# 빅 데이터 플랫폼별로 골라쓰는 분석용 데이터 시스템

**이지은 실장**

**한국IBM 소프트웨어그룹**

# Agenda

- Analytic Challenges in Big data era
- Delivering Deep Analytics
  - IBM PureData system for Analytics
- Delivering Operational Analytics
  - IBM PureData system for Operational Analytics
- Big data platform to accelerate analytics

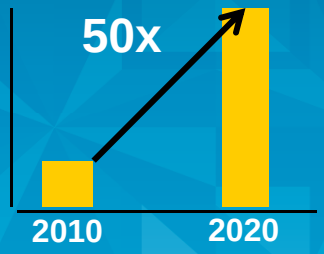
# Agenda

- Analytic Challenges in Big data era
- Delivering Deep Analytics
  - IBM PureData system for Analytics
- Delivering Operational Analytics
  - IBM PureData system for Operational Analytics
- Big data platform to accelerate analytics – different analytics

# Today's big data challenges for both transactions and analytics are increasing demands on data systems

*Mobile* *Commerce* *Social* *Analytics* *Big Data* *Cloud*

Increasing
**Volume of data**
*requires growing capacity*




**35 ZB**

by 2020

Increasing
**Velocity of data**
*requires higher performance*

**Millions of transactions per second**

Telco subscriber activity logging

Increasing
**Variety of data**
*requires new techniques*

**Billions of devices & sensors**

Smart Meters, RFIDs, GPS…

**A smarter approach to meeting data challenges is required to:**

**Reduce complexity ▪ Accelerate time to value ▪ Improve IT economics**

# Smarter Analytics should be your goal

CIOs rank Analytics as the

**#1 factor**

Contributing to an organization's competitiveness.[1]

Organizations that embrace analytics are more than

**2X**

as likely to outperform their Peers.[2]

Financial outperformers are

**64%**

more likely to use analytics to evaluate talent supply and demand on an ongoing basis.[3]

Enterprises that apply advanced analytics have

**33%**

More revenue Growth and

**12X**

more profit growth.[4]

[1] IBM CIO Study 2009
[2] IBM IBV/MIT Sloan Management Review Study 2011
[3] IBM CHRO Study 2010
[4] IBM CFO Study 2010

# Achieve Smarter Analytics by using all types of analytics against all types of data

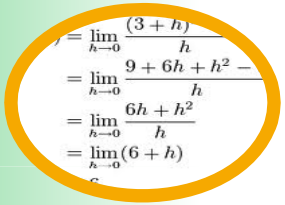
**Smarter Analytics**

*Operational reports*

*Analytic reports*

*Alerts*

*Statistical Analysis*



$$= \lim_{h \to 0} \frac{(3+h)}{h}$$
$$= \lim_{h \to 0} \frac{9 + 6h + h^2 -}{h}$$
$$= \lim_{h \to 0} \frac{6h + h^2}{h}$$
$$= \lim_{h \to 0} (6 + h)$$

