Big Data - Manage it Better

Leila Romane
Database Lead UKI
24th September 2013

Agenda

- 09-15 - 09:30 – Data Management for the Era of Big Data - Danilo Noveli
- 09:30 -10:00 - How Big Data is Transforming Industry - Chris Grote
- 10:00 – 10:30 – DB2 with Blu Acceleration – Les King
- 10:30 – 11:00 - DB2 Blu for SAP – Ferdinand Prezenski
- 11.:00 – 11:15 - Coffee break and Ask the Experts
- 11:15 - 11:30 - Update on Purescale – Les King
- 11:30 - 12:00 - PureData TX – Ferdinand Prezenski & Richard Hewitt
- 12:00 - 12:30 – Informix - NoSQL for the Enterprise = NewSQL - Steve Shoaf
- 12:30 - Lunch

Data Management for the Era of Big Data
Innovation that Matters to Your Business

Danilo Novelli
Director, WW Database Sales

without analytics

BigData

is just a bunch of data

MYTH: Big Data is only about large datasets; let's say larger than what you have

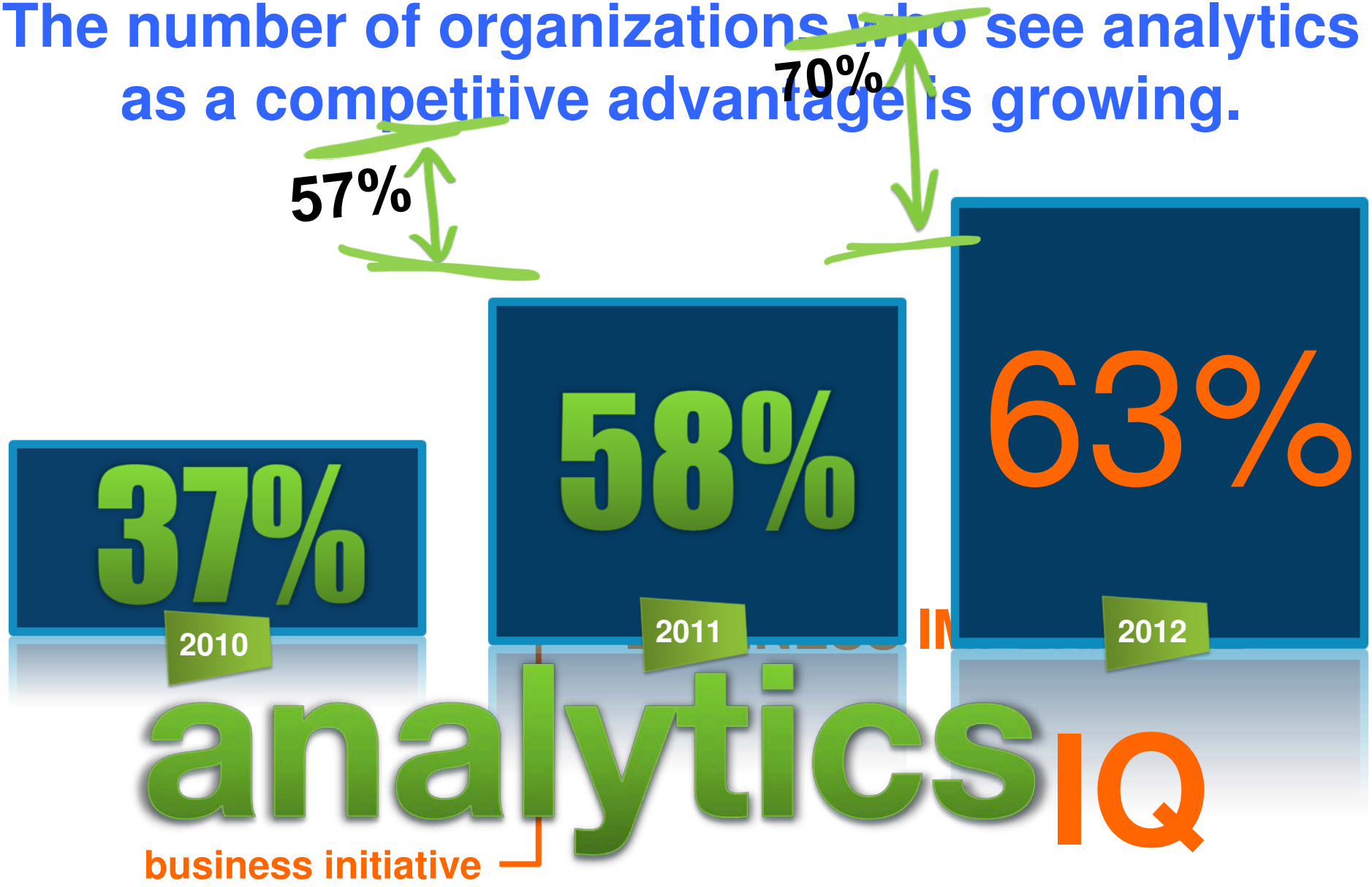
MYTH: Big Data means Hadoop..that's it

MYTH: Big Data means 'rip-and-replace', no need for RDBMS and governance

MYTH: NoSQL means no SQL

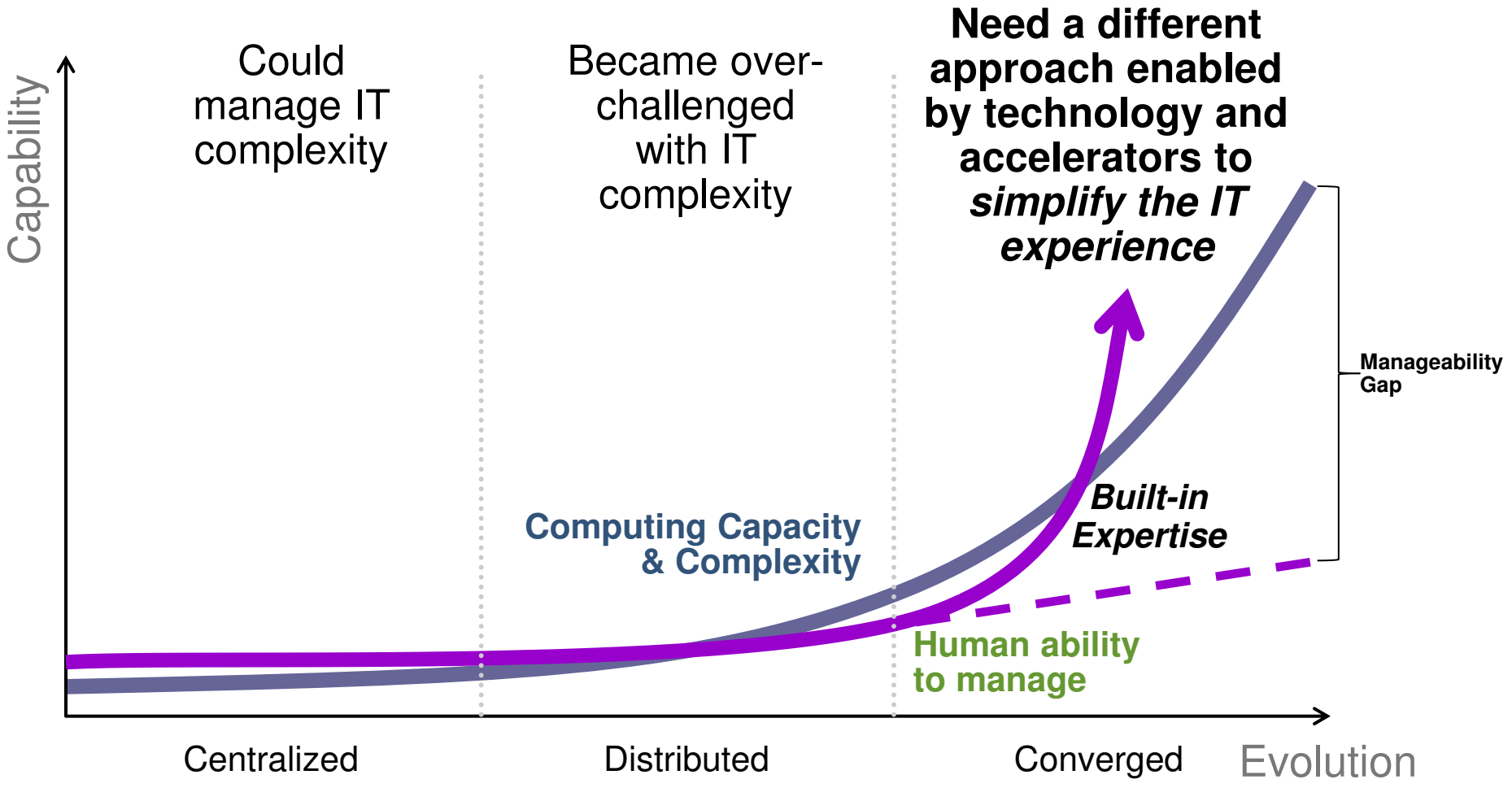
MYTH: Big Data means unstructured data for sentiment

The number of organizations who see analytics as a competitive advantage is growing.

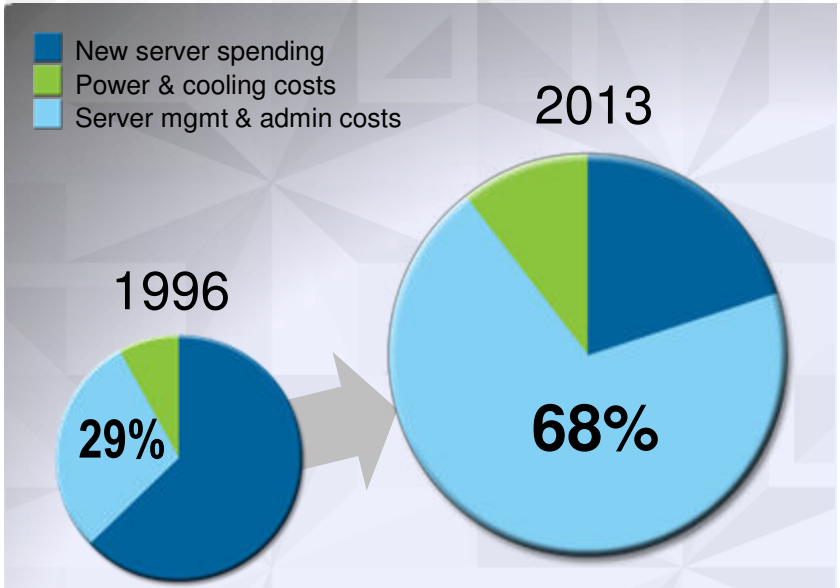


Source:
IBM IBV/MIT Sloan Management Review Study 2011

The challenge is the time and cost spent on managing data



Technology is the leading force for impacting business
Is IT ready for the challenge?



68% of IT

operating costs in 2013 will be for
management and administration

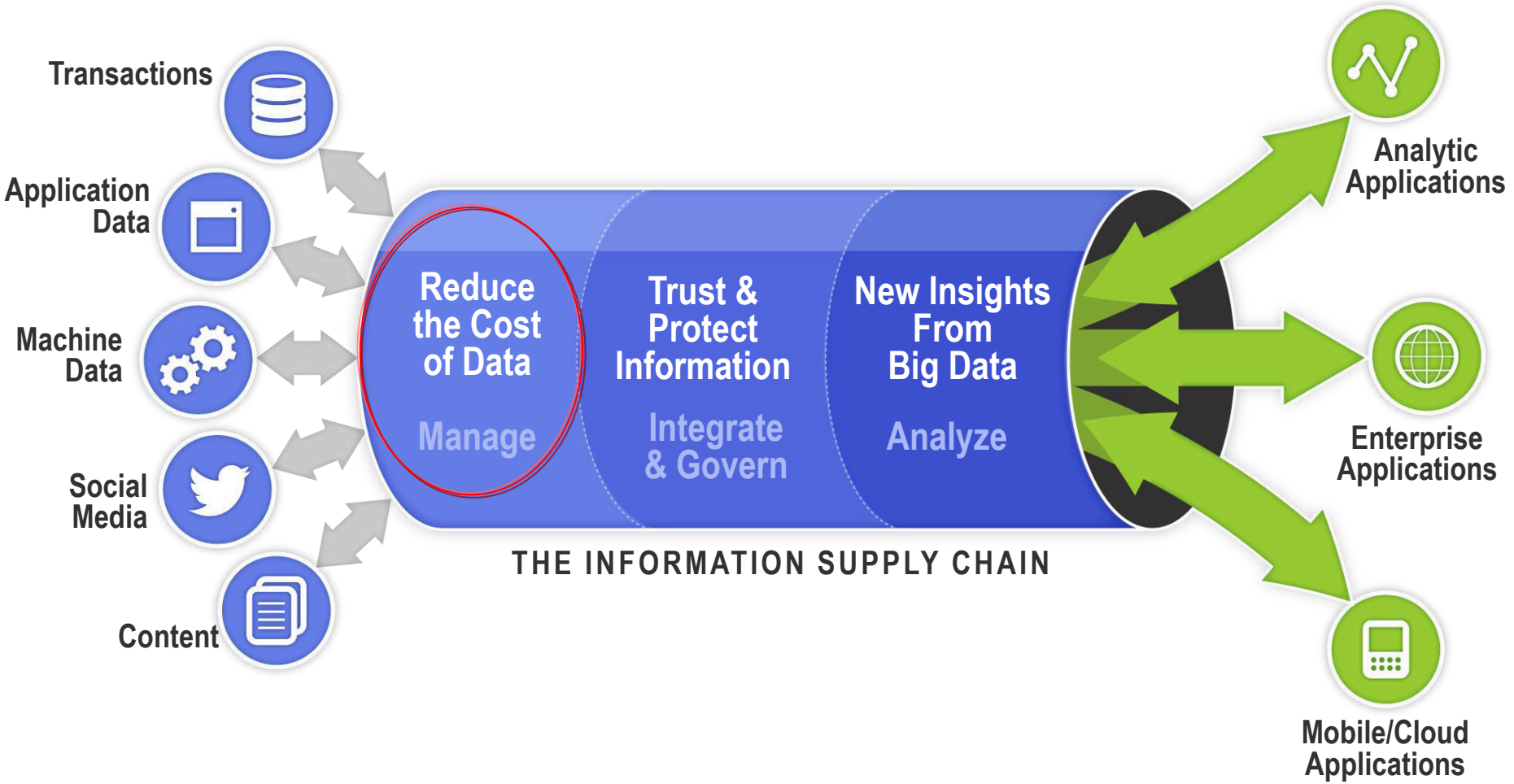
* IDC; Converged Systems: End-User Survey Results presentation; September 2012; Doc #236966

Only 1 in 5

organizations allocate more than
50% of IT budget to new projects

* 2012 IBM Global Data Center Study

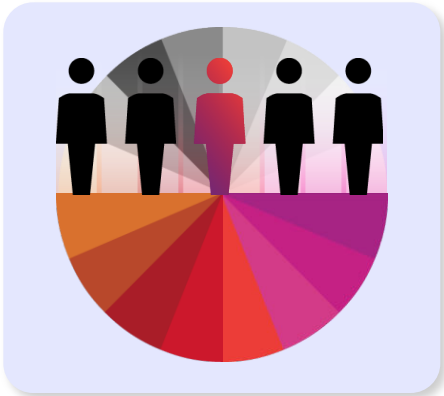
An integrated and holistic approach to harnessing big data



IBM Data Management: Reduce the cost of data

We need to improve efficiency…

*Only 1 in 5 organizations allocate more than 50% of IT budget to new projects**



* 2012 IBM Global Data Center Study

…to deliver improved outcomes

1. Future-proof for lower operational and storage costs, and gain better database performance
2. Expertise and simplicity in meeting diverse workload requirements
3. Deliver right-time insights with operational data warehousing
4. Improve application performance and lower cost by archiving data
5. Increase application efficiency of development and test

Innovation that Matters to Your Business

Expert Systems

Integrated by design, built in expertise, simplicity

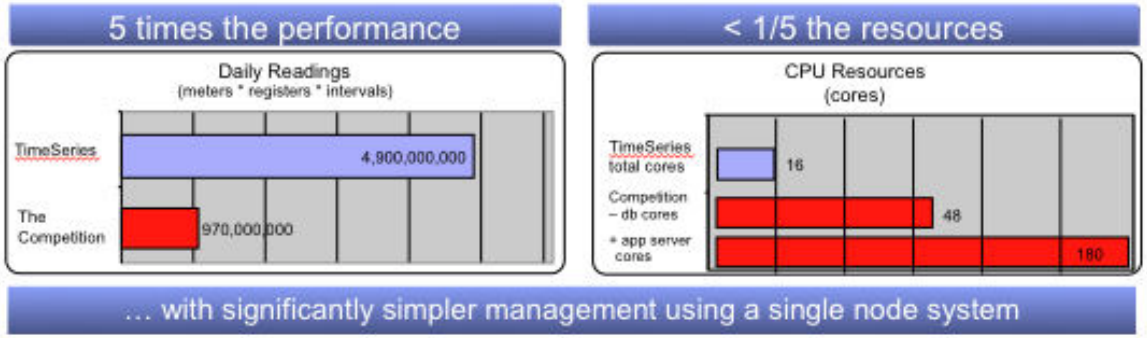
PureData (New)
System for Hadoop

PureData
System for Analytics

PureData
System for Operational Analytics

PureData
System for Transactions

Time-based Sensor Data Processing



BLU Acceleration - Simple and fast! Load and go!

Dynamic In-Memory

LIS.TEC

"Even if your data does not completely fit into memory - you still have great performance gains. In the tests we ran we were seeing queries run up to 100X faster with BLU Acceleration."
– Joachim Klassen, Consultant, LIS.TEC

Actionable Compression

mindray

"… we continued to see exceptional compression rates – our tables compressed at over 92%. But, our greatest thrill wasn't the compression rates (though we really like it), rather the improvement we found in query speed which was more than 50X faster than with row-organized tables."
- Xu Chang, Chief DBA Support - DB2 and Oracle Databases

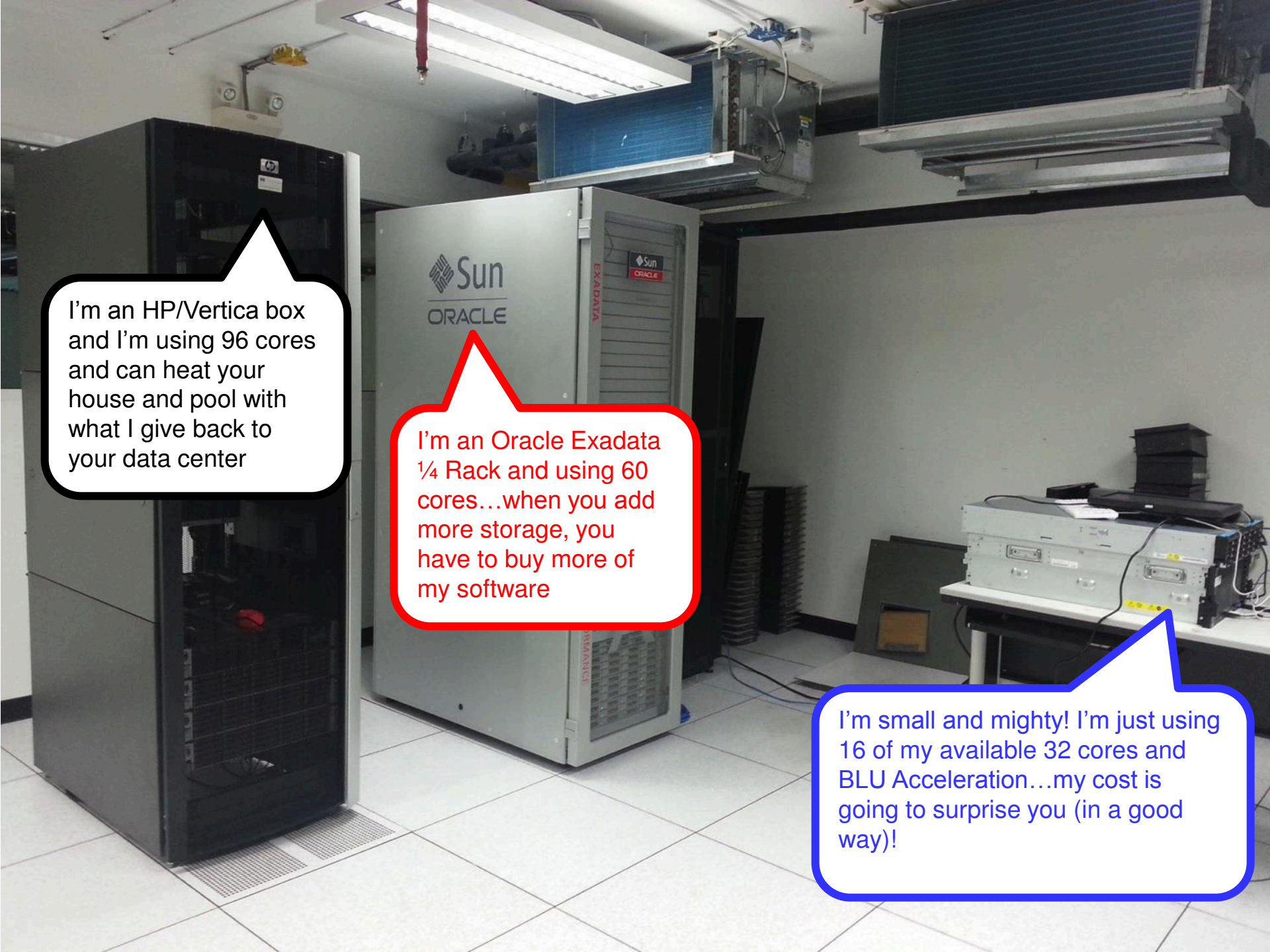
Parallel Vector Processing

"The performance of DB2 10.5 with BLU Acceleration is quite amazing. We ran our tests on a system that is about 16x less powerful than our current production system. And yet BLU was able to outperform our production system in every aspect."
–Mohankumar Saraswatipura, Lead DBA,Reckitt Benckiser

Data Skipping

BNSF RAILWAY

"It was amazing to see the faster query times compared to the performance results with our row-organized tables. The performance of four of our queries improved by over 100-fold! The best outcome was a query that finished 137x faster by using BLU Acceleration."
- Kent Collins, Database Solutions Architect, BNSF Railway


I'm an HP/Vertica box and I'm using 96 cores and can heat your house and pool with what I give back to your data center
Sun
ORACLE
EXADATA
I'm an Oracle Exadata ¼ Rack and using 60 cores…when you add more storage, you have to buy more of my software
I'm small and mighty! I'm just using 16 of my available 32 cores and BLU Acceleration…my cost is going to surprise you (in a good way)!

Data Management in Action

50%

Reduction in data management costs

60%

Cut in storage requirements – $1M

$1M

Saved in software and storage costs

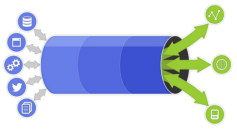
30%

Reduction in heating costs – 8k tonnes less of CO_2

1500X

Faster queries: 11 hours to 26 seconds

IBM Information Management: Committed To Client Success



Broadest and best portfolio for Big Data

Data Management, Big Data Platform, Information Integration and Governance

More delivery choices and lower TCO

Multi-platform Software, PureData Systems, Cloud Services, System z

Proven expertise and innovation that drive faster results

Gain results within 30 days or less

Get started on any information challenge and grow

Reduce the Cost of Data, Trust and Protect Information, New Insights from Big Data

Thank You!

