

**BusinessConnect**  
Era of Smart

# Accelerate Big Data & Analytics with IBM PureData System

**Marko Milek**

Data & Analytics Technical Leader , ASEAN

Software





# Today's increasing demand on data systems

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

Year	Volume of data
2010	~0.7 ZB
2020	35 ZB

50x

35 ZB  
by 2020

2010 2020

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

Millions of  
transactions  
per second

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

Billions of  
devices &  
sensors

**Smarter approach to meeting data challenges is required**  
**Reduce complexity ■ Accelerate time to value ■ Improve IT economics**

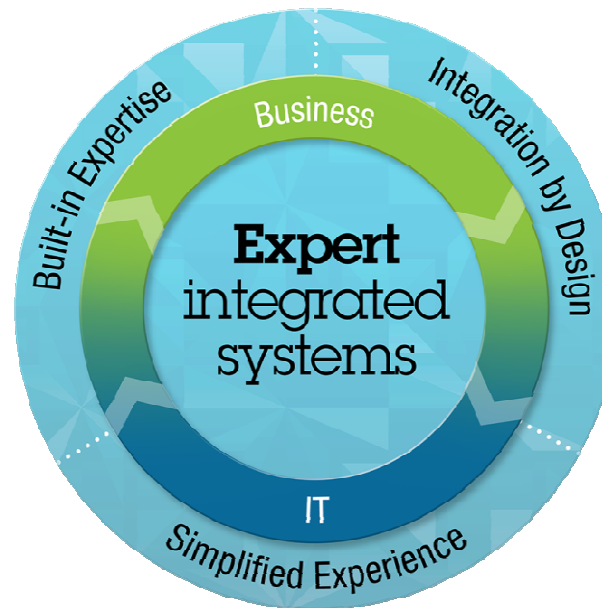


time for a new class of expert integrated systems

# PureSystems

## Built-in Expertise

*Capturing and automating what experts do*



## Integration by Design

*Deeply integrating and tuning hardware and software*