*Spatial Analysis*

*Forecasting*

*Predictive Modeling*

*Optimization*


*Social Analytics*

*Web Analytics*

*Transactional & Application Data*

*Documents*

*Enterprise Content*

*Machine & Sensor Data*

*Social Data*


*Web Data*

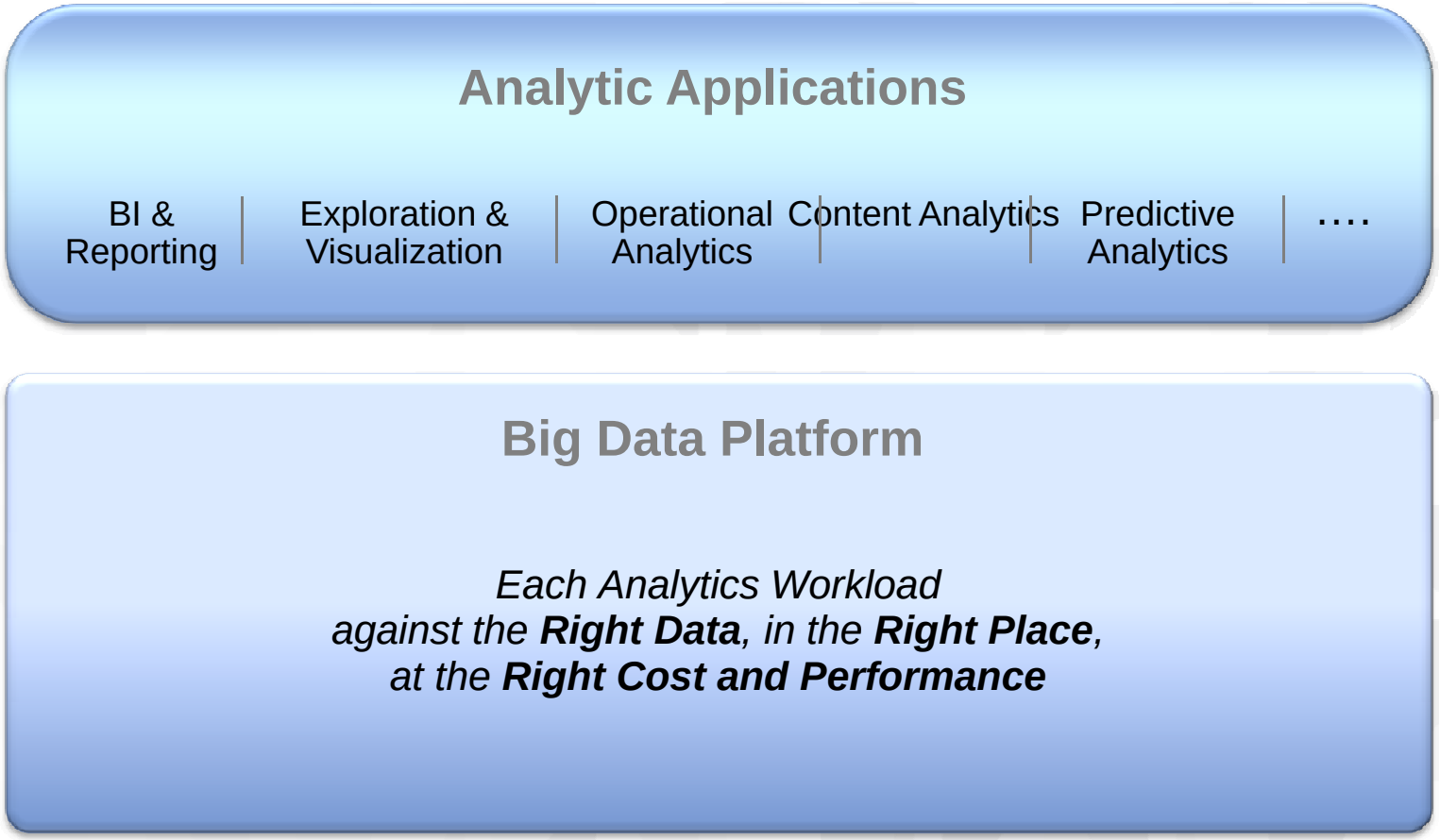
# Challenges to overcome

| *Do you face these challenges:* | *Then you need a platform that provides:* |
| --- | --- |
| Difficulty adding new data or analytic capability | Increased Agility |
| Lack of analytical insight | Accelerated Time to Value |
| Broad spectrum of workload and SLA requirements | Fit for Purpose Solutions |
| Growing data volume, variety and velocity | Tools for gaining insight from Big Data |
| Complicated system lifecycles | Reduced Complexity |
| Administration complexity | Simplicity |
| Growing costs of IT | Increased Efficiency |

# Today's Analytic Challenges Means Thinking Differently about Architecture

- Data systems optimized for analytic workloads
- Support for different types of analytics
- Simplicity as a design requirement
- Agility as a design requirement

# The Answer? A Big Data Platform



# PureData


## *for Analytics*

*Optimized system delivering data services for analytics*

## *for Operational Analytics*

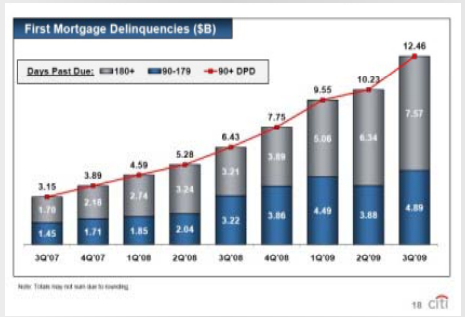
*Optimized system delivering data services for operational analytics*

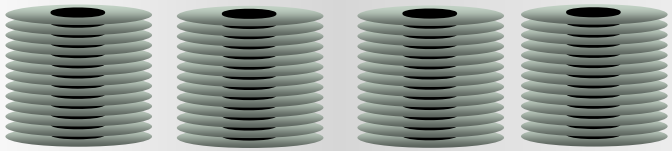
PureSystems

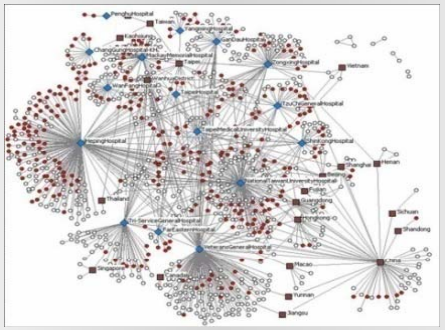
# Agenda

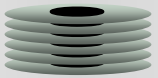
- Analytic Challenges in Big data era
- Delivering Deep Analytics
  - IBM PureData system for Analytics
- Delivering Operational Analytics
  - IBM PureData system for Operational Analytics
- Big data platform to accelerate analytics

# Big Data Meets Deep Analytics



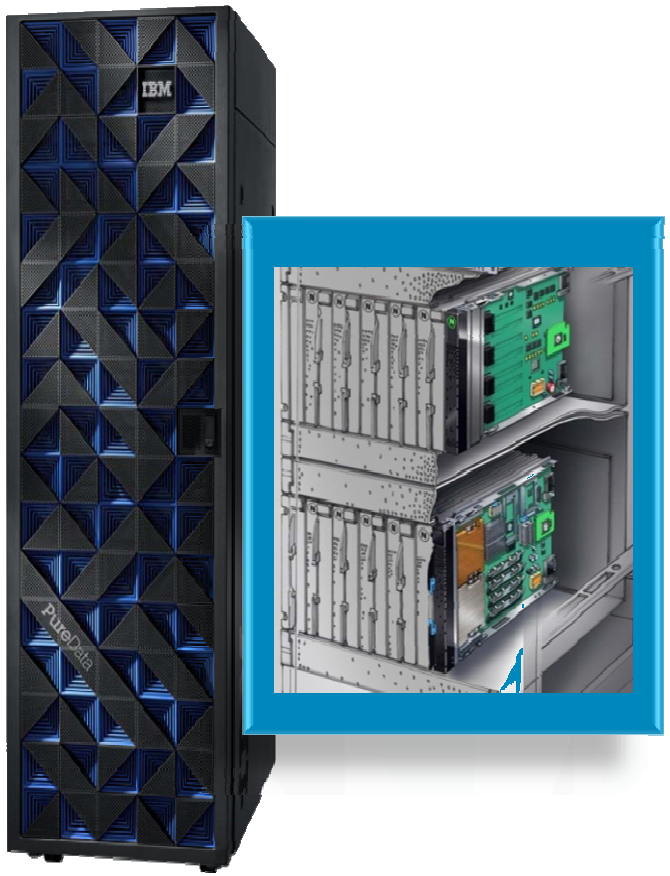


## Analytics without constraint

# IBM PureData System for Analytics

**New generation Netezza appliance**

*분석 어플리케이션을 위한 최적화된 성능 및 단순성을 제공하는 네티자 기술*




DB,서버, 스토리지 를 통합하여

모든 구성을 최적화

- High speed data ingest
- High speed analysis
- No indexing
- No tuning
- No administration

