



IBM Software

# Information Management & Analytics Forum **2013**

Return on Information: The New ROI

## New Era of Data Warehousing and Analytics

Presented by:

James Cho, PureData Systems Architect

Sameer Vaishampayan, Data Warehouse Architect



# Disclaimer

The information contained in this presentation is provided for informational purposes only.

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.

In addition, this information is based on IBM’s current product plans and strategy, which are subject to change by IBM without notice.

IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this presentation or any other documentation.

Nothing contained in this presentation is intended to, or shall have the effect of:

- Creating any warranty or representation from IBM (or its affiliates or its or their suppliers and/or licensors); or
- Altering the terms and conditions of the applicable license agreement governing the use of IBM software.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

# *Agenda*

- Introduction : Challenges Organization are facing
- Database Architectures
- Operational Analytics Strategy and Focus
- PureData for Operational Analytics
- PureData for Analytics
- DB2 10 Key Technology
- Real Time Analytics Value and Case Studies
- Questions/Discussion

Organizations are focused on delivering better outcomes

**Generate  
More Revenue**

**Reduce Risk**

**Predict Future Outcomes  
with Greater Confidence**

**Lower Costs**

IBM Software

**Information Management & Analytics Forum 2013**

Return on Information: The New ROI



## Yet All Organizations are Facing an Information Challenge

### Trust 1 in 3

Business leaders frequently make decisions based on information they don't trust, or don't have.

### Vision 83%

Of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness.

### Access 1 in 2

Business leaders say they don't have access to the information they need to do their jobs.

### Pre-integrated 35%

Of businesses will look to replace their current warehouse with a pre-integrated warehouse solution in the next 3 years, only 14% have today.

# Information Complexity Across the Organization is Increasing



**Executive**



**Business Manager**



**Line Manager**



**Casual Business User**



**Business Analyst**



**Financial Analyst**



**What is happening?**

**Why are we on/off track ?**

**What is likely to happen?**

**What should we do next ?**



Analytics-driven Organizations Can...

## Increase agility

...rapidly respond to opportunity

- **Precise customer contribution margins**
- **Reduced operating costs**



Analytics-driven Organizations Can...

## Detect patterns

...stop crime before it happens

- **Higher case closings**
- **Improved officer safety and resource allocation**



Analytics-driven Organizations Can...



## Anticipate demand

...and immediately match it

- **Infusing business analytics into every new solution**
- **Able to scale to meet the demands of internal growth**



IBM Software

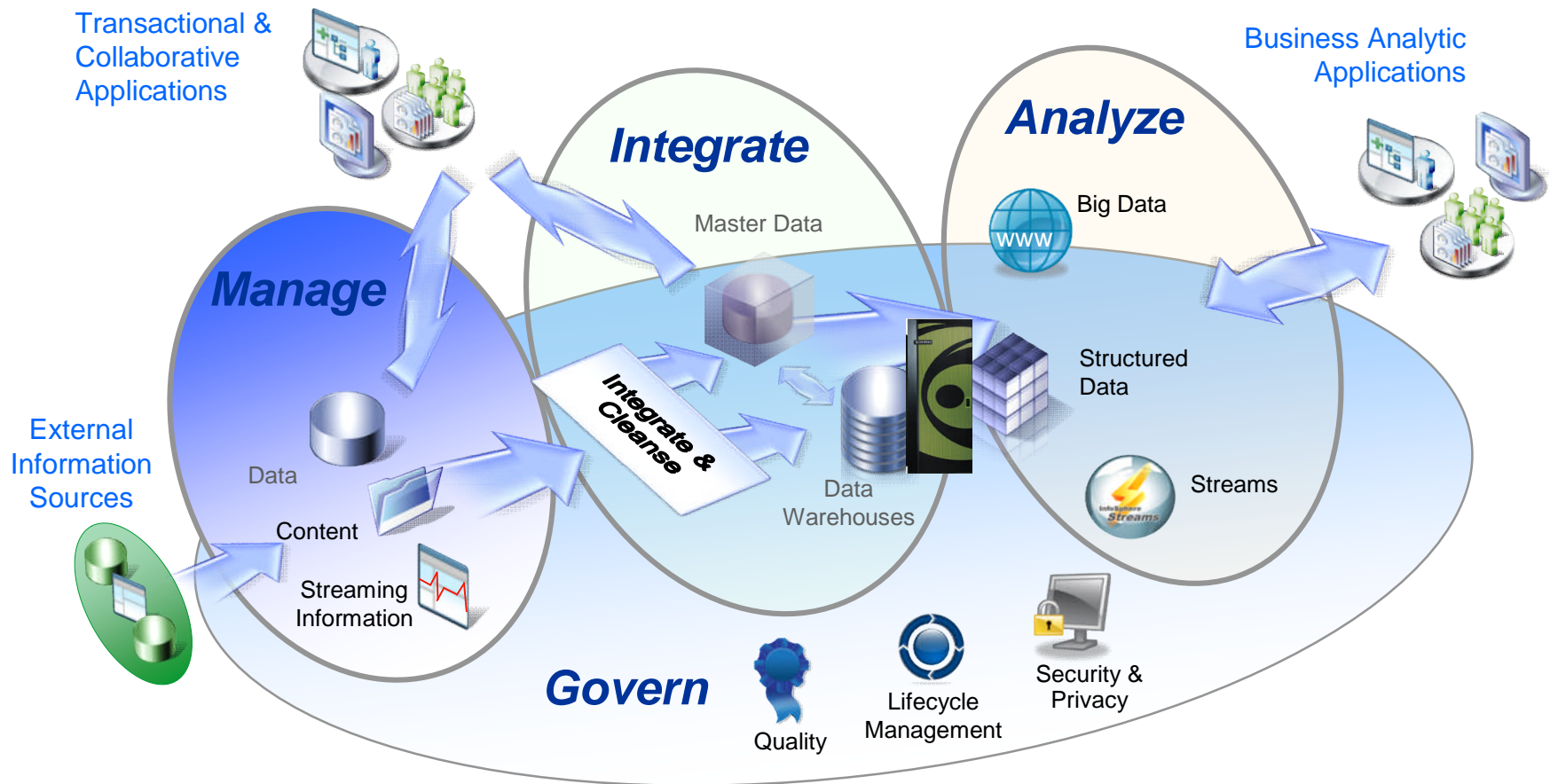
Information Management & Analytics Forum 2013

Return on Information: The New ROI

# Data Warehouse

## The Information Ecosystem

**Silo'd Data** → **Trusted Information** → **Actionable Insight**

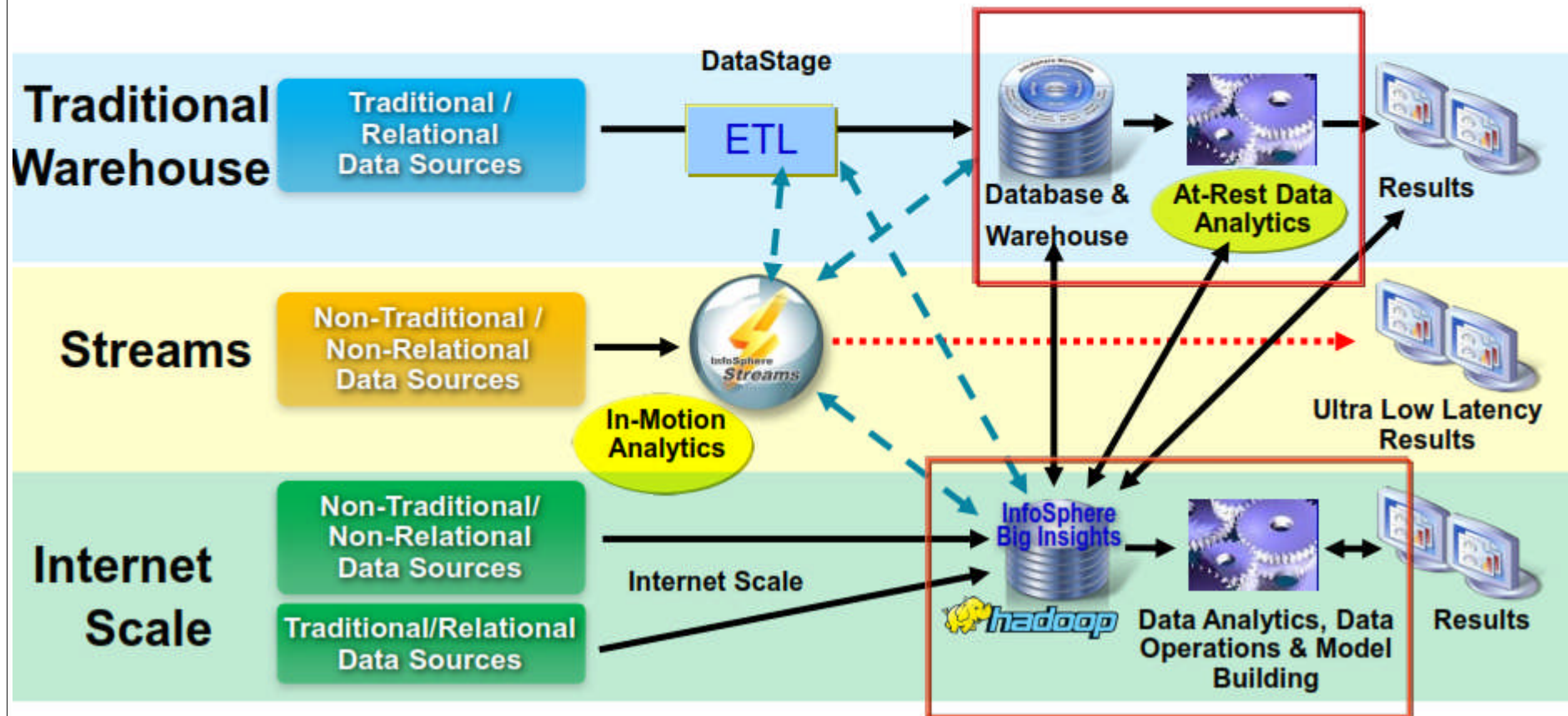




# Most Comprehensive Portfolio

## - IBM Data warehousing & Analytics

- 3 Key Swim Lanes
  - Traditional/Relational Data Sources w/ Traditional BI
  - Non-Traditional/Non-Relational Data Sources (e.g. real-time) w/ Event-based Analytics
  - Unconventional, Big Data Data Sources w/ Big Data Analytics

