High Performance Computing going mainstream...

Mujdat Timurcin

IBM Systems and Technology Group

IBM Turk

GRAPH 500 **MIRA**
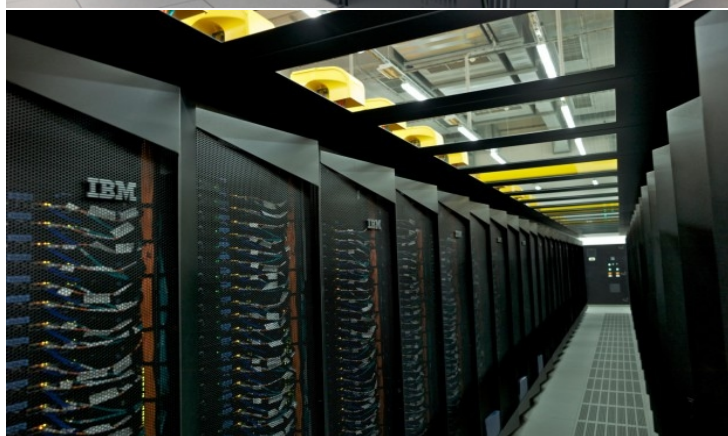
IBM's Supercomputing Chops
TOP500 list of the world's fastest supercomputers

#1 SEQUOIA IBM POWER (LLNL)
16+ petaflops of performance
Could provide a 40-fold improvement in the prediction of earthquakes and safe evacuation routes
1.55x faster and **2.49x** more energy efficient[1]

#3 MIRA IBM POWER (ANL)
8+ petaflops of performance
More than 5 billion computing hours will be available to researchers on Mira every year
Could develop new materials to stretch the charge on an electric car battery up to **500 miles**

#4 SuperMUC (LRZ)
3 petaflops
Completely water cooled making it **40%** more efficient[2]
Captured energy from SuperMUC is reused to heat campus buildings, providing **$1.25 million** in energy savings per year.

One petaflop runs a quadrillion or 10^{16} instructions per second.

IBM

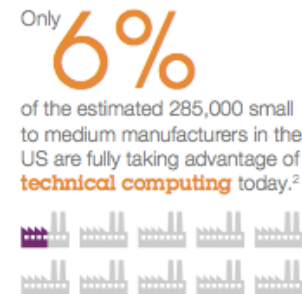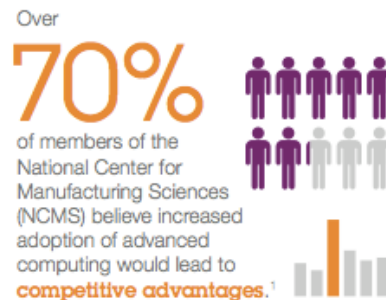[1] Compared to the next fastest supercomputer in the world.
[2] Compared to an air cooled machine.
Lawrence Livermore National Lab (LLNL), Argonne National Lab (ANL), Leibniz Supercomputing Centre (LRZ)

High Performance Computing Goes Mainstream

High-powered technical computing increasingly is used to solve practical problems in manufacturing, life sciences, oil and gas, and other industries, but many companies still aren't fully tapping its potential.

Over
70%
of members of the National Center for Manufacturing Sciences (NCMS) believe increased adoption of advanced computing would lead to **competitive advantages**.[1]

Only
6%
of the estimated 285,000 small to medium manufacturers in the US are fully taking advantage of **technical computing** today.[2]

Technical computing achievments

The Boeing Company aims to use simulations to redesign the vertical tail of a commercial jet, potentially saving

$300 million in fuel costs annually.

Using IBM technical computing, **Vestas Wind Systems** reduced their wind turbine placement analysis from weeks to less than **one hour.**

Red Bull Racing used IBM technical computing software to simulate new car designs and achieved a

20% increase
in performance and throughput, coming up with a design that reduces their cars' drag on the track.

Critical IT Trends for Technical Computing Users

| Explosion of data | How to spot trends, predict outcomes and take meaningful actions? |

| Inflexible IT infrastructures | How to manage inflexible, siloed systems and business processes to improve business agility? |

| Escalating IT complexity | How to manage IT costs and complexity while speeding time-to-market for new services? |

Introducing the new **IBM Technical Computing Portfolio**
Powerful. Comprehensive. Intuitive.