**Speed ! Simplicity ! Scalability ! Smart !**

# IBM PureData System for Analytics - Inside

## *Optimized Hardware + Software*

- Hardware accelerated AMPP
- Purpose-built for high performance analytics
- Requires no tuning

## *SMP Hosts*

- SQL Compiler
- Query Plan
- Optimize
- Admin

## *Disk Enclosures*

- User data, mirror, swap partitions
- High speed data streaming

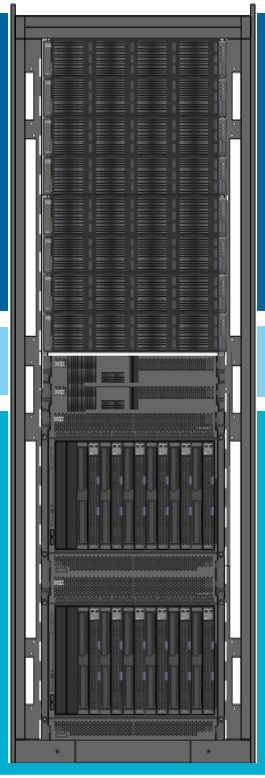
## *Snippet Blades ™*

- Hardware-based query acceleration with FPGAs
- Blistering fast results
- Complex analytics executed as the data streams from disk

# IBM PureData System for Analytics – HW overview

Scales from
¼ Rack to 10 Racks

32 TB to 1.2 PB of
User Data



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

- 2 Hosts (Active-Passive):
- 2 Quad-Core Intel 2.6 GHz CPUs
- 7x146 GB SAS Drives
- Red Hat Linux 5 64-bit

- 14 Sparta 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

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

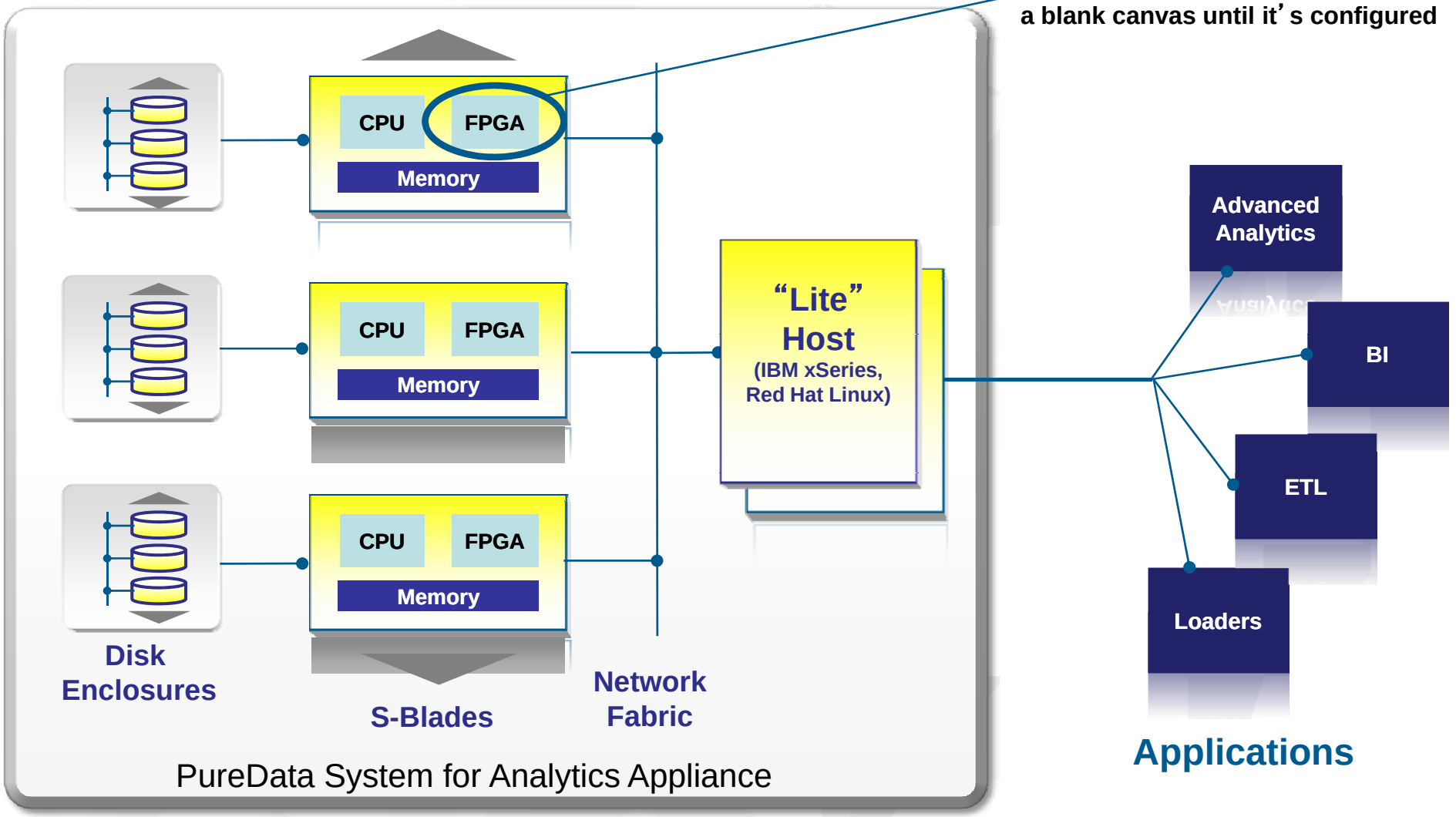
| | |
|---|---|
| • Power Requirements: | 7.6 kW |
| • Cooling Requirements: | 7.8 kW |