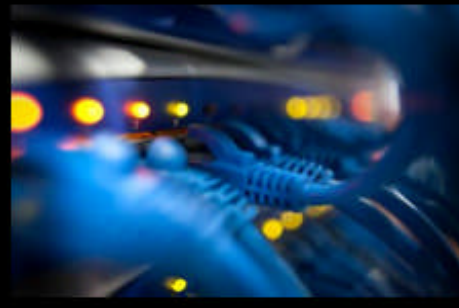


# Traditional data warehousing for today's business analytics **Can be too complex for many customers**

## A lack of optimization to meet the demands of advanced analytics

- Too complex an infrastructure
- Too complicated to deploy
- Too much tuning required
- Too inefficient at analytics
- Too many people needed to maintain
- Too costly to operate

**Too long to get answers**





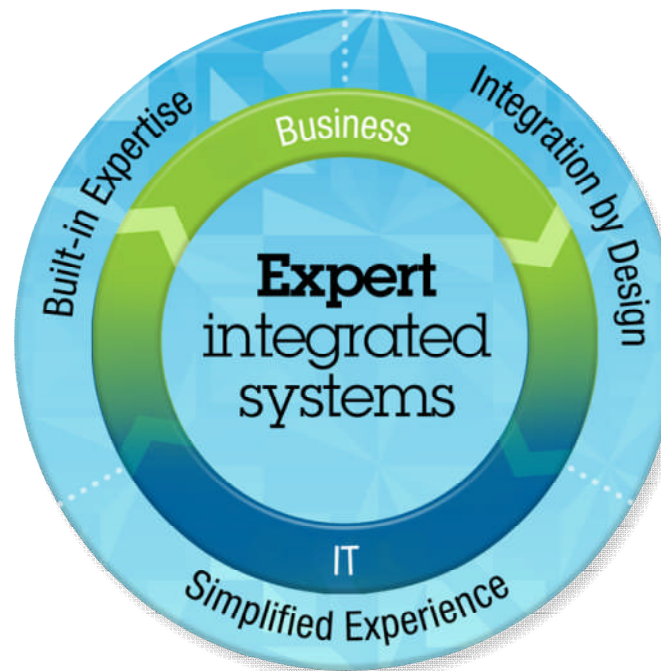
A new family of expert integrated systems

# PureSystems

*Systems with integrated expertise and built for cloud*

## **Built-in Expertise**

***Capturing and automating what experts do*** – from the infrastructure patterns to the application patterns



## **Integration by Design**

***Deeply integrating and tuning hardware and software*** – in a ready-to-go workload optimized system

## **Simplified Experience**

***Making every part of the IT lifecycle easier*** - with integrated management of the entire system and a broad open ecosystem of optimized solutions

# IBM PureSystems Family

How much flexibility, integration and workload optimization do you want **out of the box**?

## PureFlex

### Infrastructure

*Integrated and optimized infrastructure with flexibility*

*Runs your choice of applications and middleware*



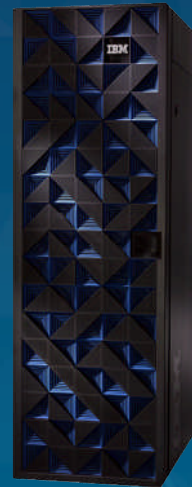
- Delivering IT infrastructure services

## PureApplication

### Application Platform

*Integrated and optimized application platform*

*Built on IBM middleware to accelerate deployment of your choice of applications*



- Delivering application platform services

## PureData

### Data Platform

*Integrated and optimized data platform*

*Delivers high performance data services to transactional and analytics applications*



- New PureSystem with models optimized exclusively for data workloads



# IBM PureData System



## PureData System for Transactions

- **Pattern based *database deployment in minutes, not hours*<sup>1</sup>**
- **Handles *more than 100 databases on 1 system*<sup>2</sup>**

## PureData System for Analytics

*powered by  
Netezza technology*

- **10-100x faster than traditional custom systems<sup>4</sup>**
- **20x greater concurrency and throughput for tactical queries than previous Netezza technology<sup>5</sup>**

## PureData System for Operational Analytics

- **Continuous ingest of operation data**
- **Handles 1000+ concurrent operational queries<sup>3</sup>**
- **Up to 10x storage savings with adaptive compression<sup>6</sup>**

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 IBM internal tests of prior generation system, and on system design for normal operations under expected typical workload. Individual results may vary.  
4. Based on IBM customers' reported results. "Traditional custom systems" refers to systems that are not professionally pre-built, pre-tested and optimized. Individual results may vary.  
5. Based on IBM internal performance benchmarking  
6. Based on client testing in the DB2 10 Early Access Program

# Database Architectures

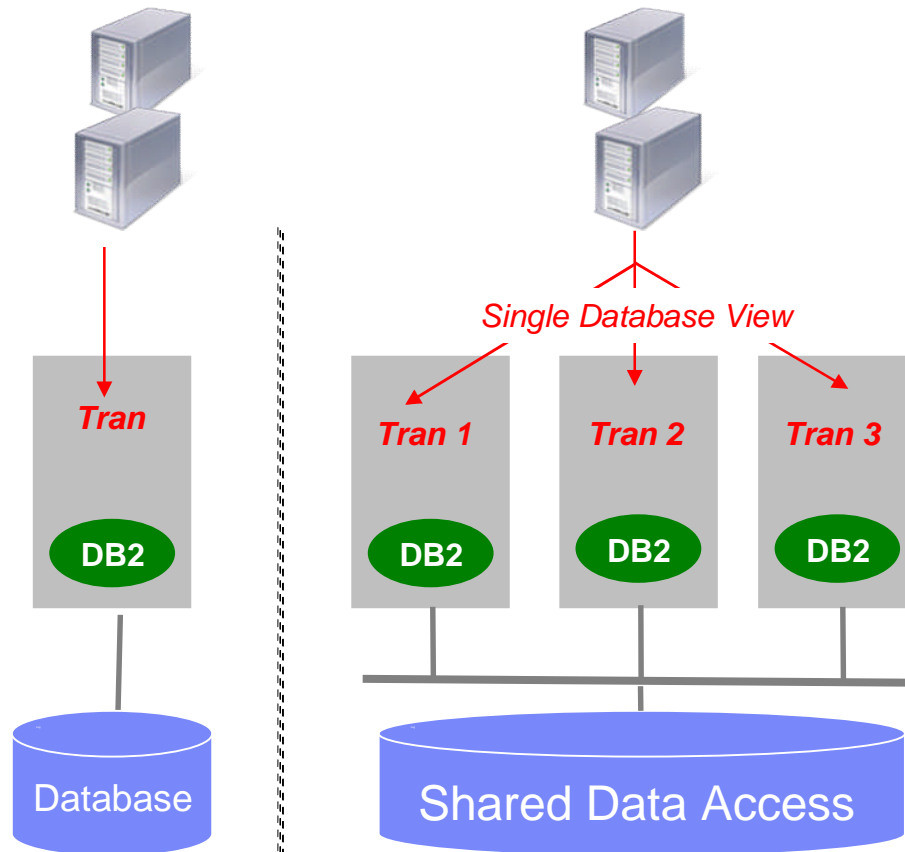
IBM Software

**Information Management & Analytics Forum 2013**

Return on Information: The New ROI

# Database Architectures: Single, Shared, Shared Nothing

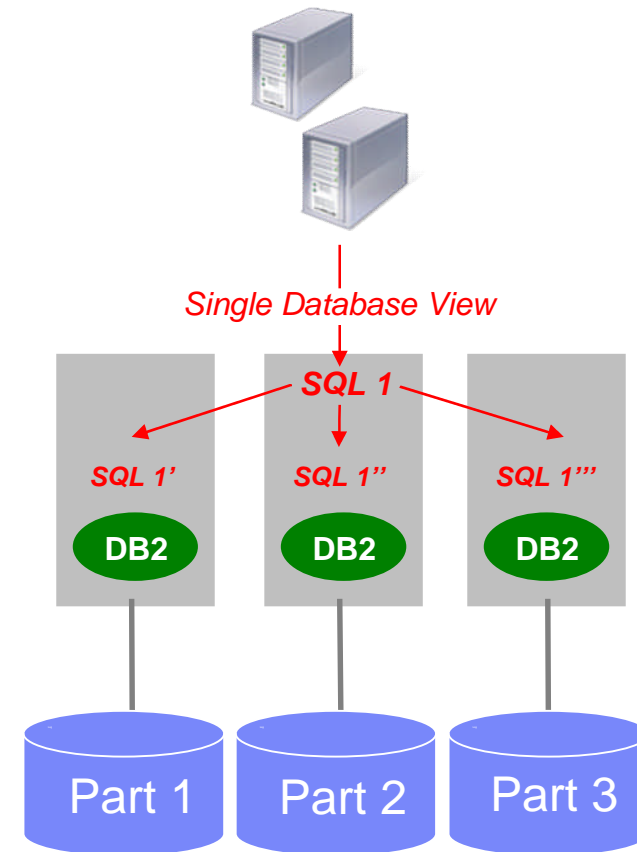
## Optimized for OLTP



- *Core DB2*
- Ideal for OLTP and data marts

- *DB2 pureScale Data Sharing*
- Ideal for active/active OLTP/ERP scale out

## Optimized for Analytics

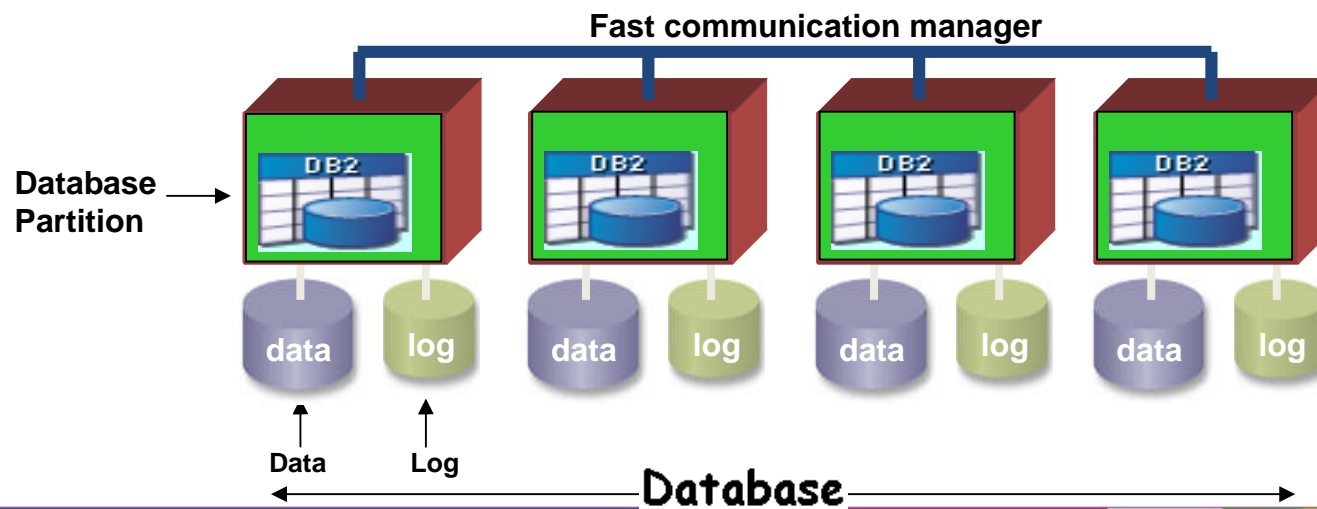


- *DB2 InfoSphere Warehouse (aka DPF)*
- Ideal for data warehousing with MPP scale out for near linear scalability and query processing