Agenda

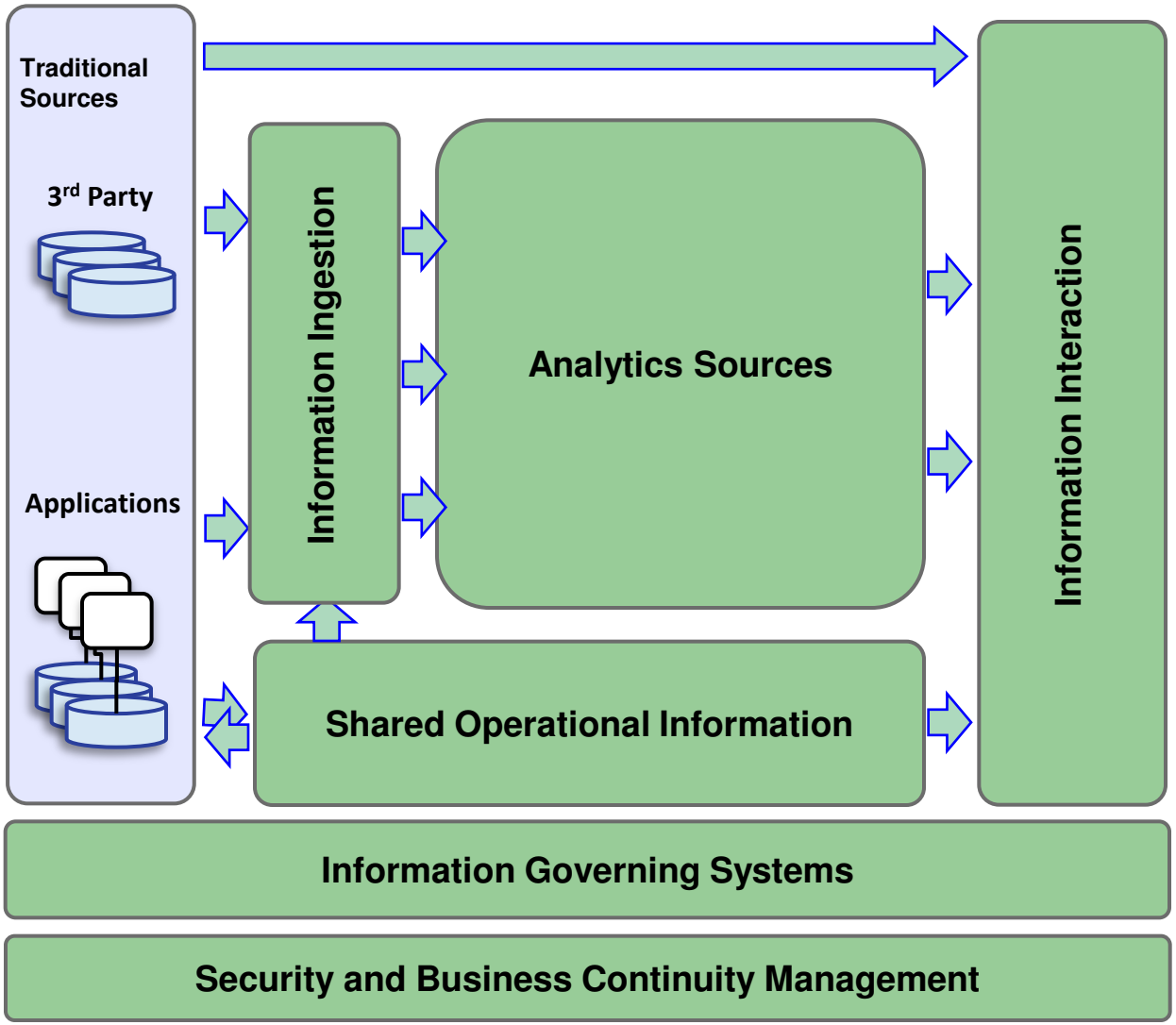
- 09-15 - 09:30 – Data Management for the Era of Big Data - Danilo Noveli
- 09:30 -10:00 - How Big Data is Transforming Industry - Chris Grote
- 10:00 – 10:30 – DB2 with Blu Acceleration – Les King
- 10:30 – 11:00 - DB2 Blu for SAP – Ferdinand Prezenski
- 11.:00 – 11:15 - Coffee break and Ask the Experts
- 11:15 - 11:30 - Update on Purescale – Les King
- 11:30 - 12:00 - PureData TX – Ferdinand Prezenski & Richard Hewitt
- 12:00 - 12:30 – Informix - NoSQL for the Enterprise = NewSQL - Steve Shoaf
- 12:30 - Lunch

Evolution of Data Management in the Big Data Landscape

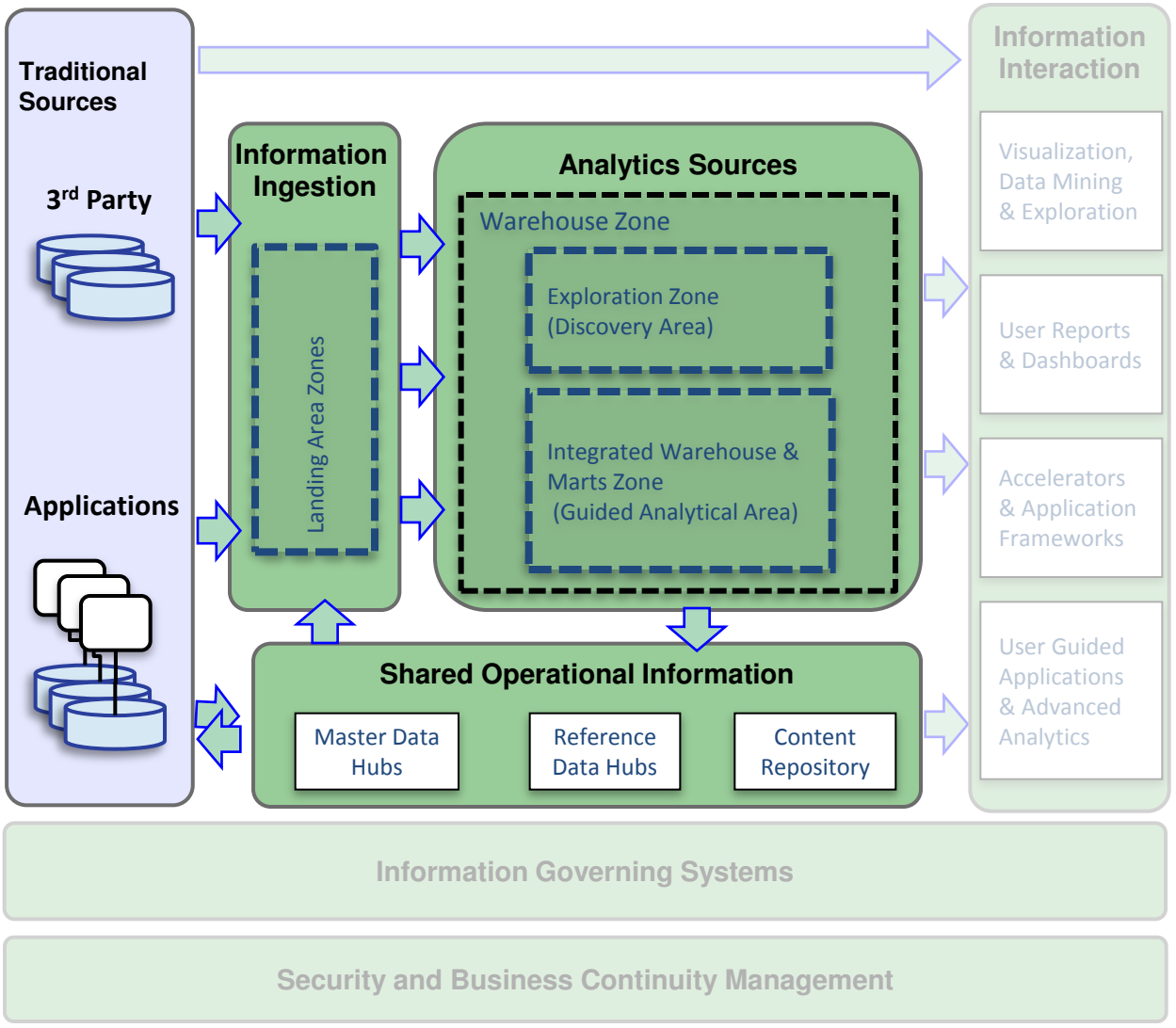
London, September 24th 2013

Chris Grote

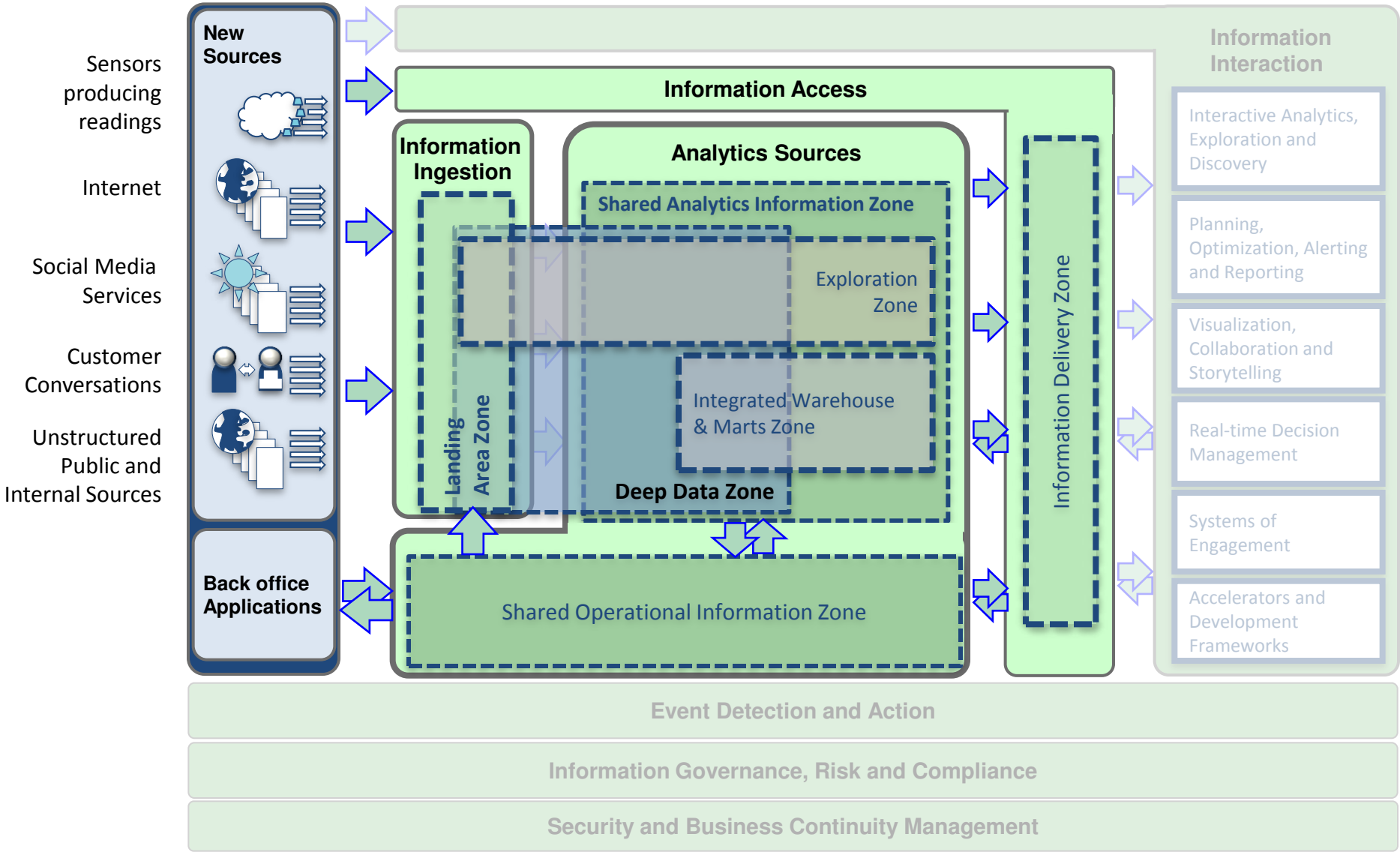
Traditional Landscape Recap



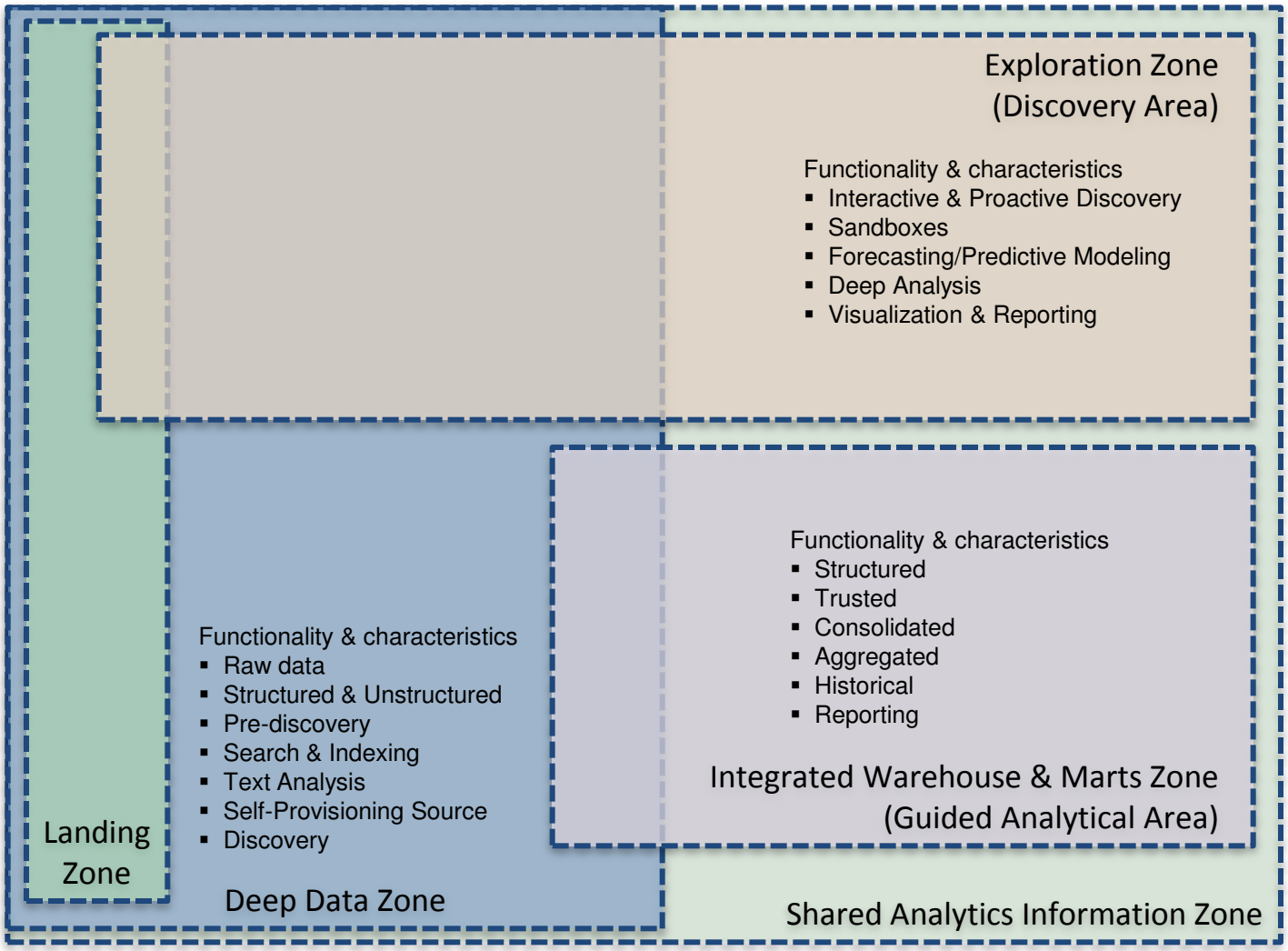
Traditional Landscape Recap – Analytical Sources (Zones View)



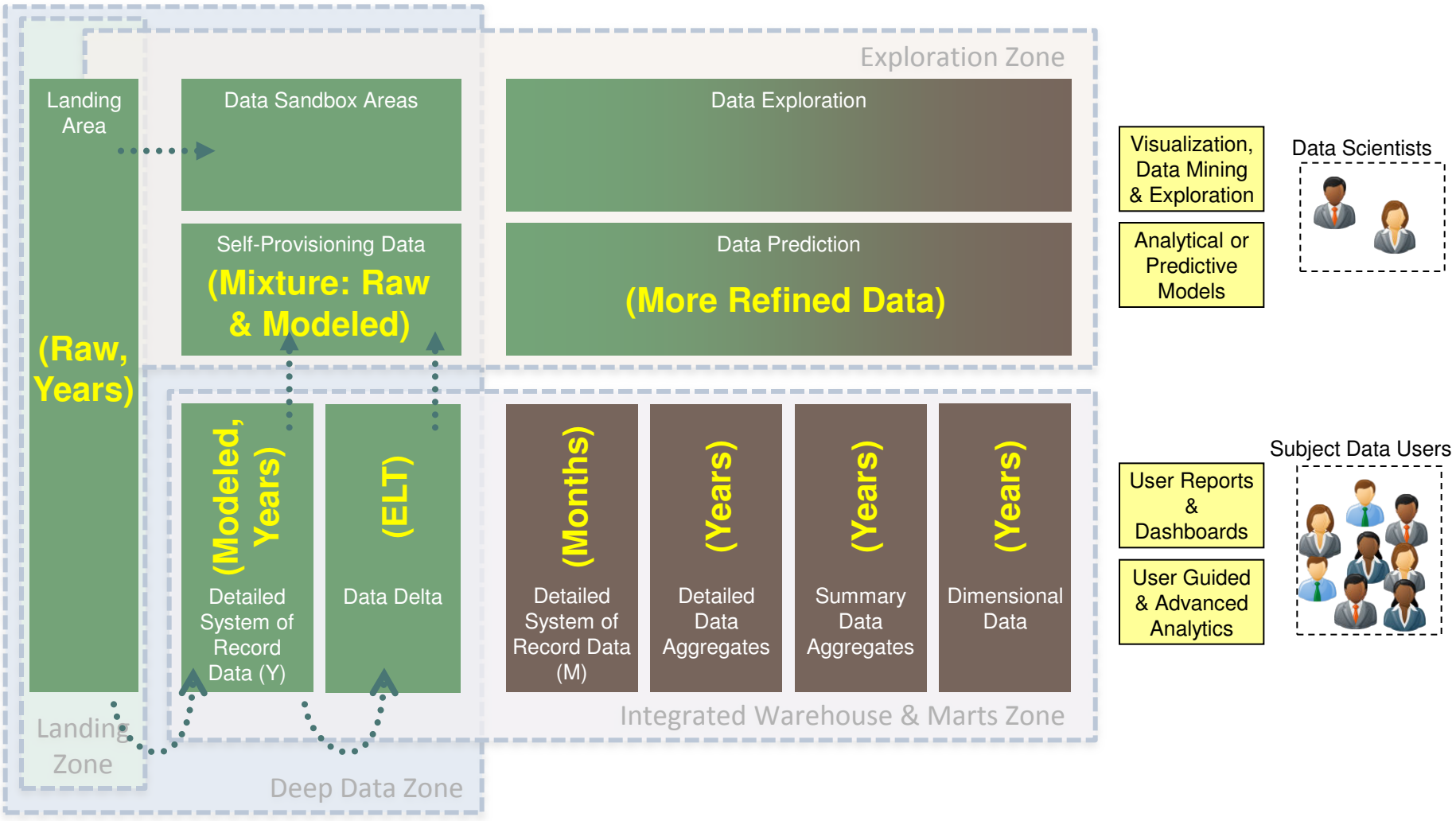
What is changing? Big Data Enhanced Analytical Zones



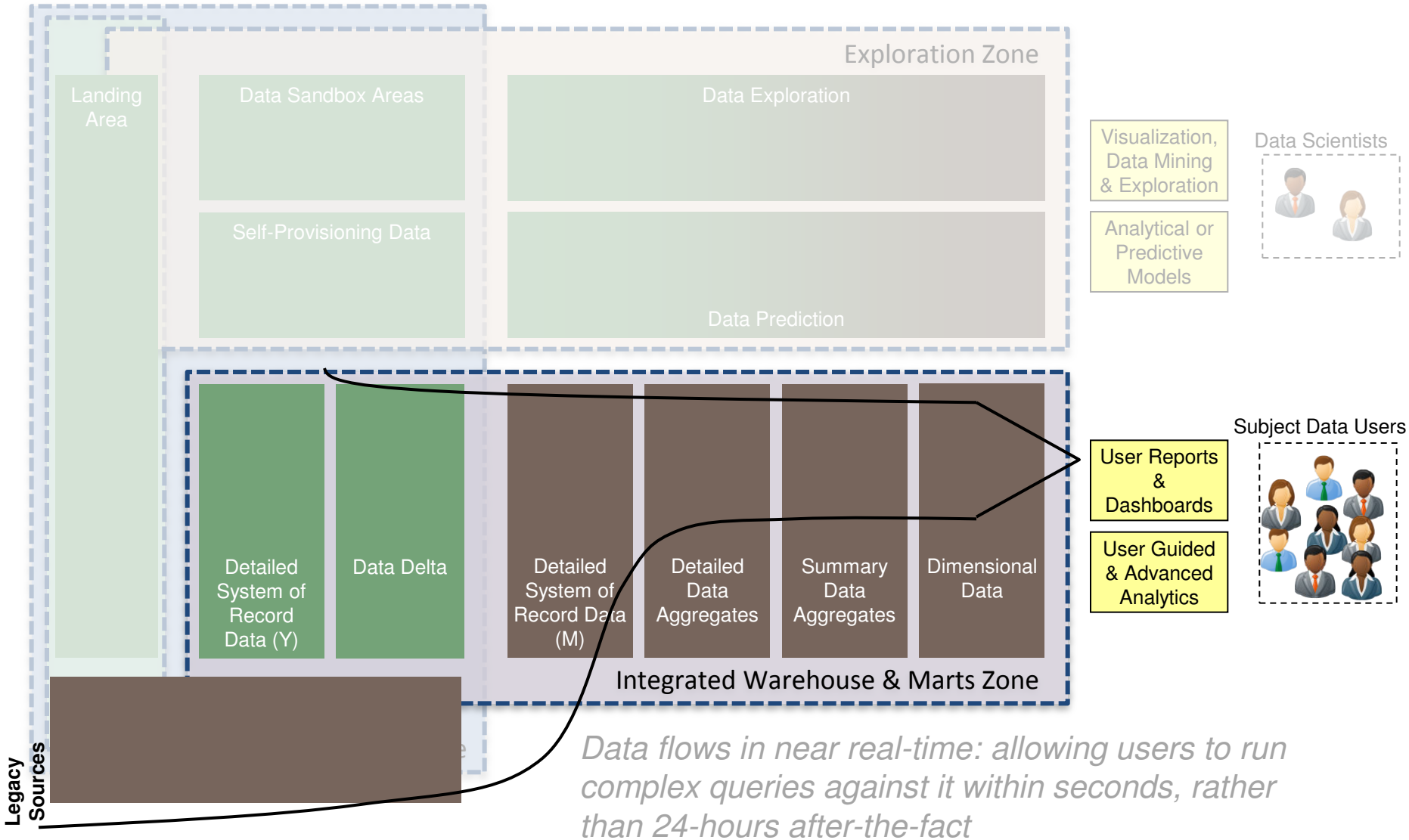
Characteristics of Big Data Enhanced Analytical Zones



Big Data Analytical Zones – Schema Areas

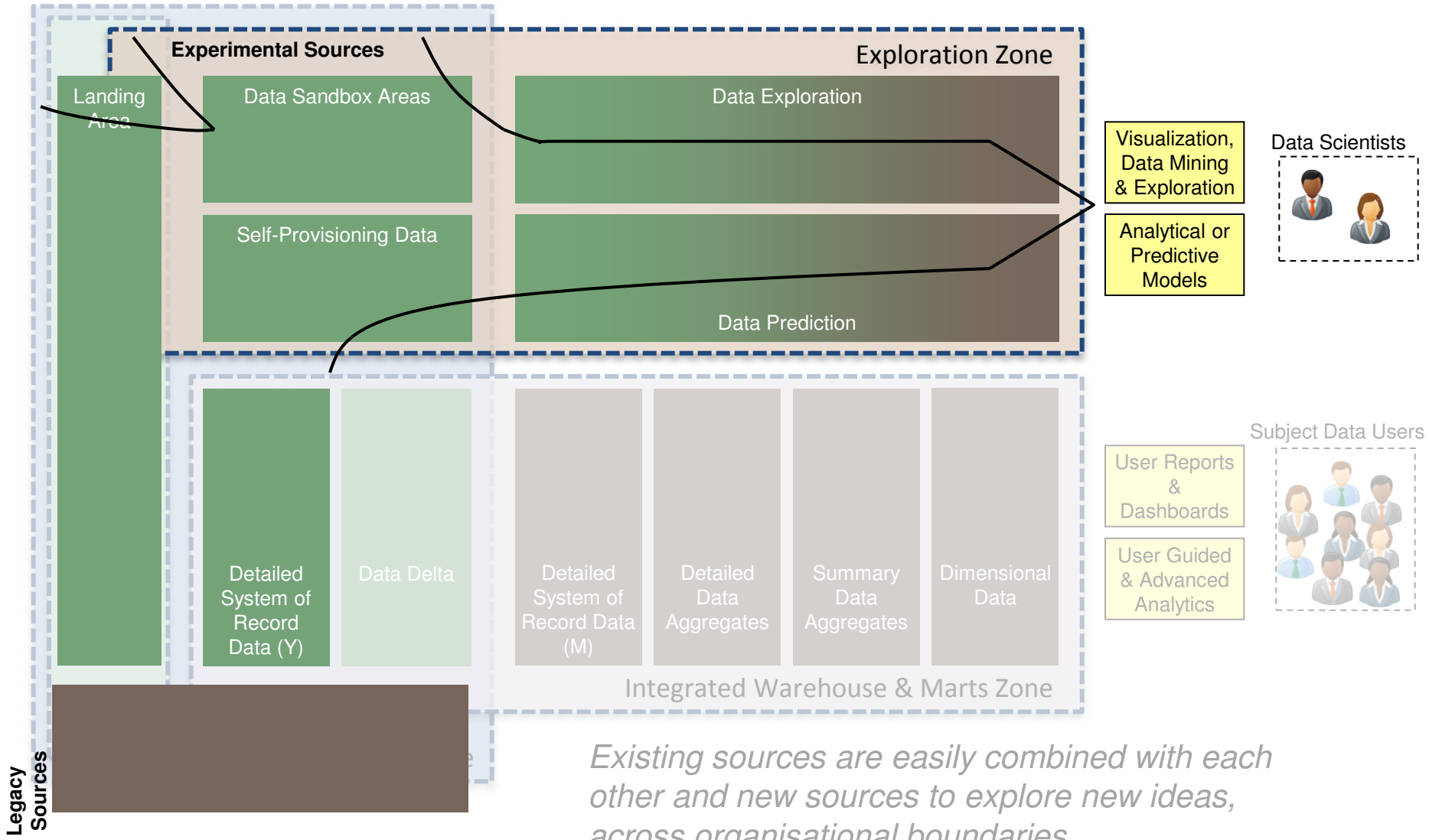


Example: optimising end-user interactivity to quickly identify fraud



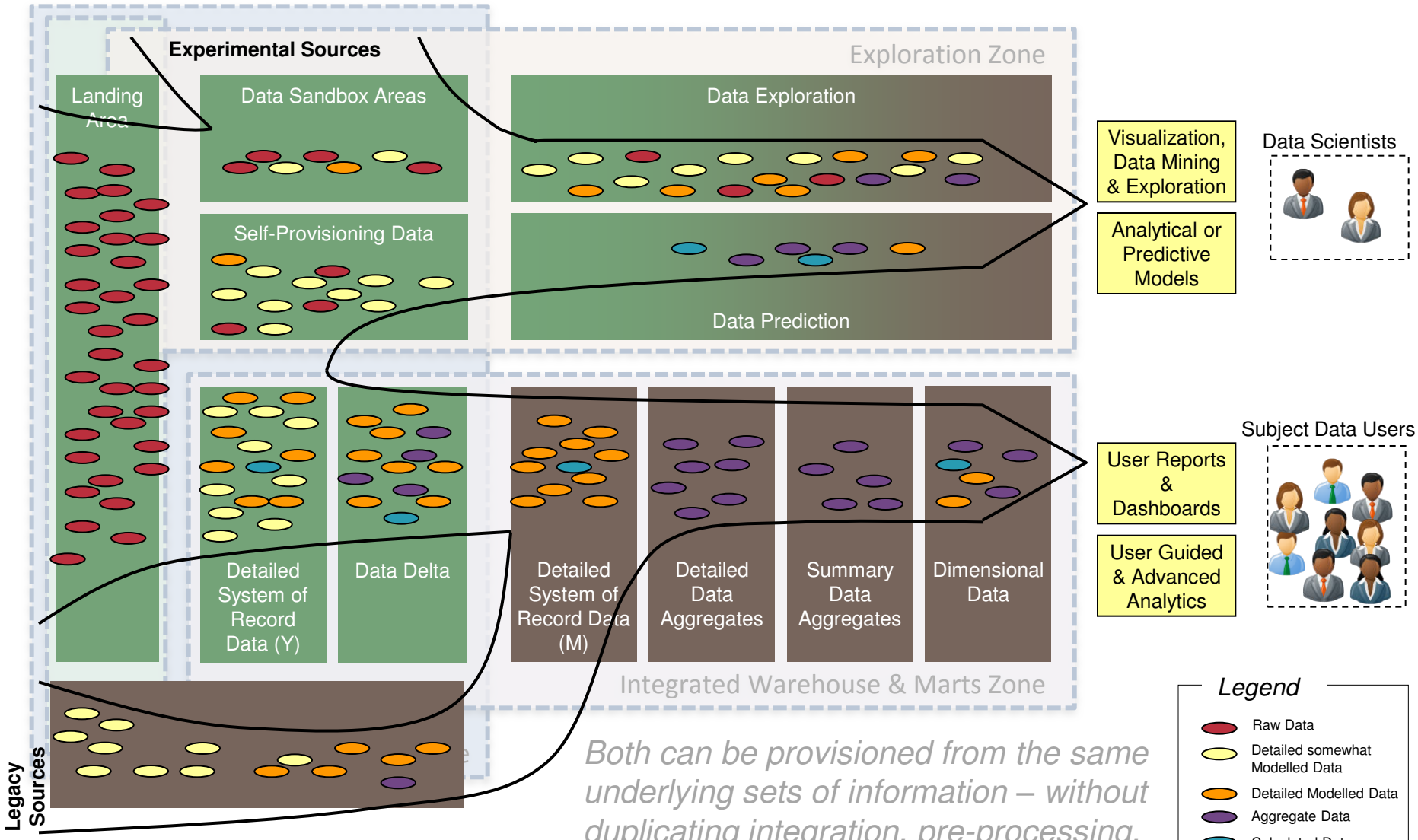
Data flows in near real-time: allowing users to run complex queries against it within seconds, rather than 24-hours after-the-fact