** 4X compression assumed

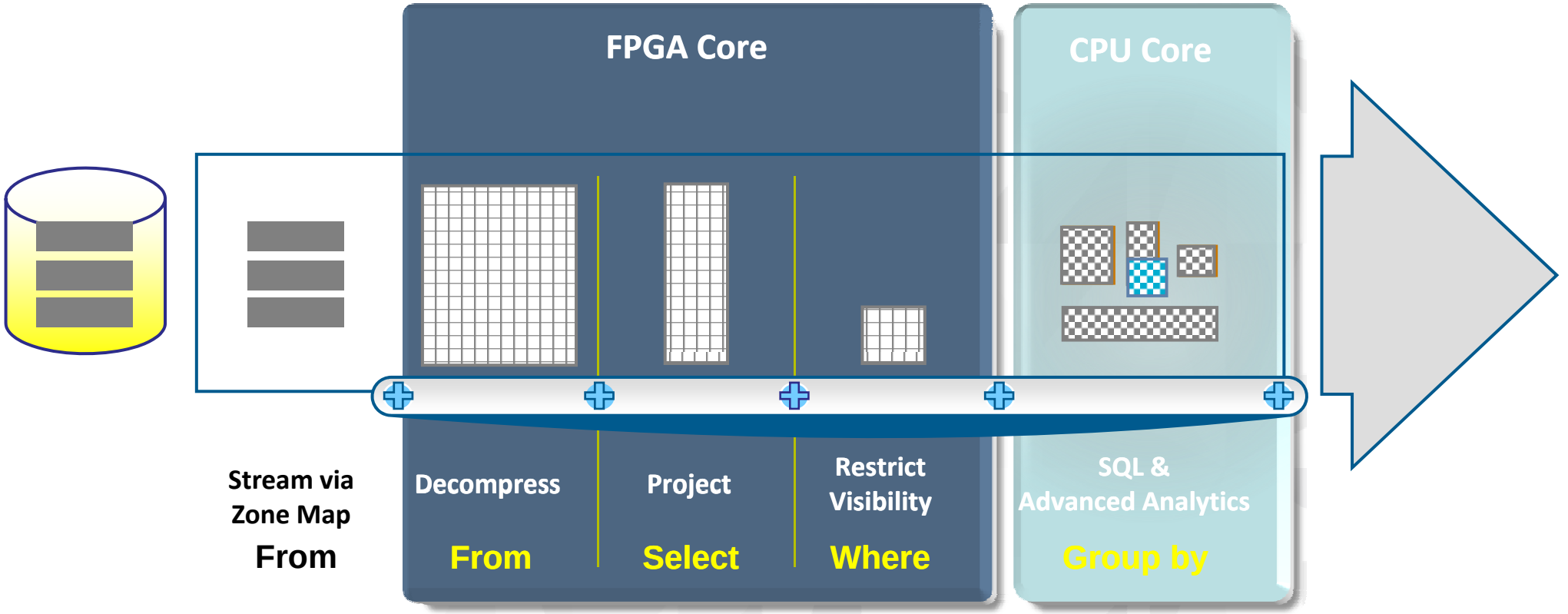
# IBM PureData System for Analytics - AMPP Architecture

*: SMP 와 MPP 의 조합*

***Field Programmable Gate Array*** **= a blank canvas until it's configured**



# IBM PureData System for Analytics – S-Blade Data stream Processing




```
Select State, Age, Gender, count(*) From MultiBillionRowCustomerTable Where BirthDate between ‘01/01/1960’
And State in (‘FL’, ‘GA’, ‘SC’, ‘NC’) Group by State, Age, Gender Order by State, Age, Gender
```

# IBM PureData System for Analytics – Enhancement

New generation Netezza appliance

## 성능

- 18X increase in concurrency for tactical queries
- Average 3X faster for mixed workloads
- Up to 100 queries/second micro analytic workloads
- Page Level Zone Maps reduces  disk scanning

## 관리 및 효율성

- More than half a dozen performance improvements in:
  - Optimizer efficiency
  - Memory management
  - Communications protocols
  - Workload management
- Faster, Better, and completely transparent to the end-user

## 장애 대응력 및 안정성

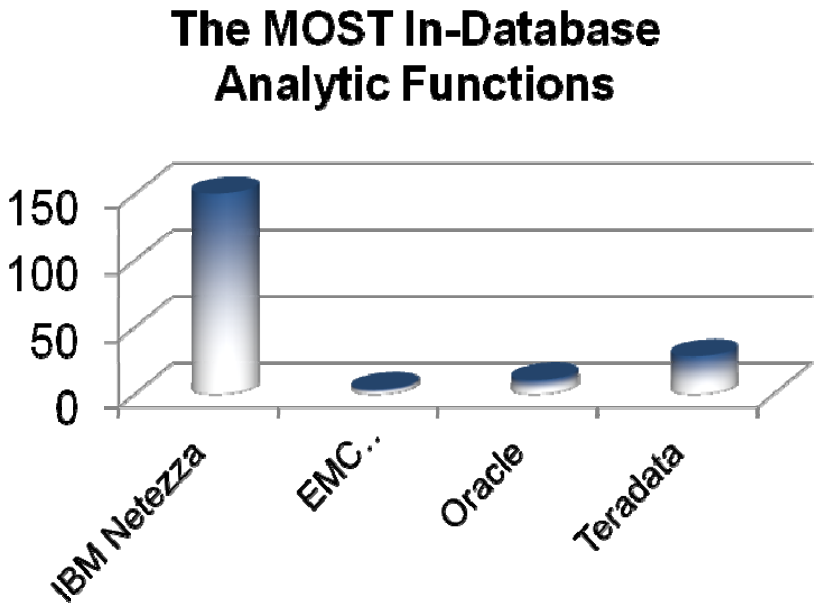
- Failover now leverages an extra blade in the chassis
- Multi-pathing for Dynamic failover allowing automatic re-direction when failure occurs