# The design of the PureData for Operational Analytics complements the DB2 shared nothing architecture to deliver a scalable platform

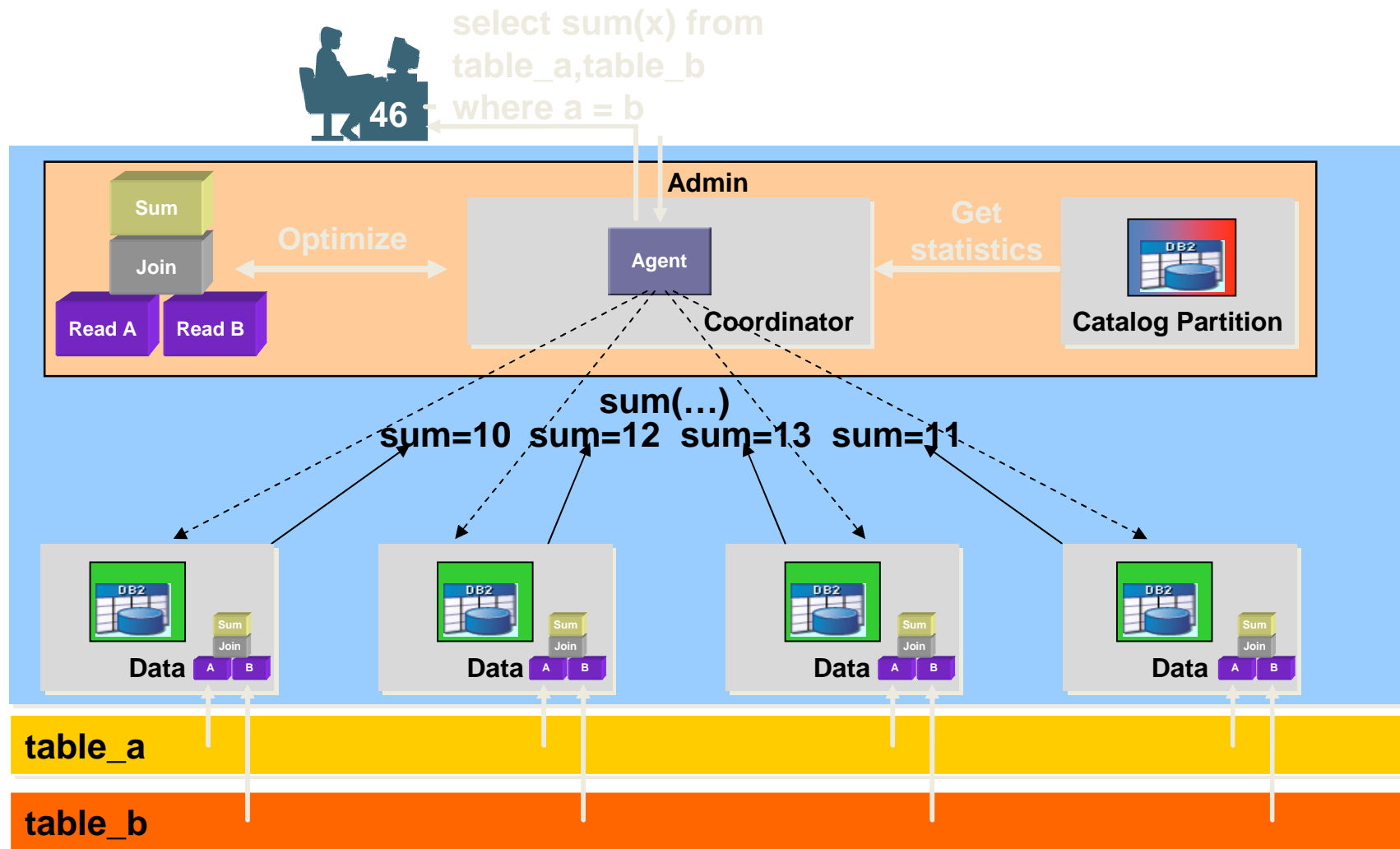
## Partitioned Database Model

- Database is divided into multiple partitions
- Database Partitions can run on different servers
- Each Database Partition has dedicated resources
  - Engine, Logging, Locking, Caches, etc.
- Parallel Processing occurs on all partitions and is coordinated by the DBMS
- **Single system image** to user and application





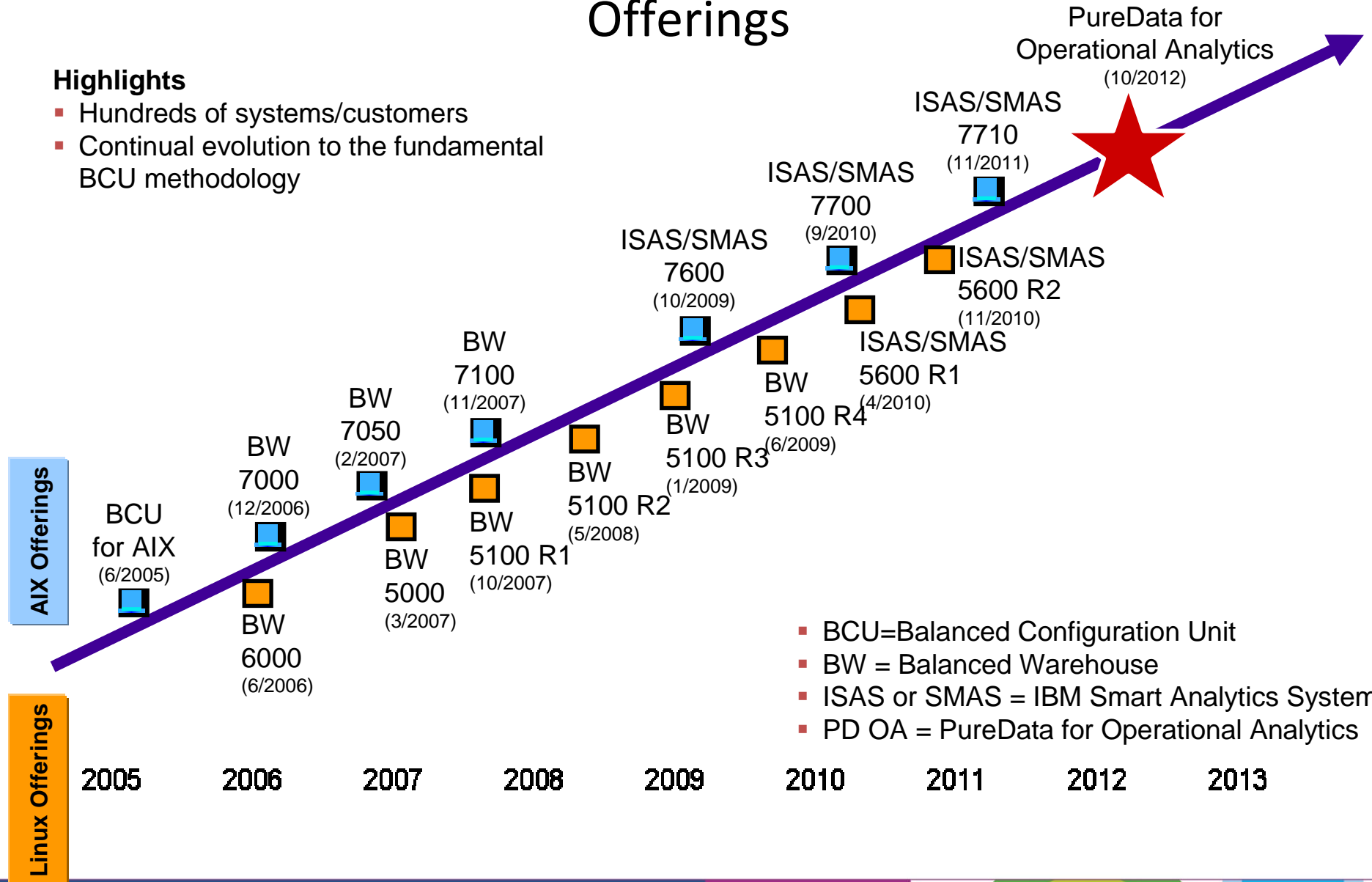
# Parallelism: Database Partitions & Query Processing



# Progression of IBM Data Warehouse Integrated Offerings

## Highlights

- Hundreds of systems/customers
- Continual evolution to the fundamental BCU methodology



- BCU=Balanced Configuration Unit
- BW = Balanced Warehouse
- ISAS or SMAS = IBM Smart Analytics System
- PD OA = PureData for Operational Analytics

# Operational Analytics Strategy and Focus

IBM Software

**Information Management & Analytics Forum 2013**

Return on Information: The New ROI

# Operational Analytics

*Extreme concurrent query volumes on real time information*



Business Users, Call Centers, Online Queries, etc  
*100s to 1,000+ Read and Update Queries*

