

Big Data

Ingo Fuchs
NetApp

*Practical Solutions That Make
A Difference Today*



© 2012 NetApp, Inc. All rights reserved.



Thoughts – What Big Data is about...

Doing Things Differently

- Understand the options
- Become a difference-maker
- Be prepared

Staying Practical

- Fits your needs
- Affordable
- Proven



Not Even to The "Peak"

5 Billion

smart
phones

35 Zettabytes

Estimated size of the
digital universe in
2020

80%

of data is
unstructured

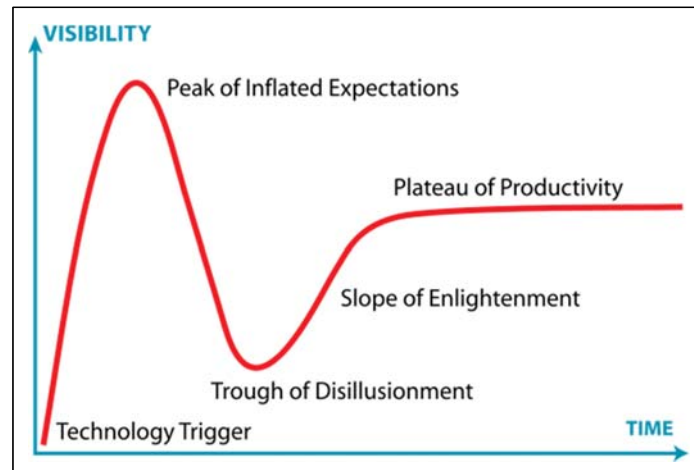


30 Billion

pieces of new content to
Facebook per month

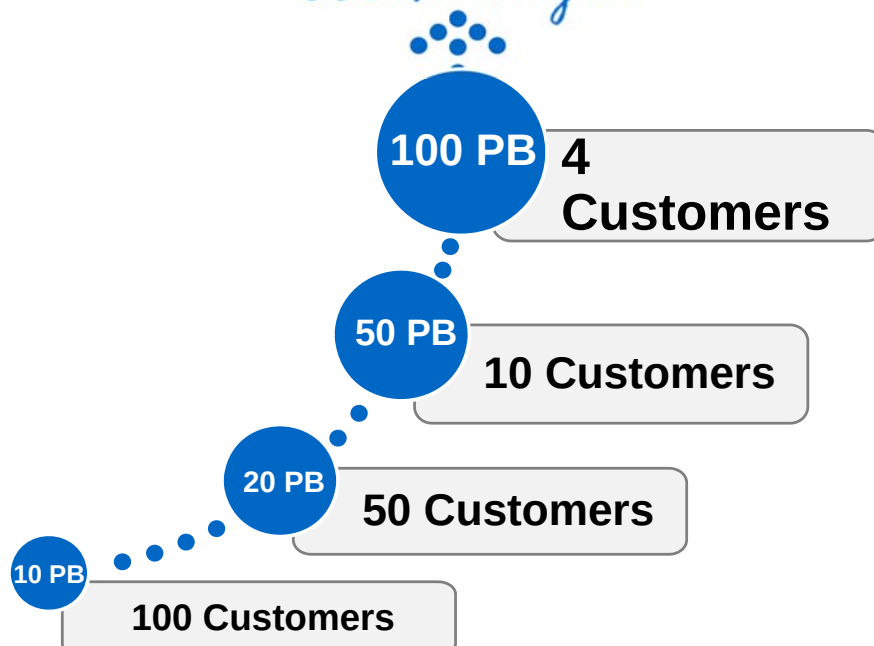


Sensors
Video
Music
Location
Weblogs



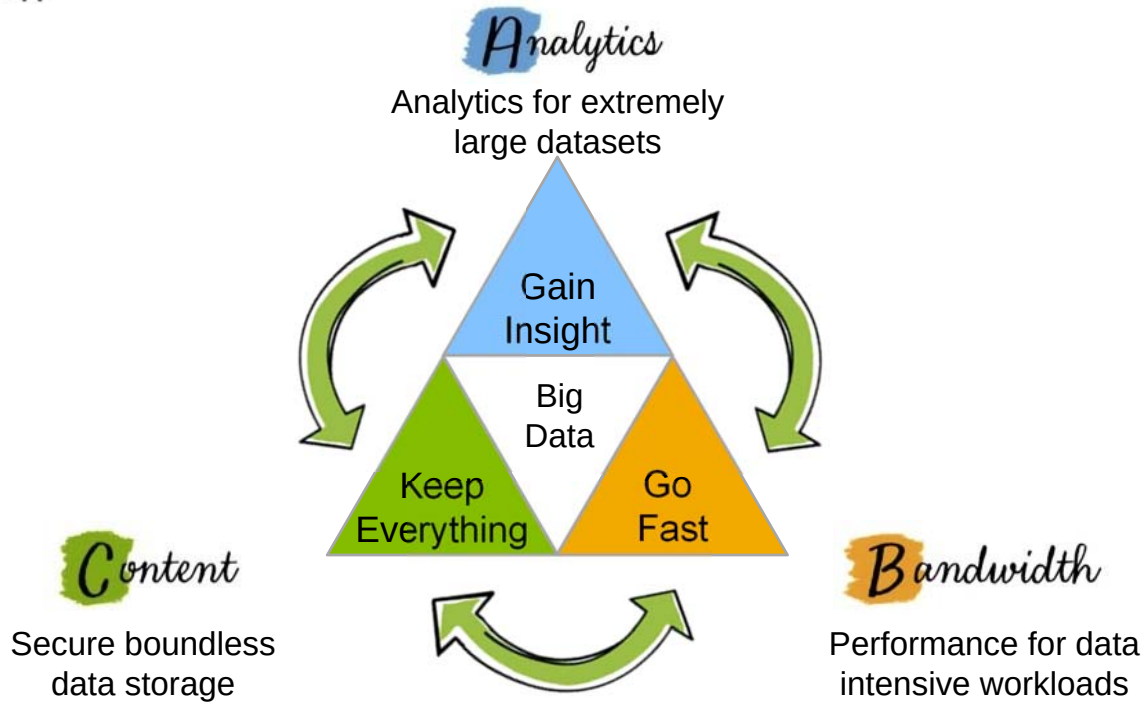
Experience Managing Data at Scale

880 Petabytes





Big Data ABCs



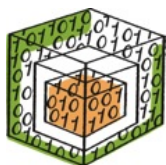
Analytics





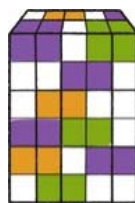
Big Analytics Challenges and Solutions

Data Warehouse



Fast, space-efficient backup and recovery with storage utilization up to 90%. Less raw capacity with modular scalability

Mixed Use Database, Cubes



Optimized for IBM, Oracle and Microsoft. Simplified data management and protection. Zero down time

Hadoop



Enterprise class Hadoop with Lower total cost of ownership and based on open standards

© 2012 NetApp, Inc. All rights reserved.

7



NetApp Open Solution for Hadoop



Enterprise-class Hadoop Solution

- Packaged ready-to-deploy modular Hadoop cluster
- Lower TCO, Higher reliability

Open Hadoop eco-system of partners

- Open standards
- Cloudera, Hortonworks distributions
- Interoperate with enterprise analytics eco-systems

© 2012 NetApp, Inc. All rights reserved.