Solutions

| Integrated Solutions | Industry Solutions | Intelligent Cluster | HPC Cloud | Big Data |

Software

Parallel Environment Runtime

Parallel Environment Developer

Platform LSF

Platform HPC

Platform Symphony

Platform MPI

Engineering and Scientific Libraries

Platform Cluster Manager

GPFS

Platform Application Center

Systems & Storage

iDataPlex Intelligent Cluster

PureFlex

System x & BladeCenter

BG/Q

P7-775

LTO Tape 3592 Automation

DS5000 DS3000

DCS3700

HPSS

SoNAS

NEW HPC Cloud Solutions

Overview

- Innovative solutions for dynamic, flexible HPC cloud environments

What's New

- New LSF add-on: IBM Platform Dynamic Cluster V9.1
  - Workload driven dynamic node re-provisioning
  - Dynamically switch nodes between physical & virtual machines
  - Automated job checkpoints and migration
  - Smart, flexible policy and performance controls
- Enhanced Platform Cluster Manager – Advanced capabilities
- New complete, end to end solutions

Use Case 1: HPC Infrastructure Management

- Self-service cluster provisioning & management
- Consolidate resources into a HPC cloud
- Cluster flexing

Use Case 2: Self-service HPC

- Self-service job submission & management
- Dynamic provisioning
- Job migration and/or checkpoint-restart
- 2D/3D remote visualization

Use Case 3: Cloud Bursting

- 'Burst' internally to available resources
- Burst externally to cloud providers

NEW Financial Risk and Crimes Solution

Overview

- High-performance, low-latency integrated risk solution stack with Platform Symphony - Advanced Edition and partner products including:
  - BigInsights and IBM Algorithmics
  - 3rd party Partner products: Murex and Calypso

What's New

- New solution stacks to manage and process big data with speed and scale
- Sales tools that highlight value of IBM Platform Symphony
  - Financial Risk: Customer testimonial videos; inclusion in SWG risk frameworks, and S&D blueprints
  - Financial Crime: with BigInsight for credit card fraud analytics
  - TCO tool and benchmarks

Use Case 1: Financial Risk including Credit Value Adjustment (CVA) analytics

- Accelerates compute intensive workloads up to 4X e.g. Monte Carlo simulations, Algorithmic Riskwatch "cube" simulations
- Integrated with IBM Algorithmics, Murex and Calypso
- High throughput: 17K

Use Case 2: Big Data for Financial Crimes

- Accelerates analyses of data for fraud and irregularities
- Supports BigInsight
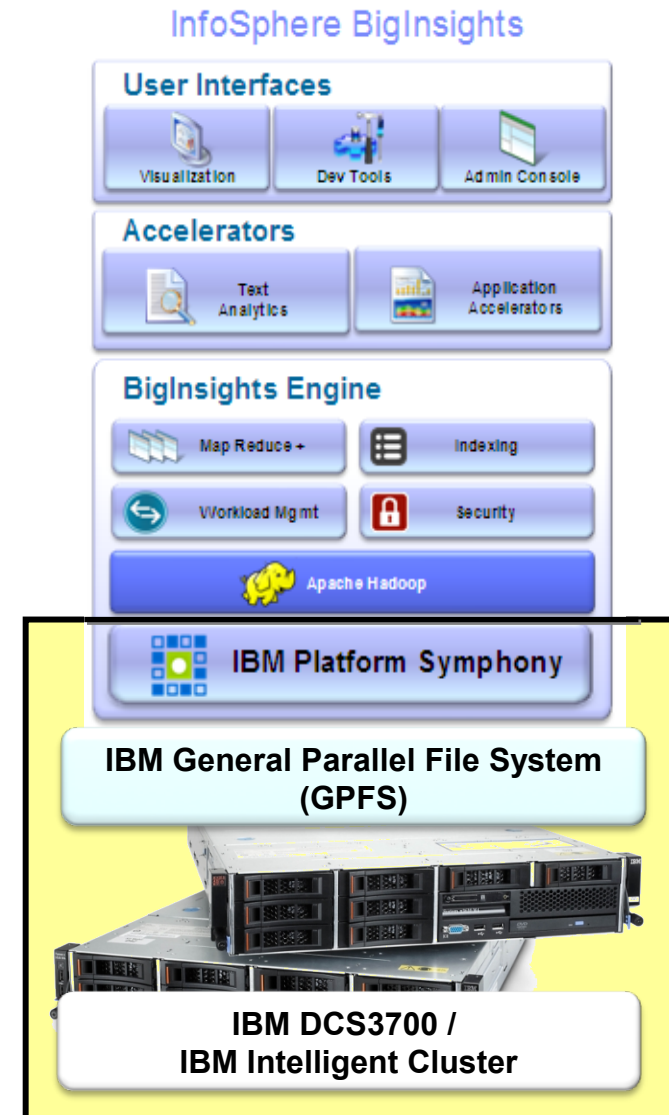- Faster than Apache Hadoop distribution

NEW Technical Computing for Big Data Solutions

Overview

• High-performance, low-latency "Big Data" solution stack featuring Platform Symphony, GPFS, DCS3700, Intelligent Cluster – proven across many industries