Business Analysts

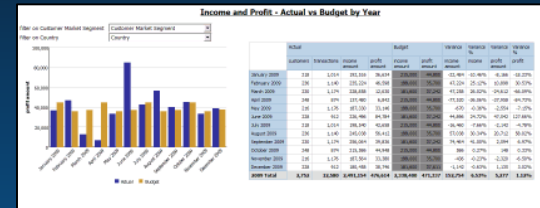


*Multiple, Concurrent Analytic Queries*



Data Warehouse

SALES  
 2010  
 2009  
 2008  
 2007  
 2006  
 2005



BI Reports and Analytics



# IBM Big Data Strategy: Operational Analytics

## *DB2 10 Distributed Platform Operational Element of our Big Data Story*

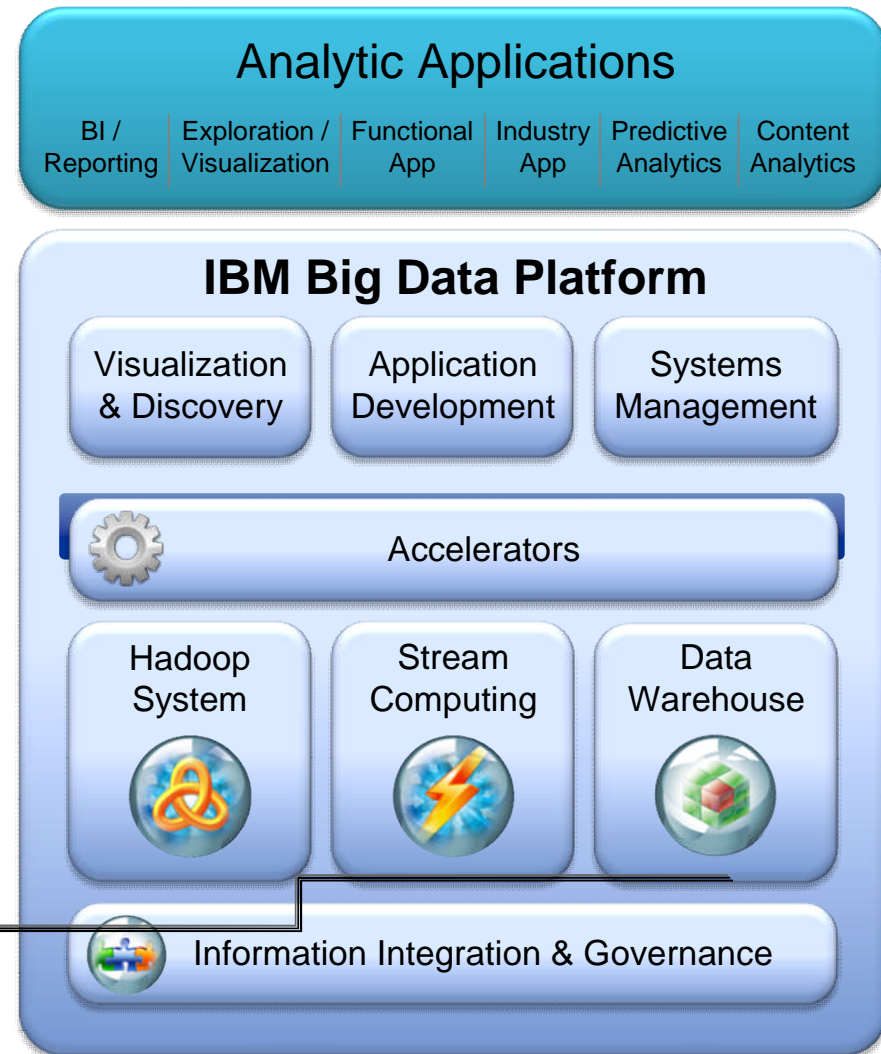
- Integrate and manage the full variety, velocity veracity, and volume of data
- Apply advanced analytics to information in its native form with the right type of system.

Relational analytical appliances announced:

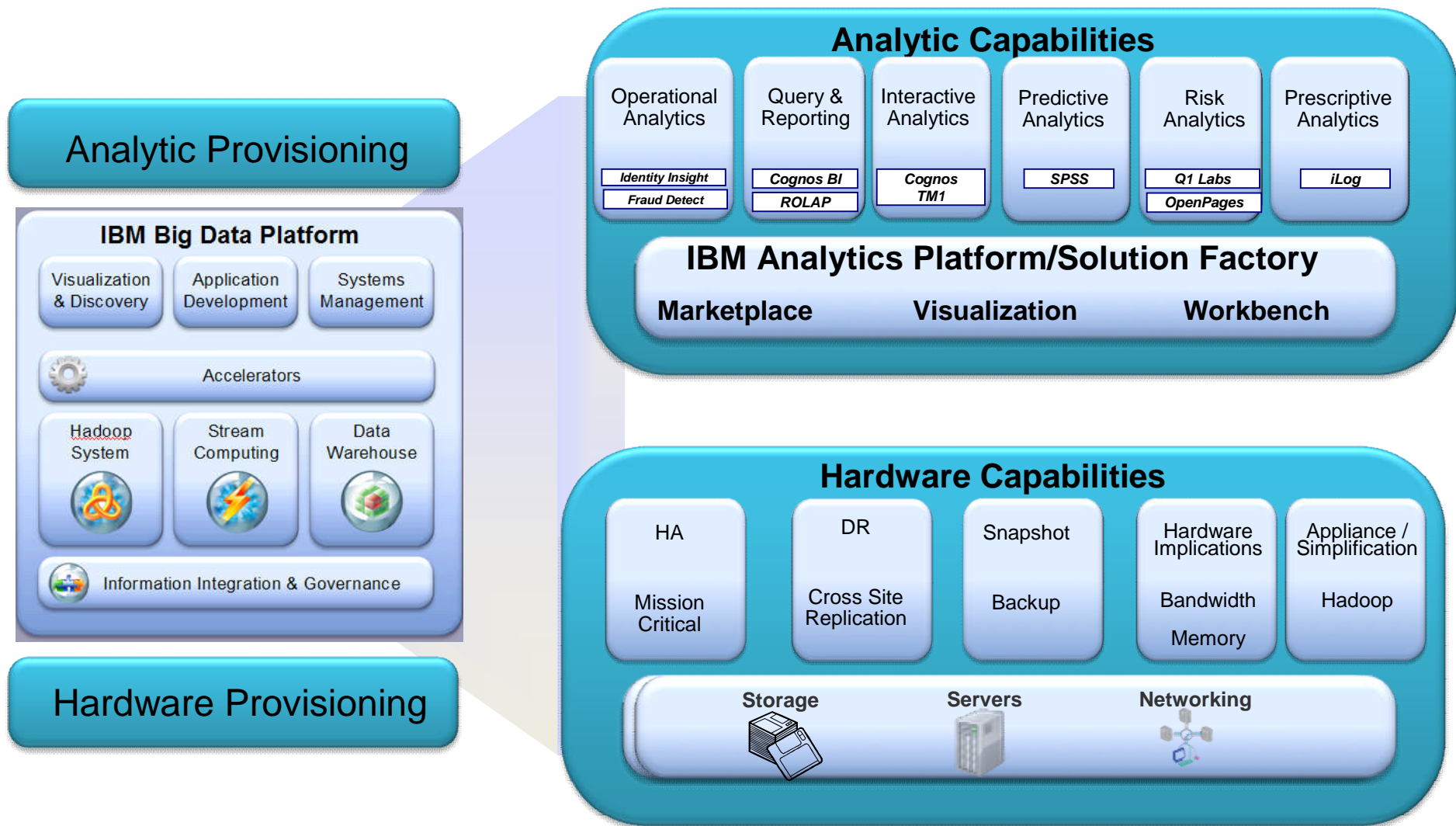
- PureData for Analytics  
*NextGen Netezza*
- **PureData for Operational Analytics**  
*NextGen IBM Smart Analytics System*

### **PureData for Operational Analytics is powered by DB2 10**

- Real-time, operational data warehousing
- Massively Parallel Processing (MPP)
- In-database mining & Hybrid OLAP
- Native XML & relational data warehouse
- Continual Data Ingest for Real-time Analysis
- Integration with Hadoop systems



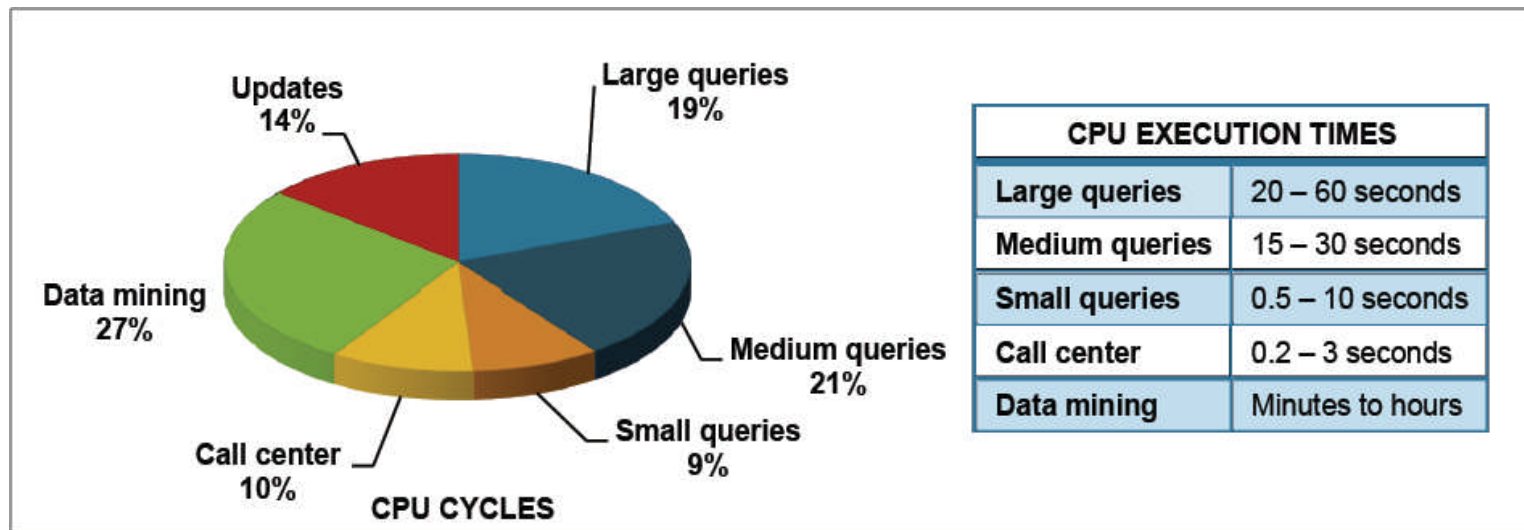
# IBM Vision: To deliver a Big Data Family that provides the integrated strengths of our Big Data Platform, Business Analytics, Optimized Storage and Global Services