# IBM PureData System for Analytics – Integrated by Design

*Netezza In-Database Analytics 2.0*

**Netezza**
**In-Database Analytics**

- Transformations
- Mathematical
- Geospatial
- Predictive
- Statistics
- Time Series
- Data Mining




- ✓ 데이터의 이동 없음
- ✓ 심도 있고 광범위한 분석
- ✓ 고성능 병렬 처리

# Agenda

- Analytic Challenges in Big data era
- Delivering Deep Analytics
  - IBM PureData system for Analytics
- Delivering Operational Analytics
  - IBM PureData system for Operational Analytics
- Big data platform to accelerate analytics

# Operational Analytics

*Extreme concurrent query volumes on real time information*

Business Users, Call Centers, Online Queries, etc

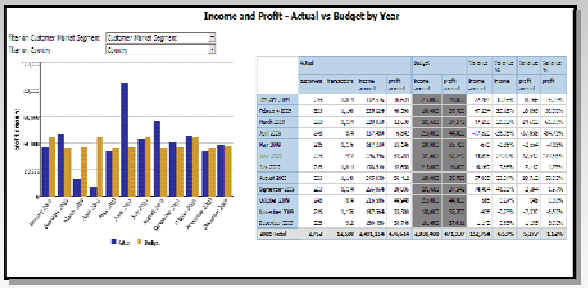
*100s to 1,000+ Read and Update Queries*

