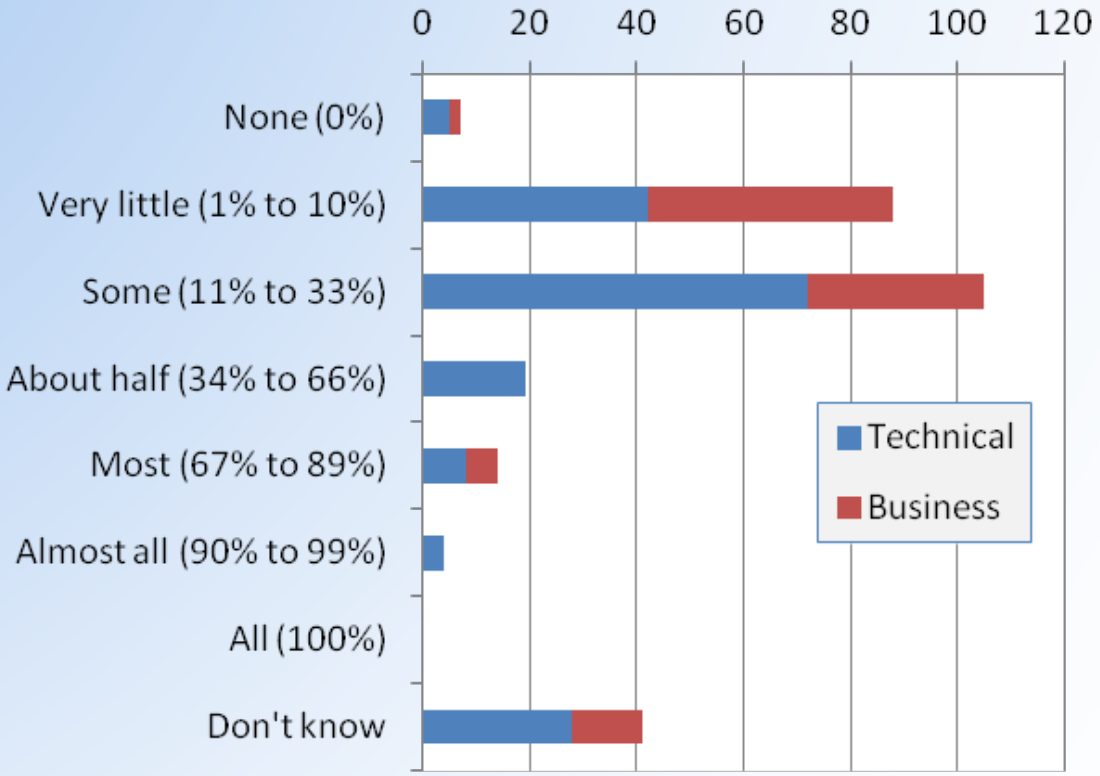
- 278 total respondents
  - 178 “Technical” (HPCwire and HPC500 user group)
  - 100 “Enterprise” (Gabriel Consulting)
  - 165 commercial, 67 academic, 46 government
- Surveys completed April – August 2013
- Builds on original survey from early 2012 (306 respondents: 204 Technical, 102 Business)
- End users discuss their environments, challenges, solutions, and “satisfaction gaps” in addressing Big Data challenges

Intersect360 RESEARCH

Insight #1: Big Data, Big Opportunity

What percent of your organization's IT budget in 2013 will be related to Big Data?

- Money is being spent on Big Data
- 60% of those responding will spend more than 10% of the **IT budget** relate to Big Data
- Use caution in describing "the Big Data market"



Intersect360 RESEARCH

What Are Big Data Applications?

- When we ask vendors to describe their Big Data solutions, they use words like *Hadoop* and *graph*
- When users describe their Big Data applications, they use words like *analyze* and *algorithm*

Application usage	
Internal	281
Purchased	144
Open source	133
Unspecified	16
Total	**574**
Number of respondents	**225**

Intersect360 Research, 2012

- About half of usage is internal applications
- Remainder split between purchased (ISV) and open source

