Big Data trends

- Almost universally about figuring out the larger context of human actors
- Sharpening relevance in moments of truth
- Through rapid contextualization of situational signal
- With slow and pervasive integration (space, time) of other dimensions of human existence
- By **context engines** capable of identifying 10 or so useful items from perhaps $10^{12}$ digital breadcrumbs
Context

The information halo that wraps all the dimensions of human existence, in particular including all our content (information-wrapped assets) we own or desire.
large context
The Changing Face of IT

 Mountains of unprocessed information
 Capacity- and cost-oriented IT operations
 People-heavy information processes and workflows
 IT-imposed look and feel in information systems
 Information must be found

 Extracted, correlated, and contextualized information
 Quality oriented optimizations
 Agency, automation and affordances
 Contextualized remapping into end user world view
 Information finds context
Where the rubber meets the road ...

- As is
  - 85h to contextualize!
  - Unforgiving infrastr.

- To be
  - Log everything! ($10^{14}$)
  - Complex MDM (<10 s)
  - Many enrichment algos
  - UX = ability to put any piece of right data into the customer’s hands at a moment’s notice, often before the query
  - Think Google Now, FB Taste, realtime MBRs

![Graph showing business value of information over time since ingest, with expectedness and right time implications.](image)
Lessons

▸ What I didn’t mention?
  • SQL vs NoSQL
  • Optimizing DBs, SEs
  • 20:20 analytics

▸ Why?
  • Moving bits around to optimize for disk overheads is no longer the game
  • New era, new primitives
    ▸ Sparsity, atomicity

▸ Custom apps, behind the firewall infrastr!

▸ How to scale out when you must
  • Processing in store path and in query path
  • Can’t add servers just to get more DRAM
  • Graceful degradation of CPU overhead @ large, low latency and low cost persistent memory
Comparative architecture models

**iT’s tools**

- Search engines & databases

**Q**

**D**

**iT’s tools**

- foraging loop (scent)

\[ Q, Q', \ldots \leftrightarrow D, D', \ldots \]

- sensemaking loop (intent)

\[ D \rightarrow R, R', \ldots \rightarrow R(D), R'(D), \ldots \rightarrow T \]

- contextual delivery

\[ D \rightarrow Q \rightarrow \text{profiles & contexts} \]

**extraction, correlation**
# Convergence: An Evolving Notion

<table>
<thead>
<tr>
<th>WHAT</th>
<th>2002-2006</th>
<th>2006-2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving Intelligence closer to Storage</td>
<td>Moving Persistence closer to Intelligence</td>
<td></td>
</tr>
<tr>
<td>WHY</td>
<td>Storage media unsuitable for bringing into CPU-memory complex</td>
<td>Solid state media increasingly suitable for memory like access</td>
</tr>
<tr>
<td>HOW</td>
<td>Storage grid • Offload content mining &amp; metadata</td>
<td><strong>Scale out server-attached NVM</strong></td>
</tr>
<tr>
<td></td>
<td>Data grid • Offload indexing &amp; query processing</td>
<td>RDMA-accessible persistent memory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• I/O Memory (PCIe)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Flash as ACM/ExtMem</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Storage-class memory</td>
</tr>
</tbody>
</table>
The ideas in play

Looking good

- Memory-semantic access
- Familiar storage management metaphor
- Media wear leveling

Questionable

- Disk-era protocols
- Disk-era optimizations
<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
<th>Company</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>Mission to consolidate memory and storage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>ioMemory technology unveiled</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>First products launched</td>
<td>IBM</td>
<td>1 million IOPS, Dell strategic investment</td>
</tr>
<tr>
<td>2009</td>
<td>HP OEMs products</td>
<td></td>
<td>IBM OEMs products, Samsung strategic investment</td>
</tr>
<tr>
<td>2010</td>
<td>Dell OEMs products</td>
<td></td>
<td>VSL introduced</td>
</tr>
<tr>
<td>2011</td>
<td>IPO on NYSE</td>
<td>ioTurbin acquired</td>
<td>&gt;120 channel and alliance partners, ioDrive2 announced</td>
</tr>
<tr>
<td>2012</td>
<td>1 Billion IOPS</td>
<td></td>
<td>2,500 customers, 50+ Petabytes shipped, Supermicro OEMs products</td>
</tr>
</tbody>
</table>