• Low Latency Hadoop stack with Platform Symphony, Advanced Edition and InfoSphere BigInsights

What's New

• New solution stacks to manage and process big data with speed and scale

InfoSphere BigInsights

User Interfaces
- Visualization
- Dev Tools
- Admin Console

Accelerators
- Text Analytics
- Application Accelerators

BigInsights Engine
- Map Reduce +
- Indexing
- Workload Mgmt
- Security
- Apache Hadoop

IBM Platform Symphony

IBM General Parallel File System (GPFS)

IBM DCS3700 / IBM Intelligent Cluster

Technical Computing Software

Simplified management, optimized performance

The backbone of Technical Computing

IBM Platform Computing
can help accelerate your application results

For technical computing and analytics distributed computing environments

Optimizes Workload Management
- Batch and highly parallelized
- Policy & resource-aware scheduling
- Service level agreements
- Automation / workflow

Service Availability

Aggregates Resource Pools
- Compute & Data intensive apps
- Heterogeneous resources
- Physical, virtual, cloud
- Easy user access

Delivers Shared Services
- Multiple user groups, sites
- Multiple applications and workloads
- Governance
- Administration/ Reporting / Analytics

SILOED APPLICATIONS   SHARED GRID

Transforms Static Infrastructure to Dynamic
- Workload-driven dynamic clusters
- Bursting and "in the cloud"
- Enhanced self-service / on-demand
- Multi-hypervisor and multi-boot

CentOS
vmware   redhat.
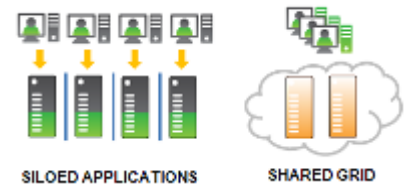
Clients span many industries

Platform Symphony

Platform LSF

"Platform Computing came to us as a true innovation partner, not a supplier of technology, but a partner who was able to understand our problems and provide appropriate solutions to us, and work with us to continuously improve the performance of our system"

- Steve Nevey, Business Development Manager
Red Bull Technology

Watch Red Bull video

European Bank

"Platform enterprise grid solution enable us to share a formerly heterogeneous and distributed hardware infrastructure across applications regardless of their location, operating system and application logic, … helping us to achieve our internal efficiency targets while at the same time improving our performance and service quality"

-Lorenzo Cervellin, Head of Global Markets and Treasury Infrastructure
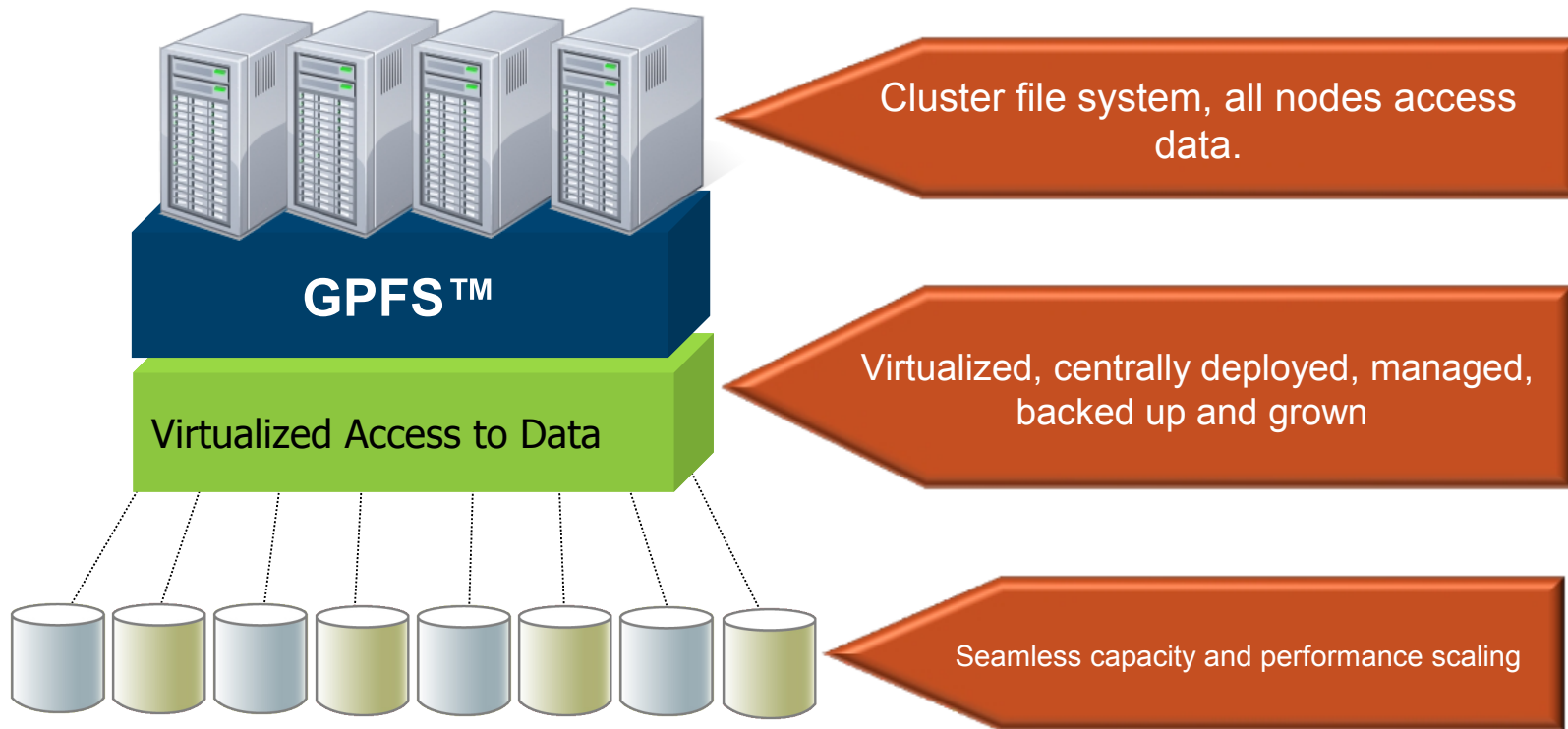**UniCredit Global Information Services**