# Performance and Mixed Workloads

## – Mixed Workloads

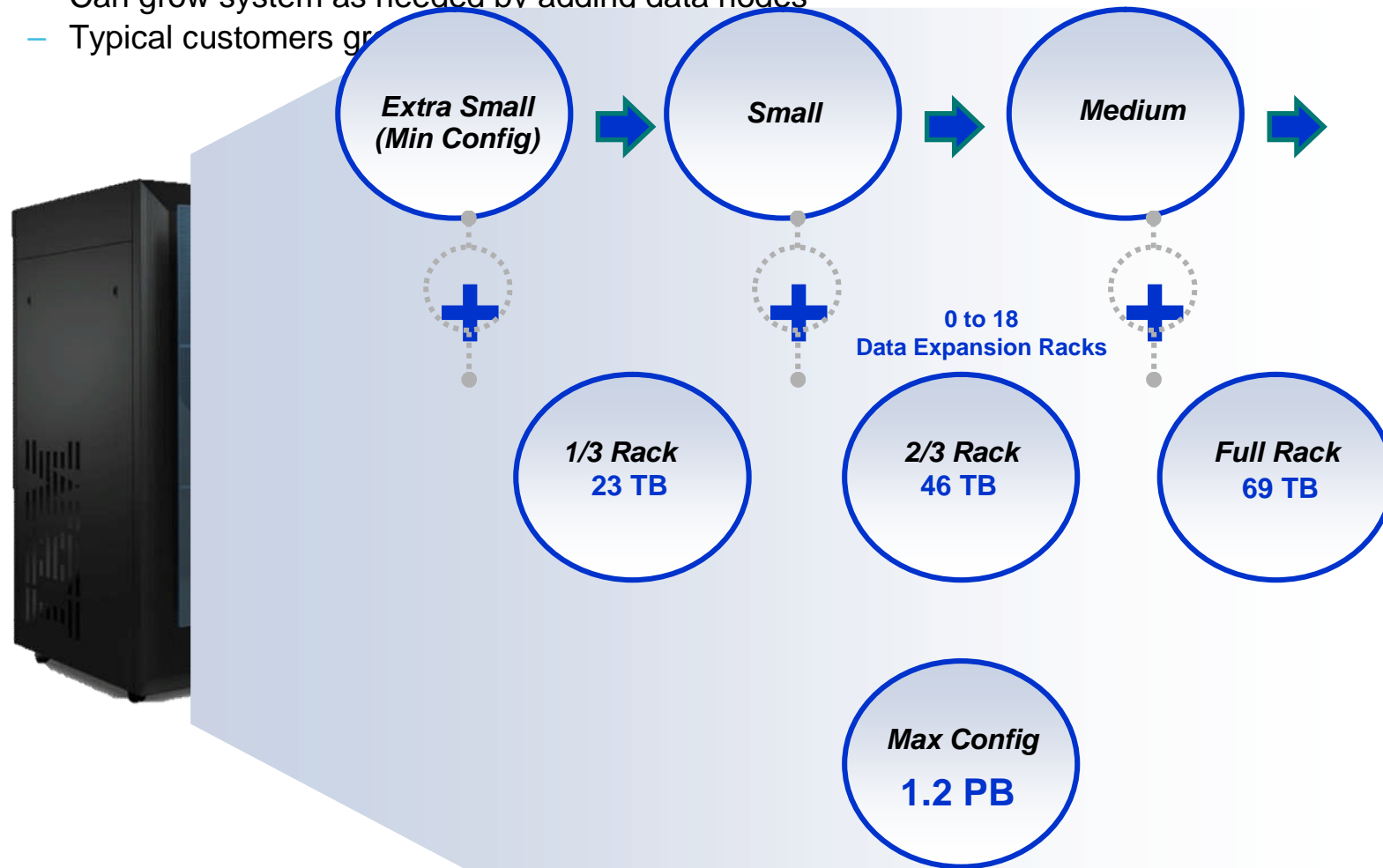
- Data Mining
- Call Center
- Small, large and extra large queries
- High concurrency
- Example:



# Proven Scalability

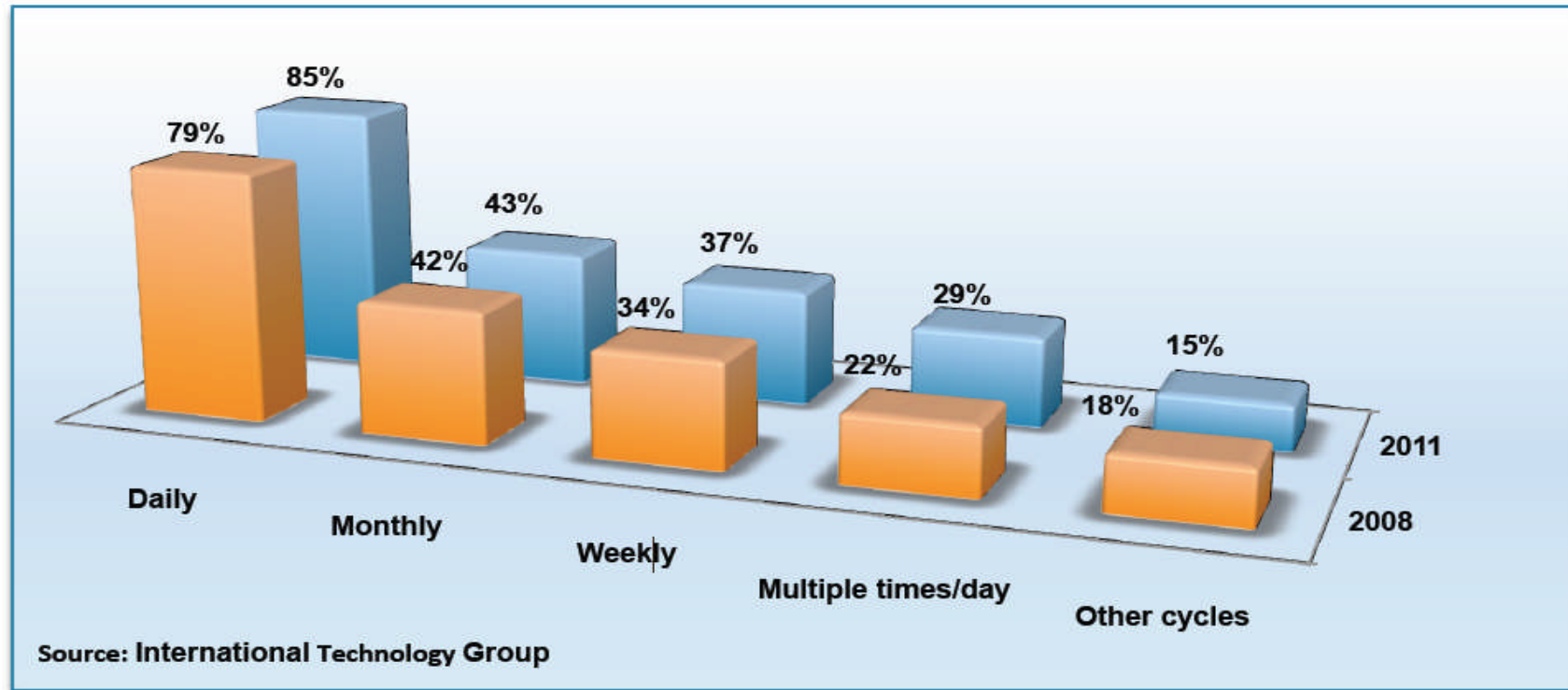
- **Value to clients**

- Can grow system as needed by adding data nodes
- Typical customers grow from 1/3 rack to full rack





# Availability and Concurrency



# IBM PureData System: Optimized exclusively for data services

## Optimized for data services:

- Transactional
- Analytics

## Expert integrated:

- Data platform
- Infrastructure
- Unified platform management
- Built-in expertise

# PureData

**New**



Data Platform

*Delivering Data Services*

***Workload optimized performance***

***Fast time-to-value***

***Integrated management***

***Single point of support***

***Automated updates for faster maintenance***

# IBM PureData System for Operational Analytics

*Optimized exclusively for operational analytic data workloads*

**PureData**  
System for Operational Analytics  
*Delivering data services for  
operational analytics*

## Speed

- Designed for 1000+ concurrent operational queries
- Continuous ingest of operational data
- MPP analytics (Massively Parallel Processing)

## Simplicity

- Fast time-to-value
- Automatic, policy-based data placement and workload management
- Integrated management and support

## Scalability

- Multiple sizes with data capacity up to a Petabyte

## Smart

- In-database analytics for leading applications
- Supports DB2 applications unchanged and Oracle Database apps with minimal change
- Clients have experienced cases of 10x storage space savings via Adaptive Compression