## Simplified Experience

*Making every part of the IT lifecycle easier*



# PureData



## Data Platform

Powering Big Data Platform Services

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

# PureSystems



# Different applications drive different data workloads

Order Management



Transaction Processing

Sales Analysis



Reporting and Analytics

Real Time Fraud Detection



Operational Analytics

Customer Analysis



Any Structure Analytics



# Systems need to be optimized for different data workloads

Order Management



Sales Analysis



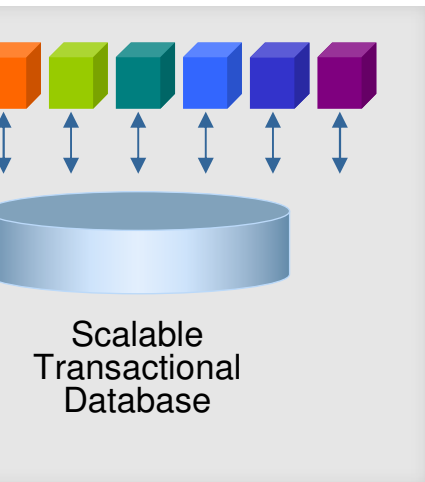
Real Time Fraud Detection



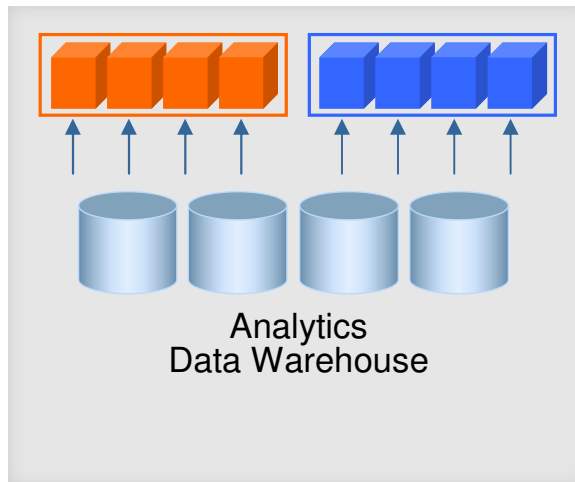
Customer Analysis



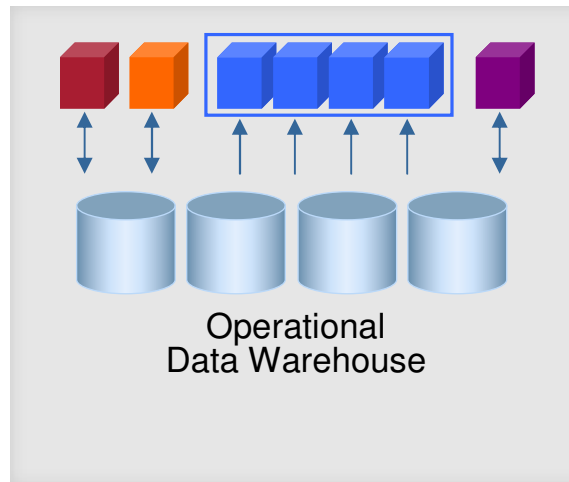
Transaction Processing



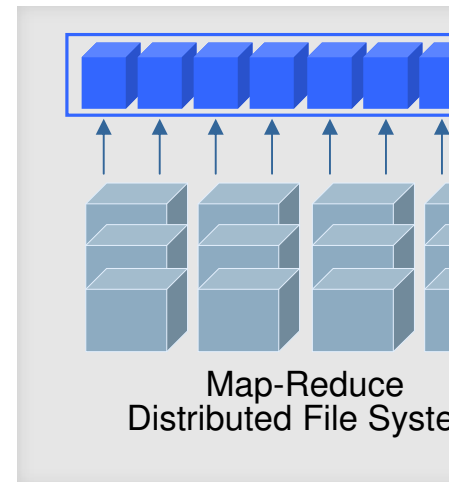
Reporting and Analytics



Operational Analytics



Any Structure Analytics





# Meeting workload optimized performance

## Meeting Big Data Challenges – Fast and Easy!

**PureData**  
System for Transactions