Platform HPC

"Platform's software was a clear leader from the beginning of the process"

-Chris Collins, Head of Research & Specialist Computing
University of East Anglia

IBM also offers the most widely used, commercially available, technical computing data management software

GPFS™

Virtualized Access to Data

Cluster file system, all nodes access data.

Virtualized, centrally deployed, managed, backed up and grown

Seamless capacity and performance scaling

IBM General Parallel File System - *Scalable, highly-available, high performance file system optimized for multi-petabyte storage management*

GPFS pioneered Big Data management

Extreme Scalability	Proven Reliability	Performance
File system	No Special Nodes	High Performance Metadata
2^{63} files per file system	Add/remove on the fly	Striped Data
Maximum file system size: 2^{99} bytes	Nodes	Equal access to data
Maximum file size equals file system size	Storage	Integrated Tiered storage
Production 5.4 PB file system	Rolling Upgrades	
Number of nodes	Administer from any node	
1 to 8192	Data replication	

IBM innovation continues with **GPFS Active File Management** (AFM) for global namespace

GPFS introduced concurrent file system access from multiple nodes.

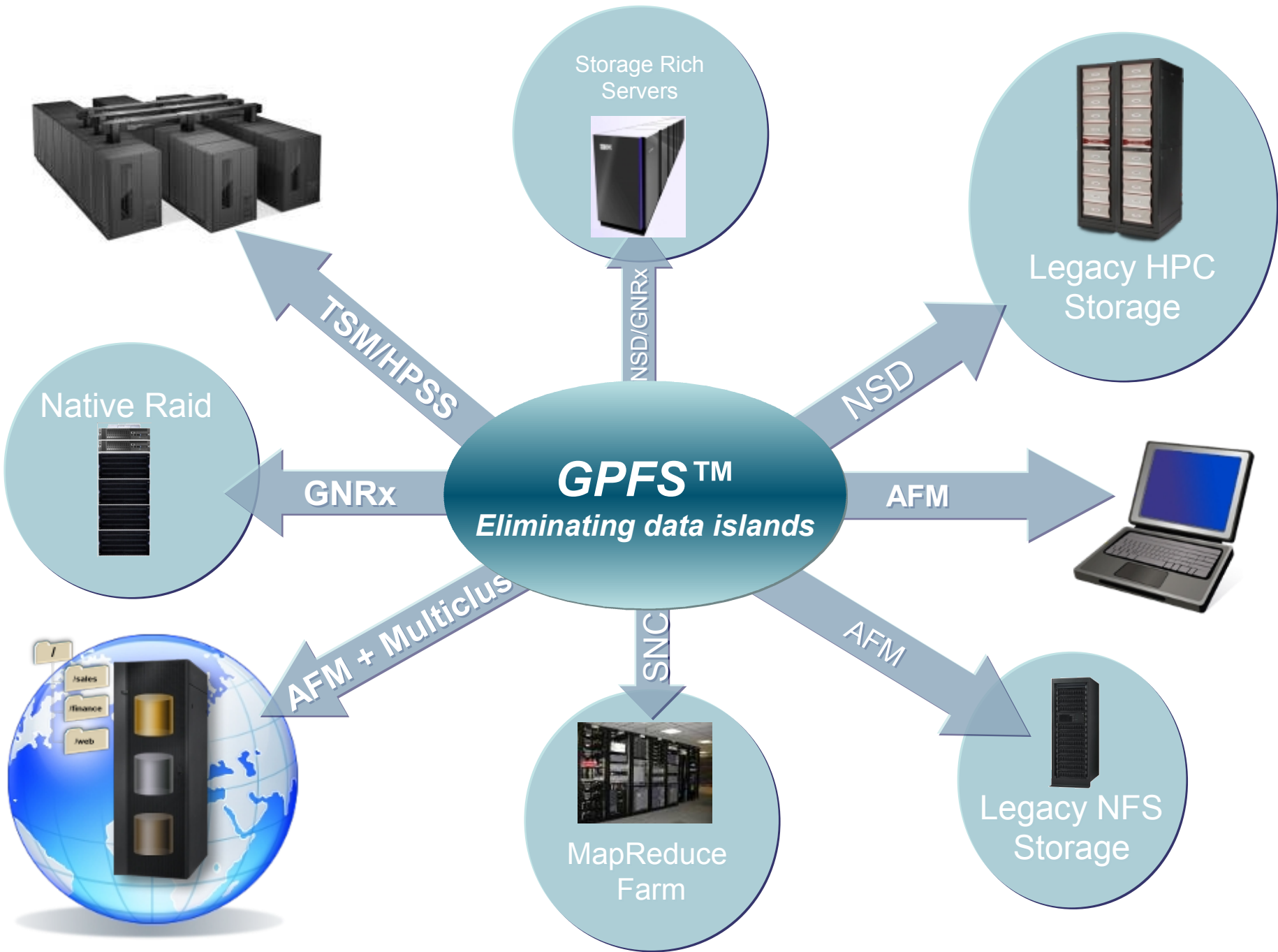Multi-cluster expanded the global namespace by connecting multiple sites