Use Case Example: NetApp Auto Support



Phone home data representing information about the status NetApp storage controllers

- Correlate disk latency (hot) with disk type
 - 24 billion records
 - 4 weeks to run query
 - Hadoop implementation 10.5 hours
- Bug detection through pattern matching
 - 240 billion records – Too large to run
 - Hadoop implementation 18 hours



Bandwidth





Big Bandwidth Challenges and Solutions

Full Motion Video



Scalable density and performance to ingest and simultaneously analyze UAV and satellite video data

Video Storage for Surveillance



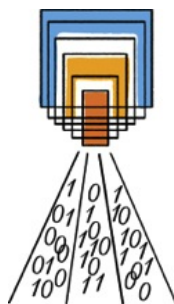
High bandwidth & density supporting hundreds or thousands of HD cameras

Media Content Management



High ingest & play-out rates with support for media and entertainment workflows

HPC: Lustre & Seismic



Massively parallel distributed file system for large scale cluster computing and O&G Seismic Processing



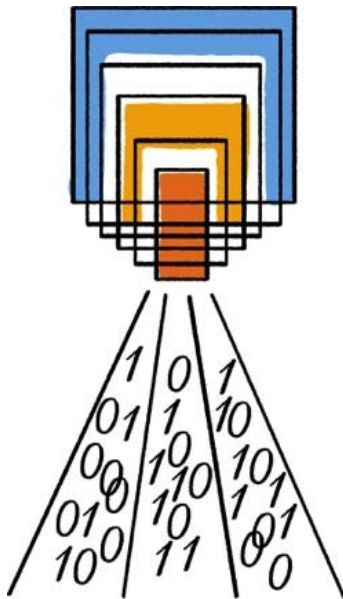
Bandwidth HPC: Lustre

- Performance to meet the needs of the world's fastest Supercomputers
- High Bandwidth & Density
 - 1.8PB & 30GB/s per 40U rack
- Highly available
 - No Single points of failure
 - Extensive RAS features
- NetApp provided 7x24 Lustre Support
- NetApp Professional services





Lawrence Livermore National Lab



Sequoia – announced as the fastest supercomputer and storage combination on the planet at ISC 2012

- Supercomputer storage to support twenty thousand trillion arithmetic operations per second with access speeds up to 1 TB/sec
- 55PB of usable storage
- Simulations for nuclear weapons viability
- Counter Terrorism
- Energy Security
- Understanding Climate Change

Press Release: <http://www.netapp.com/us/company/news/news-rel-20110928-990734.html>



Content





Big Content Challenges and Solutions

File Services



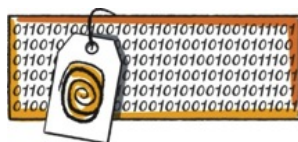
Multi-application workloads
Non-disruptive operation
Integrated data protection, efficiency

Enterprise Content Repository



Infinite container
Fixed content
Non-disruptive operation
Integrated data protection, efficiency

Distributed Content Repository



Large, multi-site repository
Policy based data management
Metadata-enabled object storage



Enterprise Content Repositories

ONTAP Cluster Mode with Infinite Volume



Single large content repository

- Scales to PBs and billions of files across cluster
- Native storage efficiency



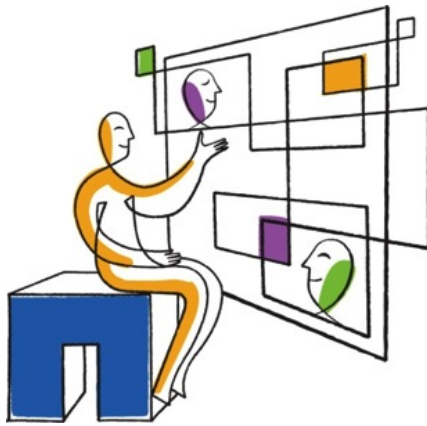
Simplified operations

- Simplifies application workflows
- Load balances data at ingest
- Starts small, grow granularly



High availability

- Protects against disk and hardware failures
- Snapshots & Replication for quick recovery
- Manage & Upgrade non-disruptively



- Managed Teaching/Learning Environment
- Offices in NA, Europe, Asia, Australia
- Millions of users daily from K-12 to Universities and Businesses
- 4+PB managed by 5 IT staff
- Provisioning reduced 12 hours to 5 mins
- Restores reduced 72 hours to 5 mins

“When you consider that the average company manages less than 100TB of data per full-time employee, NetApp is helping us achieve a business breakthrough. Our team manages approximately 500TB of data per full-time employee on NetApp® storage, all while maintaining near-100% storage infrastructure availability.”