Example: incorporating additional sources to explore new fraud



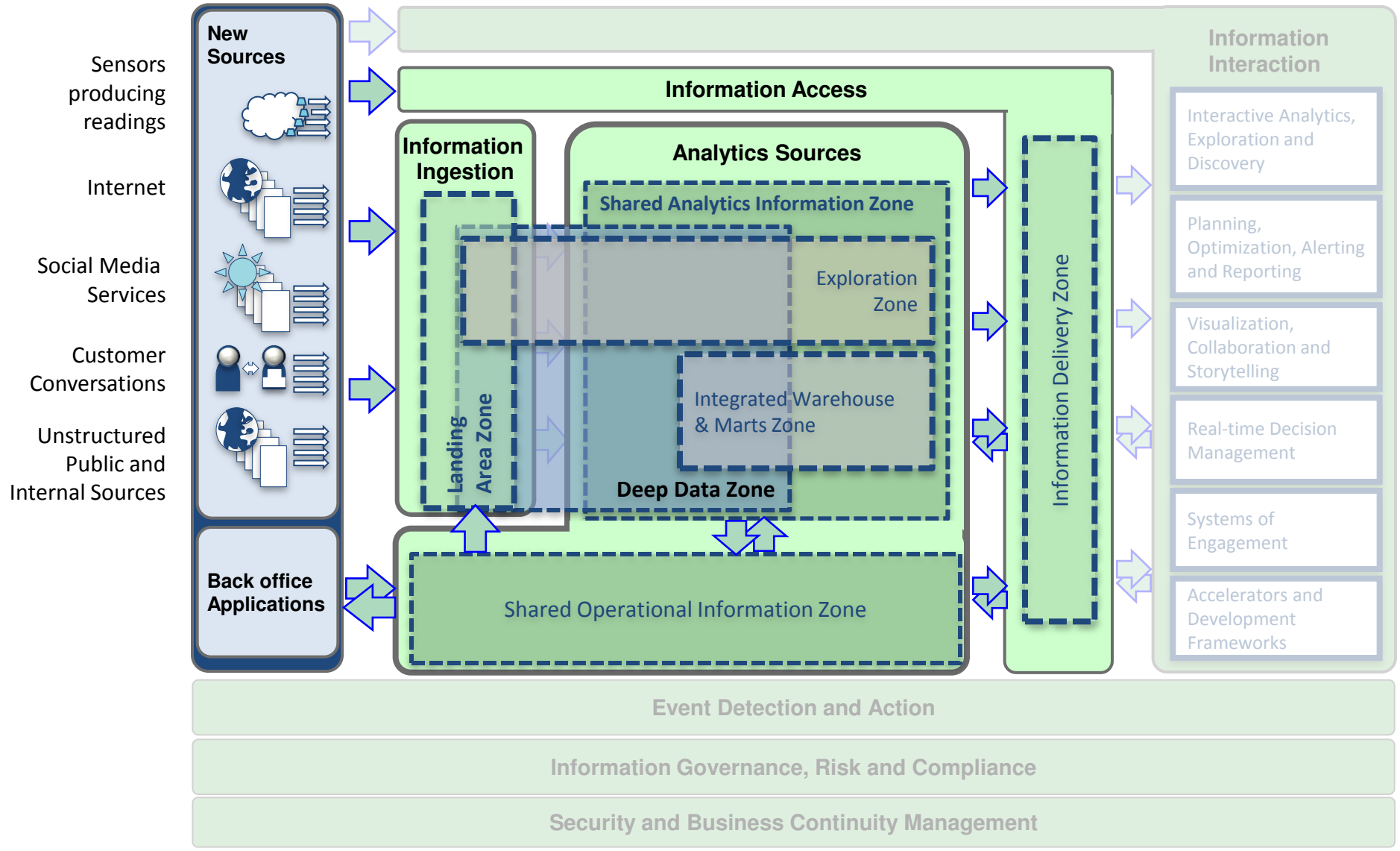
Existing sources are easily combined with each other and new sources to explore new ideas, across organisational boundaries

Big Data Analytical Zones – Schema Areas



Both can be provisioned from the same underlying sets of information – without duplicating integration, pre-processing, etc

Big Data and the Broader IT Landscape

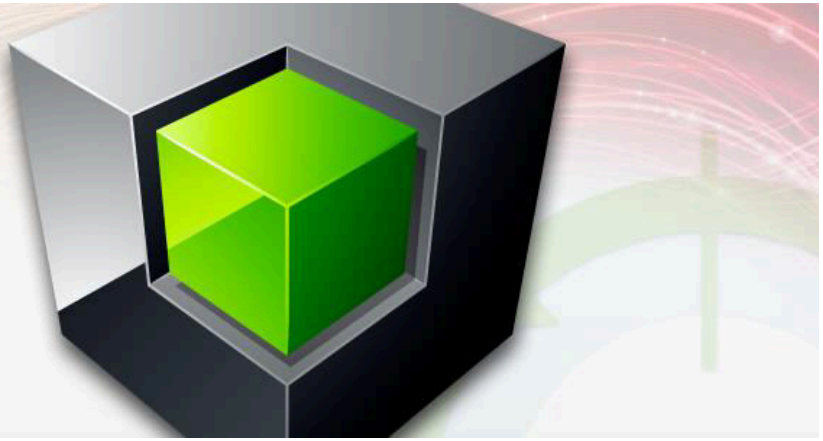


Agenda

- **09-15 - 09:30 - Key Note - Danilo Noveli**
- **09:30 -10:00 - How Big Data is Transforming Industry - Chris Grote**
- **10:00 – 10:30 – DB2 with Blu Acceleration – Les King**
- **10:30 – 11:00 - DB2 Blu for SAP – Ferdinand Prezenski**
- **11.:00 – 11:15 - Coffee break and Ask the Experts**
- **11:15 - 11:30 - Update on Purescale – Les King**
- **11:30 - 12:00 - PureData TX – Ferdinand Prezenski & Richard Hewitt**
- **12:00 - 12:30 – Informix - NoSQL for the Enterprise = NewSQL - Steve Shoaf**
- **12:30 - Lunch**

DB2® 10.5

with BLU Acceleration



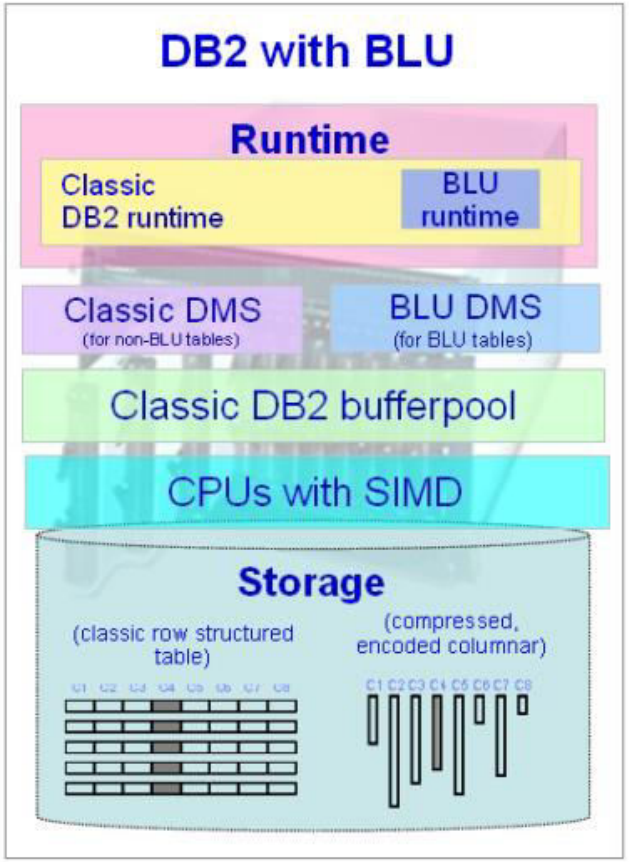
Les King
Director, Information Management
lking@ca.ibm.com
August, 2013

Customer Video

http://www.youtube.com/watch?v=5T6f74gYu1Y&noredirect=1

What is DB2 with BLU Acceleration?

- **New technology for analytic queries in DB2 LUW**
  - DB2 column-organized tables add columnar capabilities to DB2 databases
    - Table data is stored column organized rather than row organized
    - Using a vector processing engine
    - Using this table format with star schema data marts provides significant improvements to storage, query performance, ease of use, and time-to-value
  - New unique runtime technology which leverages the CPU architecture and is built directly into the DB2 kernel
  - New unique encoding for speed and compression
    - This new capability is both main-memory optimized, CPU optimized, and I/O optimized



Target Use Cases for BLU Acceleration

1. **Analytics and Reporting**
   - Single Server implementation
   - Targeting environments with <20TB of active data
   - Some DPF environments can now be single partition

2. **Mixed Workload Environments**
   - 40%-50% of OLTP environments have analytics and reporting activity
   - Accelerate the analytic and reporting workloads
   - Reduce pressure on the OLTP workload

3. **Cognos, OLAP, Dynamic Cubes**
   - Would provide same value in any other similar type environment

4. **SAP BW**
   - In beta now
   - Expect certification by end of 3Q

How Fast Is BLU Acceleration?

Customer	Performance Gains
BNSF	Up to 137x
Handelsbanken	7x – 100x
Triton Consulting	46x
Yonyou	40x
Coca-Cola Bottling	4x - 15x

~25x
speedup
is average

BNSF RAILWAY

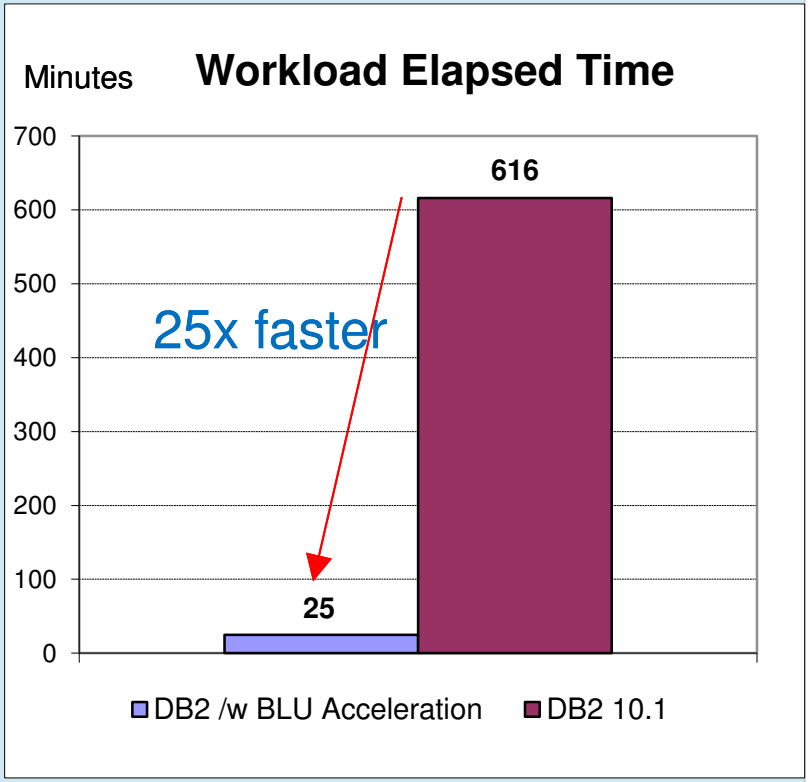
"It was amazing to see the faster query times compared to the performance results with our row-organized tables. ***The performance of four of our queries improved by over 100-fold! The best outcome was a query that finished 137x faster by using BLU Acceleration."***

- Kent Collins, Database Solutions Architect, BNSF Railway

BLU Acceleration Performance

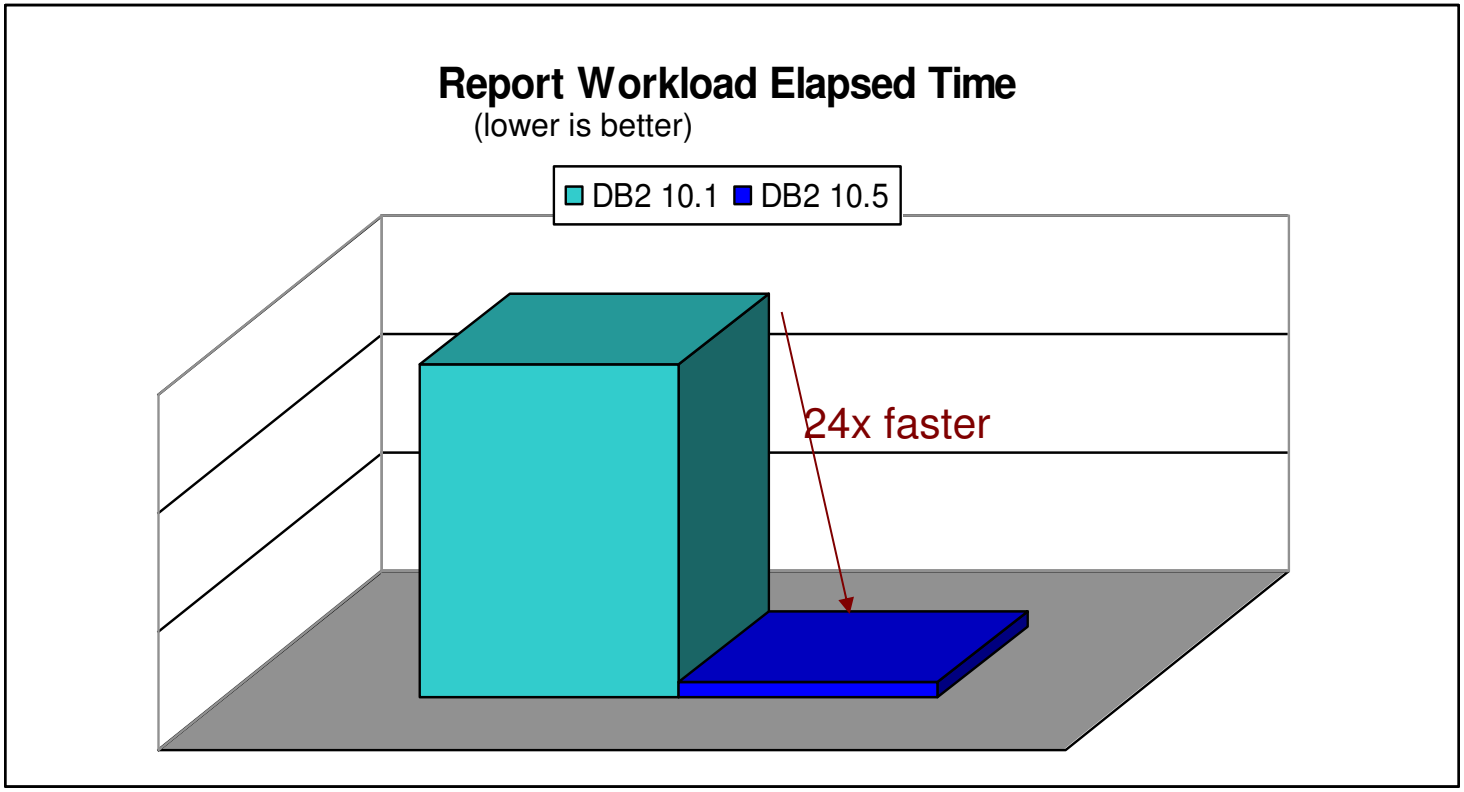
New

- **POPS (Proof of Performance and Scalability)**
  - Derived from Redbrick performance test
  - Classic sales analytics
  - 5.5years of data (2000 days) for 63 stores
    - ~4TB of raw data
    - 2 fact tables
    - 5 dimension tables
  - Broad range of queries with varying selectivity / aggregation
- **Substantial Storage Savings with BLU Acceleration**
  - 2.5x less space than DB2 10.1
- **Massive Performance Gains**
  - 25x speedup over DB2 10.1
  - Maximum query speed up over 400x



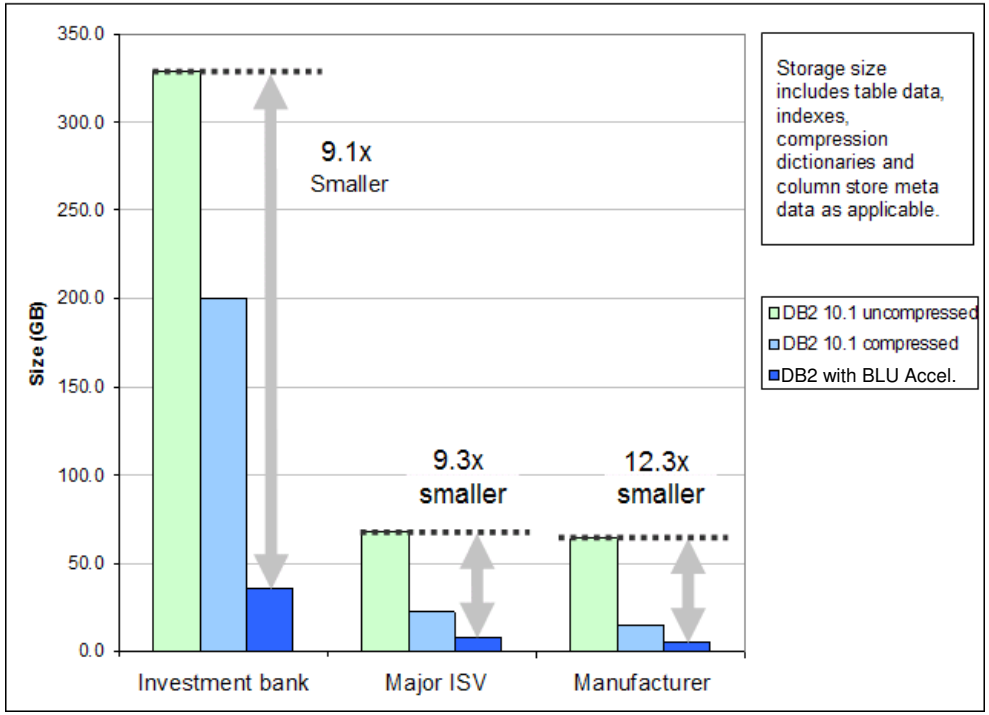
Intel® Xeon® Processor E5-4650

DB2 10.5 and Cognos BI 10.2 Dynamic Cubes on Power 7+

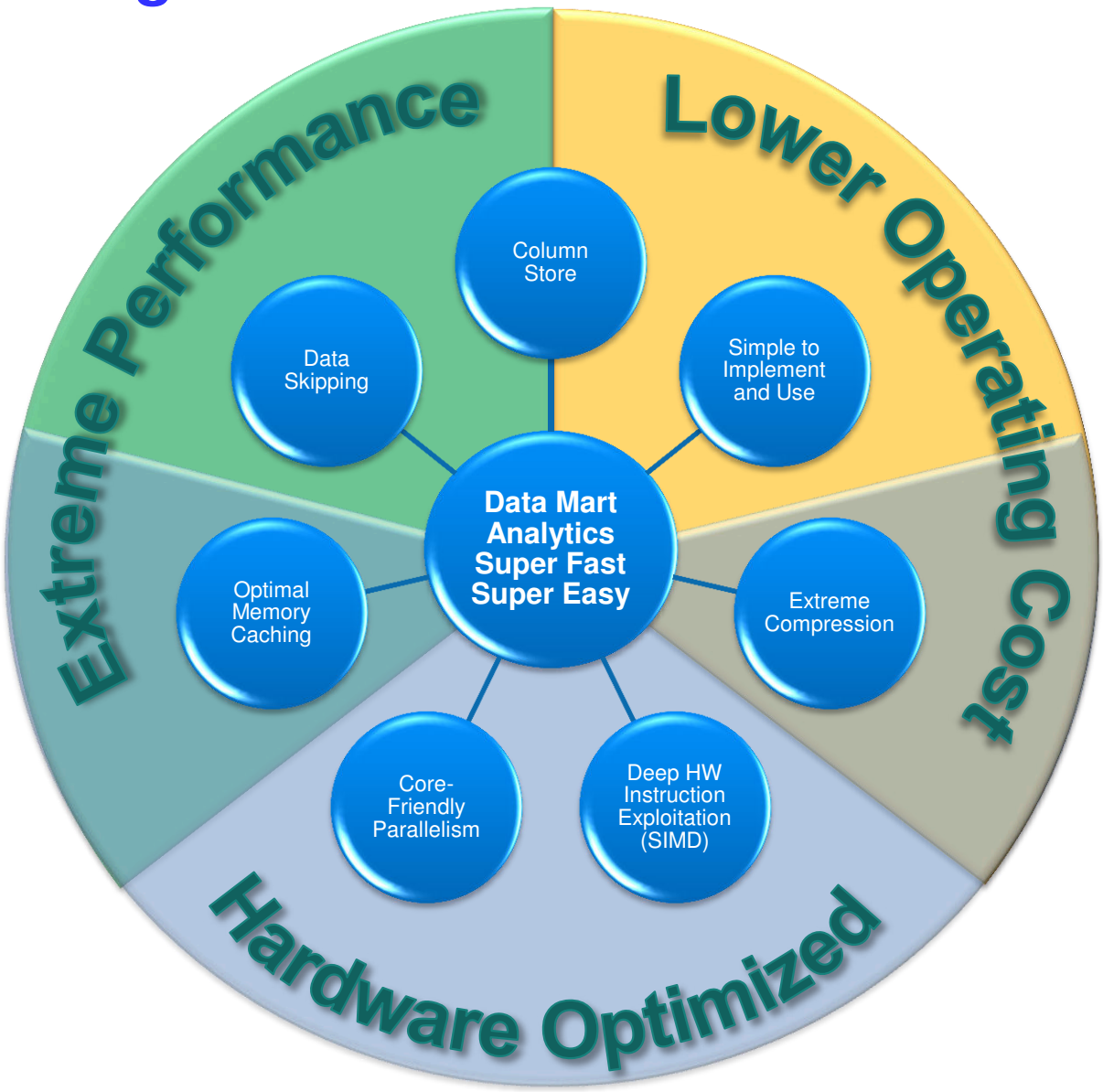


Storage Savings

- **Multiple examples of data requiring substantially less storage**
  - 5% of the uncompressed size
  - Fewer objects required
- **Multiple compression techniques**
  - Combined to create a near optimal compression strategy
- **Compression algorithm adapts to the data**



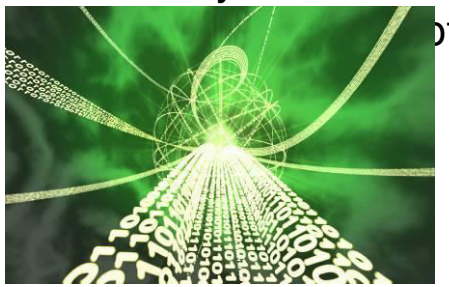
The Seven Big Ideas of DB2 with BLU Acceleration



7 Big Ideas: Our secret sauce

1

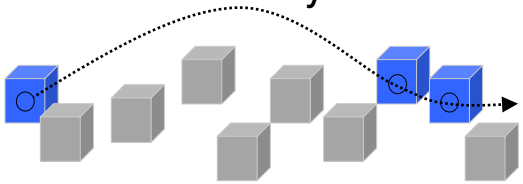
- **Compute-friendly encoding & compression**
  - Massive compression with approximate Huffman encoding
  - Evaluation while compressed!
  - Register-friendly encoding dramatically improves efficiency



of I/O,

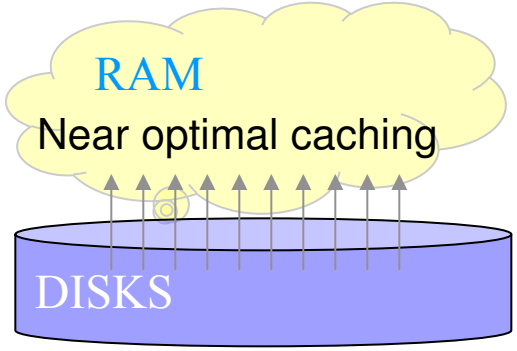
2

- **Automatic Strata Maps**
  - Synopses enable automatic detection of large sections of data that can be ignored by a query
  - Order of magnitude savings in <u>all</u> of I/O, RAM and CPU
  - No DBA action to define or use – truly invisible.



3

- **Scan-friendly memory caching**
  - New algorithms cache in RAM effectively
  - High percent of interesting data fits in memory
  - Data can be larger than RAM



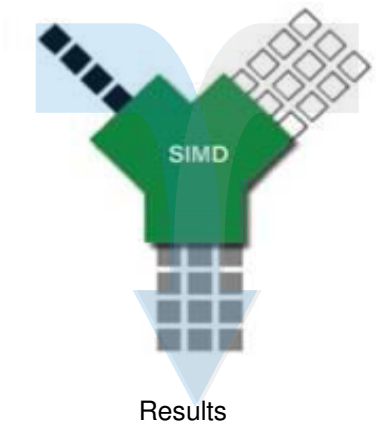
7 Big Ideas: Our secret sauce

4

- **Core-friendly parallelism**
  - KIWI: Kill It With Iron. Uniquely leverage multi-core
  - Careful attention maximizes CPU cache, cacheline efficiency.
  - Many have tried, few have succeeded

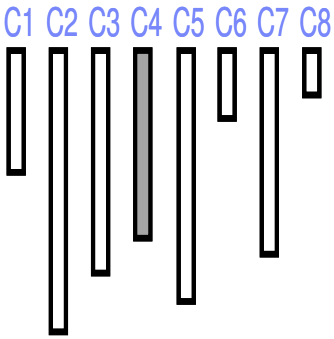
5

- **Multiply the power of the CPU**
  - Single Instruction Multiple Data (SIMD)
  - Compute with 1 instruction what may have taken >4.
  - CPU vector processing
  - Scanning, joins



6

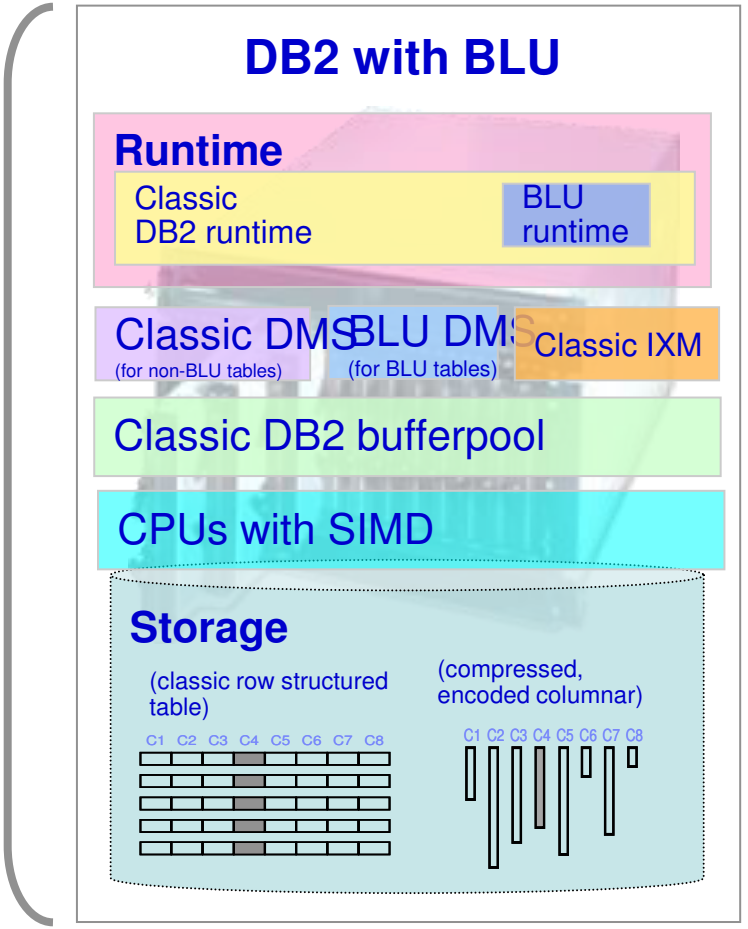
- **Column Store**
  - Minimal I/O
  - Improved compression
  - Favors scan based processing
  - L2 efficiency



7 Big Ideas: Our secret sauce

7

- **Single mature database system**
  - **BLU is part of DB2**
    - Reuse DB2 process model, storage
  - **Rich function and admin tools**
    - All core SQL capabilities
      - DB2 classic runtime compensates for any missing pieces. (e.g. some scalar functions, sorting, etc)
    - Rich existing DB2 admin utilities and tools
  - **Reliable**
    - Majority of the system is mature DB2 code
    - New code is validated with **hundreds of thousands** of existing DB2 test cases, stress systems, and performance measurements.