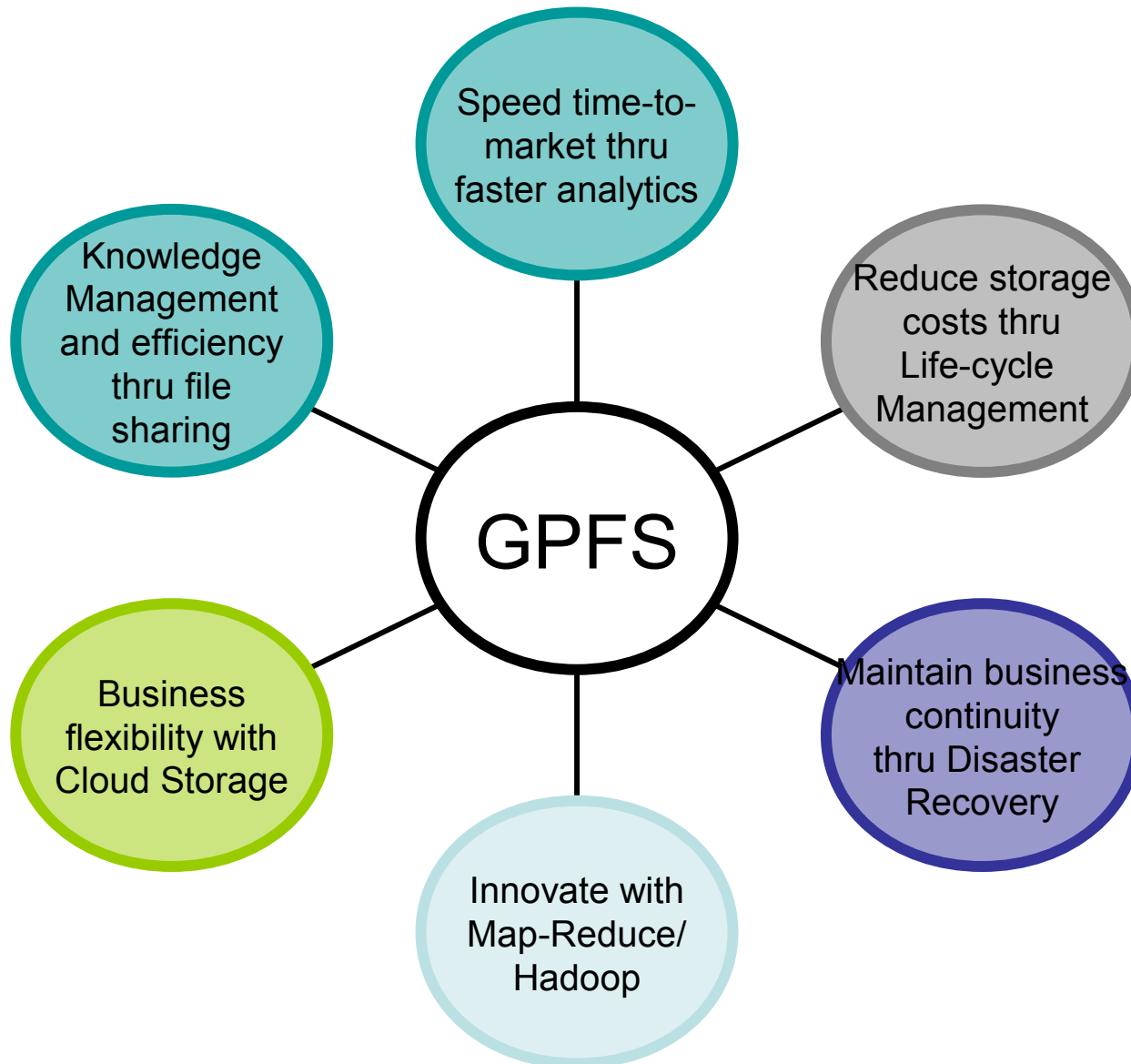AFM takes global namespace truly global by automatically managing asynchronous replication of data