IBM Software

Information Management & Analytics Forum 2013

Return on Information: The New ROI

# Configuration Overview – IBM PureData System for Operational Analytics

## Server

- 730IOC 2U server
- Dual socket 16 CPU cores @ 3.xx GHz

## Storage:

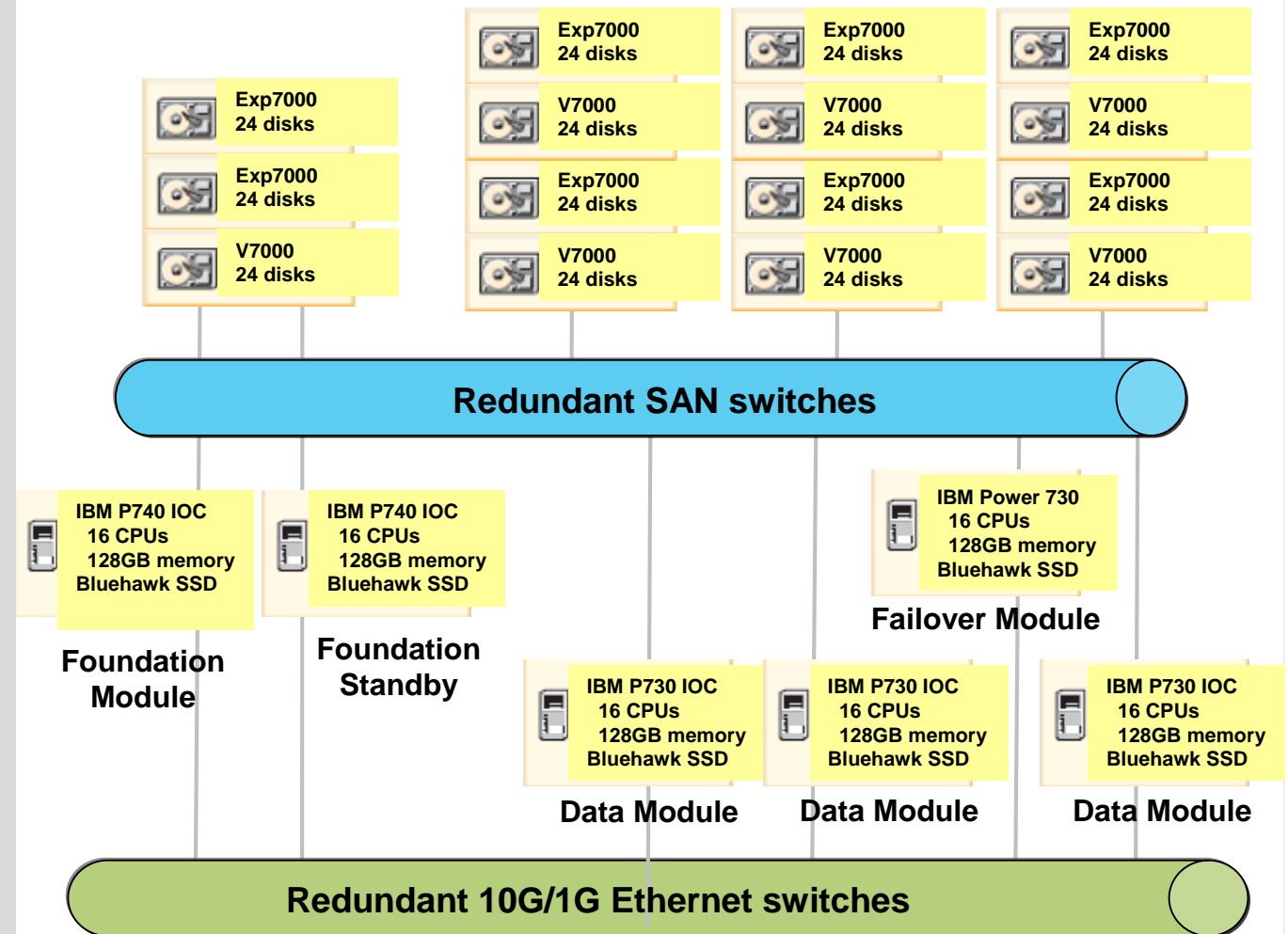
- V7000 controllers
- 96 disks per data module
- 8 Gbps FC
- Disks
  - 900GB 10k RPM
- BlueHawk SSD drawer
  - 6 SSD

## Software:

- AIX 7.1 TL1
- DB2 DB2 10/9.7 FP5

## DB2 logical design / module

- 8 LDPs, each with:
  - 2 3.55 GHz p7+ cores
  - 12 disks
  - 16 GB memory
- HA Group: 3 active servers with Standby



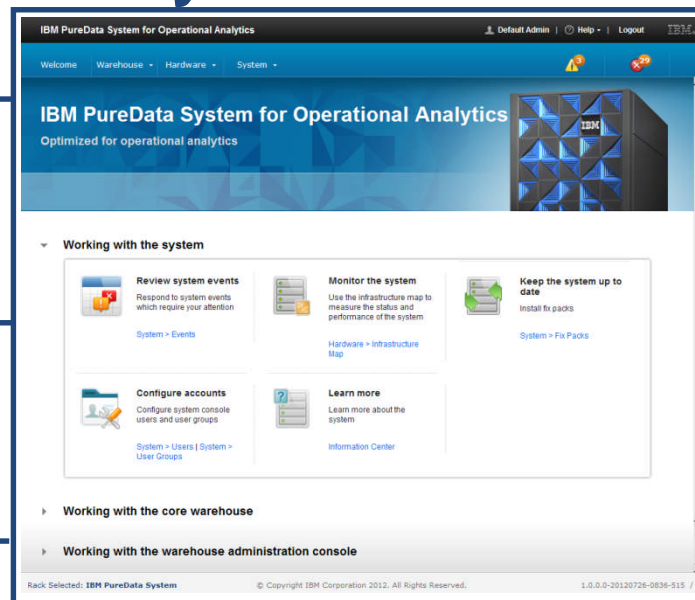
# PureData System for Operational Analytics System Console

## System Console

**Management**

**Monitoring**

**Maintenance**



- Provide “Easy to Use”, “Common” interfaces for management of Pure System Family

### Unified Appliance UI

**Alerting via SNMP & Email**  
(Single System Software Status, Hardware)

**Platform Monitoring**  
(Hardware and non workload specific)

**Maintenance Wizard**  
(Launch point integration)

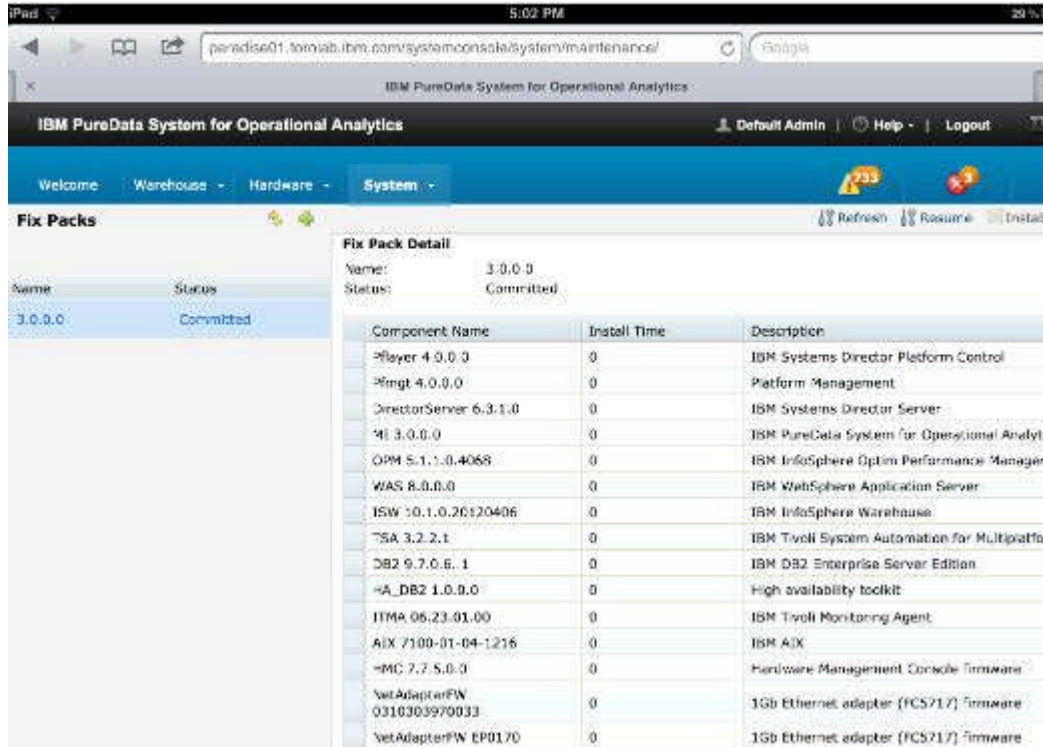
**User Authentication**  
(For console roles/access only)

**Workload Monitoring**  
OPM (SSO, Debranding, LIC)

**License Acceptance**  
On initial interface access



# Simplified maintenance with pre-integrated fixes



The screenshot shows the IBM PureData System for Operational Analytics console. The main navigation bar includes 'Welcome', 'Warehouse', 'Hardware', and 'System'. The 'Fix Packs' section is active, displaying a table of installed fix packs. The 'Fix Pack Detail' table is expanded, showing the following components:








Component Name	Install Time	Description
Pfayer 4.0.0.0	0	IBM Systems Director Platform Control
Pfmgmt 4.0.0.0	0	Platform Management
DirectorServer 6.3.1.0	0	IBM Systems Director Server
MI 3.0.0.0	0	IBM PureData System for Operational Analytics
OPM 5.1.1.0.4066	0	IBM InfoSphere Optim Performance Manager
WAS 8.0.0.0	0	IBM WebSphere Application Server
ISW 10.1.0.20120406	0	IBM InfoSphere Warehouse
TSA 3.2.2.1	0	IBM Tivoli System Automation for Multiplatfor
DB2 9.7.0.6.1	0	IBM DB2 Enterprise Server Edition
HA_DB2 1.0.0.0	0	High availability toolkit
ITMA 06.23.01.00	0	IBM Tivoli Monitoring Agent
AIX 7100-01-04-1216	0	IBM AIX
HMC 7.7.5.0.0	0	Hardware Management Console Firmware
NetAdapterFW 0310303970033	0	1Gb Ethernet adapter (FC5717) Firmware
NetAdapterFW EP0170	0	1Gb Ethernet adapter (FC5717) Firmware

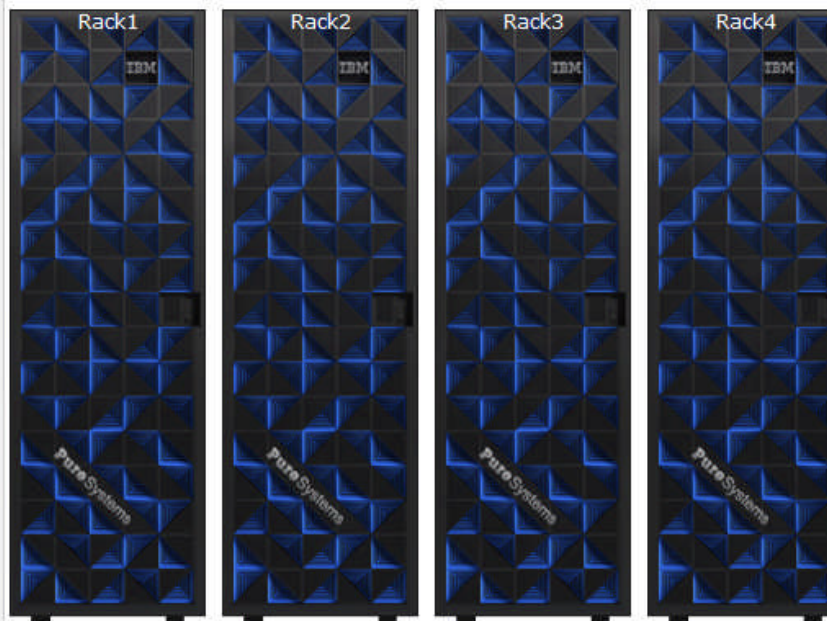
- Single point of contact for support
- Automated updates for faster maintenance
- All hardware firmware and OS software patches integrated and tested together at the factory

Infrastructure Map (Graphics View) Cluster

 Refresh  Switch to Tree View

 Default  Status  LED  Temperature  Performance Show Component Name

- Legend**
- All
  -  Critical
  -  Warning
  -  CPU utilization
  -  Memory utilization
  -  Storage utilization
  -  Hosts
  -  Volumes



Rack Selected: IBM PureData System

© Copyright IBM Corporation 2012. All Rights Reserved.