### For apps like Management

cluster services  
for transactional  
throughput and scalability

**PureData**  
System for Analytics

### For apps like Sales Analysis

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

**Powered by  
Netezza Technology**

**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

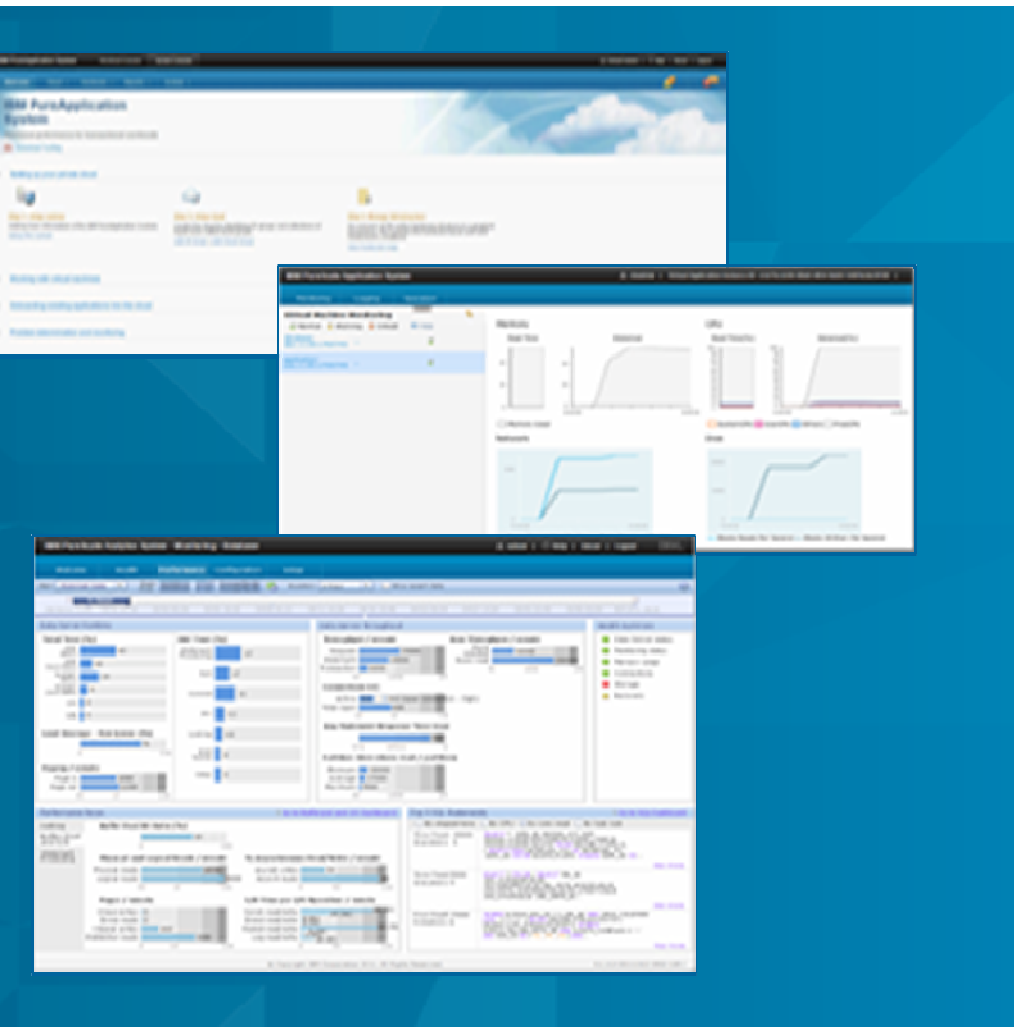
**PureData**  
System for Hadoop

### For apps like Big Data Exploration

Hadoop services optimized for  
exploration of large volumes of data  
with any type of structure; and a  
queryable archive  
to augment traditional  
data warehousing



# Simplified, integrated system management



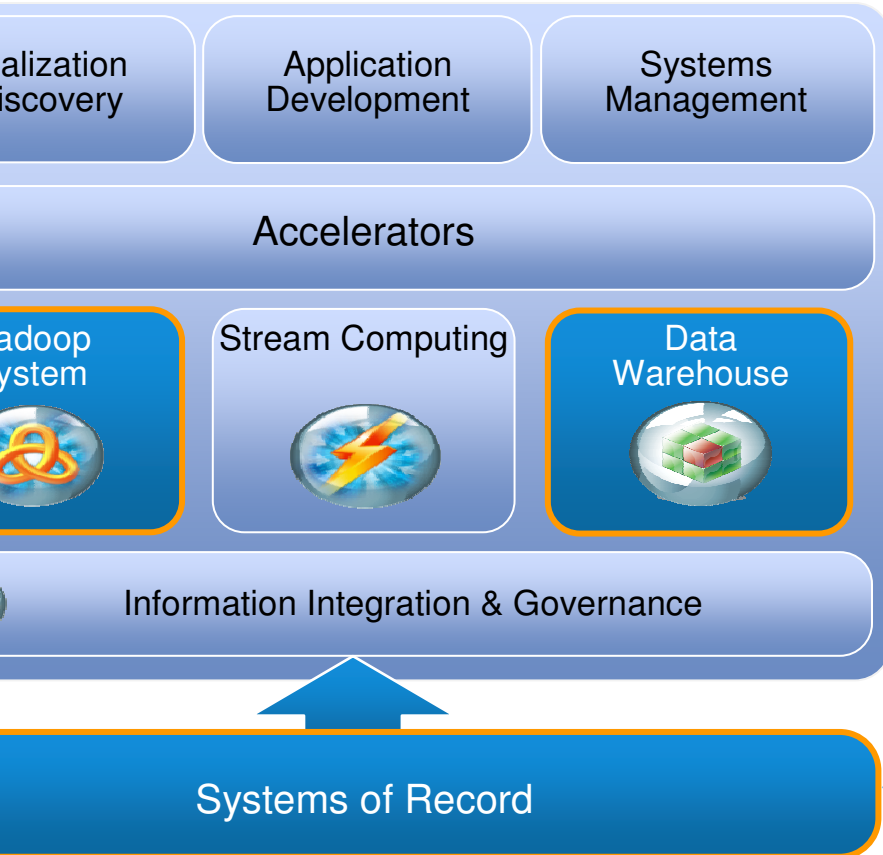
- Consistent IBM PureSystem console to manage all resources and workloads
- Role-based security and tasks
- System firmware and OS updates applied with no planned downtime
- Single point of contact for support