BLU Acceleration : Memory optimized. CPU optimized. I/O optimized

Memory Optimized

- Memory latency optimized for
  - Scans
  - Joins
  - Aggregation
- More useful data in memory
  - Data stays compressed
  - Scan friendly caching
- Less to put in memory
  - Columnar access
  - Late materialization
  - Data skipping

CPU Optimized

- CPU acceleration
  - SIMD processing for
    - Scans
    - Joins
    - Grouping
    - Arithmetic
- Keeping the CPUs busy
  - Core friendly parallelism
- Less CPU processing
  - Operate on compressed data
  - Late materialization
  - Data skipping

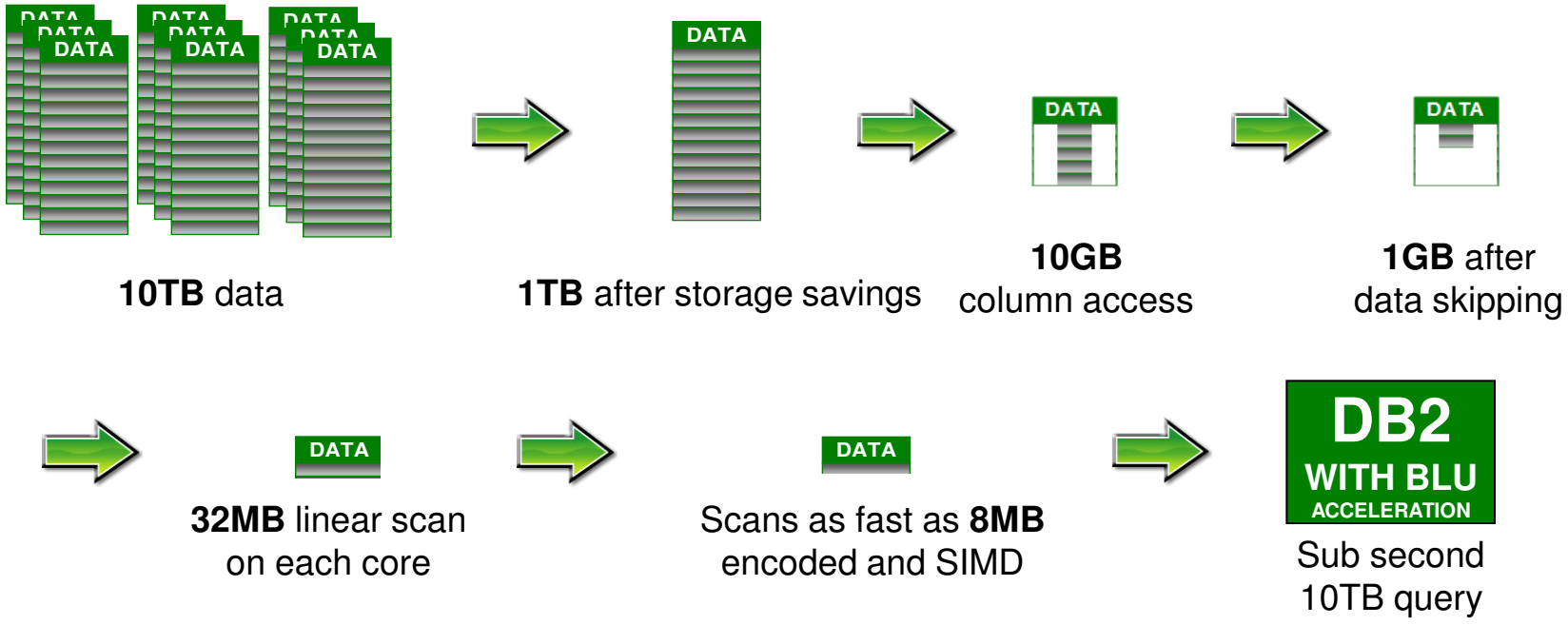
I/O Optimized

- Less to read
  - Columnar I/O
  - Data skipping
  - Late materialization
- Read less often
  - Scan friendly caching
- Efficient I/O
  - Specialized columnar prefetching algorithm

7 Big Ideas: How DB2 with BLU Acceleration Helps

~Sub second 10TB query – An Optimistic Illustration

- **The system – 32 cores, 10TB table with 100 columns, 10 years of data**
- **The query: `SELECT COUNT(*) from MYTABLE where YEAR = '2010'`**
- **The optimistic result: sub second 10TB query! Each CPU core examines the equivalent of just 8MB of data**



More Customer Experience

"It was amazing to see the faster query times compared to the performance results with our row-organized tables. The performance of four of our queries improved by over 100-fold! The best outcome was a query that finished 137x faster by using BLU Acceleration."
- Kent Collins, Database Solutions Architect, BNSF Railway

"We were very impressed with the performance and simplicity of BLU. We found that some queries achieved an almost 100x speed up with literally no tuning!"
- Lennart Henäng, IT Architect, Handelsbanken

"I have now reviewed IBM's new Big Data effort, BLU Acceleration, and my opinion is this: yes, it will deliver major performance enhancements in a wide variety for specific Big Data cases, and yes, I do view their claim of 1000x acceleration in some cases as credible."
– Wayne Kernochan, President

Infostructure Associates
Leveraging Information for Competitive Advantage

The performance of DB2 10.5 with BLU Acceleration is quite amazing. We ran our tests on a system that is about 16x less powerful than our current production system. And yet BLU was able to outperform our production system in every aspect. We were truly running our analytics on the DB2 10.5 column organized tables at the speed of thought."
–Mohankumar Saraswatipura, Lead DBA.Reckitt Benckiser

"With my analytic query workload running 45x times faster with BLU Acceleration in DB2 10.5, I no longer have an excuse for my usual coffee run!"
- Iqbal Goralwalla, Head of DB2 Managed Services, Triton

DB2 with BLU Acceleration Early Customer Quotes

*"**Using DB2 10.5 with BLU Acceleration, our storage consumption went down by about 10x** compared to our storage requirements for uncompressed tables and indexes. In fact, I was surprised to find a 3x increase in storage savings compared to the great compression that we already observed with Adaptive Compression on the DB2 10.5 server."*

- Kent Collins, Database Solutions Architect, BNSF Railway

*"One of the things I really like about BLU Acceleration is that it enables me to put **column-organized tables beside row-organized tables in the same database.** In our mixed environment, we realized an amazing **10-25x reduction in the storage** requirements for the database when taking into account the compression ratios, along with all the things **I no longer need to worry about: indexes, aggregates, and so on**."*

-Andrew Juarez, Lead SAP Basis and DBA

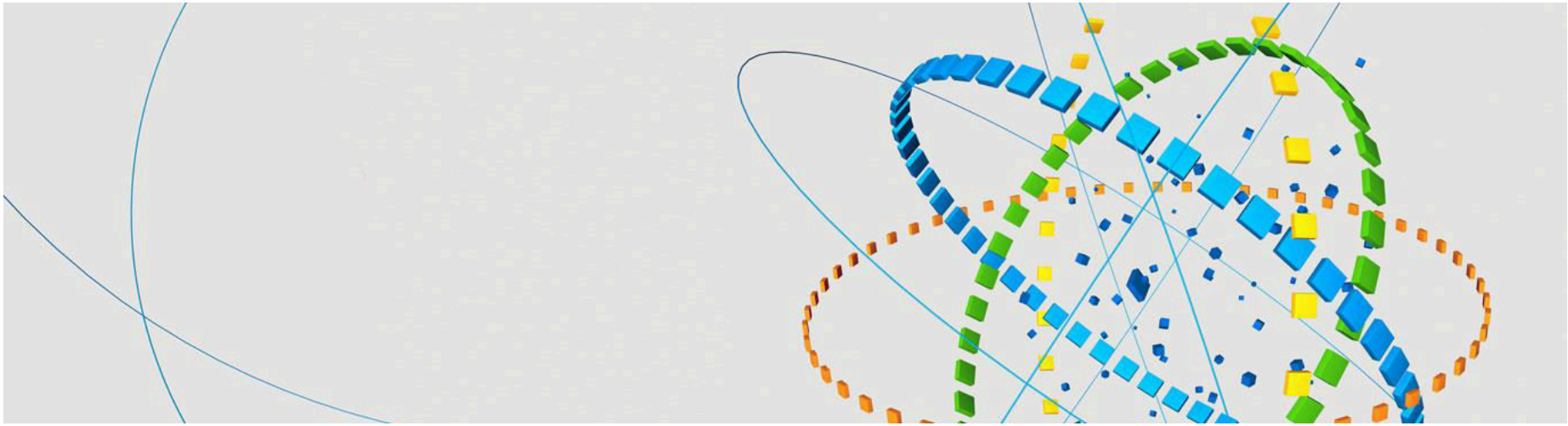
Thank You!

Agenda

- **09-15 - 09:30 - Key Note - Danilo Noveli**
- **09:30 -10:00 - How Big Data is Transforming Industry - Chris Grote**
- **10:00 – 10:30 – DB2 with Blu Acceleration – Les King**
- **10:30 – 11:00 - DB2 Blu for SAP – Ferdinand Prezenski**
- **11.:00 – 11:15 - Coffee break and Ask the Experts**
- **11:15 - 11:30 - Update on Purescale – Les King**
- **11:30 - 12:00 - PureData TX – Ferdinand Prezenski & Richard Hewitt**
- **12:00 - 12:30 – Informix - NoSQL for the Enterprise = NewSQL - Steve Shoaf**
- **12:30 - Lunch**

DB2 BLU for SAP

Ferdinand Prezenski– IBM Europe Director of Database



DB2 TCO benefit for SAP Customer

DB2 compared to any other SAP supported database

reduces annual OPEX costs

by > 20%

improves SAP performance

by > 30 %

TCO Reduction through DB2 for SAP - Savings Potential

Cost aspect	DB2 Benefit	Typical Savings Potential
SW-Cost	▪ Reduction of database-license and maintenance cost (compared to Oracle) through attractive DB2 prices ▪No cost for additional database management tools based on comprehensive DB2 product bundle	~ 25 - 40% (Maintenance 60%)
Storage	▪ Reduction of storage cost through DB2 compression -Smaller database size -Less I/Os -Smaller backup volume and faster backups	~ 40 -80%
Server	▪ Reduction of server cost through better performance and scalability -Efficient use of RAM due to compressed data in DB2 buffers	~10 -15% at database server
Operation / Administration	-Simplified administration, better patch/release-planning -Better 24x7 HA&DR solution easier achievement of SLAs	
TCO	▪**Sum of all DB2 benefits**	**~ 20-40%**

Extract of SAP customers who migrated to DB2

OPEX Cost Comparison: German Utility Company 5200 SAP user

	Oracle 10	IBM DB2 V10	ORACLE V11 (incl. Advanced Compression)
IT Service	**2,7 MIO EUR**	**2,6 MIO EUR**	**2,7 MIO EUR**
Storage	**1,4 MIO EUR**	**0,3 MIO EUR**	**1,0 MIO EUR**
Release upgrade DB(*)	**0,1 MIO EUR**		**0,1 MIO EUR**
Total per year	**4,2 MIO EUR**	**2,9 MIO EUR**	**3,8 MIO EUR**
Reduction per year		**30 %**	**9,5%**

(*) Oracle DB upgrade every 4 years, separate project, project cost apportioned by year.

- Operational Benefits

1st-year cost-avoidance Oracle licenses	**$250,000**
Database size reduction	**40%**
1st-year (additional) storage cost savings	**$100,000**
Annual license, storage, maintenance savings	**$175,000**
Database response time improvement	**5% to 10%**
ROI	**+205%**
ROI breakeven	**8 months**
5-year internal rate of return	**+133%**

SAP and DB2 – Strong Partnership

- **Deep Exploitation over a Decade of Joint Development**
  - All new database technologies will require time until they reach the same level of integration and maturity that DB2 has with SAP
  - DB2 has a history track of success with SAP
  - Deep Exploitation with DB2 and SAP
- **Partnership Continues**
  - SAP certified PureData System in November 2012
  - SAP DB2 10.5 certified in August 2013
  - SAP BLU Certification is on track for BW – target date this week
  - Joint development roadmap in place for 2015+
- **SAP Certification of BLU and BW**
  - Staged Delivery across BW capabilities
  - Stage 1 - Standard InfoCubes, Non-Cumulative InfoCubes, DB2 Near-Line Storage
  - Stage 2 – DSOs, Master Data, Flat InfoCubes, Transactional InfoCubes, InfoSets, Persistent Staging Area (PSA)

DB2-SAP: Strategic Technology Alignment and Support

Database version	Database GA	SAP DB GA	Delay between database and SAP GA (in months)	SAP DB support until
DB2 8.2	29th April 2005	3rd June 2005	1	31.12.2015*
DB2 9.1	28th July 2006	31st August 2006	1	31.12.2017*
DB2 9.5	31st October 2008	20th Dec 2008	2	31.12.2017*
DB2 9.7	26th June 2009	28th August 2009	2	31.12.2022*
DB2 10	***4th April 2012***	***July 2012***	***3***	***31.12.2022****
Oracle 9i	June 2001	Q1/2003	21	31.07.2008
Oracle 10g	January 2004	Q3/2006	31	31.07.2011
Oracle 11g	July 2007	Q2/2010	33	January 2015

* DB2 follows SAP's maintenance strategy 7+2,
Status: 2nd November 2011
Source: SAP Marketplace, SAP hint 1168456, SAP hint 1174136

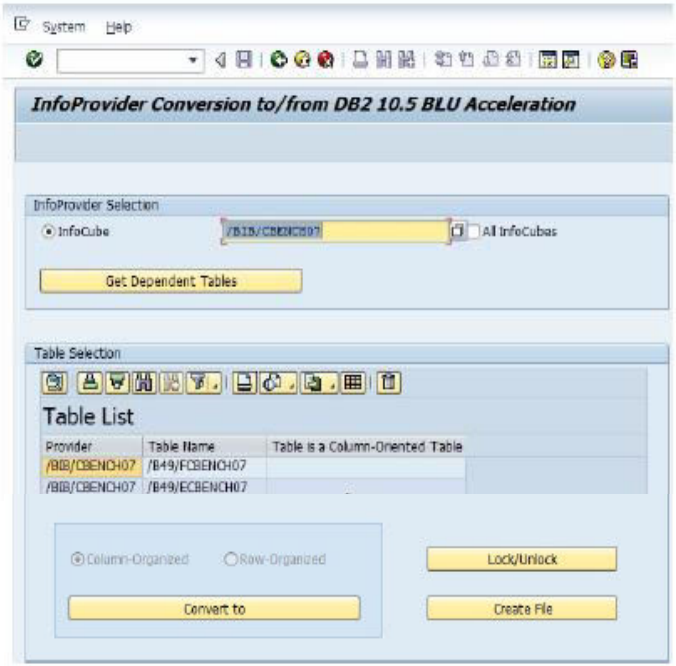
Joint DB2 & SAP development: Early tests during the implementation in Toronto + Walldorf

- DB2 between 1-2 months delay supported by SAP
- DB2 supports existing database version for a long time (DB2 8.2 support ends after Oracle 11.2g de-support)

Relaxed long-term project planning combined with usage of most current DB2 technology

SAP BW: Information about BLU

- Support for SAP BW and NLS
  - SAP BW 7.00 and higher (expected Sep. 2013)
  - DB2 10.5 FP1 and higher
  - DB2 10.5 BLU extensions will be delivered with SAP BW support packages
  - First wave: standard & non-cumulative info-cube
  - Second wave: DSO and Master Data
- DB6CONV – SAP ABAP tool for Online/Offline table move
  - report SAP_CDE_CONVERSION_DB6
  - Calls DB2 Admin Move Table (AMT) to move
    - Non-BLU -> BLU online
    - BLU -> BLU / Non-BLU in read-only mode
- db6_update_db script
  - Enables WLM concurrency threshold
- SAP ABAP Dictionary extension to support BLU tables as new table type
- DBA Cockpit
  - Support new performance metrics for BLU tables



DB2 Optimized for SAP - Roadmap

2005

NW 2004
- Streamlined install
- Streamlined admin

Version 8.2.2
- Automatic storage
- Deployment optimized for SAP

2006

NW 7.0
- Embedded database
- Reduced storage costs
- Self tuning
- Minimal admin
- Enhanced DBA Cockpit

2007

Version 9.1
- Storage limits removed
- Autonomic / TCO features
- Compression

2008

NW 7.0 SR3
- Turn-key HA solution
- Turn-key compression
- Integrated MDC advisor
- Deferred Table Creation

Version 9.5
- Integrated & automatic HA+DR
- Integrated Flash Copy
- Threaded Architecture
- DPF Scaling Improvements

2009

NW 7.x
- Integrated Nearline Storage
- Top-Down, Revamped Monitoring
- Next generation DBA Cockpit
- Improved End-To-End Support
- Database Performance Warehouse
- Integrated Workload Management
- Initial integration of DB2 pureScale
- MDC Advisor Stage 2

Version 9.7
- Deeper Deep Compression
  - 40+% reduction in Indexes
  - 40+% reduction in Temps
  - Compressed LOBs
- Near-0 Storage Admin
  - Sparse MDC tables for simple space reclaim
  - Easy table space reclamation
  - Ease the path to Automatic Storage
- Extending Online Operations
  - Change schema definitions online
  - Reorganization improvements
- Full 360° Monitoring

2010

NW 7.x
- Optimized DSO activation for DB2 DPF
- Enhanced near-line storage
- Separation of duties
- ERP partitioning engine
- Integrated HA for SAP
- Graceful maintenance

DB2 9.8 pureScale
- Continuous availability
- Seamless OS and hardware maintenance
- OLTP scale out

2011

2012

DB2 10.1
- 2 X warehouse performance improvement
- Adaptive compression
- Extended multi core support
- No touch space reclamation
- Geographically dispersed pureScale cluster
- Multiple temperature storage
- Multiple standby with time delay
- Extended transparent fail over

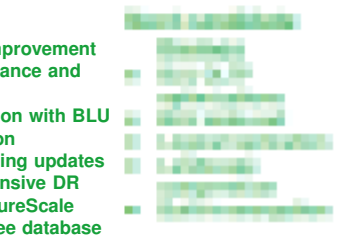
2013

NW 7.x
- Multi-temperature management
- Deep integration of DB2 pureScale
- Seamless space reclamation
- Identity Management with DB2
- Cloud readiness

2014

DB2 10.5
- Factors improvement in performance and columnar compression with BLU acceleration
- Online rolling updates
- Comprehensive DR solution pureScale
- REORG-free database

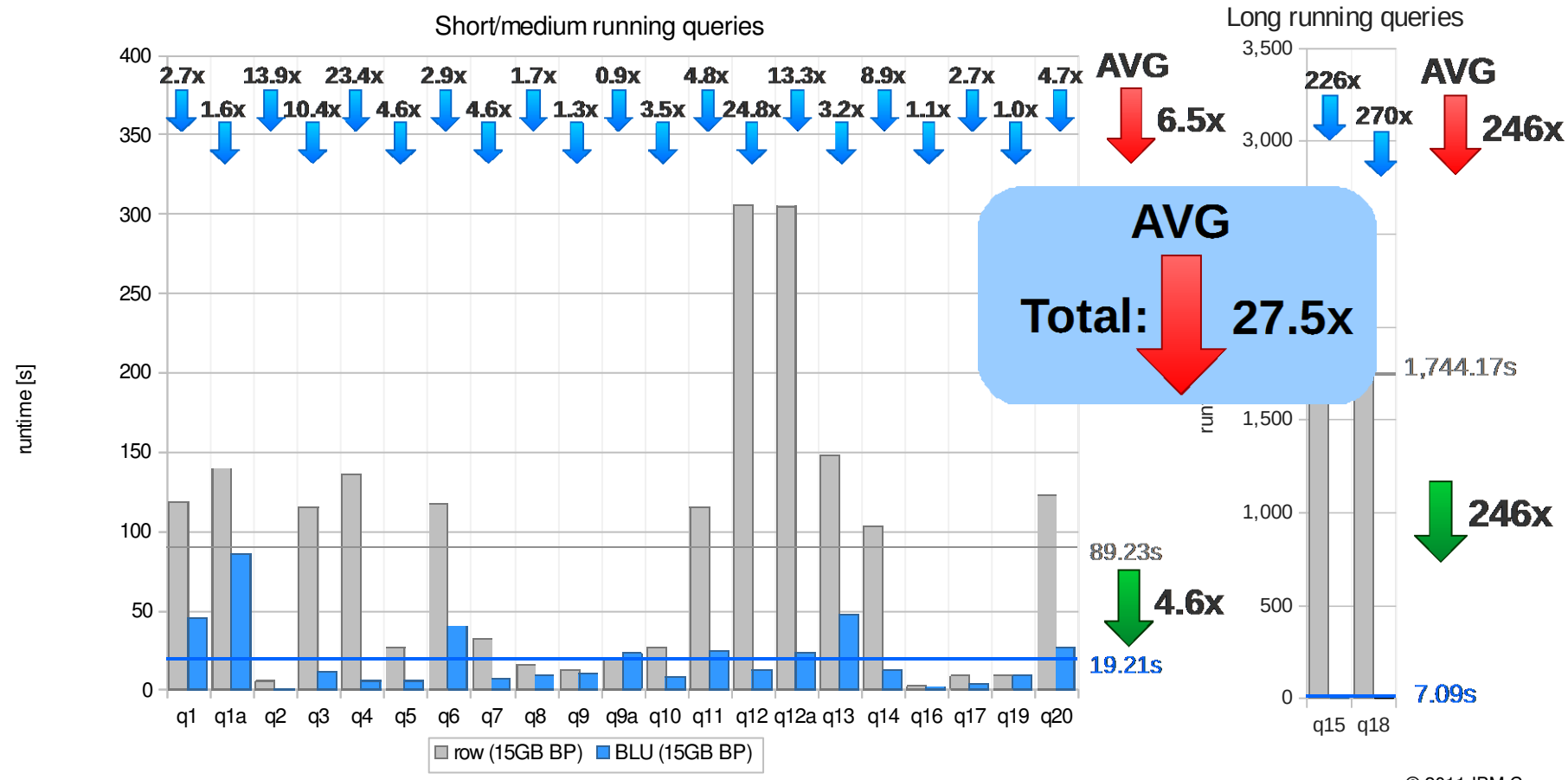
2015



BW Support on DB2 10.5 BLU – BW Query Performance (1)

BW InfoCube scenario
- Fact table with 438 Million records
- Size (table + indexes)
  - Row based (adaptive compr.): 33 GByte
  - **BLU: 12 GByte**
- DB2 parallel query degree switched ON

Test environment
- 2 Intel® Xeon® processors (8 cores in total), 2.14GHz
- 32 GB RAM
- SUSE Linux Enterprise Server V11
- SAP BW 7.30
- DB2 10.5 pre GA, **15 GB Bufferpool**



SAP BW Support on DB2 10.5 BLU – Compression

- **BW InfoCube E-fact table with 438 Million rows**
  - **Row table: 7 secondary single integer indexes, 1 compound unique index with 7 integers**
  - **BLU table: 1 compound unique index with 7 integers**

Lab Tests

F- fact table	Table size	Index Size	Total
Row, 10.5 Without Compr.	80,4 GB	42,9 GB	123,3 GB
Row, 10.5 Adapt.Compr.	26,7 GB	15.0 GB	41,7 GB
BLU, 10.5	12,3 GB	7,8 GB	20,1 GB

6,1x smaller than uncompressed
2,1x smaller than adaptive compressed

BW Support on DB2 10.5 BLU – BW Query Performance

➔ By factors faster SAP BW queries

- Fast query run time on InfoCubes without BW aggregates
- Fast query run time on DSOs

➔ No time consuming SQL query tuning

- Fast “Out of the box” performance

➔ Better prediction of BW query run time

IBM Database Offering for SAP Applications

SAP's offering:
-BW ➔ Hana
-BS ➔ HANA/Sybase ASE
- NLS ➔ Sybase IQ

Average diaglog response time 0,2 - 0,8 sec

DB2 LUW

SAP Business-Suite, Industry Solutions

OLTP workload

Average diaglog response time 0,4 – 2+ sec

DB2 LUW

SAP BW
Big Data &
op. Analytics

OLAP workload

DB2 LUW

NLS for BW
NLS for HANA

Transactional

Analytical

Near-line Storage

DB2 10.5 / BLU – Data Center Excellence
including Virtualization and Consolidation

Criteria compared with uncompressed source system	**DB2 10.5 / BLU**	**SAP HANA**
OS support	AIX, Linux, *Win*	Suse Linux
SAP release	7.0 and higher	7.30 and higher
Virtualisation (incl. production)	Yes	No
2-Tier support	Yes	No
3-Tier support	Yes	Yes
NLS (nearline storage) support as underlying database	Yes	No
Number of patches per year ****	~2	60 (SPS06)
Non-disruptive HW / IT support	Yes, utilize existing HW & IT concept	No, new HW & IT concept
Migration to and from possible with available tools	Yes	No
Percentage of source storage *	~10 to 30%	~150 to 300%
Percentage of source RAM **	~100 to 200%	~1000+%
License	OEM: based on SAV*** (8%) Direct: #users, #cores	BS/BW: 15% / 8% of SMVB RAM size based

* DB2 compression reduces storage capacity by 70-95%, depending on share of column-store objects
** DB2 9.7/10.1 requires usually 2-5% of the database size, with DB2 10.5 approx 4-10% expected
*** SAV-SAP Application Value, SMVB-SAP Maintenance Value Base
**** based on experiences with DB2 V8, V9 and DB2 10.1

Status: July 2013

SAP‘s statement about Consolidation & Virtualization

- SAP note 1681092 - Multiple SAP HANA databases on one appliance
  - SAP **does not support** running multiple SAP HANA databases (SIDS) on a single **production** SAP HANA appliance
  - SAP **does support** running multiple SAP HANA databases on a single **non-production** (DEV, QA, test)
  - running multiple DBs on one SAP HANA appliance may impact performance of various types of operations, as contention for memory resources may occur
  - SAP support will address the performance issue only if it can be shown to exist when only one DB is running on the SAP HANA appliance (...you may stop all but one of the DBs and see if the issue persists)