Business Analysts

*Multiple, Concurrent Analytic Queries*

SALES

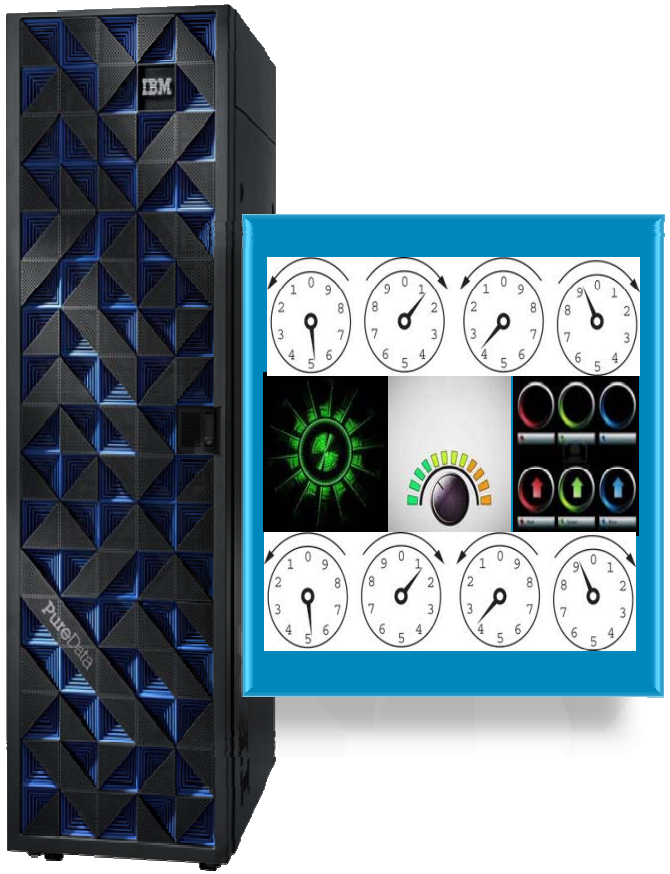
2010
2009
2008
2007
2006
2005




Data Warehouse

BI Reports and Analytics

# IBM PureData System for Operational Analytics

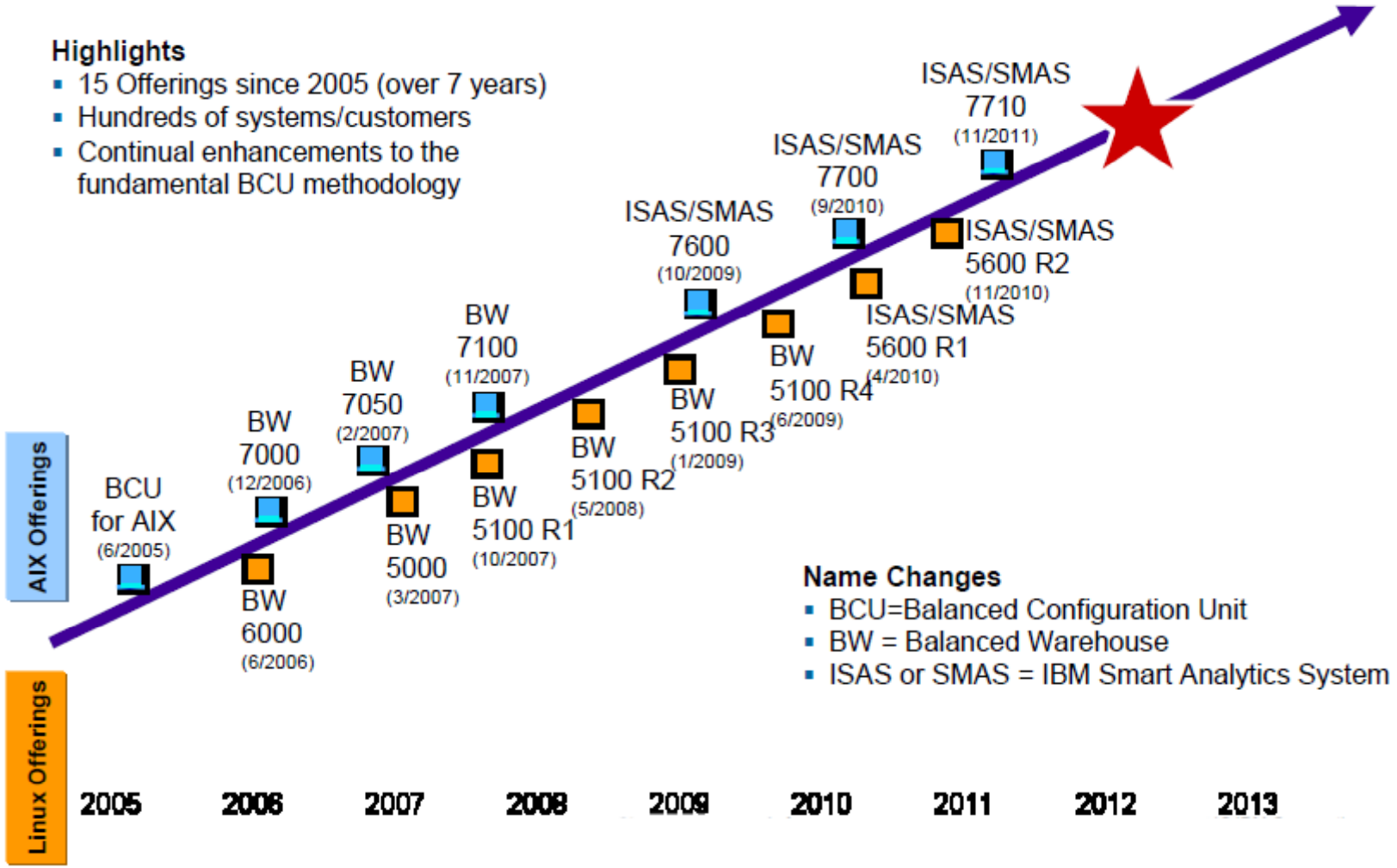



## *Optimized for a mix of interactive and analytic queries*

- Preset and configured for top performance, throughput, and efficient resource utilization
- Continuous ingest of operational data
- Balanced throughput and performance through dynamic self-tuning
- Policy-based automatic workload management
- Automated multi-temperature data management

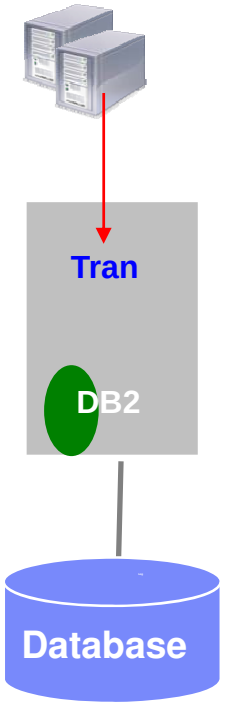
# IBM PureData System for Operational Analytics - History

## Progression of IBM Data Warehouse Integrated Offerings



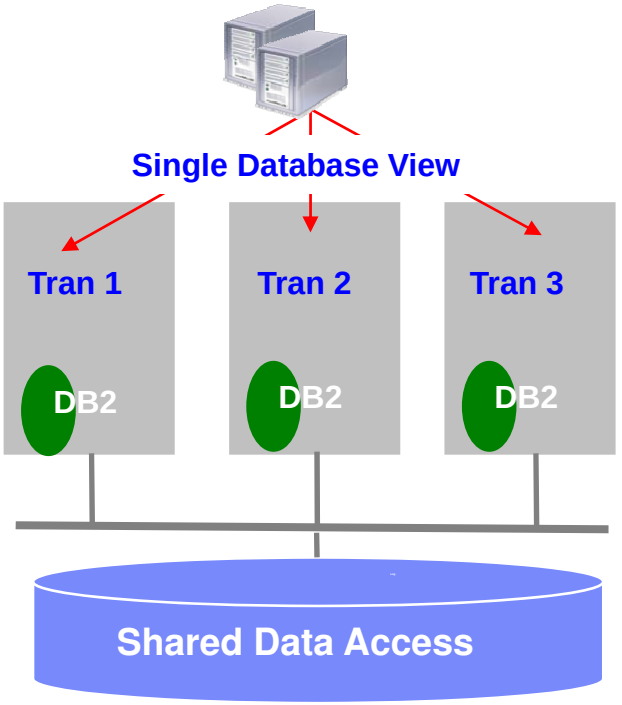
# IBM PureData System for Operational Analytics – DB Architecture