# PureData is an expert integrated system for big data



## IBM Big Data Platform



## PureData

### For Hadoop

Optimized system to accelerate big data analytics and online archive with appliance simplicity



### For Analytics

Optimized system delivering data services for analytics & reporting

### For Operational Analytics

Optimized system delivering data services for operational analytics

### For Transactions

Optimized system delivering data services for transactions



## PureData System for Transactions

### *Optimized exclusively for transactional data workloads*

- **Handles 100+** databases on 1 system
- **Up to 30x** database scalability
- **Up to 10x** storage space savings
- Database node recovery in seconds
- Supports DB2 applications unchanged
- Supports Oracle Database applications with minimal change



# nt examples



**Global Leader in Payment Processing**

- DB2 Cluster provides advantages over competition
- Lower cost payment routing improves competitiveness
- Easily handles transaction volumes during shopping peaks
- Extreme high availability at the required scale



**Leading Research Institution**

- Simplified IT infrastructure for capturing, managing and applying scientific data
- System fully loaded, tested and deployed in hours
- Database clusters developed in minutes



**Transportation Giant**

- Consolidated server environment by approx. 90%
- Able to scale easily to manage increased workloads
- Gained powerful database platform and increased ability of IT systems to manage large workloads efficiently



# Simplified deployment of high availability

## Uninterrupted access to data with consistent performance

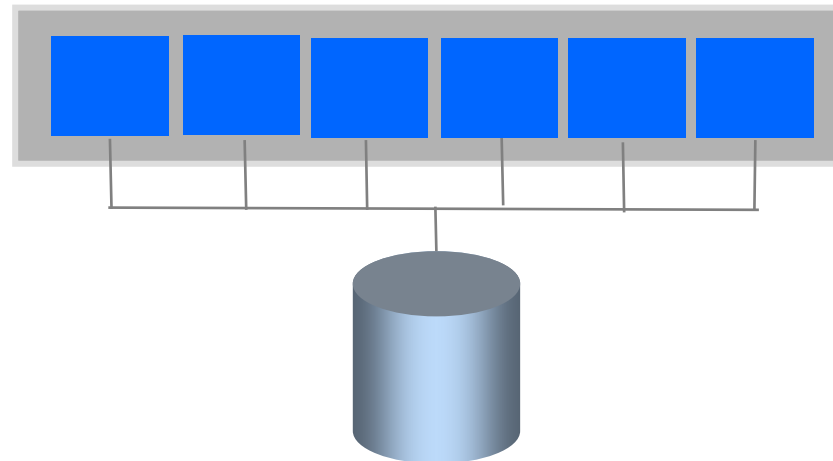
Additional systems –  
and it yourself

PureData System  
with built-in exper

over several days/weeks:

- Define High Availability topology
- Configure HW/SW/Network
- Set up storage pools
- Install multiple operating systems
- Install database instances
- Set up primary and secondary management systems
- Set up database members
- Set up backup processes
- Test, tune, reconfigure

6-node database cluster



In minutes,

1. Just specify database description and topology pattern



# Simplified deployment of multiple databases

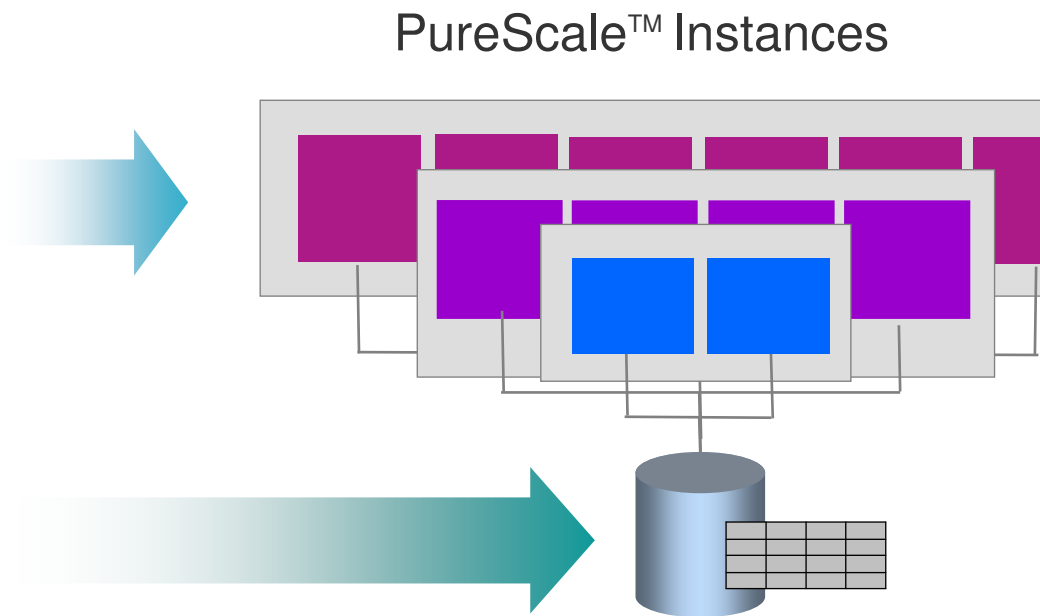
## Deploy topology and databases in minutes using patterns