- SAP note 1788665 - SAP HANA running on VMware vSphere VMs
  - For **non-production** SAP HANA instances use only
  - Multiple virtual machines can be deployed on a single SAP HANA appliance. Each SAP HANA instance deployed on a virtual machine is recommended to be sized the same as SAP HANA deployed on "bare metal" SAP HANA appliances

Status: 14th July 2013

Green IT: DB2 on POWER vs HANA on Intel

Customer runs DB2 on POWER

- 180 systems, 48 production
- 26 HA (LPM*) + 26 DR (PowerHA)
- 2 data centers

➔ 4 POWER servers

Possible HANA implementation **

- 180 systems, **48 production**
- **26 HA + 26 DR clusters**
- 2 **BIGGER** or more data centers

- 48 HANA servers for production
- 52 HANA servers for HA+DR clusters
- up to 48 HANA servers for test/QA
- up to 48 HANA servers for dev
- up to 36 HANA servers for rest

➔ 101-232 HANA servers

4 POWER servers versus **101+ HANA servers !!!**

* **LPM - AIX live partition mobility**

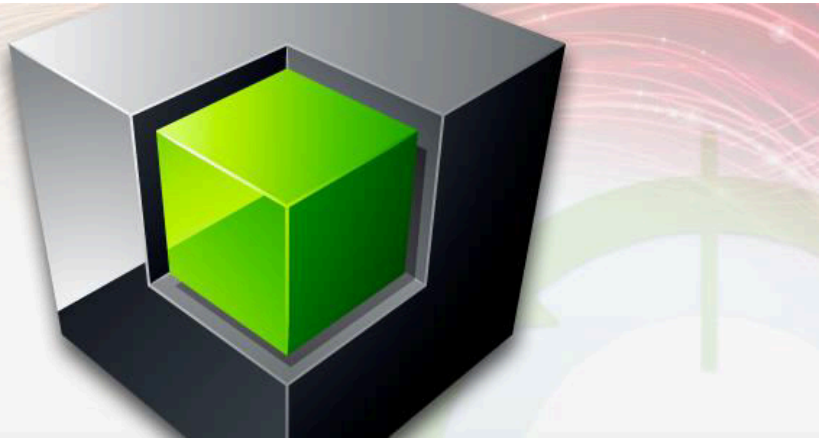
** **SAP note 1788665, 1681092**: No virtualization and no multiple SAP HANA databases on a production SAP HANA appliance

Agenda

- **09-15 - 09:30 - Key Note - Danilo Noveli**
- **09:30 -10:00 - How Big Data is Transforming Industry - Chris Grote**
- **10:00 – 10:30 – DB2 with Blu Acceleration – Les King**
- **10:30 – 11:00 - DB2 Blu for SAP – Ferdinand Prezenski**
- **11.:00 – 11:15 - Coffee break and Ask the Experts**
- **11:15 - 11:30 - Update on Purescale – Les King**
- **11:30 - 12:00 - PureData TX – Ferdinand Prezenski & Richard Hewitt**
- **12:00 - 12:30 – Informix - NoSQL for the Enterprise = NewSQL - Steve Shoaf**
- **12:30 - Lunch**

DB2® 10.5

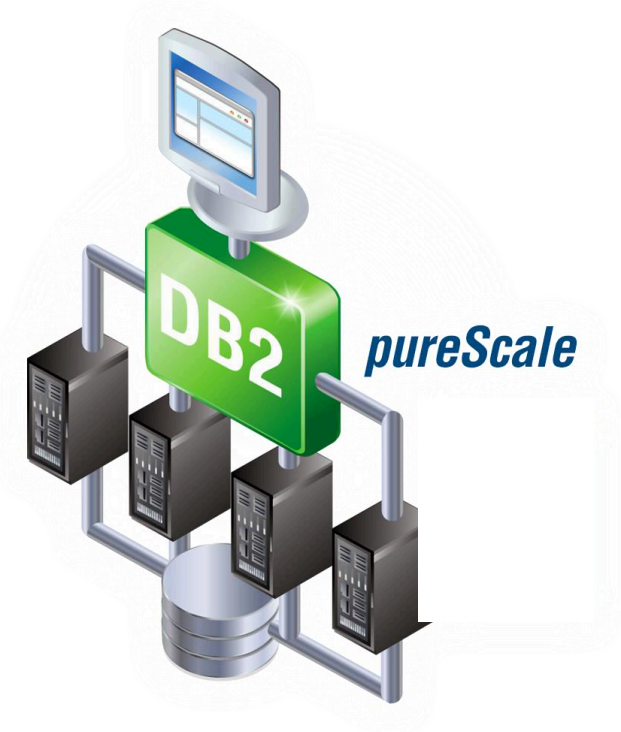
with BLU Acceleration



DB2 10.5 Delivers 'Always Available' Transactions
99.999% Up Time, Optimized for OLTP Workloads

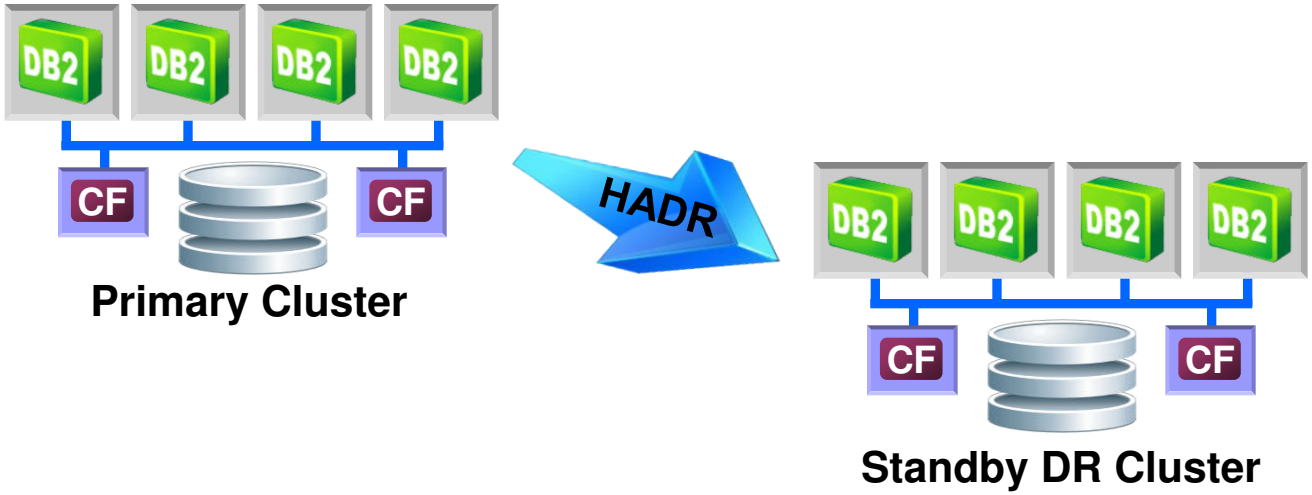
- **DB2 pureScale**
  - Clustered, shared-disk architecture
  - Provides improved availability, performance, and scalability
  - Complete application transparency
  - Scales to >100 members
  - Leverages z/OS cluster technology

- **New DB2 10.5 pureScale enhancements**
  - Rich disaster recovery capabilities with HADR
  - Rolling fix pack updates
  - Online add member
  - Online table REORG
  - Support for any x86 server platform



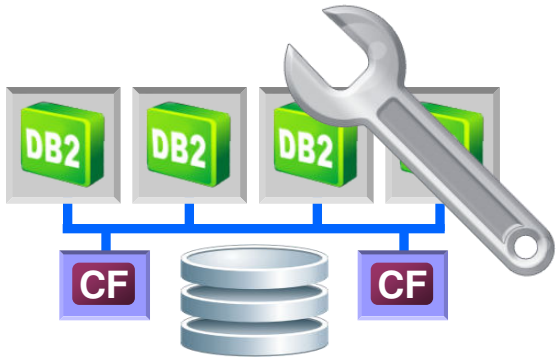
HADR in DB2 pureScale

- **Integrated disaster recovery solution**
  - Simple to setup, configure, and manage
- **Support includes**
  - `ASYNC` and `SUPERASYNC` modes
    - `SYNC/NEARSYNC` under development
  - Time delayed apply
  - Log spooling
  - Both non-forced (role switch) and forced (failover) takeovers



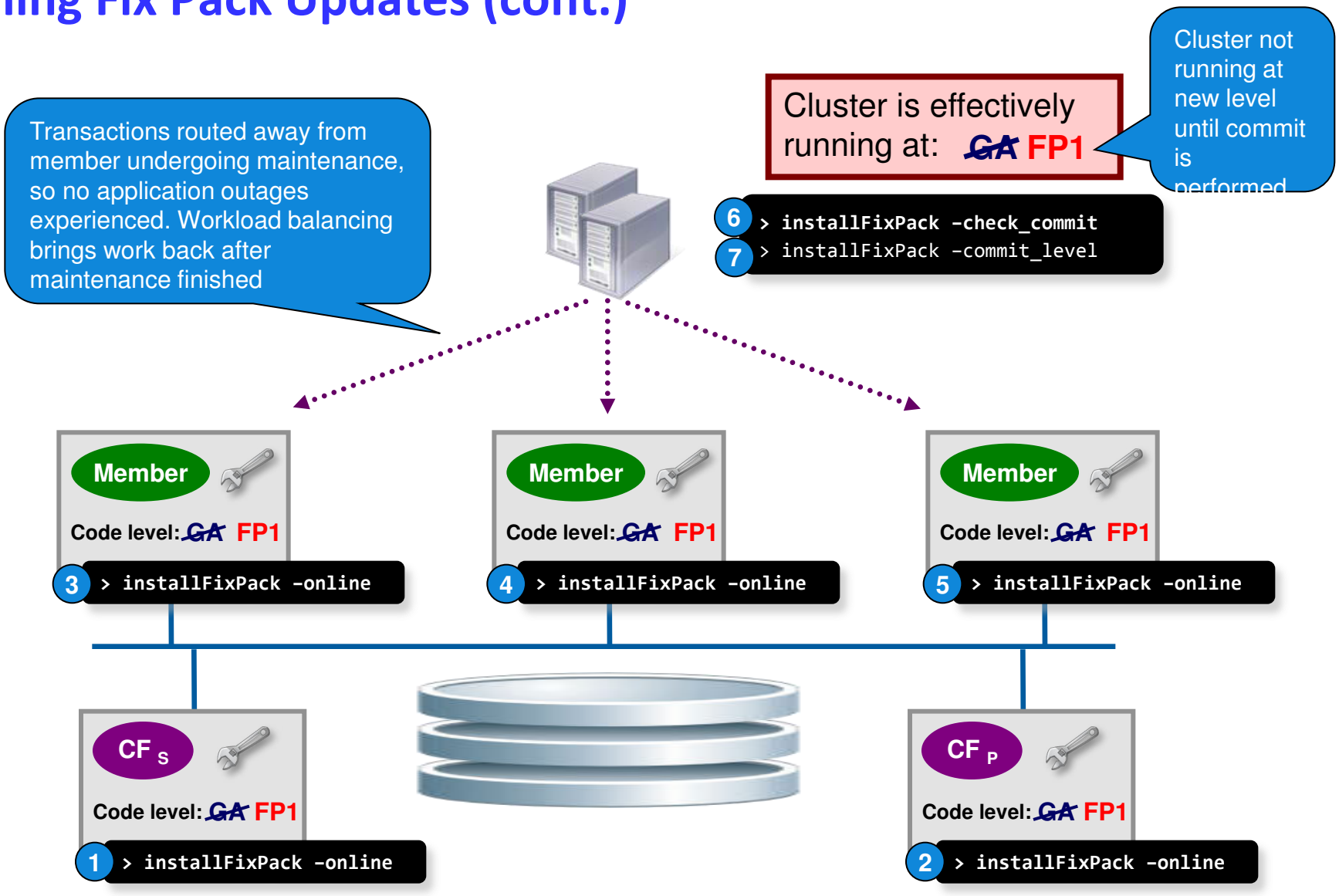
Rolling Database Fix Pack Updates

- **Transparently install pureScale fix packs in an online rolling fashion**
- **No outage experienced by applications**
- **Single `installFixPack` command run on each member/CF**
  - Quiesces member
    - Existing transactions allowed to finish (configurable timeout, default is 2 minutes)
    - New transactions sent to other members
  - Installs binaries
  - Updates instance
    - Member still behaves as if running on previous fix pack level
  - Unquiesces member



- **Final `installFixPack` command to complete and commit updates**
  - Instance now running at new fix pack level

Rolling Fix Pack Updates (cont.)



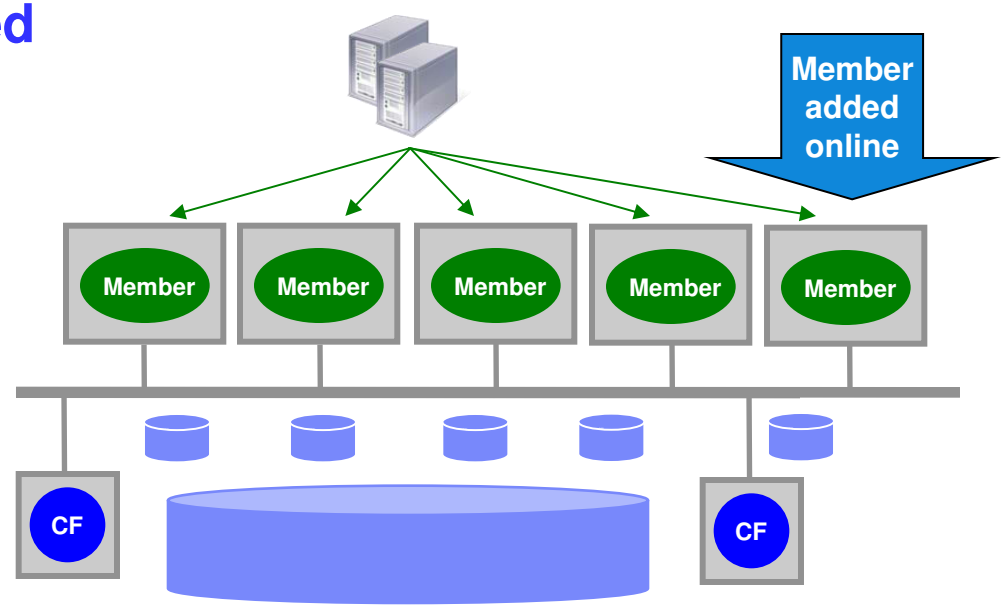
Online Add Member

- **New members can be added to an instance while it is online**
  - No impact to workloads running on existing members
  - Previously, required an outage of the entire instance to add a new member

- **No change in add member command**

```
db2iupdt -add -m <newHost> -mnet <networkName> <instance>
```

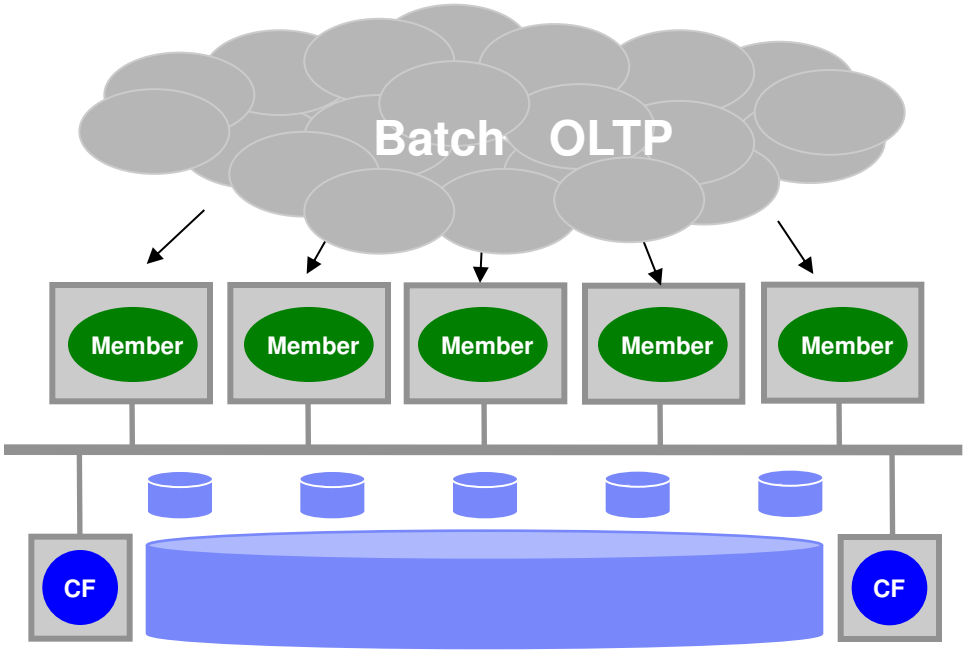
- **Offline backup no longer needed after adding new members**



Multi-Tenancy: Member Subsets

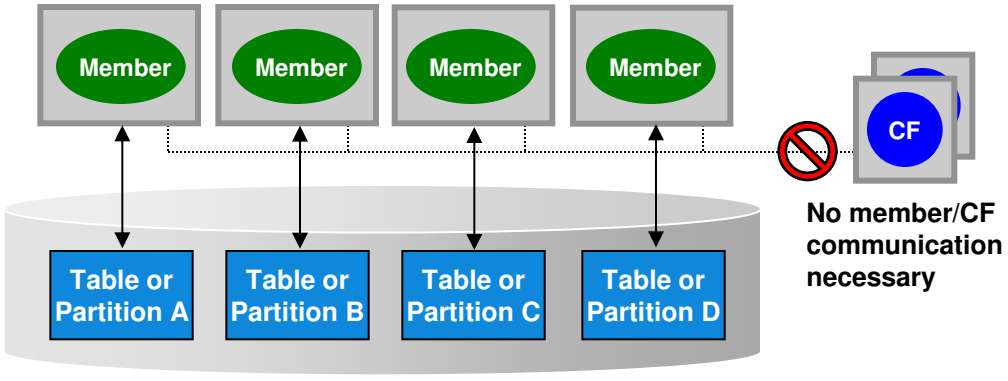
- **Previously, an application/tenant could only be configured to run**
  1. On one member (client affinity) or
  2. Across all members in cluster (workload balancing)

- **Can now point applications to subsets of members which enables**
  - Isolation of batch from transactional workloads
  - Multiple databases in a single instance to be isolated from each other



Multi-Tenancy: Explicit Hierarchical Locking (EHL)

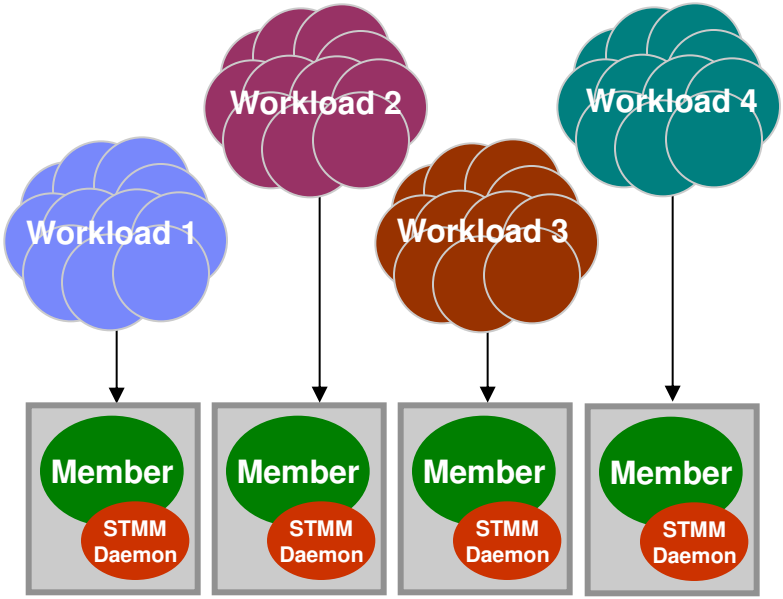
- **Designed to remove data sharing costs for tables/partitions that are only accessed by a single member**
  - Avoids CF communication if object sharing not occurring
- **Target scenarios**
  - Workload affinitization
  - Workload consolidation and application affinitization
- **Enabled via new `OPT_DIRECT_WRKLD` database configuration parameter**
  - Detection of data access patterns happens automatically and EHL will kick in when data is not being shared after configuration parameter set



Multi-Tenancy: Self-Tuning Memory Management (STMM)

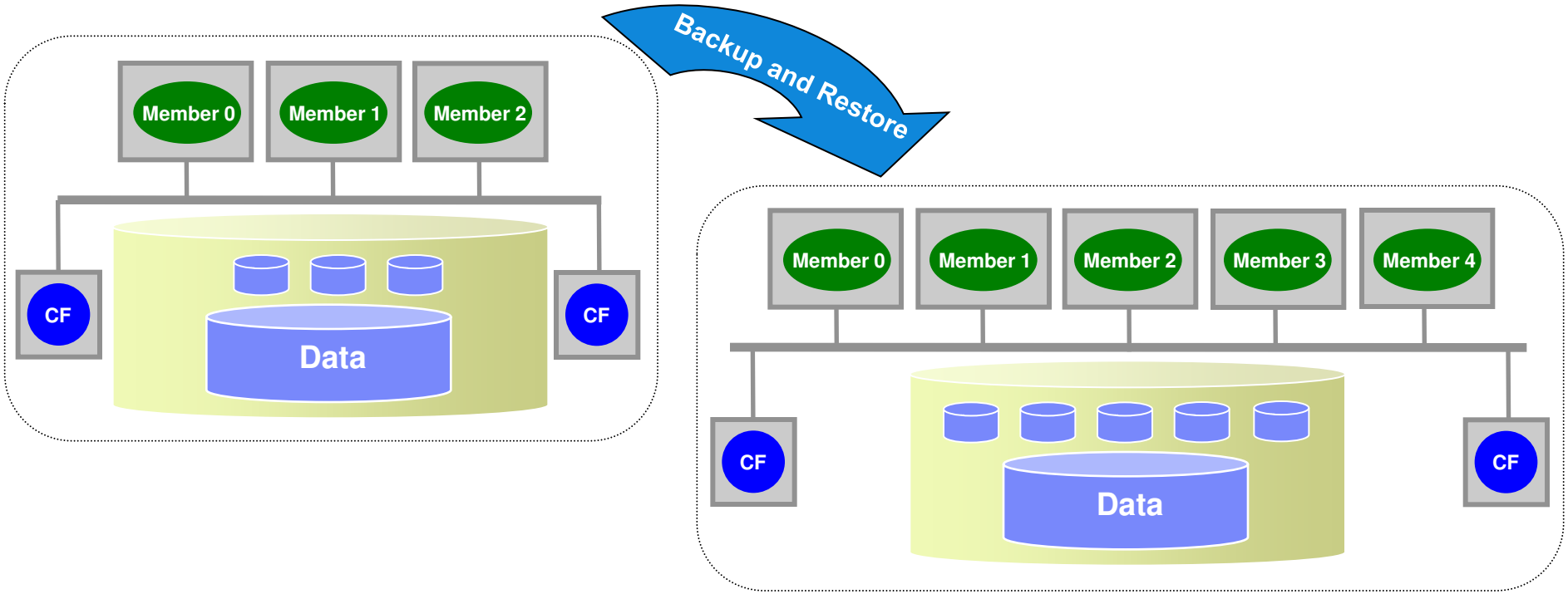
- **Prior DB2 pureScale STMM design**
  - Single tuning member makes local tuning decisions based on workload running on that member
    - Other member becomes tuning member in case of member failure
  - Broadcasts tuning decisions to other members
  - Works well in single homogeneous workload scenarios

- **DB2 pureScale now allows per-member STMM tuning**
  - Workload consolidation
  - Multi-tenancy
  - Batch workloads
  - Affinitized workloads



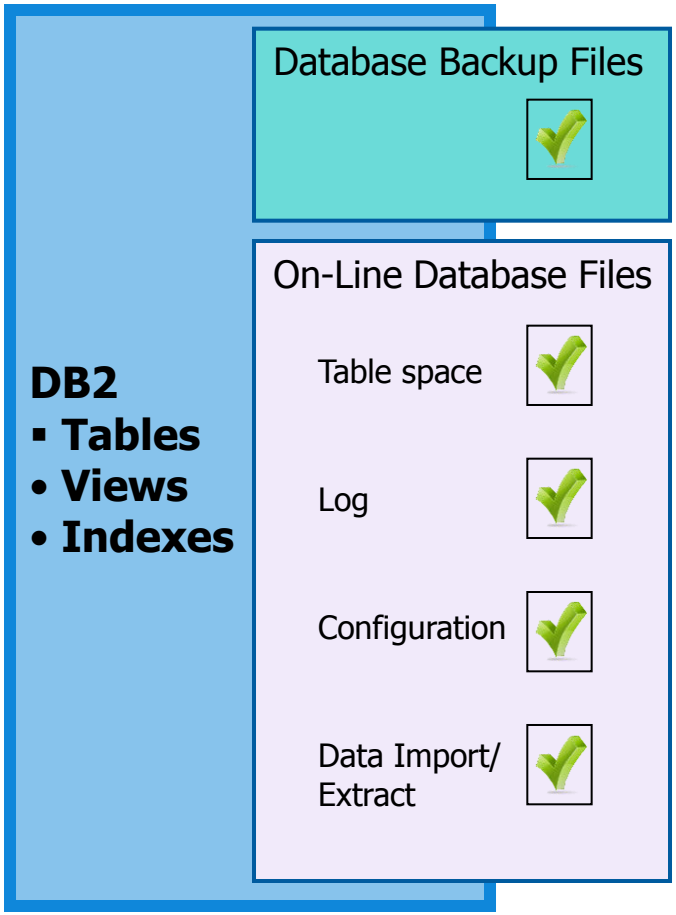
Topology-Changing Backup and Restore

- Backup and restore between topologies with differing numbers of members
- Backup and restore from DB2 pureScale to non-DB2 pureScale (and vice-versa)



DB2 pureScale Database Encryption

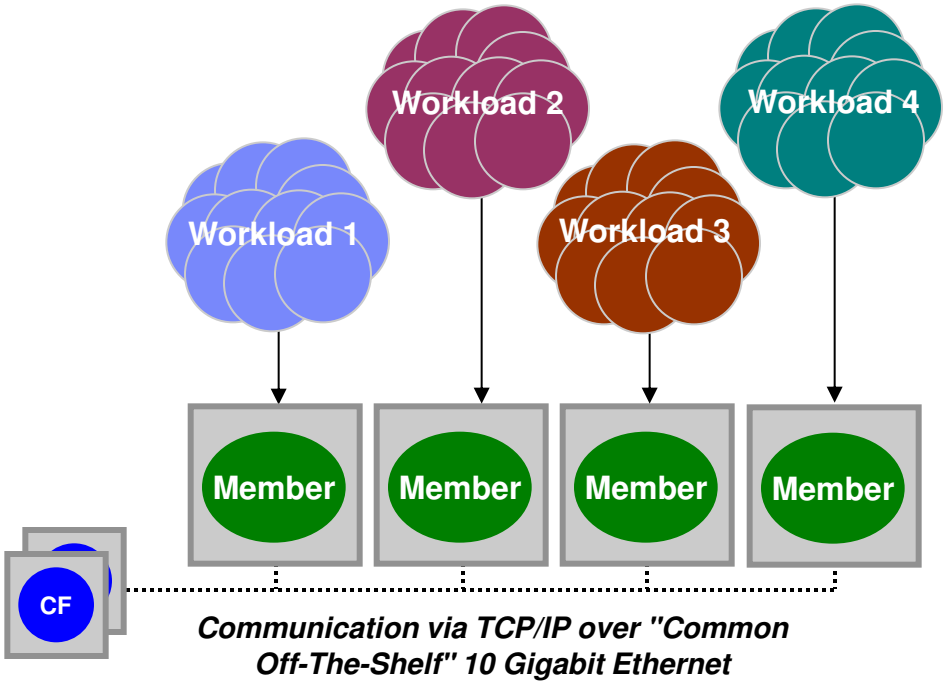
- **Encrypt DB2 backups**
  - Audit and prevent unauthorized restores
- **Selectively encrypt DB2 files**
  - Control decryption by user, process
  - Audit unauthorized access attempts
  - Control privileged OS users
- **Automatic key management**
  - Transparent to existing applications



Simplifying Development and Test: TCP/IP Sockets

- **Support TCP/IP sockets protocol**
  - Utilizes COTS (Common Off-The-Shelf) Ethernet
    - 10 Gigabit Ethernet required
  - Simpler to implement with reduced Hardware requirements

- **Target scenarios**
  - Workload consolidation
  - High read/write ratio workloads
    - For example, retail workloads
  - Affinitized workloads
  - Virtualized environments (cloud)
    - VMware, KVM
  - Test/development



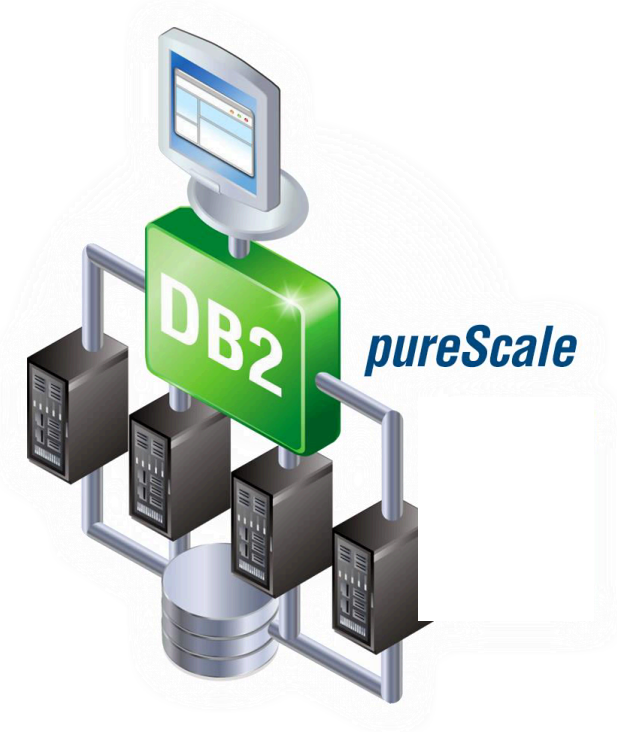
DB2 10.5 – Achieving 99.999% Availability with pureScale

- **High Availability & Workload Balancing built into architecture**

- **Avoiding Planned Outages**
  - Online OS and Hardware upgrades
  - Rolling DB2 fix pack updates
  - Online add member
  - Online table REORG

- **Avoiding Unplanned Outages**
  - Online recovery
  - Some disaster recovery capabilities with GDPC
  - Rich disaster recovery capabilities with HADR
  - Support for any x86 server platform

- **Other Capabilities**
  - On-disk encryption
  - Flexible server support and pricing models
  - Multi-tenancy capabilities
  - WLM capabilities



Thank You!