## Optimized for OLTP




### Core DB2

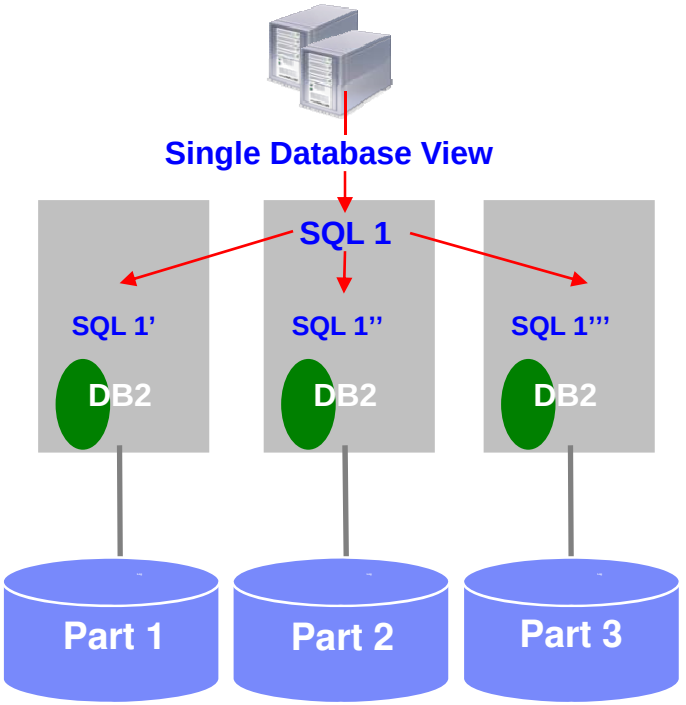
**Ideal for OLTP and data marts**




### DB2 pureScale Data Sharing

**Ideal for active/active OLTP/ERP scale out**

## Optimized for Analytics




### DB2 InfoSphere Warehouse

**Ideal for data warehousing with MPP scale out for near linear scalability and query processing**

# IBM PureData System for Operational Analytics - hardware overview


- IBM POWER7 P740 & P730 16 Core servers @ 3.55GHz
- IBM Storwize® V7000 with 900GB drives
- Ultra SSD I/O Drawers, each with six 387GB SSD
- Blade Network Technologies 10G and 1G Ethernet switches
- Brocade SAN switches (SAN48B-5)

Scales to PB+ capacity*

| Scalable to PB+* | | | |
|---|---|---|---|
| Extra Small | Small | Medium | Large |
| 64.8 TB* | 151.2 TB* | 237.6 TB* | 324 TB* |

*Unformatted raw disk capacity

# IBM PureData System for Operational Analytics - Enhancement

## 비용측면의 효율성

- Clients have experienced cases of 10x storage space savings via Adaptive Compression
- Integrated backup for rapid backup and restore without moving data on or off the system

## 실시간 분석력

- Faster, more accurate decision making with real-time operational analytics
- Continuous Ingest of data
- Built-In Time Travel query enabling faster historical and trend analytical queries

## 생산성 및 관리 용이성

- Integrated and simplified system monitoring and maintenance
- Simplified support for multi-tenant operational warehouses
- Row and Column access controls to support multiple tenant operational warehouses

# PureData System for Operational Analytics – key capabilities

## Data management

- Continuous Data Ingest
- Multi-Temperature Data Management
- Storage Compression with Adaptive Compression
- Row and column Access Controls
- Workload Manager

## Data movement

- SQL Warehousing Tool
- InfoSphere Federation Server

## Operating System

- IBM AIX

## Analytics

- Cubing Services
- Cognos Business Intelligence * (5 user entitlements)
- Intelligent Miner

## Tooling

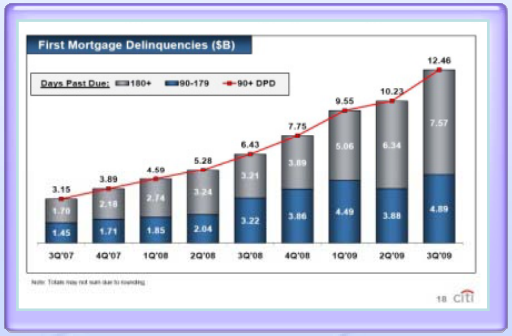
- PureData System Console
- Design Studio
- Optim Development Studio

# Agenda

- Analytic Challenges in Big data era
- Delivering Deep Analytics
  - IBM PureData system for Analytics
- Delivering Operational Analytics
  - IBM PureData system for Operational Analytics
- Big data platform to accelerate analytics

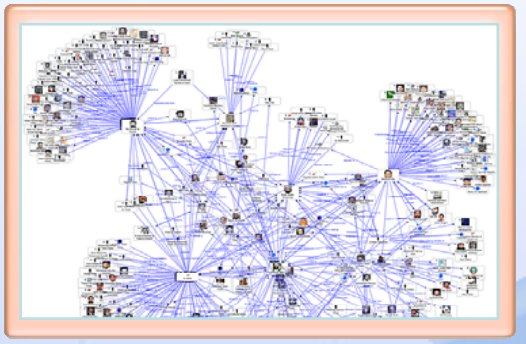
# PureData System for Reporting and Beyond as big data platform

## Reporting, BI and Ad-Hoc Analysis



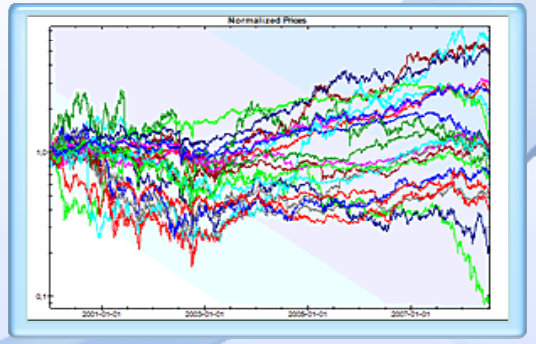

- What happened?
- When, where, how much?