### Topology patterns

Automatically creates, configures and deploys database system topology with built-in redundancy and high performance

### Database patterns

Automatically creates, configures and deploys 1 or client-specified databases optimized for transactional workloads





# PureData

System for Analytics



## *Optimized exclusively for analytic data workloads*

- **10-100x** faster than traditional custom systems
- **20x** concurrency and throughput for tactical queries
- **Patented MPP** hardware acceleration
  
- No indexing, No tuning, No storage administration
- High speed data ingest
- Richest set of in-database analytics

Powered by  
Netezza Techno



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."*

*Edwin Marcial, Chief Technology Officer*



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*



**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*



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."*

*Dr. Murali Ramanathan, SUNY Buffalo*





# IBM PureData System for Analytics – new model *fastest performing Netezza technology to date!*

## *Accelerate Performance of Analytic Queries*

- **3X faster performance<sup>1</sup>**  
*for Big Data analytics*
- **128 GB/sec effective scan rate per rack<sup>2</sup>**  
*to tackle Big Data faster*

## *Increase Efficiency of your Data Center*

- **50% greater data capacity per rack<sup>3</sup>**  
*helps optimize data center efficiency*
- **More capacity and less power per rack**  
*than both Oracle and Teradata*

## *Simplicity and Ease of Administration*

- **Improved system management and resilience**  
*to spend less time managing and more time delivering value*
- **70% fewer service calls**  
*with more spare drives and faster disk regeneration*

<sup>1</sup> Based on a comparison of the IBM PureData System for Analytics N2001 to the IBM PureData System for Analytics N1001. The performance speed refers to the query times on both macro-analytic and mixed workload tests as conducted in IBM engineering lab benchmarks. The N2001 average of 3x faster than those of the N1001. Individual

<sup>2</sup> 128 GB/sec scan rate assuming an average of 4x compression across the system. Individual

<sup>3</sup> Capacity of IBM PureData System for Analytics N2001 compared to previous generation IBM PureData System