Agenda

- **09-15 - 09:30 - Key Note - Danilo Noveli**
- **09:30 -10:00 - How Big Data is Transforming Industry - Chris Grote**
- **10:00 – 10:30 – DB2 with Blu Acceleration – Les King**
- **10:30 – 11:00 - DB2 Blu for SAP – Ferdinand Prezenski**
- **11.:00 – 11:15 - Coffee break and Ask the Experts**
- **11:15 - 11:30 - Update on Purescale – Les King**
- **11:30 - 12:00 - PureData TX – Ferdinand Prezenski & Richard Hewitt**
- **12:00 - 12:30 - Informix NoSQL for the Enterprise = NewSQL - Steve Shoaf**
- **12:30 - Lunch**

Deliver transactional data services with speed, simplicity and lower cost

IBM PureData System for Transactions

PureData

Data Platform

- *Workload optimized performance*
- *Data load ready in hours*
- *Integrated management*
- *Automated maintenance*
- *Single point of support*

PureSystems

Delivering Big Data Platform Services

IBM PureData System

Meeting Big Data Challenges – Fast and Easy!

PureData
System for Transactions

For apps like E-commerce…

Database cluster services optimized for transactional throughput and scalability

---

PureData
System for Analytics

Powered by Netezza technology

For apps like Customer Analysis…

Data warehouse services optimized for high-speed, peta-scale analytics and simplicity

---

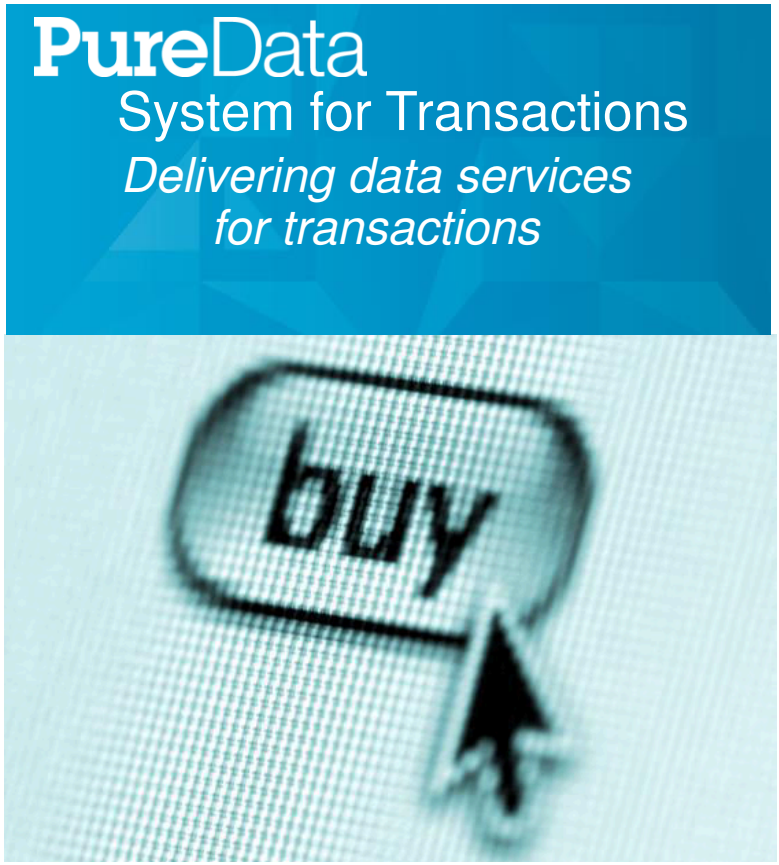
PureData
System for Operational Analytics

For apps like Real-time Fraud Detection…

Operational data warehouse services optimized to balance high performance analytics and real-time operational throughput

IBM PureData System for Transactions highlights

Optimized exclusively for transactional data workloads



Speed

- Industry leading DB2 performance
- Database node recovery in seconds[1]

Simplicity

- Database deployment in minutes, not hours[1]
- Capable of running multiple database software versions
- Handles more than 100 databases on 1 system[2]
- No planned system downtime for firmware / OS upgrades[1]

Scalability

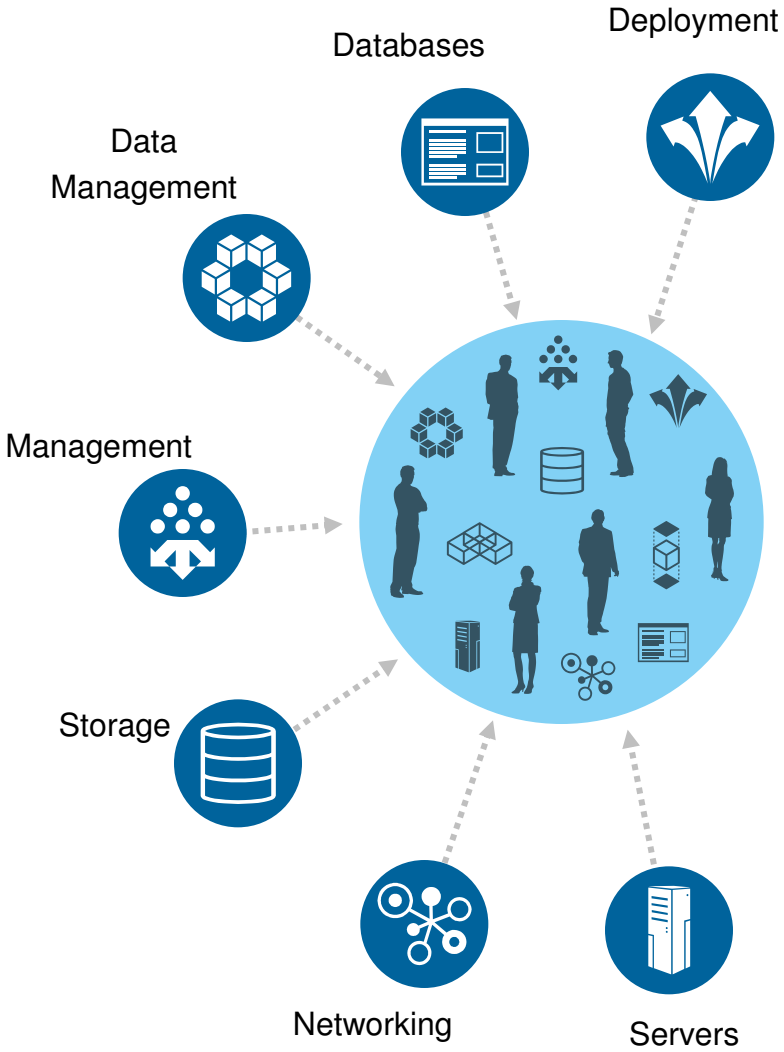
- Scaling up to 30x[3]
- Designed to expand from small to medium & medium to large configuration with no planned system downtime required

Smart

- Supports Oracle Database apps with minimal change; supports DB2 applications unchanged
- Clients have experienced cases of 10x storage space savings via Adaptive Compression[4]

Footnotes:
1. Based on IBM internal tests and system design for normal operation under expected typical workload. Individual results may vary.
2. Based on one large configuration.
3. Based on the designed minimum and maximum processor and memory resources required for a single database.
4. Based on client testing in the DB2 10 Early Access Program.

Reduce your system integration costs



Optimized solution stack

- Factory integrated and optimized
  - *Server, storage, networking and software*
- High data availability
  - *Automatic failure detection and online recovery*
- Solid State exploitation
  - *Automatic management of hot, warm and cold data for faster performance*
- Optimized database patterns
  - *Database patterns pre-tuned and pre-configured for performance*
- Integrated fixes
  - *Zero down time for system maintenance*

IBM PureData System for Transactions helps reduce downtime costs

Downtime costs have grown 38% since 2010

Yearly Cost Metric	Best-In-Class (Top 20%)	Industry Average (Middle 50%)	Laggards (Bottom 30%)
# of business interruption events	.3	2.3	4.4
Length of event (hours)	.1	1.0	9.0
Total disruption (hours)	.03	2.3	39.6
Average cost per hour of disruption	$101,600	$181,770	$99,150
Total cost of business interruptions	$ 3,048	$418,071	$3,926,340

'Datacenter Downtime – How much does it really cost?', Aberdeen Group, March 2012, 134 organizations

Simplified deployment with high availability

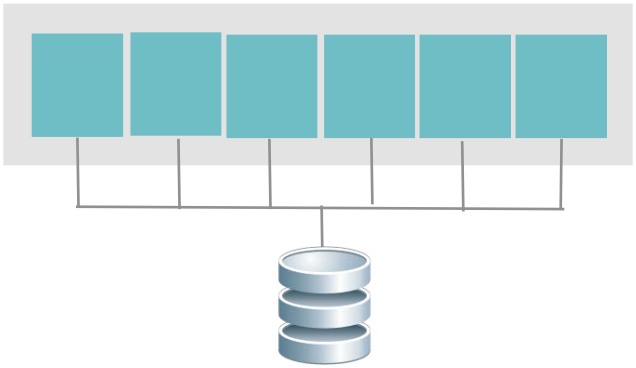
Uninterrupted access to data with consistent performance

Traditional systems - **build it yourself**

Over several days/weeks:

1. Define High Availability topology
2. Configure HW/SW/Network
3. Set up storage pools
4. Install multiple operating systems
5. Install database instances
6. Set up primary and secondary management systems
7. Set up database members
8. Set up backup processes
9. Test, tune, reconfigure...

6-node
database cluster instance



PureData System for Transactions - **built-in**

In minutes,

1. Just specify cluster name, description and topology pattern

Simplified deployment of multiple databases

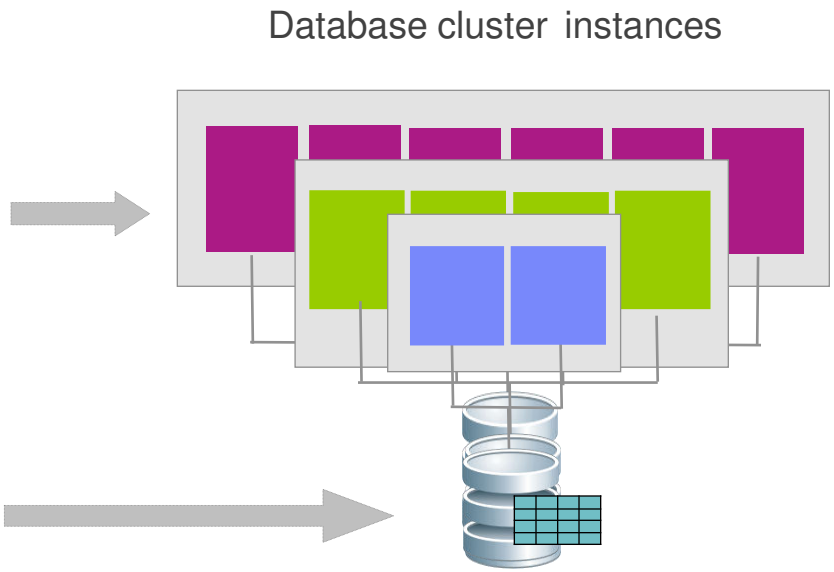
Deploy topology and databases in minutes using patterns

Topology patterns

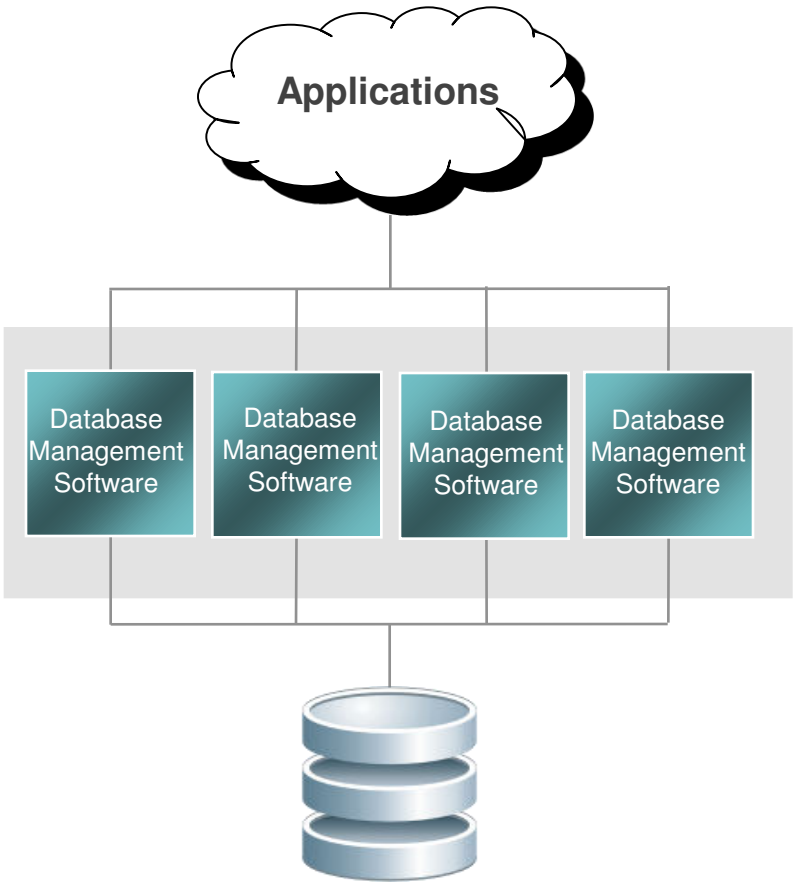
Automatically creates, configures and deploys a DB2 pureScale™ database system with built-in redundancy and high performance

Database patterns

Automatically creates, configure and deploys IBM or client-specified databases optimized for transactional workloads



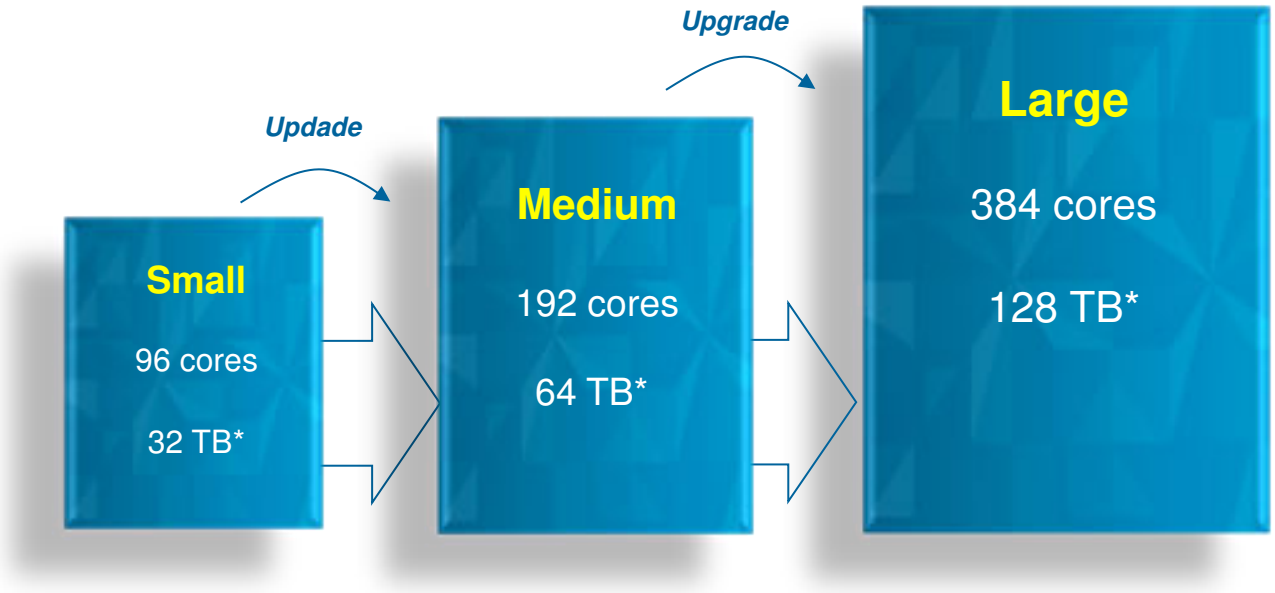
Simplified application development



- **Higher scalability** provided by adding more nodes with no application changes required
- **Dev/Test/Staging/Production database** repeatability through database patterns
- **Integrated data movement tools** to speed creation of test databases
- **Built-in Oracle compatibility** mode for minimal to no application changes
- **Higher utilization and lower costs** provided through shared resource management

Simplified capacity upgrades

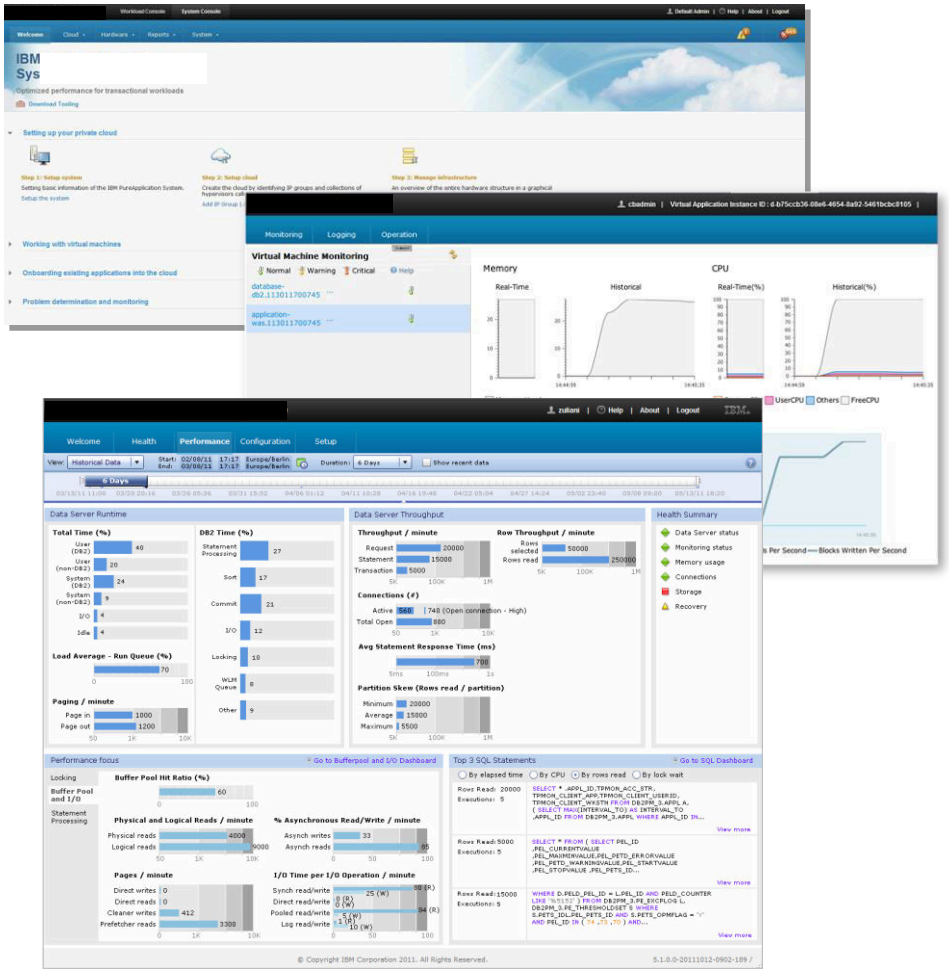
- Each configuration arrives with
  - High capacity, high scale computing
  - Integrated solid data and disk storage
  - High speed networking
  - Scalable database management
  - Integrated systems management
- Easy, non-disruptive upgrade to larger systems



Simplified database administration

Self-balancing	Data access requests automatically load balanced for optimal performance
Self-tuning	Memory management dynamically balances resources
Self-optimizing	Best data placement and access automatically selected based on usage statistics for optimal performance
Self-monitoring	Based on thresholds and alerts, system will monitor and automatically make changes as needed to improve performance
Self-healing	Failed database nodes are isolated and recovered automatically

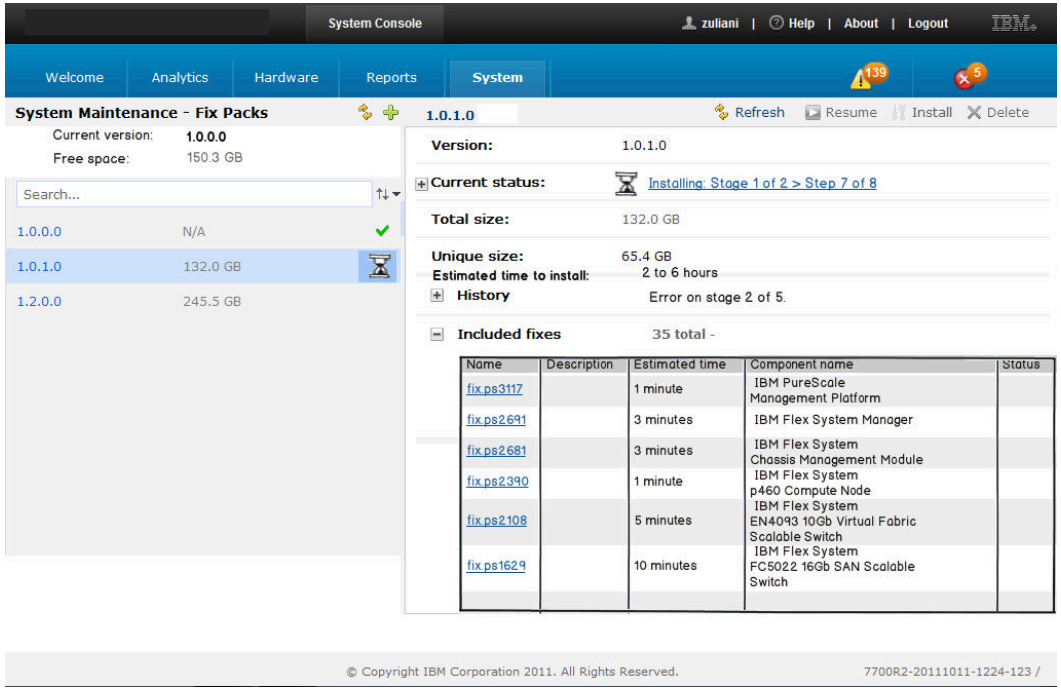
Simplified and integrated system management



- Single console to manage all resources and work running on the system
- Role-based security and tasks
  - management
  - monitoring
  - maintenance
- Easy integration with broader enterprise monitoring tools and processes
- Consistent IBM PureSystems console

Simplified maintenance with pre-integrated fixes

Reduce risk and eliminate manual errors when applying maintenance



- All hardware firmware and OS software patches integrated and tested together at the factory
- Can apply hardware and OS maintenance with zero downtime
- Single line of support
  - Integrated stack support

Under the covers

External Network Connectivity

- Dual 10Gb Ethernet Switches

Balanced HDD & SSD storage

- 48-disk module (12 SSD + 36 HDD) to optimize for performance & cost.
- RAID 10 for extreme storage reliability
- Up to 9.5 TB SSD and 64 TB HDD

Storage System

- Storewize v7000 chassis and expansion units.
- Holds up to 192 disks
- 1:3 balanced SSD to HDD ratio

Flex System™ Compute Node

- Up to 24 per rack

Flex System Chassis

- Up to 2 per rack
- Up to 14 compute nodes each
- Fully redundant Network (En) and Storage (SAN) connectivity

PureSystem™ Manager

- 2 per rack for redundancy
- Integrate management for all system resources

IBM GTS Assessment of PureData Potential Benefits over 5 years

Comparing **PureData for Transactions** vs BAU* for deploying & managing 90 transactional databases with Active-Active HA

21% lower Technology costs

server, storage, software

40-55% lower Labor costs

database, server, storage, network, infrastructure, user account management

74% lower Energy & Cooling costs

*Contact IBM for a complimentary assessment of business value and impact on **your** environment*

IBM Global Technology Services (GTS) is amongst the largest IT strategic outsourcing providers. GTS uses non-IBM and IBM technology. This GTS assessment compares a Build As Usual (BAU) environment for 90 Active-Active High Availability DB2 databases to using PureData for Transactions Large for the same number and profile of databases over a 5 year period. Comparisons are based on 26 tests of PureData over a three month period vs the standard technology, labor and energy & cooling costs for the BYO configuration. In both scenarios storage is assumed to grow 15% per year, with 500 concurrent users on average per database. Results will vary based on the specific client environment.