1993   **2005**   **2011**

Storage Rich Servers

Legacy HPC Storage

Native Raid

TSM/HPSS

NSD/GNRx

NSD

GNRx

GPFS™
Eliminating data islands

AFM

AFM + Multiclus

SNC

AFM

MapReduce Farm

Legacy NFS Storage

/sales
/finance
/web

Speed time-to-market with faster analytics

- Issue:
  - We are in the era of "Smarter Analytics"
    - Data explosion makes I/O a major hurdle.
    - Deep analytics result in longer running workloads
    - Demand for lower-latency analytics to beat the competition

- GPFS was designed for complex and/or large workloads accessing lots of data:
  - Real time disk scheduling and load balancing ensure all relevant information and data can be ingested for analysis
  - Built-in replication ensures that deep analytics workloads can continue running should a hardware or low level software failure occur.
  - Distributed design means it can scale as needed

Reduce storage costs thru Life-cycle Management

- Issue:
  - Increasing storage costs as dormant files sit on spinning disks
  - Redundant files stored across the enterprise to ease access
  - Aligning user file requirements with cost of storage

- GPFS has policy-driven, automated tiered storage management for optimizing file location.
  - ILM tools manage sets of files across pools of storage based upon user requirements
  - Tiering across different economic classes of storage: SSD, spinning disk, tape – regardless of physical location.
  - Interface with external storage sub-systems such as TSM and HPSS to exploit ILM capability enterprise-wide.

Maintain business continuity thru disaster recovery

- Issue:
  - Need for real-time or low latency file access
  - File data contained in geographic areas susceptible to downtime
  - Fragmented file based information across a wide geographic area

- GPFS has inherent features that are designed to ensure high availability of file-based data
  - Remote file replication with built-in failover
  - Multi-site clustering enables risk reduction of stored data via WAN
  - Space efficient point-in-time snapshot view of the file system enabling quick recovery

Innovate with Big Data or Map-Reduce/Hadoop

- Issue:
  - Unlocking value in large volumes of unstructured data
  - Mission critical applications requiring enterprise-tested reliability
  - Looking for alternatives to the Hadoop File System (HDFS) for map-reduce applications

- As part of a Research project, there is an active development project called GPFS-SNC to provide a robust alternative to HDFS
  - HDFS is a centralized file system with a single point of failure, unlike the distributed design of GPFS
  - GPFS Posix compliance expands the range of application that can access files (read, write, append) vs HDFS which cannot append or overwrite.
  - GPFS contains all of the rich ILM features for high availability and storage management, HDFS does not.

Business flexibility with cloud storage

- Issue:
  - Requires highly scalable storage - petabytes of data
  - Reliability to handle frequently-occurring failures in large systems
  - Resource efficiency to maintain the economics of cloud computing

- GPFS core features make it an ideal infrastructure for cloud computing storage requirements.
  - Distributed design of I/O servers is highly scalable to support billions of files.
  - Fault-tolerance provides needed failover to ensure QoS requirements
  - File tiering and ILM enable cost effective storage options to match user requirements.

Knowledge management and efficiency thru file sharing

- Issue:
  - Geographically dispersed employees need access to same set of file based information
  - Supporting "follow-the-sun" product engineering and development processes (CAD, CAE, etc)
  - Managing and integrating the workflow of highly fragmented and geographically dispersed file data generated by employees

- GPFS global name space support and Active File Management provide core capabilities for file sharing
  - Global namespace enables a common view of files, file location no matter where the file requestor, or file resides.
  - Active File Management handles file version control to ensure integrity.
  - Parallel data access allows for large number of files and people to collaborate without performance impact.