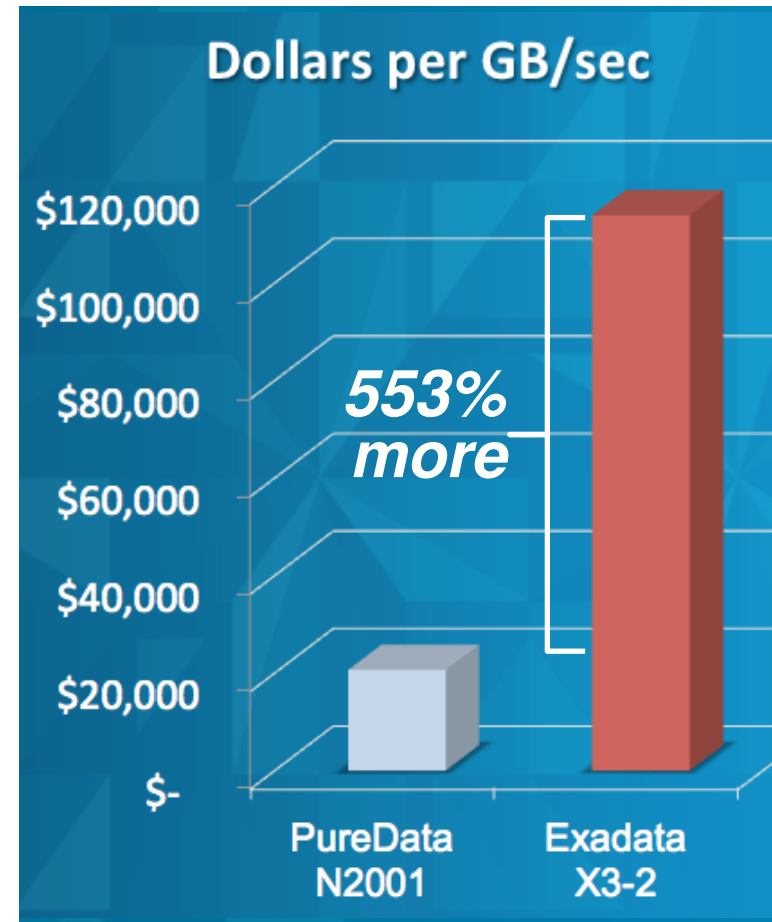


# Spending less on PureData than Oracle Exadata Faster performance!

*Accelerate Performance  
of Analytic Queries*

Comparing dollars per GB per second...

**Exadata will cost you more than 500%  
to scan the same gigabyte of data!**<sup>1</sup>



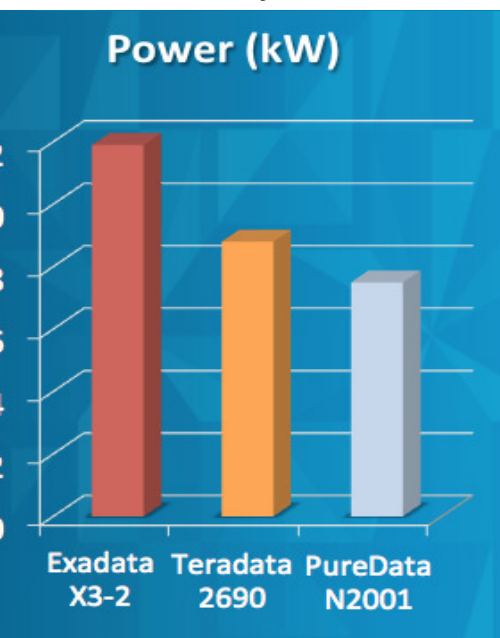
<sup>1</sup> IBM PureData System for Analytics N2001 = \$2,637,300 list price/128 GB per sec. = \$20,603.91/GB/sec. 128 GB/sec scan rate assumes an average of 4x out of the box compression across the system. Oracle list price /100 GB/ per sec. = \$113,842.80/GB/sec. Oracle scan rate from Oracle X3 datasheet at <http://www.oracle.com/us/products/database/exadata-db-machine-x3-2-1851253.pdf> and <http://www.oracle.com/us/corporate/pricing/exadata-pricelist-070598.pdf> and <http://www.oracle.com/us/corporate/pricing/technology-price-list-070617.pdf>. Individual results may vary. This comparison is configurati



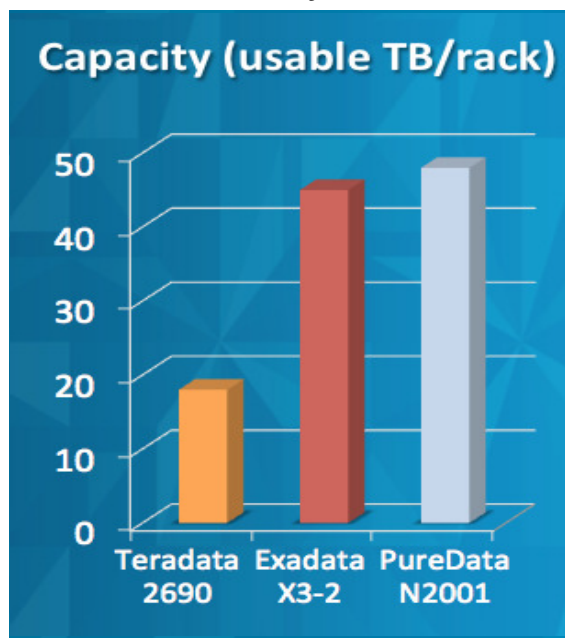
# Increase data center efficiency faster, more efficient systems