Fusion-io First Mover Milestones

May 22, 2013
ioMemory Portfolio

**ioDrive II**
- Up to 1.2TB

**ioDrive II Duo**
- Up to 2.4TB

**Octal**
- Up to 10.24TB maximize performance for large data sets

**ioFX**
- 420GB of workstation acceleration

**ioDrive**
- 160GB to 640GB of Fusion’s industry-leading ioMemory

**ioDrive Duo**
- 160GB to 1.28TB of Fusion’s industry-leading ioMemory

**ioScale™**
- 3.2TB to 12.8TB for Web & Cloud Environments
Evolution of Enterprise Flash

FLASH + DISK
SSD

FLASH AS DISK
SSD

FLASH BEYOND DISK

FLASH AS MEMORY
Evolution of Enterprise Flash

**FLASH AS DISK**
- Application source code converts native data structures into block I/O
- Conventional I/O Access
- Block I/O
- Proprietary Storage OS

**FLASH BEYOND DISK**
- Application source code does I/O with native data structures
- Native: Enhanced I/O
  - Atomic I/O Transaction
  - User-Defined Object Transaction
  - Key-Value Transaction
- Open Interface Layer

**FLASH AS MEMORY**
- Application source code manipulates native data structures directly in persistent memory
- Native: Persistent Memory
  - High-speed Logging
  - Checkpointed Memory
  - Memory Transaction
- Open Interface Layer
Hybrid Memory Storage Hierarchy

Accessed Like Memory and Managed Like Storage

- L1, L2, L3 CPU Caches
- DRAM
- Persistent Memories
- Flash
- Hard Drive

ACCESS DELAY

- Nanoseconds
- Microseconds
- Milliseconds

CYCLES TO WAIT

- 2 cycles
- 4 million cycles

May 22, 2013
Fusion-io Accelerates

<table>
<thead>
<tr>
<th>DATABASES</th>
<th>VIRTUALIZATION</th>
<th>SEARCH</th>
<th>ANALYTICS</th>
<th>BIG DATA</th>
<th>COLLABORATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORACLE</td>
<td>VMware</td>
<td>ORACLE Text</td>
<td>AccessData</td>
<td>hadoop</td>
<td>Exchange</td>
</tr>
<tr>
<td>MySQL</td>
<td>Hyper-V</td>
<td></td>
<td>Autonomy</td>
<td>mongoDB</td>
<td>Lotus</td>
</tr>
<tr>
<td>Sybase</td>
<td>Windows Server</td>
<td></td>
<td>MarkLogic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INGRES</td>
<td>XenDesktop 5</td>
<td></td>
<td>LexisNexis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PostgreSQL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM DB2</td>
<td>KVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INFORMIX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HPC</th>
<th>MESSAGING</th>
<th>WORKSTATION</th>
<th>DEVELOPMENT</th>
<th>CACHING</th>
<th>SECURITY/LOGGING</th>
<th>WEB</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLUENT</td>
<td>IBM MQ</td>
<td>Autodesk</td>
<td>PERFORCE</td>
<td>ArcSight</td>
<td>splunk</td>
<td>LAMP</td>
</tr>
<tr>
<td>Magmasoft NX</td>
<td></td>
<td>SolidWorks</td>
<td></td>
<td>Powered by</td>
<td></td>
<td>.NET</td>
</tr>
<tr>
<td>NASTRAN</td>
<td></td>
<td>Adobe</td>
<td></td>
<td>SOA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUSTRE</td>
<td></td>
<td>software as</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM GPFS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

16
## Comprehensive Customer Success

<table>
<thead>
<tr>
<th>FINANCIALS</th>
<th>WEB</th>
<th>TECHNOLOGY</th>
<th>RETAIL</th>
<th>MANUFACTURING/GOVERNMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Logos]</td>
<td>[Logos]</td>
<td>[Logos]</td>
<td>[Logos]</td>
<td>[Logos]</td>
</tr>
</tbody>
</table>

- **5x** FASTER DATA ANALYSIS
- **30x** FASTER DATABASE REPLICATION
- **40x** FASTER DATA WAREHOUSE QUERIES
- **15x** QUERY PROCESSING THROUGHPUT
- **15x** FASTER QUERIES

30+ case studies at [http://fusionio.com/casestudies](http://fusionio.com/casestudies)
THANK YOU