PureExperience Program: Let us prove it at no charge

1. Guided analysis of business value
2. PureSystems Technology Demonstration
3. On-Site Trial & Support
   - No charge execution of on-site service engagement
   - Trial period use of PureSystems
   - Access to a technical advocate for usage questions and advice
   - Single point of IBM support and maintenance

http://www.ibm.com/software/data/pureexperience/

Questions?

Thank you!!!

Agenda

09-15 - 09:30 - Key Note - Danilo Noveli

09:30 -10:00 - How Big Data is Transforming Industry - Chris Grote

10:00 – 10:30 – DB2 with Blu Acceleration – Les King

10:30 – 11:00 - DB2 Blu for SAP – Ferdinand Prezenski

11.:00 – 11:15 - Coffee break and Ask the Experts

11:15 - 11:30 - Update on Purescale – Les King

11:30 - 12:00 - PureData TX – Ferdinand Prezenski & Richard Hewitt

12:00 - 12:30 - Informix NoSQL for the Enterprise = NewSQL - Steve Shoaf

12:30 - Lunch

Informix 12.1: NoSQL for the Enterprise = NewSQL

Delivering Native, mobile and hybrid apps though NoSQL, JSON, RDBMS

Acknowledgements and Disclaimers

Availability. References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates.

The workshops, sessions and materials have been prepared by IBM or the session speakers and reflect their own views. They are provided for informational purposes only, and are neither intended to, nor shall have the effect of being, legal or other guidance or advice to any participant. While efforts were made to verify the completeness and accuracy of the information contained in this presentation, it is provided AS-IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this presentation or any other materials. Nothing contained in this presentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.

Industry-wide: NoSQL is gaining traction because…..

- **Non-traditional needs driven by Web 2.0 interactive applications**
  - Document stores, key-value stores, graph and columnar DBMS
  - Lower development costs – DevOps deployment
- **The Three Vs:**
  - Velocity – high frequency of data arrivals
  - Volume – BigData
  - Variability – unstructured data, continuous change requires rapid / immediate response
- **Scale-out requirements across heterogeneous environment – Cloud computing**
  - Low cost commodity platforms
  - Immediate extensibility
  - Global access

What is a NoSQL Database?

- **Not Only SQL or NOt allowing SQL**
- **A non-relational database management systems**
  - Does not require a fixed schema
  - Avoids join operations
  - Scales horizontally
  - No ACID (eventually consistent)
- **Good with distributing data and prototype project**
- **Big with web developers**

Provides a mechanism for storage and retrieval of data while providing horizontal scaling.

Basic NoSQL Terms

Term	Description
NoSQL	A class of database management systems that use some API other than SQL as the primary language. Two common features in such databases are a flexible schema, and automatic sharding and query routing across distributed nodes.
JSON	Acronym for JavaScript Object Notation – It is a text-based standard for data representation and interchange. The JSON format is often used for serializing and transmitting structured data over a network connection. It is used primarily to transmit data between a server and web application, serving as an alternative to XML.
BSON	A standardized binary representation format (see bsonspec.org) for serializing JSON documents. It allows for faster traversal of the document than when using the textual representation.

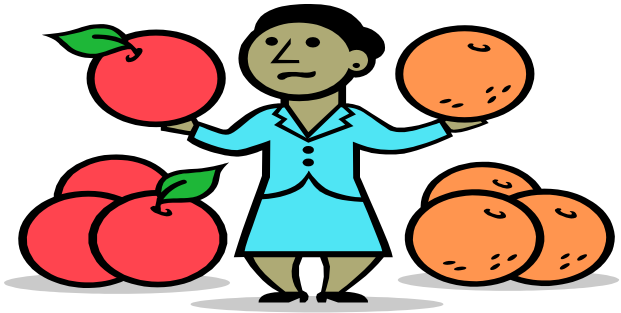
Basic Terms Translation

Mongo/NoSQL Term	Informix Term
Database	Database
Collection	Table
Document or BSON document	Row
Field	Column
Embedded documents and links	Table joins
Aggregation framework	Group by with aggregation functions

Apples and Oranges

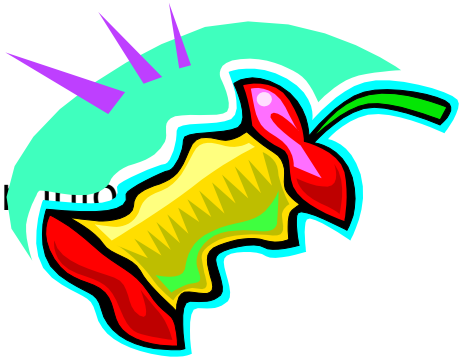
Relational systems and non-relational systems solve different problems and have different philosophies on server responsibility.

Informix – Relational Database	MongoDB - Document Store
Scales within node and by adding nodes	**Scales by adding nodes**
Suite of data protection capabilities	Minimal security
Transactional	No multi-statement transactions
Guaranteed writes	Write concern levels
Consistency of data	Eventual consistency
DB schema defines app structures	**App structures define DB data**



Informix Core Themes to a NoSQL Solution

- **Invisible and Easy to Install and Administer**
  - Support for Mongo Data Base is now part of Informix
  - JSON/BSON – NoSQL and SQL bi-directional function
    - Can, but do not have to combine both data organization

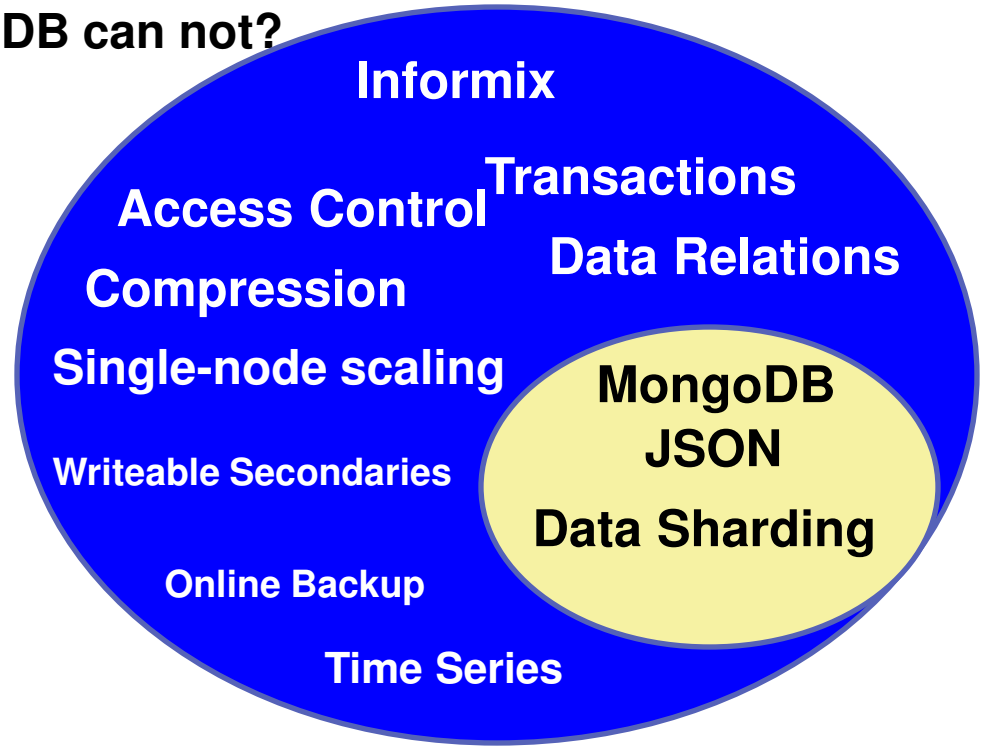


- **Dynamic Elasticity**
  - Simple to Scale Up
  - Easy to Scale-out
  - Adding and removing nodes is simple

- **Informix Value Add Propositions**
  - Hybrid functionality (combined NoSQL and Relational)
    - Relational tables and NoSQL collections co-existing in the same database
    - Join between NoSQL and Relational tables
    - Joins utilize indexes on both Relational and NoSQL
  - Enterprise level functionality

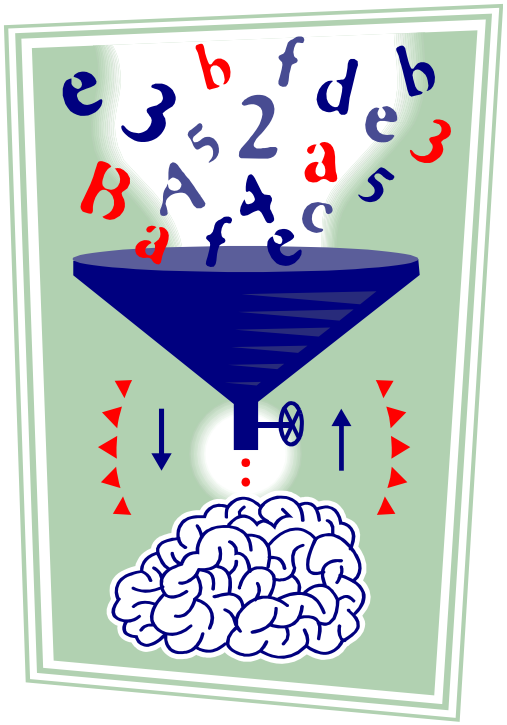
Major Capability Differences

- **What can MongoDB and Informix both do?**
  - Handle structured data in JSON format
  - Distribute (shard) query execution between server nodes
- **What can Informix do that MongoDB can not?**
  - Relationships between entities
  - Transactions
  - Access Control
  - …a great many things



Informix 12.1 New Release with New Functionality

- **Add three new built-in data-types**
  - Longlvarchar
  - JSON
  - BSON
- **New data types are native to all databases**
  - Automatically convert JSON to BSON document
  - Automatically converts BSON to JSON
- **Add new Built-in BSON Functions**
- **Complete the Sharded Operations**
  - Query in 12.10.UC1
  - Insert, Delete, Update
- **Add Simplification**
  - Installation
  - Resource Allocation



Informix's Unique Value and Capabilities

Benefit from NoSQL capabilities, using MongoDB APIs, to exploit the world-class strengths of Informix

- **Modern Interface providing JSON and BSON native support**
  - Flexible Schema support allows rapid delivery of application
  - Compatible with all MongoDB programming interfaces
  - Connect the same application developed for MongoDB to Informix with minimal/no application changes
  - Access traditional relational data from NoSQL/MongoDB application
- **Super scale out**
  - Simplify the ability to scale out to multiple nodes, multiple versions, multiple copies
  - Provided diskless and disk based scale out at the individual node with automatic failover
  - Provided Sharded Insert, Update, Delete and Query operations
  - Cloud and Virtualized environment supportability

Informix ~~NoSQL~~ NewSQL – The Hybrid Solution
Best of Both Worlds

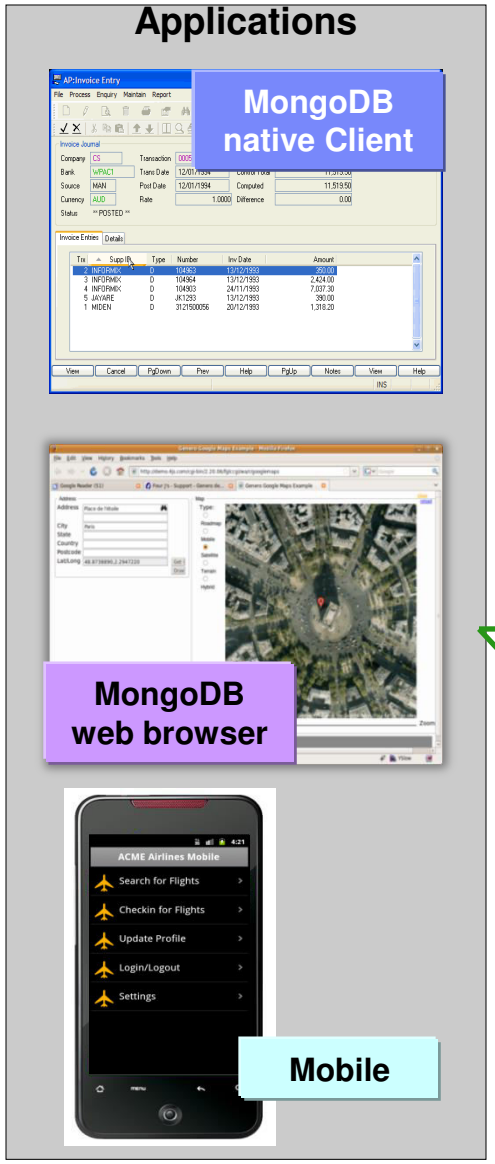
- **Relational and non-relational data in one system**
  - JSON (BSON) as first-class citizen data type
- **NoSQL/MongoDB Apps can access Informix Relational Tables**
- **Distributed Queries**
- Multi-statement Transactions
- Enterprise Proven Reliability
- Enterprise Ready Security
- Enterprise Level Performance

Informix provides the capability to leverage

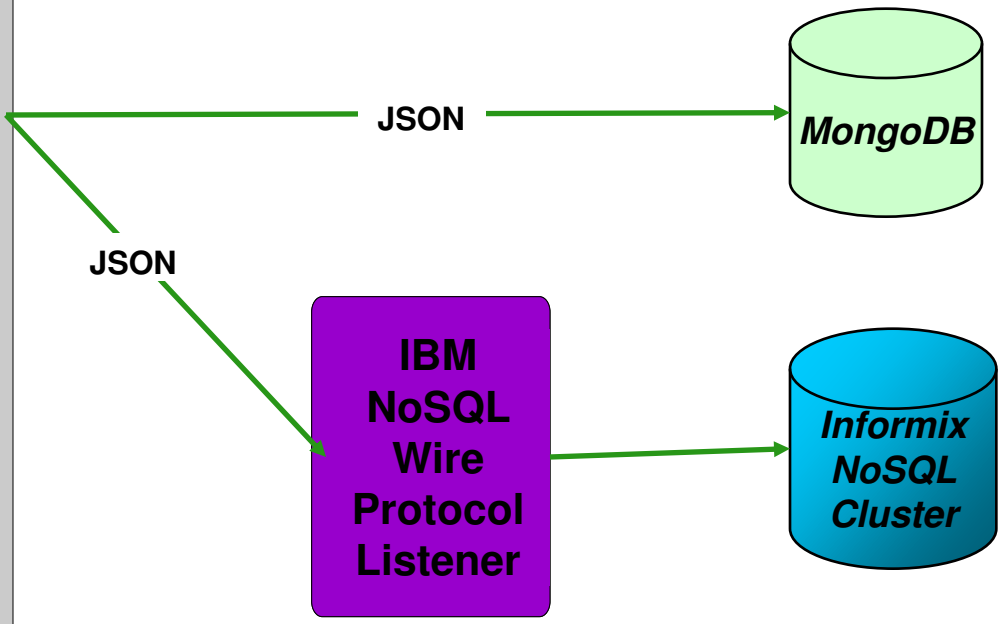
the abilities of both relational DBMS and document store systems.

MongoDB does not. It is a document store system lacking key abilities like transaction durability.

High Level Architecture



- **New Wire Protocol Listener supports existing MongoDB drivers**
  - Simple port change allows applications written for MongoDB to be intercepted by wire listener
  - Compatible with all MongoDB programming interfaces
    - Java, PHP, Python, Javascript, etc.
- **The wire listener combines MongoDB messages and BSON documents to perform actions against a distributed data store**



Where Informix NewSQL Wins

Easy application development, in popular, new languages, to exploit the traditional strengths and new NoSQL capabilities of Informix

- **In Retail Enterprises**
  - Front-end order processing
  - Product to Location demand patterns/predictions
- **Hospitability**
  - Reservation system
  - Targeted loyalty program benefits, services delivery
- **Insurance**
  - Manage documents, records, claims, using the JSON/BSON support
- **Internet of Things/Everything (IoT/IoE)**
  - 10s of billions of internet/web enabled devices – a sensor driven world
  - Services, platforms, delivered on the strengths of Informix TimeSeries and NoSQL capabilities

Some Typical NoSQL Use Cases - Mostly Interactive Web/Mobile

- **Online/Mobile Gaming**
  - Leaderboard (high score table) management
  - Dynamic placement of visual elements
  - Game object management
  - Persisting game/user state information
  - Persisting user generated data (e.g. drawings)
- **Display Advertising on Web Sites**
  - Ad Serving: match content with profile and present
  - Real-time bidding: match cookie profile with ad inventory, obtain bids, and present ad
- **Dynamic Content Management and Publishing (News & Media)**
  - Store content from distributed authors, with fast retrieval and placement
  - Manage changing layouts and user generated content
- **E-commerce/Social Commerce**
  - Storing frequently changing product catalogs
- **Social Networking**
  - Feeds
  - Extractions
- **Communications**
  - Device provisioning
  - Session control
- **Logging/message passing**
  - Drop Copy service in Financial Services (streaming copies of trade execution messages into (for example) a risk or back office system)

Enterprise Level NoSQL Operational Requirements

- **Consistent low latency, even under high loads**
  - Ability to handle thousands of users
  - Typically millisecond response time
- **Schema flexibility and development agility**
  - Application not constrained by fixed pre-defined schema
  - Ability to handle a mix of structured and unstructured data
- **Continuous availability**
  - 24x7x365 availability
  - Online maintenance operations
  - Ability to upgrade hardware or software without down time
- **Dynamic Elasticity**
  - Rapid horizontal scalability
  - Ability to add or delete nodes dynamically in the grid
  - Application transparent elasticity
- **Low cost infrastructure**
  - Commonly available hardware (Windows & Linux,…)
- **Reduced need for database administration and maintenance**

Scalability

- **Better performance on multi-core, multi-session scenarios**
  - Architecture has finer grain locking – not just entire database as with MongoDB
  - Better concurrency because less resources locked
- **Document Compression**
  - 60% to 90% observed
- **Bigger documents – 2GB maximum size**
  - MongoDB caps at 16MB
- **Informix has decades of optimization on single node solution**

Better utilization of enterprise system resources means less need to shard, for Informix

MongoDB has higher space requirements for same data

Security

- **Encryption**
  - Protects data from access in transit and on disk
- **Auditing**
  - Records who has accessed data
- **Discretionary Access Control**
  - Verifies that a user is authorized to do what they are trying to do – roles, etc

Informix has decades of solving customer security requirements

With MongoDB

- Security mostly responsibility of the application
  - Every application has to code for security
    - Consistent implementation of policies?

Support and Maintenance

- **IBM Informix Support**
  - Consistently highly rated (#1 at VendorRate 2009)
  - Simple offering
  - Severity and level of response determined by impact to customer

- **Informix reliability second to none**
  - Greater than five 9s uptime
  - Possible to manage 1000s of seats per DBA

- **MongoDB Support**
  - Various support offerings
  - Level of response determined by subscription

Enterprise Version Comparisons

Informix	MongoDB
Replicas Unlimited Writeable – local node updates	Replicas 12 per replica set Read-only
Complete, easy, automated, online backup/restore	Partial solution requiring index rebuild, or file system only
Suite of structured data extensions TimeSeries, Spatial, Text, Video…	Primitive spatial and text search capabilities
Mobile/Remote Administration (OAT)	3rd Party
Security Auditing, Kerberos, encryption, role and fine grain access control…	Security Kerberos, role access control
Reduced storage requirements - data compression	Not available

Informix and MongoDB Have Free Editions

Editions	Informix	MongoDB
Free	Developer Innovator-C	Standard
For Purchase	Express, Workgroup, Advanced Workgroup, Enterprise, Advanced Enterprise	Enterprise

MongoDB Subscriptions

	Basic	Standard	Enterprise
Edition	MongoDB	MongoDB	MongoDB Enterprise
Support	9am-9pm local, M-F	24x7x365	24x7x365
License	AGPL	Commercial	Commercial
Emergency Patches	Not Included	Included	Included
Price	$2,500 / Server / Year	$5,000 / Server / Year	$7,500 / Server / Year

Subscription information obtained from 10Gen site, June 26, 2013.

Price Point Comparison Estimate, 3-year cost

Dual Core Intel Nehalem	Innovator-C	Express (4 core, 8 GB, 2 ER nodes)	Workgroup (16 core, 16 GB, unlimited nodes)
Product Cost	$0	$8,540	$19,740
Support Subscription Year 1 24 x 7 x 365 Production System Down Development Call Emergency Patches Free Upgrades	$1,680	Included	Included
Support Renewal Year 2	$1,680	$1,708	$3,948
Support Renewal Year 3	$1,680	$1,708	$3,948
Total	$5,040	$11,956	$27,636

MongoDB Enterprise, 3-year cost*: $22,500*

Retail prices subject to change, valid as of June 26, 2013.

Informix 12.1 - Analytic Access Through NewSQL

Cognos included in Advanced Enterprise and Advanced Workgroup Editions

- Includes Cognos BI 10.2 license
  - Five user license
  - Provides powerful BI capability to the product, synergistic with IWA

SPSS included in Advanced Enterprise Edition

- SPSS Modeler and Statistics
  - Single user license
  - Provides predictive analytic capabilities, synergistic with IWA and Cognos

Informix NoSQL Answers for Mobile Requirements

- **Consistent low latency, even under high load**
  - Informix is an enterprise, industrial strength DBMS of handling thousands of users
  - Brings core DBMS functional, operational, and administrative capabilities to NoSQL based apps – Mobile or Web

- **Schema-less Flexibility and Development Agility**
  - Provides JSON & BSON functionality by default
  - Adopted core MongoDB API functionality
  - Leverages Informix's history of "keeping it simple" for JSON and BSON support
  - Provides the ability to integrate relational and NoSQL data
    - Allow indexed joins between relational and NoSQL data!

Informix NoSQL
is
Informix NewSQL
is
Simply Powerful

- Knowledge is power…..
- Capture and analyze the “interactive” element of your customer’s relationship with your business
- Change the business view from “what happened?” to “what is happening?”
- Create, learn, adjust, and move forward

Thank You!

shoafs@us.ibm.com
+1 925 899 8747

The
following slides are technical details and
deep dive questions
not part of the presentation

Informix NoSQL Comparisons

Description	Informix NoSQL	MongoDB
Sharded Queries, Inserts, Deletes and Updates	YES	YES
Leverage durable transactions and eventually consistent transactions in the same database session	YES	NO
Leverage enterprise grade compliance and security ▪Built in ability to track who modifies or read data ▪Selectively provide read and/or write access to data by collection ▪Simple consistent permission model across all shards ▪Native encryption of data on disk, network and backups	YES	NO
Ability to utilize multiple indexes on a single query	YES	NO
Commodity High-Availability solutions with active-active failover nodes	YES	NO
Point-in-Time Online backup and restore built into the product ▪Fully scheduled unattended online backup to Disk, Cloud, Storage manager ▪Point-in-Time restore of a single NoSQL collection or document	YES	NO
Simple Analytics Solution on NoSQL collection – Cognos	YES	NO
Lock Capabilities •Database Level •Collection Level •Document Level	 Yes Yes Yes	 Yes No No
Scale a single instance/node to fully utilize all of a computer	YES	NO
Automatic collection/document compression	YES	NO
Maximum document size	2 GB	16 MB
High performing architecture on vmware and cloud environments	Yes	NO

Informix NoSQL Hybrid Comparisons

Description	Informix NoSQL	MongoDB
Combine the power and performance of a traditional relational database with modern NoSQL capabilities	YES	NO
Join data between NoSQL collections and traditional relational data	YES	NO
Durable and/or Eventually Consistent transactions spanning multiple SQL statements	YES	NO
Combine the NOSQL indexes with Traditional Relational indexes	YES	NO
Industrial strength query optimizer when combining NOSQL with Traditional Relational models	YES	NO
Leverage durable transactions and eventually consistent transactions in the same database session	YES	NO
Online backup and restore built into the product • Fully scheduled unattended online backup to Disk, Cloud, Storage manager • Point-in-Time restore of a single NoSQL collection or document	YES	NO
Enterprise class NoSQL timeseries solution utilized by sensor (SCADA) oriented industries (see benchmark data)	YES	NO
Not new to low DBA requirements when dealing with distributed/replicated environments ▪ Wal-Mart ▪ Over 15,000 servers with 6DBA;350 system are fully replicated ▪ Cisco VOIP ▪ Completely embed, with no DBAs; Every system has update anywhere environment with 20 node maximum	YES	NO

NoSQL Requirements driven by Use Cases

Description	Informix
Consistent Low Latency, even under high load • Ability to handle thousands of users • Typically millisecond response time	Yes

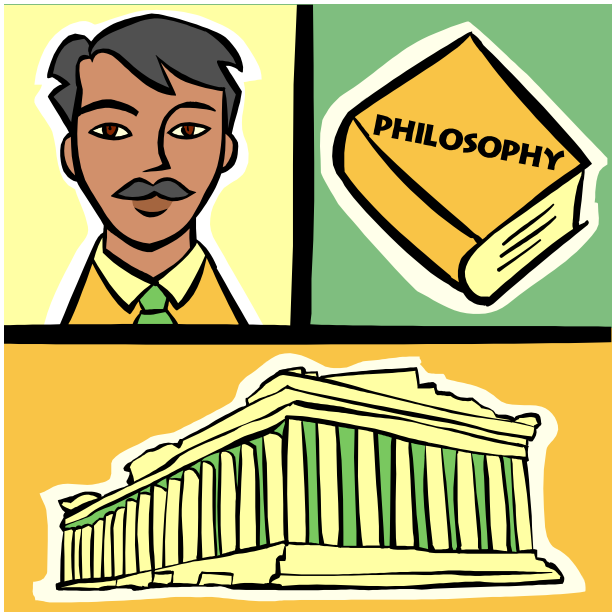
Schema Flexibility & Development Agility
• Application not constrained by fixed pre-defined schema
  • Application drives the schema
  • Ability to develop a minimal application rapidly, and iterate quickly in response to customer feedback
  • Ability to quickly add, change or delete “fields” or data-elements
• Ability to handle a mix of structured and unstructured data
  • Easier, faster programming -> Faster time to market, quick to adapt
	Yes

Continuous Availability • 24x7x365 availability • (Today) Requires data distribution and replication • Online Maintenance Operations • Ability to upgrade hardware or software without any down time	Yes
Dynamic Elasticity • Rapid horizontal scalability • Ability to add or delete nodes dynamically • Application transparent elasticity (e.g. automatic (redistribution of data, if needed) • Cloud compatibility	Yes
Low cost infrastructure • Commonly available hardware (Windows & Linux,…) • Lower cost software (open source or pay-per-use in cloud)	Yes
Low/No Admin Reduced need for database administration, and maintenance	Yes

Full ACID (Atomicity, Consistency, Isolation, Durability) NOT a requirement

NoSQL Database Philosophy Differences

- **No ACID**
  - No ACID (Atomicity, Consistency, Isolation, Durability)
  - An eventual consistence model
- **No Joins**
  - Generally single row/document lookups
- **Flexible Schema**
  - Rigid format



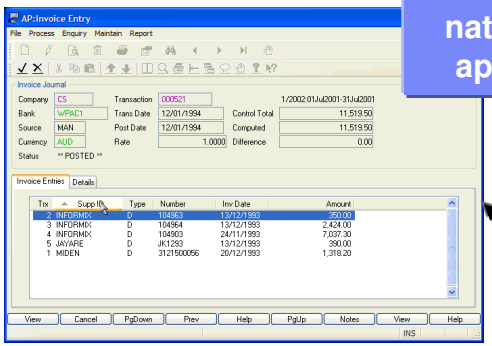
New Functionality

Simple, Simple, Simple

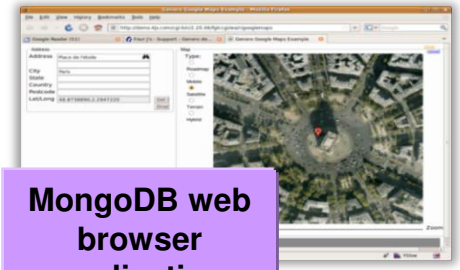
Description
Auto tuning of CPU VPS
Auto Table Placement
Auto Buffer pool tuning
Auto Physical Log extension
Auto Logical Log Add
Asynchronous Sharded Deletes
Asynchronous Sharded Updates
Asynchronous Sharded Inserts
Easy Install

Client Applications

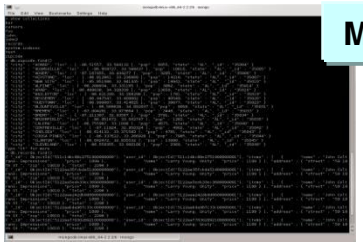
- New Wire Protocol Listener supports existing MongoDB drivers
- Connect to MongoDB or Informix with same application!