– Jonas Hirshfield, VP, Infrastructure Technology, Blackboard



Distributed Content Repositories

StorageGRID



Large content repository for big, unstructured data

- Billions of data sets, dozens of petabytes



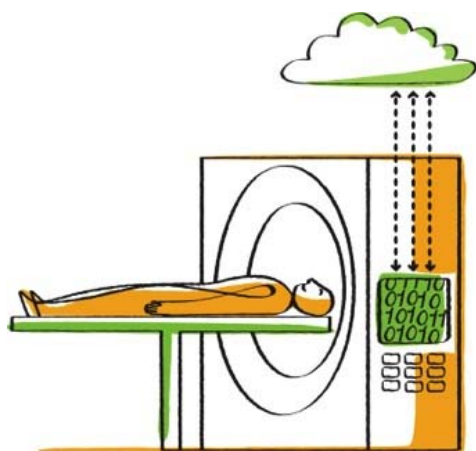
Create, manage and consume content globally

- Predictable access to data independent of location
- Policy-controlled data stores at each site



Intelligent data classification and access

- Metadata-based management



- STaaS offering for healthcare providers
- Medical Image Archive Cloud
 - Two sites with ~1PB each
 - 2TB+ local cache at each edge site
 - 8x growth in capacity last 12 months
 - 100% uptime since start of service
 - “Forever” retention policies
 - ~60% of customers use hybrid cloud model
- Solution offers a proven 100% up-time with automated data movement from on-premise to off-premise public clouds with “keep forever” retention policy and indefinite growth

Press Release: <http://www.netapp.com/us/company/news/news-rel-20111128-36413.html>



Benefits for our customers

- **Abstraction of physical storage** location(s) from application clients
- **Single global namespace** enables flexibility in growth and workloads across geographically distributed sites
- Cost efficient management of distributed disk and archival/tape media
- **Automatic policy-based data management** at the object level, for:
 - data protection
 - data recovery
 - cost reduction
 - workflow optimization
- Efficient locating and retrieval of data through use of **metadata**
- Proactive data integrity protection for **long-term archive** & preservation



CDMI Standardization Update

- The CDMI standard effort is now three years old
 - CDMI 1.0 released in late 2010 (Initial release)
 - CDMI 1.0.1 released in late 2011 (First errata release)
 - CDMI 1.0.2 released in mid 2012 (ISO errata release)
- CDMI 1.1 is under active development
 - 12 Extensions from multiple vendors and end users
 - Contributors include IBM, Netapp (Chair), Cisco, Cleversafe, Dell, EMC, HP, HDS, IBM, Mezeo, Oracle (Chair), and VMWare
- The SNIA has submitted CDMI 1.0.2 to ISO for international standardization as ISO/IEC 17826
 - Initial ballot received positive votes from most national bodies, concerns from the UK and Chinese delegations have been resolved
 - Final ballot is scheduled for June, 2012



CDMI Adoption – Gaining Momentum

- EU FP7 SIENA Initiative
- EU FP7 Vision Cloud
- VENUS-C
- NIST SAJACC
- StratusLab
- UK GridPP
- UK NGS Cloud
- SARA
- OpenNebula
- GICTF JP
- TOSCA
- Italian DigitPA



What's Next?

- CDMI - Federated Repositories
- HPC, Analytics and Content Repositories
- Open Rack/Stack designs
- XaaS architectures (internal and managed)



Next Steps – Team with the Experts



- Visit stand G16:
 - Play **Big Data Bingo**
 - Daily prize draws and more
- Go online:
www.netapp.com/au/company/leadership/big-data
- Follow us: @NetAppANZ
- Follow me: @ingofuchs
- Contact:
 - Andy.Hunne@netapp.com
(0437 043 628)
 - Matt.Swinbourne@netapp.com
(0499 499 150)

Thank you