Intersect360 RESEARCH

Insight #2: Not Just Hadoop

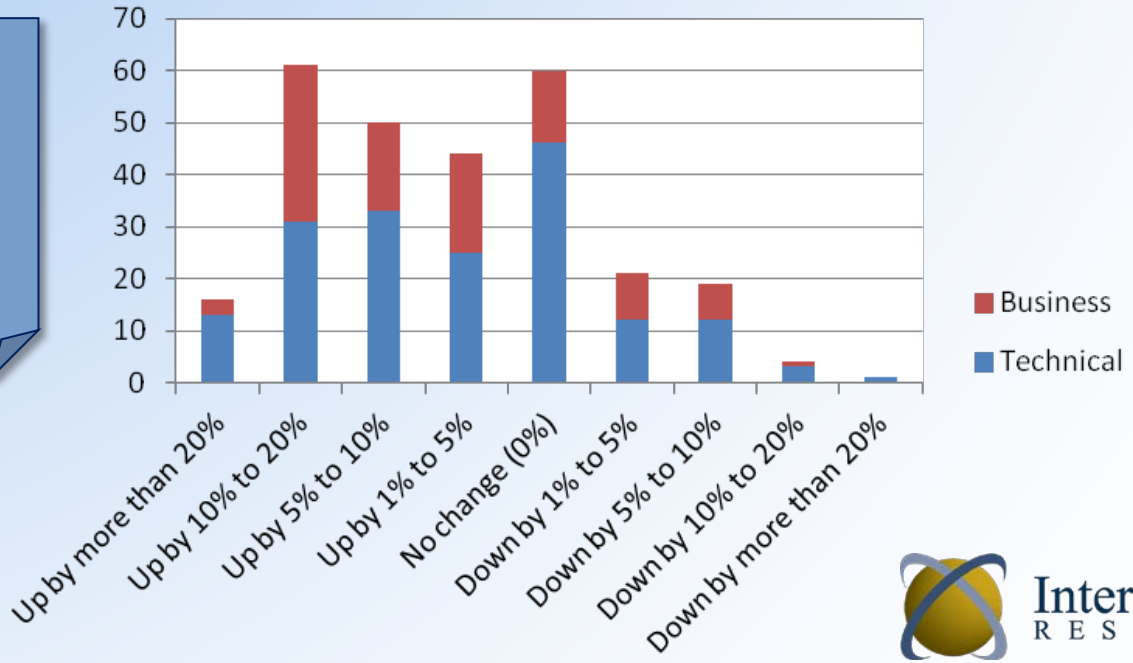
- In 2012, only 17% of respondents mentioned Hadoop when describing their Big Data applications. **In 2013, this went down**, driven by Enterprise respondents.
- Deployments might be based on Hadoop, but the majority of Big Data implementations are on **in-house applications and algorithms**.
- Most common source of data is also “in-house.”
- ISV software for Big Data is thinly scattered.

Intersect360 RESEARCH

Defining the Opportunity

- How much added spending is there because of Big Data? Look at IT budget growth (below)
- What technologies have expanded opportunities? Look at “satisfaction gaps”

Expected two-year change in IT budget



Intersect360 RESEARCH

Insight #3: Performance Counts

Satisfaction Gaps for Storage Solutions

Satisfaction Gap = Importance Score - Satisfaction Score

Technical

1. I/O performance (+.94)
2. Storage capacity (+.67)
3. RAS (+.58)

Enterprise

1. I/O performance (+.77)
2. Storage capacity (+.55)
3. RAS (+.52)

- Metrics of performance show up as key factors in Enterprise as well as Technical
- Big Data will be a driver for expanded usage of HPC, IF they can still meet enterprise requirements
- But end users might not want to think of it as HPC

Cloud + Big Data: When Trends Collide

I like sushi. I like ice cream.
Therefore I like sushi-flavored ice cream.

- Cloud is a major business computing trend. Big Data is a major business computing trend. Therefore …
- But the barriers to Big Data in cloud are the same as HPC in cloud (security, data movement, etc.)
- Not as simple as offloading everything to Amazon
- If cloud is a priority, invest in management software to coordinate workloads across public and private

Intersect360 RESEARCH

Technologies in the Discussion

- “Storage” beyond spinning disk
  - Flash / Solid-state (max I/O)
  - Tape (max capacity)
- Parallel file systems
- High-speed fabrics (e.g. InfiniBand)
- MPI
- Large shared memory spaces
- Accelerators (e.g. GPU, FPGA, Intel Xeon Phi)

Advice on Big Data

- To end users:
  - There is a competitive advantage to performance
  - Open your datacenter to new ideas
  - Invest in the technology and skill to scale
- To vendors:
  - Don’t assume you know what the problem is. Ask.
  - Don’t assume you know what the solution is. Think.
  - Discuss high performance in enterprise context

For More Information and Results

- Technology Vendors: You can buy it. Reasonable prices. See me or email sales@intersect360.com
- HPC Users: Join our HPC user group, HPC500, to get free access to research studies.
  - We will ask you to participate in surveys
  - We will keep your organizations anonymous if desired
  - www.hpc500.com

HPC500
CONNECTING HPC LEADERS

Intersect360
RESEARCH

Worlds Colliding:
Why Big Data Changes How To Think About Enterprise Storage

Intersect360
RESEARCH

Addison Snell
addison@intersect360.com

Intersect360
RESEARCH