1.0.0.0-20120729-0848-744 /



Overview

View: **Historical Data** End Time: **08/08/12 16:54** Duration: **1 Hour**  Automatic Refresh 42 sec Baseline: **None**

Learn about the time controls

06/29/12 12:00 07/02/12 04:19 07/04/12 20:39 07/07/12 12:56 07/10/12 05:18 07/12/12 21:38 07/15/12 13:57 07/18/12 06:17 07/20/12 22:37 07/23/12 14:56 07/26/12 07:15 07/28/12 23:36 07/31/12 15:55 08/03/12 08:15 08/06/12 00:35 08/08/12 16:54

America/Chicago  
0008/12 1554 - 0008/12 1654

Overview Dashboard: **BCUDB** [All Members] **BCUDB** Disconnected All Members

**Health Summary**

System: ✔ Memory Usage ✔ Monitoring Status ✔ Data Server Status Database: -- Storage ✔ Locking ✘ Workload -- Recovery ⚠ Connections

**Data Server Runtime**

**OS Time Breakdown (%)**

**DB2 Time Breakdown (%)**

**Workload Throughput per minute**

**Open Connections**

**Load Average (Average Run Queue)**

**Paging (KB per minute)**

**Row Throughput per minute**

**Average Statement Response Time (ms)**

**Member Skew (Rows read / member)**

**Performance Focus** [Go to Buffer Pool and IO dashboard](#) **Top 3 SQL Statements** [Go to SQL Statements dashboard](#)

View by: **Buffer Pool Hit Ratio (%)**

Locking: **Buffer pool**

IO: **Buffered Page Reads per minute**

SQL Processing: **Physical reads**

**Asynchronous Page IO per minute**

**Direct Page IO per minute**

Direct writes

Direct reads

**Time per IO Operation (ms)**

Pooled write

Pooled read

By Elapsed Time By CPU Time **By Rows Read** By Lock Wait Time

Rows Read: 81 Executions: 6

```
SELECT t1.TBSP_ID, SUBSTR(t1.TBSP_NAME,1,20) as TBSP_NAME, t1.TBSP_TYPE, t1.TBSP_STATE
```

[View More](#)

Rows Read: 30

```
SELECT COLNAME, TYPENAME FROM SYSCAT.COLUMNS
```

# Pure Data Analytics

IBM Software

**Information Management & Analytics Forum 2013**

Return on Information: The New ROI



# IBM PureData System for Analytics

*The Simple Appliance for Serious Analytics*

## Built-in Expertise

- No indexes or tuning
- Data model agnostic
- Fully parallel, optimized In Database Analytics

## Integration by Design

- Server, Storage, Database in one easy to use package
- Automatic parallelization and resource optimization to scale economically
- Enterprise-class security and platform management

## Simplified Experience

- Up and running in hours
- Minimal ongoing administration
- Standard interfaces to best of breed Analytics, BI, and data integration tools
- Built-in analytics capabilities allow users to derive insight from data quickly
- Easy connectivity to other Big Data Platform components



IBM Software

Information Management & Analytics Forum 2013

Return on Information: The New ROI



# PureData System for Analytics



## Transforms the User Experience

---

- ✓ **Purpose-built analytics engine**
- ✓ **Integrated database, server and storage**
- ✓ **Standard interfaces**
- ✓ **Low total cost of ownership**

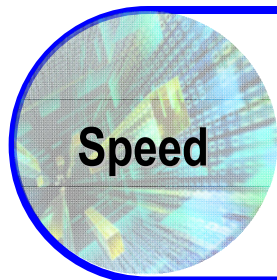
**Speed:** 10-100x faster than traditional system

**Simplicity:** Minimal administration and tuning

**Scalability:** Peta-scale user data capacity

**Smart:** High-performance advanced analytics

# What Makes PureData System for Analytics Different?



## Speed

**Up to 2000X** faster than before  
Growing by **30% every month**

*"Netezza has allowed us to reduce the complexity of regulatory reporting and processing of exchange data from days down to minutes."*



## Simplicity

Up and running **6 months**  
before having any training  
**200X** faster than Oracle system  
ROI in less than **3 months**

*"Allowing the business users access to the Netezza box was what sold it."*

- Steve Taff, Executive Dir. of IT Services

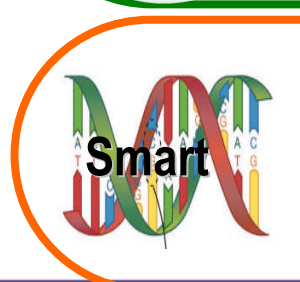


## Scalability

**1 PB** on Netezza  
**7 years** of historical data  
**100-200%** annual data growth

*"NYSE ... has replaced an Oracle IO relational database with a data warehousing appliance from Netezza, allowing it to conduct rapid searches of 650 terabytes of data."*

- ComputerWeekly.com



## Smart

SUNY Buffalo researchers reduced the time to perform **quintillions of computations** from **27 hours** to **12 minutes**

*"Once we had the data on Netezza we were able to do the same analysis and much more complex analysis in minutes. The research draws on medical records, lab results, MRI scans, and patient surveys."*

- Dr. Murali Ramanathan, SUNY Buffalo

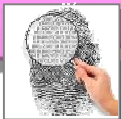


# Pre-Built In-Database Analytics



## Statistics

- Descriptive Statistics+
- Distance Measures\*
- Hypothesis Testing\*
- Chi-Square & Contingency Tables\*
- Univariate & Multivariate Distributions+
- Monte Carlo Simulation\*



## Data Mining

- Association Rules+
- Clustering+
- Feature Extraction+
- Discriminant Analysis\*



## Transformation

- Data Profiling / Descriptive Statistics+
- General Diagnostics
- Statistics+
- Sampling
- Data prep



## Predictive

- Linear Regression+
- Logistic Regression+
- Classification
- Bayesian
- Sampling
- Model Testing



## Time Series

- Autoregressive+
- Forecasting\*



## Geospatial

- Geospatial Data Type
- Geometric Functions
- Geometric Analysis



## Mathematical

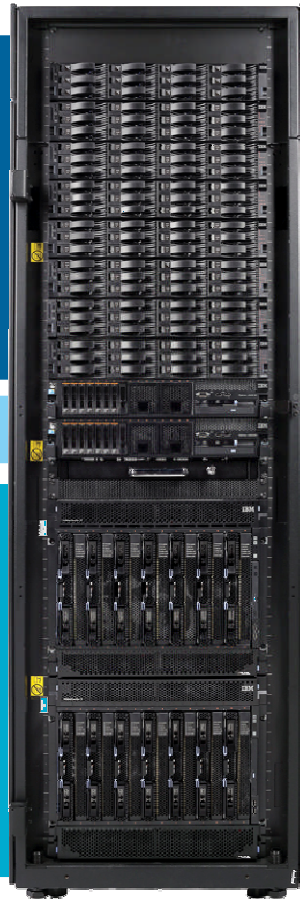
- Basic Math\*
- Permutation and Combination\*
- Greatest Common Divisor and Least Common Multiple\*
- Conversion of Values\*
- Exponential and Logarithm\*
- Gamma and Beta Functions
- Matrix Algebra+
- Area Under Curve\*
- Interpolation Methods\*

\* Fuzzy Logix DB Lytix capabilities

+ Netezza Analytics and Fuzzy Logix DB Lytix capabilities



# PureData System for Analytics Hardware Overview



- 8 Disk Enclosures
- 96 1TB SAS Drives (4 hot spares)
- RAID 1 Mirroring

- **Wire speed Compression-Decompression engine**
- **Smart Data filtering at storage level.**
- **Advanced Caching Techniques ensuring minimal data movement**

- 14 PureData for Analytics S-Blades™
- 2 Intel Quad-Core 2+ GHz CPUs
- 4 Dual-Engine 125 MHz FPGAs
- 24 GB DDR2 RAM
- Linux 64-bit Kernel

Scales from  
¼ Rack to 10 Racks

32 TB to 1.2 PB of  
User Data

- **User Data Capacity:** 128 TB\*\*
- **Data Scan Speed:** 145 TB/hr\*\*
- **Load Speed (per system):** 5+ TB/hr

- **Power Requirements:** 7.6 kW
- **Cooling Requirements:** 7.8 kW

IBM Software

**Information Management & Analytics Forum 2013**

Return on Information: The New ROI

\*\* : 4X compression assumed



# DB2 10 Key Technologies

IBM Software

**Information Management & Analytics Forum 2013**

Return on Information: The New ROI

# Infosphere warehouse 10

*Taking High Performance Operational Warehousing to the next level*

**NEW**

## Over 3X Performance Improvements

- Zig-Zag Join technology for Hi-Performance BI Queries
- New Multi-threading Enhancements for Higher efficiency and query concurrency
- Faster Index Access for enhanced Query performance

**NEW**

## Multi-Temperature Data Management

*Increase Ability to Meet SLAs; Postpone Hardware Upgrades*

**NEW**

## Time Travel Query

*Easily Analyze Historical Trends and Predict Future Demand*

**NEW**

## Row and Column Access Control

*Easy Compliance with Privacy and Sensitive Data Requirements*



ITGAIN

"The multi-temperature database management feature of DB2 V10.1 is great because the hardware world is not just RAM and hard disks. There are many types of storage options with different I/O speeds and prices....." –Thomas Kalb, CEO ITGAIN GmbH

"The introduction of the Time Travel Query feature greatly simplifies the development and maintenance of time-aware application code, resulting in up to **9 times coding cost savings.**"

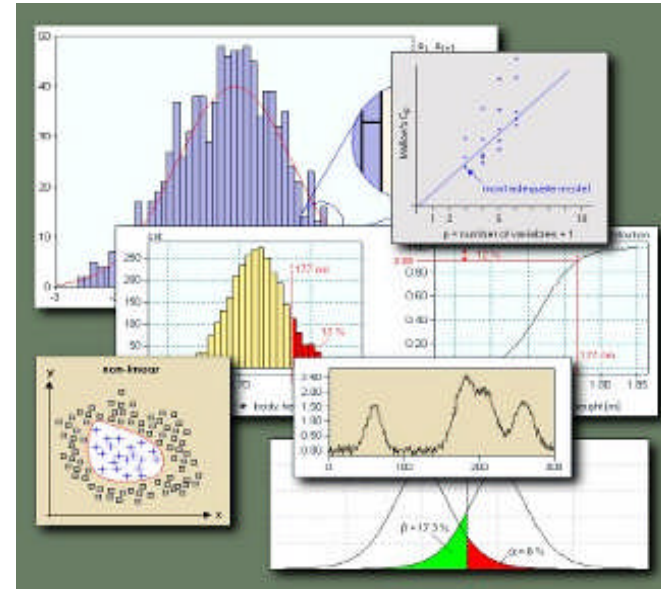
–Pradeep Naik, Wipro Technologies

"Row and Column Access Control helps us to improve data confidentiality and security in production environments."