## Predictive Analytics



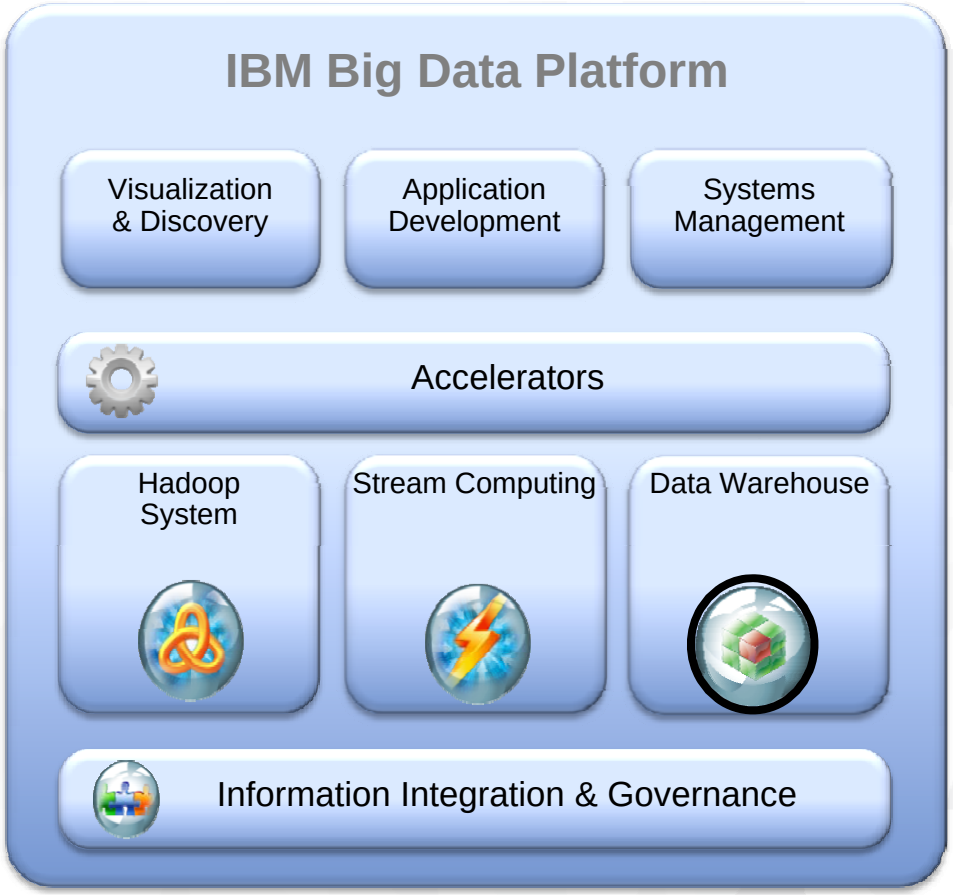
- What will happen?
- What will the impact be?

## Operational Analytics




- What is the best choice right now?

# IBM Big Data Platform to Accelerate Analytics

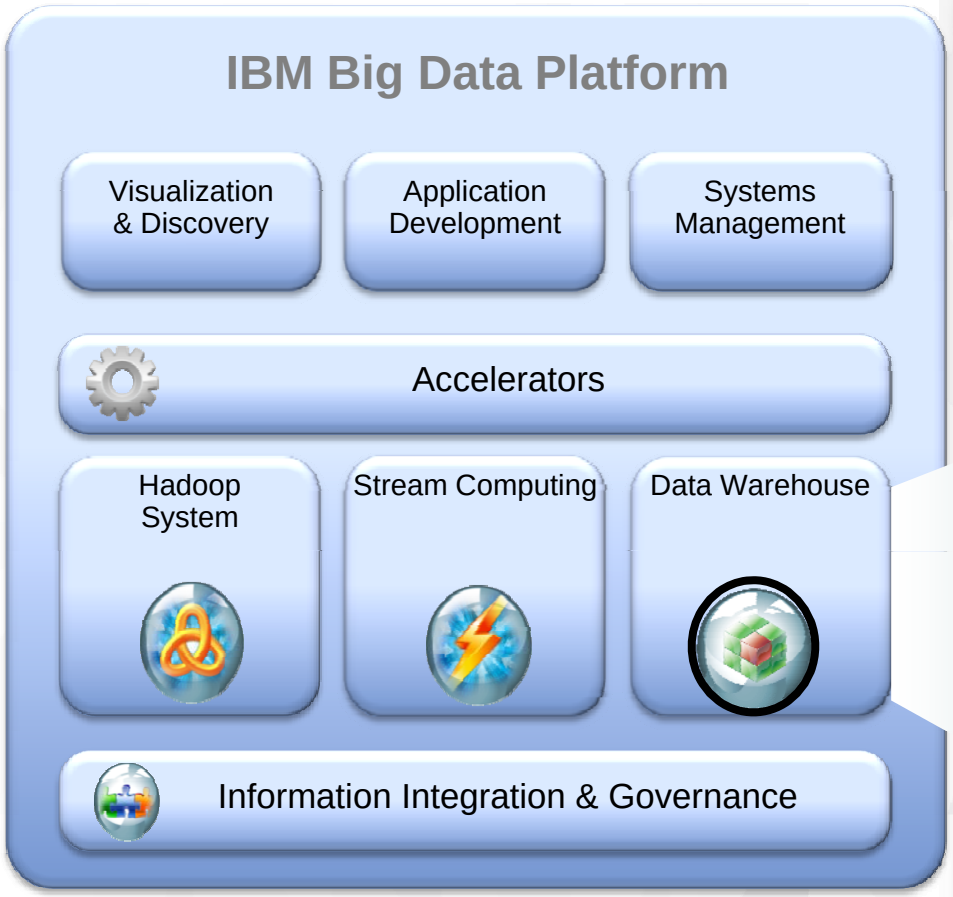



## Delivering

- Increased agility
- Improved efficiency
- Reduced complexity
- Deep analytic expertise
- Accelerated time to value

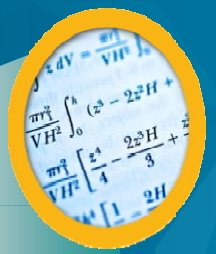
# IBM Big Data Platform to Accelerate Analytics – different analytics

**: IBM PureData System for Analytics**
**: IBM PureData System for Operational Analytics**




**Different characteristics to meet different service requirements**

- Data loading / continuous ingest
- Response latency
- Concurrency



**Analytics & Reporting**

**Operational Analytics**