Increase Efficiency  
of your Data Center

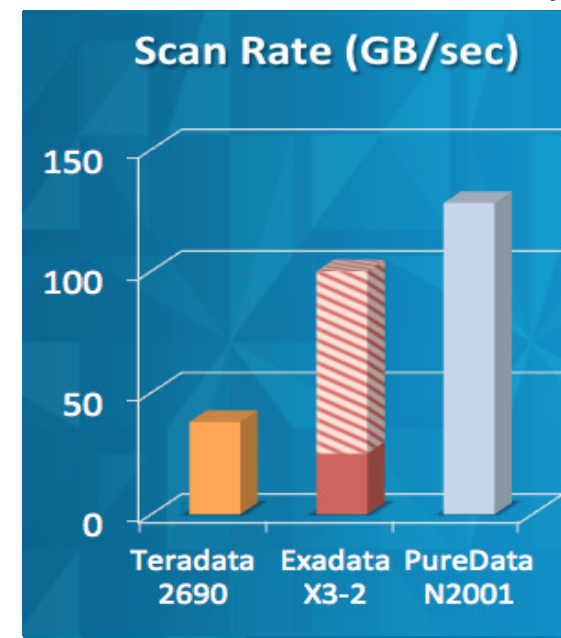
PureData uses **Less Power**  
than other systems<sup>1</sup>



PureData has **More Capacity**  
than other systems<sup>2,3</sup>



PureData has **“Out of the box”  
Faster Scan Rates** than other systems

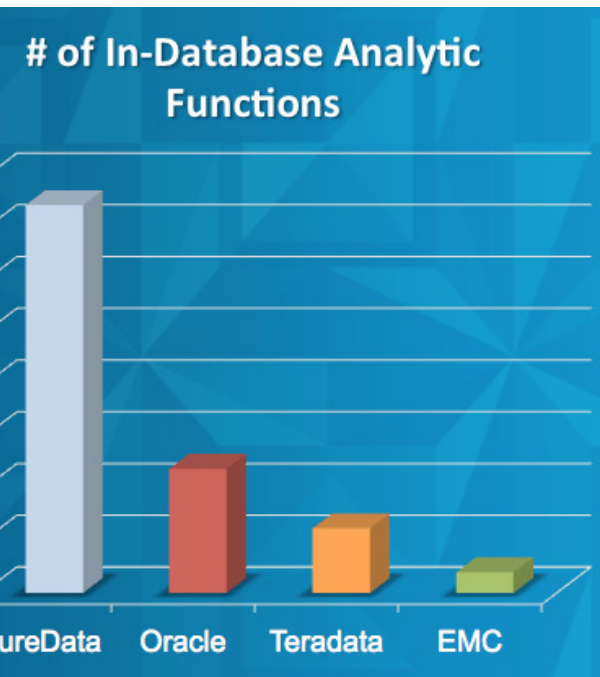


- PureData System for Analytics**
  - Offers more than 160% greater capacity than Teradata 2690
  - Has more than a 200% faster scan rate than Teradata!
  - Nearly 30% faster scan rate than Oracle X3-2
  - Uses 40% less power than Oracle X3-2
  - Less floor space
  - Less power/cooling per data center tier

<sup>1</sup> Teradata 2690 data sheet claims 8.8kW power per rack. IBM PureData System for Analytics N2001 is 3.0kW per rack. Claims are based on out of the box configuration.  
<sup>2</sup> Oracle high performance drives provide 45TB per rack before compression per Oracle X3-2 datasheet. IBM PureData System for Analytics N2001 pre-compression capacity is 48 TB per rack. Claims are based on out of the box configuration.  
<sup>3</sup> 300 GB drive capacity for Teradata 2690 is 18.2 TB before compression per <http://www.teradata.com/News-Releases/2011/Fifth-Generation-Teradata-Data-Warehouse-Appliance-Performance-and-Triple-the-Data-Capacity/>. IBM PureData System for Analytics N2001 pre-compression capacity is 48 TB per rack. Claims are based on out of the box configuration. See Slide 10.