–Jessica Tatiana Flores Montiel, DAFROS Multiservicios

# Index Management Re-defined

- Jump Scan
- Smart Index Pre-fetching
- Smart Data Pre-fetching
- Predicate Evaluation Avoidance
- **Higher performance**
  - Faster index performance
- **Lower costs**
  - Fewer indexes to maintain
  - Dramatic reduction in index reorgs



“Jump Scan optimizes buffer usage by 75 to 80%, resulting in very good improvement in overall performance and saving the CPU cycles.”  
—Shanmukhaiah D, Cognizant Technology Solutions.

# Breakthrough Savings with Adaptive Compression

*Lower Storage Costs; Lower Administration Costs*

- Higher performance
  - More efficient operation
  - Reduced maintenance windows
- Lower costs
  - Postpone upcoming storage purchases
  - Lower ongoing storage needs
  - Easier administration with reduced need for table re-orgs



“Our migration from Oracle Database to DB2 resulted in a 40% storage savings. Upgrading to DB2 9.7 and index compression brought our average savings to 57%. Now adaptive compression brings our **average savings to 77%**, dramatic savings!”

—Andrew Juarez, Lead SAP Basis / DBA, Coca Cola Bottling Company.



# Continuous Data Ingest for Operational Workloads

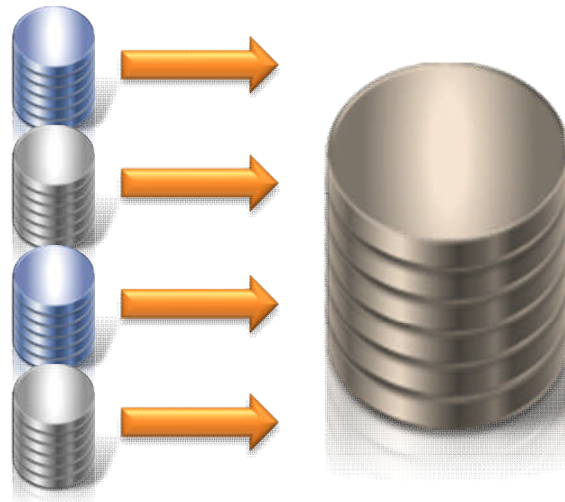
Continuous Data Ingest - new feature enabled in the underlying database

Optimized, continuous loading of data, not just periodically

**Reduced downtime & Up-to-date data helps feed the warehouse**

**Leads to faster, accurate tactical decision making.**

**Up to 39X Faster Data Loads !**



“Replacing DB2 Import with Continuous Data Ingest we reduced data ingest time by 94% for a table with 1.7GB of data.”

—Chunguang Yuan, China MinSheng Banking Corp.



# DB2 Workload Management Enhancements

*Increase Ability to Meet SLAs; Postpone Hardware Upgrades*



- CPU limits

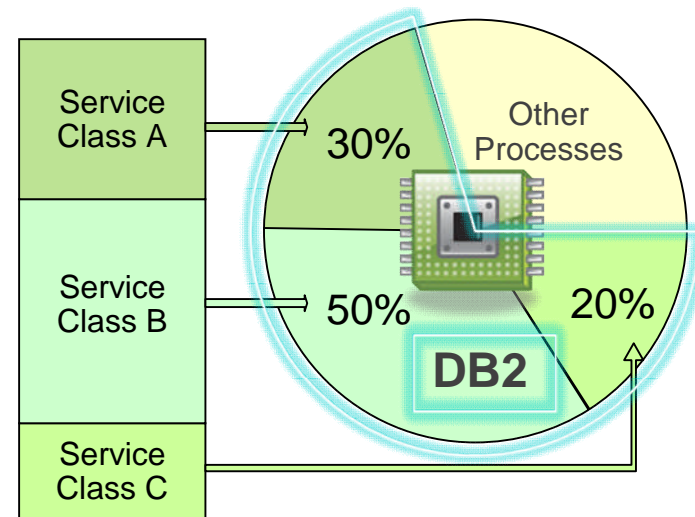
- % of resources DB2 can consume

- CPU shares

- % of limit a service class can consume
  - Hard shares and soft shares

- Higher performance

- Prioritize important workloads
  - More efficient distribution of workloads



“DB2 10 Workload Manager offers nearly **30% improvement** in performance and **40% savings** in cost by using system resources to efficiently manage priority for demanding queries.”

–Juvenal Garcia Cuevas, SBRT Computing

## Time Travel Query

*Easily Analyze Historical Trends and Predict Future Demand*



- Enable better business decisions
- Historical trend analysis
- Higher performance
  - Native support for fast performance
- Lower costs
  - Eliminate need to maintain and update custom temporal implementations
  - Easy to administer (simply turn on for any table)



“The introduction of the Time Travel Query feature greatly simplifies the development and maintenance of time-aware application code, resulting in up to **9 times coding cost savings.**”

–Pradeep Naik, Wipro Technologies

# Row and Column Access Control

## *Easy Compliance with Privacy and Sensitive Data Requirements*



- Fine-grained access control (policy-driven)
  - Hide rows from unauthorized users
  - Mask the value of columns for unauthorized users
- Does not require classification
- Higher performance
  - Less data duplication than using “Views” to mask data
  - More secure than using “Views” to mask data
- Lower cost
  - Easier to implement and maintain
  - Easier compliance with privacy and sensitive data requirements
  - Easier to maintain than using application code to mask data



“Row and Column Access Control helps us to improve data confidentiality and security in production environments.”

–*Jessica Tatiana Flores Montiel, DAFROS Multiservicios*

# Real Time Analytics Value and Case Study

IBM Software

**Information Management & Analytics Forum 2013**

Return on Information: The New ROI

Large telecommunications provider



**Billion's of calls  
made by 100M  
subscribers each**

*Continuous, 24 x 7, feeding of  
data into the Enterprise Data  
Warehouse*

*165 database partitions located  
across a cluster of 22 AIX servers*

*Improved customer service  
productivity by 42%*





# 3 UK

*Telecommunications company*



## Challenge

- **Needed to deliver a network intelligence solutions to understand:**
  - **Subscriber experience for both voice calls and mobile broadband**
  - **Network performance**
  - **Subscriber behaviour / segmentation**
  - **Churn & predict churn**

## Solution

- **IBM Telecom Industry Data Model**
- **IBM InfoSphere Warehouse software and the Balanced Configuration Unit (renamed IBM Smart Analytics System) provide a comprehensive BI platform for customer information**
- **InfoSphere DataStage and QualityStage software to standardize data and eliminate data inconsistencies**

## Business Benefits

- **Right time Data Availability**
- **Can understand the Voice and MBB Usage by a number of dimensions**
  - **By Dropped Call Rate (Voice)**
  - **By UL Volume, DL Volume, UL Throughput & DL Throughput (MBB)**
  - **By Subscriber. Location, Time of Day, Device**
- **Subscribers are segmented using the in-built data mining capabilities against the usage patterns. (Customer demographic data is not the answer.)**
- **Behavioural Segmentation**
  - **Understand what subscribers have done before they have left us, and identify trends and patterns**
  - **Used to detect Fraudulent patterns - faster than the existing Fraud System**

# Large North American Financial Institution

Identifying patterns in real-time to help prevent fraud

## The Need:

To help it fight fraud and identify money-laundering schemes, a large North American financial institution implemented enterprise data warehouse solution.

Because customer data was stored in multiple source systems across the bank, with names and dates in ten or more different formats, it made it more difficult to access and interpret the necessary information.

## The Solution:

IBM helped the bank create an enterprise data warehouse (EDW) that delivers consistent, universal data in real time to battle financial crime.

The EDW integrates information about every customer and every transaction from diverse sources and provides real-time analytics to help rapidly identify anomalous behavior suggestive of financial crime.

***“The 360-degree view of the customer that the EDW provides makes us a truly customer-centric organization, helping us to understand our customers, improve our service to them and secure their protection against fraud.”***

***— Director, Enterprise Architecture,  
A Large North American Financial  
Institution***

## Benefits:

- Ability to identify patterns in real-time helps the bank identify large scale attacks, which may affect multiple customers in several locations, and take action quickly
- Implementation of a single crime fighting infrastructure has helped to reduce the number of fraud attacks and financial losses
- Integration of information reduced 20+ data marts to one for significant economies of scale and reduced technology investment and maintenance costs

# Indian National Bank

*Mining transactions by the terabyte gives managers better insights,  
control and decision making*

## The need:

One of the largest public sector banks in India wanted to gain a competitive advantage in a shifting business dynamic and regulatory environment.

They needed to mine the vast number of transactions they processes daily to give management better insights for decision making as well as greater knowledge and control over the business.

## The solution:

They asked Tata Consultancy Services to build an Enterprise wide Data Warehouse (EDW) using IBM hardware, software and services.

They pulled 1.5 terabytes of data together from many remote systems to give management ready access to vital business information.

By mining and leveraging the data, they achieved operational efficiencies, enhanced decision support, and increased profitability.

*“To continue winning in the marketplace, we need to be fast, agile and incisive – based on insights from really knowing our business. We have that now.”*

## Benefits:

- Gaining valuable new insights into customer behavior, retail banking operations and profitability by mining all 1.5 terabytes of EDW data.
- they gains valuable new insights into customer behavior as well as new knowledge and control over its business.
- \$2 million in savings by automating the Management Information System (MIS).
- A value of \$180 million from new customer leads.

# Communities

- **On-line communities, User Groups, Technical Forums, Blogs, Social networks, and more**
  - Find the community that interests you ...
    - Information Management [bit.ly/InfoMgmtCommunity](http://bit.ly/InfoMgmtCommunity)
    - Business Analytics [bit.ly/AnalyticsCommunity](http://bit.ly/AnalyticsCommunity)
    - Enterprise Content Management [bit.ly/ECMCommunity](http://bit.ly/ECMCommunity)
- **IBM Champions**
  - Recognizing individuals who have made the most outstanding contributions to Information Management, Business Analytics, and Enterprise Content Management communities
    - [ibm.com/champion](http://ibm.com/champion)