Intelligent Cluster System x iDataPlex

Optimized platforms to right-size your Technical Computing operations

IBM leadership for a new generation of **Technical Computing**

Technical Computing is no longer just the domain of large problems

- Businesses of all sizes need to harness the explosion of data for business advantage
- Workgroups and departments are increasingly using clustering at a smaller scale to drive new insights and better business outcomes
- Smaller groups lack the skills and resources to deploy and manage the system effectively

*IBM brings experience in supercomputing to **smaller workgroup and department clusters** with IBM Intelligent Cluster™*

- Reference solutions for simple deployment across a range of applications
- Simplified end-to-end deployment and resource management with Platform HPC software
- Factory integrated and installed by IBM
- Supported as an integrated solution
- *Now even easier with IBM Platform Computing*

IBM Technical Computing expertise

IBM intelligence for clusters of all sizes!

IBM Intelligent Cluster™ – it's about faster time-to-solution

Take the time and risk out Technical Computing deployment

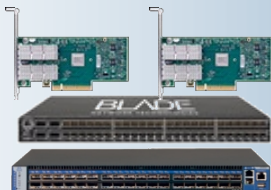Building Blocks: Industry-leading IBM and 3rd Party components

Cluster Management

OS

Management Servers

Compute Nodes

Networking

Storage

Design Build Test Install Support

IBM Intelligent Cluster

Factory-integrated, interoperability-tested **system** with compute, storage, networking and cluster management tailored to your requirements and supported as a solution**!**
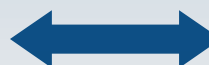
Allows clients to focus on their business not their IT – that is backed by IBM

IBM Intelligent Cluster simplifies large and small deployments

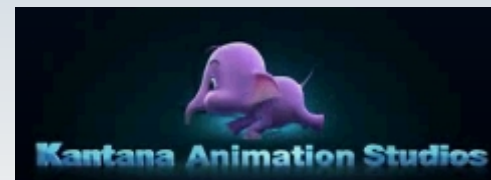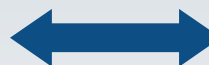**Large**   **Small**

Research

LRZ SuperMUC
Europe-wide research cluster
9,587 servers, direct-water cooled

University of Chile
Earthquake prediction and astronomy
56 servers, air-cooled

Media

Illumination Entertainment
3D Feature-length movies
800 iDataPlex servers
Rear-Door Heat eXchanger cooled

Kantana Animation Studios
Thailand television production
36 iDataPlex servers, air-cooled