MongoDB native Client application



MongoDB web browser application



MongoDB shell

MongoDB driver

MongoDB Wire Protocol

MongoDB

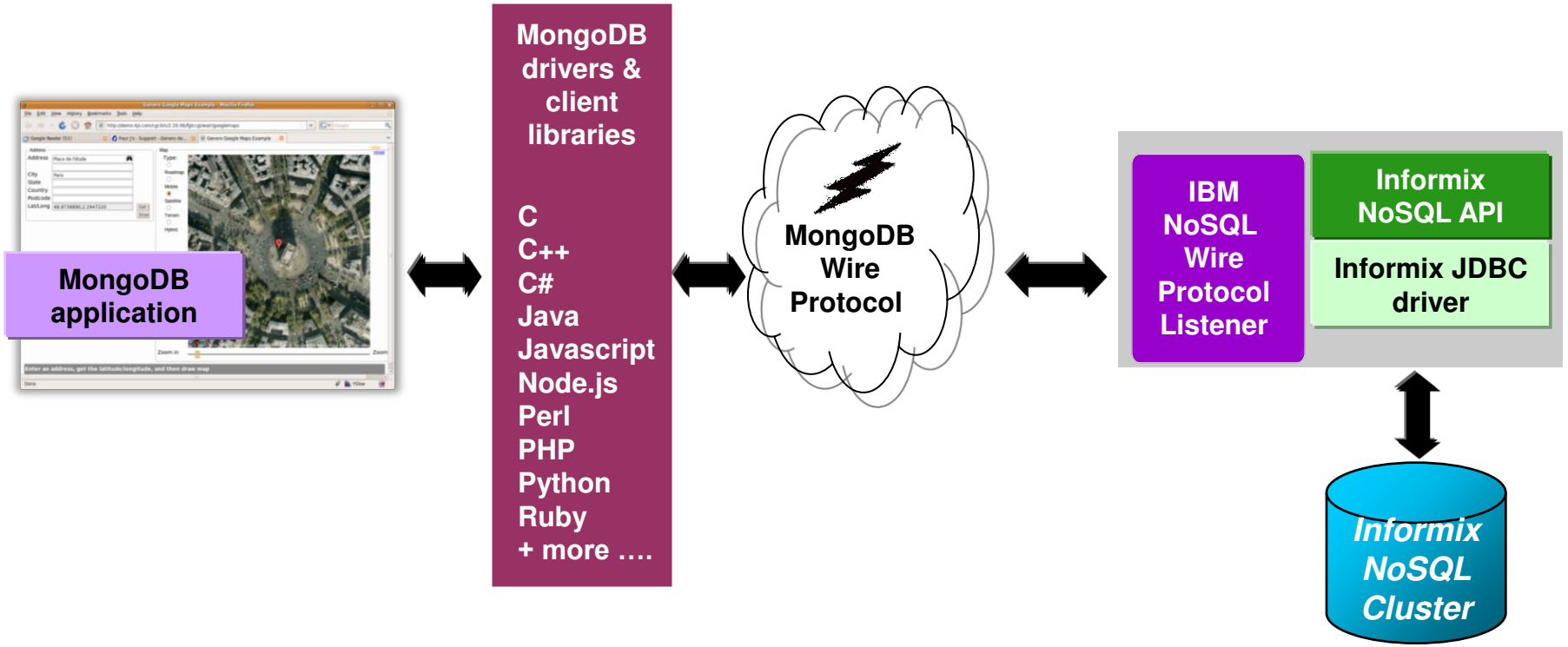
IBM NoSQL Wire Protocol Listener

Informix NoSQL Cluster

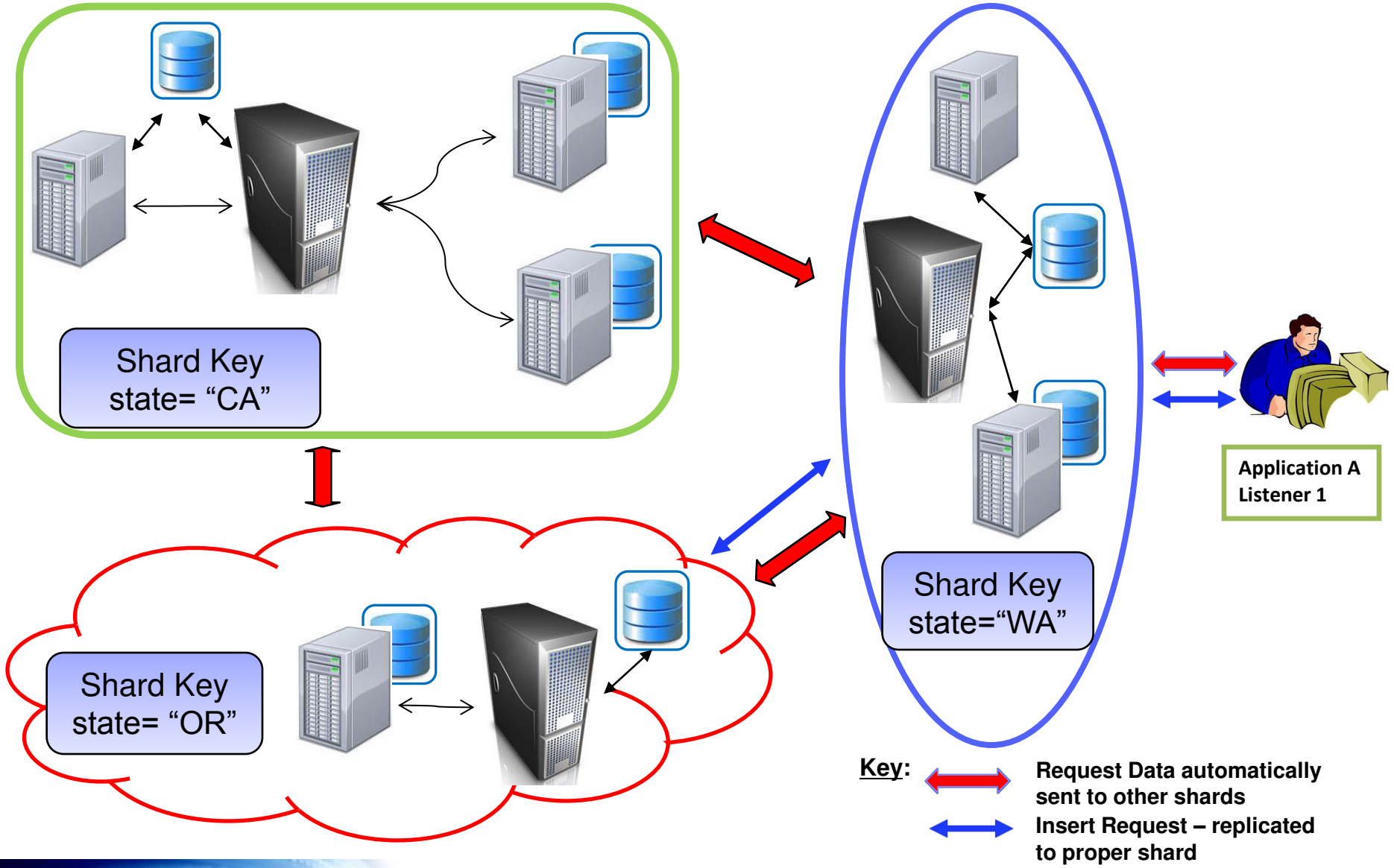
IBM Wire Protocol Listener logic shared with DB2 and Extreme Scale

Client Applications - Details

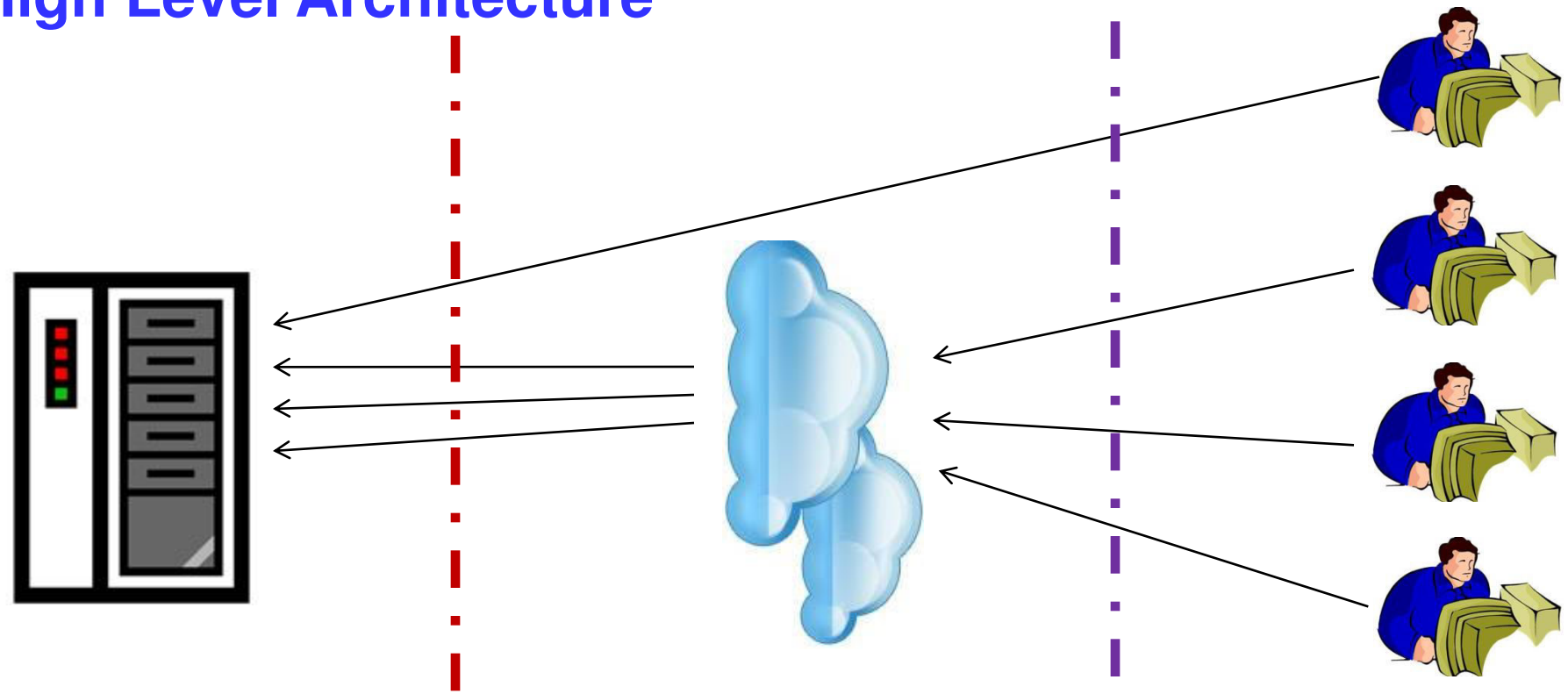
- **NoSQL Wire Protocol Listener works with existing drivers using standard MongoDB client-server protocol**
  - Java, PHP, Python, Javascript, etc.
  - MongoDB supported or community supported drivers
- **Uses new Informix NoSQL API functions**
- **Connectivity to Informix via JDBC**



Informix NoSQL Cluster Architecture Overview



High Level Architecture



Informix NoSQL Cluster Solution

Automatic Routing, Failover Coordination

Client Applications

JSON: JavaScript Object Notation

- **What is JSON?**
  - JSON is lightweight text-data interchange format
  - JSON is language independent
  - JSON is "self-describing" and easy to understand

- **JSON is syntax for storing and exchanging text information much like XML. However, JSON is smaller than XML, and faster and easier to parse.**

```
{
"name":"John Miller",
"age":21,
"count":27,
"employees": [
    { "firstName":"John" , "lastName":"Doe" },
    { "firstName":"Anna" , Middle":"Marie","lastName":"Smith" },
    { "firstName":"Peter" , "lastName":"Jones" }
                ]
}
```

BSON is a binary form of JSON.

Understanding Informix BSON Indexes

- **Indexes are created on BSON data and support**
  - Arrays
  - Composite Indexes
  - Unique Indexes (enforced at a single node level)
  - Primary Key (enforced across all nodes)

```
{
"fname":"Sadler",
"lname":"Sadler",
"company":"Friends LLC",
"age":21,
"count":27,
“phone": [ “408-789-1234”, “408-111-4779” ],
}
```

```
create index fnameix1 on customer(bson_value(bson,"fname")) using bson;

create index lnameix2 on customer(bson_value(bson,"lname")) using bson;

create index phoneix3 on customer(bson_value(bson,"phone")) using bson;
```

New Built-in BSON Expressions

```
bson_value_double(lvarchar doc, lvarchar key) returns float
bson_value_lvarchar(lvarchar doc, lvarchar key) returns lvarchar as string
bson_value_document(lvarchar doc, lvarchar key) returns lvarchar as BSON object
bson_value_array(lvarchar doc, lvarchar key) returns lvarchar as BSON array
bson_value_binary(lvarchar doc, lvarchar key) returns lvarchar as BSON binary
bson_value_objectid(lvarchar doc, lvarchar key) returns lvarchar as string
bson_value_boolean(lvarchar doc, lvarchar key) returns boolean
bson_value_date(lvarchar doc, lvarchar key) returns datetime
bson_value_code(lvarchar doc, lvarchar key) returns lvarchar as string
bson_value_int(lvarchar doc, lvarchar key) returns bigint
bson_value_bigint(lvarchar doc, lvarchar key) returns bigint
bson_value_timestamp(lvarchar doc, lvarchar key) returns datetime
bson_key_exists(lvarchar doc, lvarchar key) returns boolean
```

Understanding Informix BSON Indexes

```
create index fnameix1 on customer(bson_value(bson,"fname")) using bson;
create index lnameix2 on customer(bson_value(bson,"lname")) using bson;
create index phoneix3 on customer(bson_value(bson,"phone")) using bson;
```

```
select  * from customer where  bson_value(bson,"fname") = "Ludwig";
        -- use fnameix1
select  * from customer where  bson_value(bson,"lname") = "Sadler";
        -- use lnameix2
select  * from customer where  bson_value(bson,"phone") = "408-789-8091";
        -- use phoneix3
select  * from customer where  bson_value(bson,"phone") = "415-822-1289" OR
                               bson_value(bson,"phone") = "408-789-8091";
        -- use phoneix3
select  * from customer where  bson_value(bson,"company") = "Los Altos Sports";
        -- no index use sequential scan
```

Basic Data Distribution/Replication Terms

Term	Description	Informix Term
Shard	A single node or a group of nodes holding the same data (replica set)	Instance
Replica Set	A collection of nodes contain the same data	MACH Cluster
Shard Keys	The field that dictates the distribution of the documents. Must always exist in a document.	Shard Keys
Sharded Cluster	A group shards were each shard contains a portion of the data.	Grid/ER
Slave	A server which contains a second copy of the data for read only processing.	Secondary Server Remote Secondary

Basic MongoDB Operations Conceptual Operations

Mongo Action	Informix Action
db.customer.insert({ name: “John", age: 21 })	INSERT INTO customer (name, age) VALUES (“John”,21)
db.customer.find()	SELECT * FROM customer
db.customer.find({age: { $gt:21 } })	SELECT * FROM customer WHERE age > 21
db.customer.drop()	DROP TABLE customer
db.customer.ensureIndex({ name : 1, age : -1 })	CREATE INDEX idx_1 on customer(name , age DESC)
db.customer.remove({age: { $gt:21 } })	DELETE FROM customer where age > 21
db.customer.update({ age: { $gt: 20 } }, { $set: { status: “Drink" } }, { multi: true })	UPDATE customer SET status = “Drink" WHERE age > 20

JSON Details

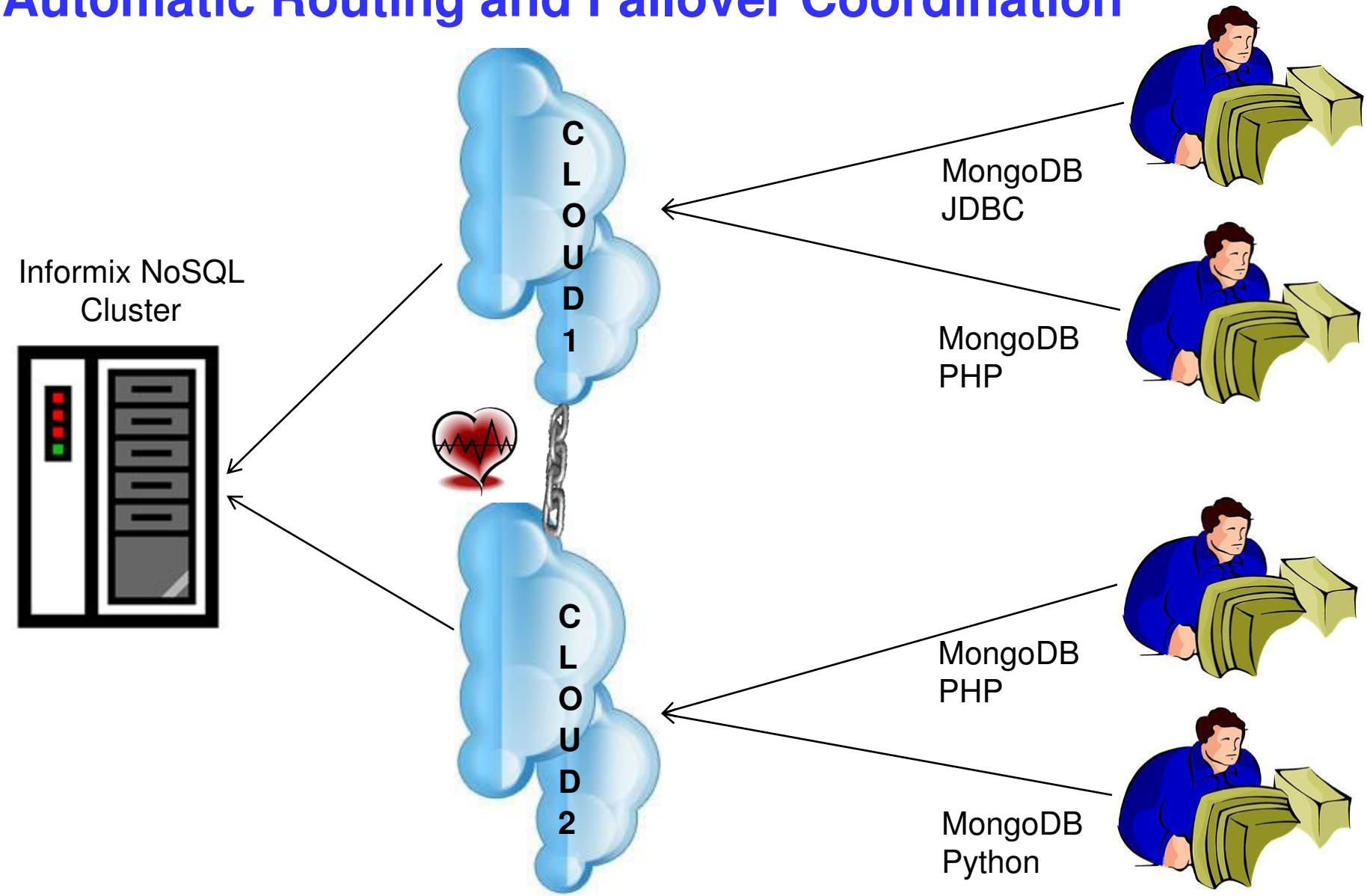
- **JSON Syntax Rules**
  - JSON syntax is a subset of the JavaScript object notation syntax:
  - Data is in name/value pairs
  - Data is separated by commas
  - Curly braces hold objects
  - Square brackets hold arrays

- **JSON Name/Value Pairs**
  - JSON data is written as name/value pairs.
  - A name/value pair consists of a field name (in double quotes), followed by a colon, followed by a value:

```
"name":"John Miller"
```

- **JSON Values can be**
  - A number (integer or floating point)
  - A string (in double quotes)
  - A Boolean (true or false)
  - An array (in square brackets)
  - An object (in curly brackets)
  - Null

Automatic Routing and Failover Coordination



Why RDBMS do not meet the new Web 2.0 Requirements

- **Consistent Low Latency, even under high load**
  - ACID requirements inherently introduce write latency
  - There is no latency-consistency tradeoff knobs available
  - Requirement can be met, but at a much higher cost (hardware, software or complexity)
- **Schema Flexibility & Development Agility**
  - Relational schemas are inherently rigid
  - Database design needs to be done upfront
  - Different rows cannot have a different structuree
  - Database design needs to be done before application is developed
  - Data modeling based on domain objects, which may not be well understood upfront
- **High Availability**
  - Requirement can be met, but at a significant cost
  - Typically hardware and software upgrades require some downtime
  - Typically rolling version upgrades are complex in clustered RDBMS
- **Dynamic Elasticity**
  - Not a natural fit for RDBMS, due to requirement for strong consistency
  - Scale-out requires partition tolerance, that increases latency
- **Low Cost**
  - Distributed RDBMS typically require specialized hardware to achieve performance
  - Popular relational databases typically require several DBAs for maintenance and tuning

Informix Answers to Mobile Requirements

- **Continuous availability**
  - Informix Grid Replication
    - Supports servers running of different
      - Database server version
      - Operating system version
      - Machine architecture
    - Automatic resynchronization for troubled nodes
    - All functionality exists commodity hardware and software
  - Connection Manager provides
    - Connections based on policy or workload
    - Automatic re-direction for down servers
- **Dynamic Elasticity**
    - Provides a one setup for new nodes
    - When the MACH component is integrated within a Grid's node
      - Provides ondemand diskless horizontal scaling
      - Failover redundancy
- **Low cost infrastructure**
  - History of many customer running thousands of systems which exceedingly high up time and little to no DBAs

Informix NoSQL Capabilities

- Key-value Stores
  - Hashed Key value model
  - Capable of massive data storage
- Column Family Stores
  - Keys that point to multiple columns and column families
- Document Data Base
  - JSON / BSON Formatted documents
  - Collections of Documents with Key-value stored versions
- Graph Databases
  - Nodes / relationship between nodes
  - Properties of nodes
  - Data values
  - Supports SPARQL ?

Use Cases for Informix NoSQL

- Session Store
  - High volume ingestion / low latency access requirements
  - Linkage to a specific user provides immediate recognition of customer preferences
  - Session restore after “break” is immediate to last functions performed increase customer confidence and connection to their desire(s)
- User Profile Store
  - Customer Profiles, orders and shipment status is immediate and searchable
  - Fast access for Authentication and preferences
  - Historical access and click streams increase personalization & targeting
- Content and Metadata Store
  - Document database for digital content, articles, instruction manuals
  - Key-value linkages provide linkage to customer profiles and history
  - Multiple image/data type support provide fast access to different data type

Use Case for Informix NoSQL (cont)

- Mobile Apps
  - Ability to store content and user data in schema-less formats increase development and deployment speed and changes to existing apps
  - Supports multi-format data storage associated with differing device types
  - Scalable storage provides for document, image, text oriented storage and retrieval of apps data / information
- Third Party Aggregation
  - High speed ingestion of data feed from multiple sources:
    - Retail store loyalty programs, social media, marketing information, purchase histories by groups, person, industries
  - Ease of format management and analysis
- High Availability Cache
  - Storage for popular / frequent search results
  - Session information
  - High Frequency Ads tailored for User Profiles, Locations, searches
  - Dual function cache for data store and fast response time for retrieval

Informix NoSQL Use Case (cont)

- Globally Distributed Data Repository
  - Scalable nodes and access across distributed systems
  - Location affinity for workload optimization
  - Capture of differing formats and languages to provide global and discrete views of the business and analytics of searches, purchases, etc.
- eCommerce
  - Ability to handle purchase spikes / peak time through scalable system architecture and low cost commodity hardware
  - Fast / rapid application deployment through schema-less development model
  - Ability to combine transaction data with user action reduces cost of upgrading business apps for transaction management
- Social Media & Gaming
  - Rapid app development , deployment, and change implementation increase ability to grow customer base and respond to trends – or create them
  - Fast access to user profile for authentication and preferences / historical information
  - Real time Analytics and trend identification

Informix NoSQL Use Cases (cont)

- Ad Targeting
  - Fast access to user profile and histories permit well managed Ad placement increasing revenue and buy decision
  - Click history and Pyschographic based suggestions
  - Ability to ingest other feeds and quickly relate to a specific customer
    - Loyalty programs
    - Social Media
    - Associations (LinkedIn, Dating sites, Industry Groups etc.)

Cognos BI 10.2

Delivers business intelligence through reports, analysis, dashboards, and scorecards

- **Reports** equip users with the information they need to make fact-based decisions.
- **Dashboards** help users access, interact and personalize content in a way that supports how they make decisions.
- **Analysis** capabilities provide access to information from multiple angles and perspectives so you can view and analyze it to make informed decisions.
- **Collaboration** capabilities include communication tools and social networking to fuel the exchange of ideas during the decision-making process.
- **Scorecarding** capabilities automate the capture, management and monitoring of business metrics so you can compare them with your strategic and operational objectives

SPSS Modeler Professional bundled with Informix

- Employ a wide range of data mining algorithms with many advanced features to get the best possible results from your Informix based data and information
  - Classification Algorithms: Predictions & forecasts
  - Association Algorithms: Group people, things, events & detect patterns
  - Association Algorithms: Discover links & sequences
  - Time Series & forecasting: statistical modeling
- SPSS Modeler Professional delivers
  - Data Mining
  - Text Analytics
  - Build Predictive Models
  - Visual Interface – visual data
  - Deploy results into Cognos BU
- Useful for
  - Customer analytics
  - Fraud Detection
  - Risk Management
  - Sales Forecasts