Technical Computing Storage

Complete, scaleable, dense solutions

from a single vendor

IBM System Storage® for Technical Computing

Storage

Middleware/Tools

LTO Tape

DCS 3700

HPSS

GPFS

SONAS

DS5000

Tivoli. Software
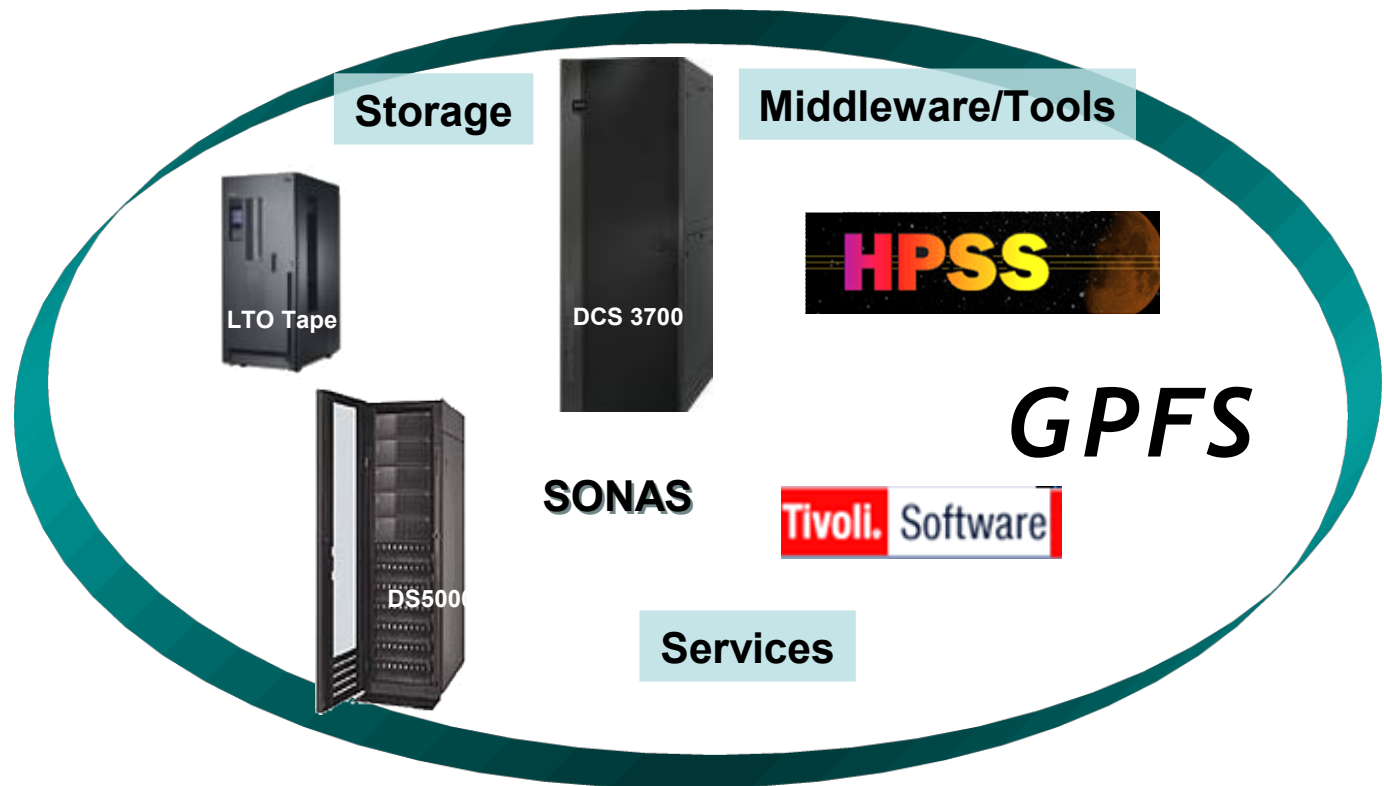
Services

▪Complete, scaleable, integrated solutions from single vendor

▪Scaling to the multi-petabyte and hundreds gigabyte/sec

▪Industry leading data management software and services

▪Big Green features lower overall costs

▪Worldwide support and service

Expanded Capabilities of IBM's Densest Storage Solution...

IBM System Storage DCS3700 now with Performance Module Option
6Gb/s x4 SAS-based storage system

Expandable performance, scalability and density starting at entry-level prices

- **New** DCS3700 Performance Controller

- High density storage system designed for General Purpose Computing and High Performance Technical Computing applications

- IBM's densest disk system: 60 drives and dual controllers in 4U now scales to over 1PB per system with 3TB drives

- **New** Dynamic Disk Pooling feature enables easy to configure Worry-Free storage reducing maintenance requirements and delivering consistent performance

- **New** Thin Provisioning, ALUA, VAAI, Enhanced FlashCopy features deliver increased utilization, higher efficiency, and performance

- Superior serviceability and easy installation with front load drawers

- Bullet-proof reliability and availability designed to ensure continuous high-speed data delivery

The DCS3700 Can Scale In clusters…with **IBM GPFS™**

- Combining IBM's GPFS clustered file management software and DCS3700, creates an extremely scalable and dense file-based management system

- Using a flexible architecture, "building blocks" of DCS3700+GPFS can be organized

	Single Building Block	Two Building Blocks
Configuration	2 GPFS x3650 Servers 3 DCS3700	4 GPFS x3650 Servers 6 DCS3700
Capacity: Raw Usable	 360TB 262TB	 720TB 524TB
Streaming Rate: Write Read	 Up to 4.8 GB/s Up to 5.5 GB/s	 Up to 9.6 GB/s Up to 11.0 GB/s
IOP Rate (4K trans.) Write Read	 3,600 IOP/s 6,000 IOP/s	 7,200 IOP/s 12,000 IOP/s

Customer Success Stories

Applying IBM technology and experience
to solve real-world issues and deliver value

IBM Watson and Citigroup

The Need:

Now, Watson is turning its attention to finance. At Citi, Watson will help the bank by deploying its "deep content analysis and evidence based learning capabilities" to help "**analyze customer needs and process vast amounts of up-to-the-minute financial, economic, product and client data."**

"We are working to rethink and redesign the various ways in which our customers and clients interact with money," Don Callahan, Citi's chief operations and technology officer, said in a statement. "We will collaborate with IBM to explore how we can use the Watson technology to **provide our customers with new, secure services designed around their increasingly digital and mobile lives**."

The Solution:

Citi said it would pair a new "customer interaction solution" with Watson to help "assist decision makers in identifying opportunities, evaluating risks, and exploring alternative actions" for its clients.

Potential Solution components:

- IBM Power

- IBM Watson

- Deep Q&A

- IBM General Parallel File Systems

THANK YOU for listening!