# by the richest set of in-database analytics



## *PureData in-database analytics*

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

- *SDK for developing your own analytics*
- *3<sup>rd</sup> party analytics*
- *Speed development and reuse of analytics*

- ✓ **No data movement**
- ✓ **Analyze deep and wide data**
- ✓ **High performance, parallel computation**



# PureData

System for Operational Analytics



## *Optimized exclusively for operational analytic data workload*

- **Handles 1000+** concurrent operational queries
- **Continuous ingest** of operational data
- **Up to 10x** storage space savings
  
- In-database analytics for leading applications
- Policy-based workload management
- Supports DB2 applications unchanged and Oracle Database applications with minimal change



# nt examples



**Pakistan's Leading  
Natural Gas Provider**

- **Real time** tracking & predicting patterns of supply and demand for **2.2 million** customers
- **Automatic and instantaneous** refresh of reports replaced a month-long reporting cycle
- Enables **quick identification** of leaks and pilferage

**Leading Korean  
Retail Chain**

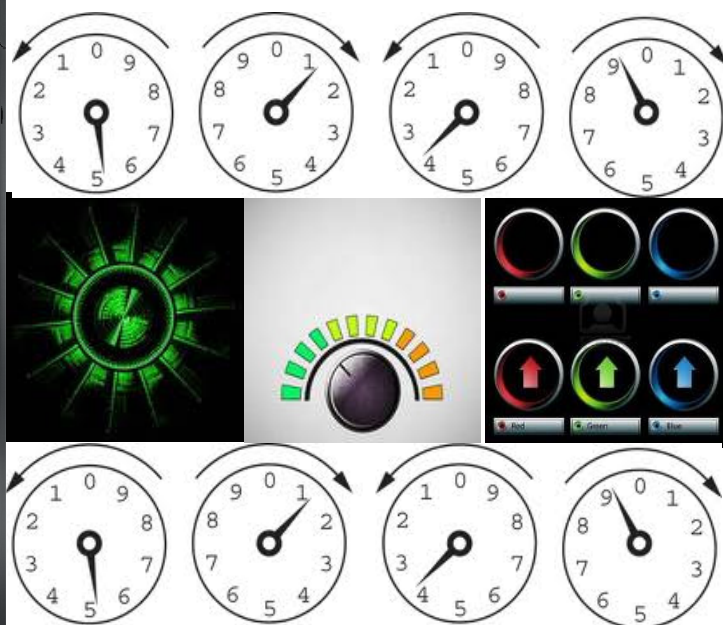
- **Market basket analysis** enables cross-selling and up-selling
- **Hundreds of concurrent** operational queries & batch processing
- **60% data reduction** with compression for storage savings
- **30% reduction in TCO** and **2.5x faster** batch processing

**A Leading Global  
Credit Card Provider**

- **Real-time** fraud assessment, customer care and business analytics
- **80 million** transactions on peak days
- **>2500 avg.** concurrent users
- **>5 TB of data** loaded per day – including throughout peak hours



# 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
- Up to 10x storage space savings via adaptive compression



## ureData System for Hadoop



## *Optimized exclusively for any structure analytic data worklo*

- **Accelerate Hadoop analytics with appl  
simplicity**
- **Integrated system management**
- **Application development accelerators**

- Easily leverage data across the big data platform
- Enterprise security, governance and high availa



# Simplify Big Data ...

## From custom and complex ... To organized simplicity



***Designed to...***

- Simplify the building, deploying and management of a Hadoop cluster
- Speed the time-to-value for Hadoop and unstructured data
- Maximize the overall analytic ecosystem
- Provide enterprise security and platform management



<sup>1</sup> Based on IBM internal testing and customer feedback. "Custom built clusters" refer to clusters that are not professionally pre-built, pre-tested and optimized. Individual results may vary.





# Benefits of IBM PureData System for Hadoop

IBM Expertise

**Accelerate Big Data  
Time to Value**

- **Deploy 8x faster** *than custom-built solutions*<sup>1</sup>
- **Built-in visualization** *to accelerate insight*
- **Built-in analytic accelerators**<sup>2</sup>  
*unlike big data appliances on the market*

IBM Experience

**Simplify Big Data  
Option & Consumption**

- **Single system console** *for full system administration*
- **Rapid maintenance updates** *with automation*
- **No assembly required, data load ready in hours**

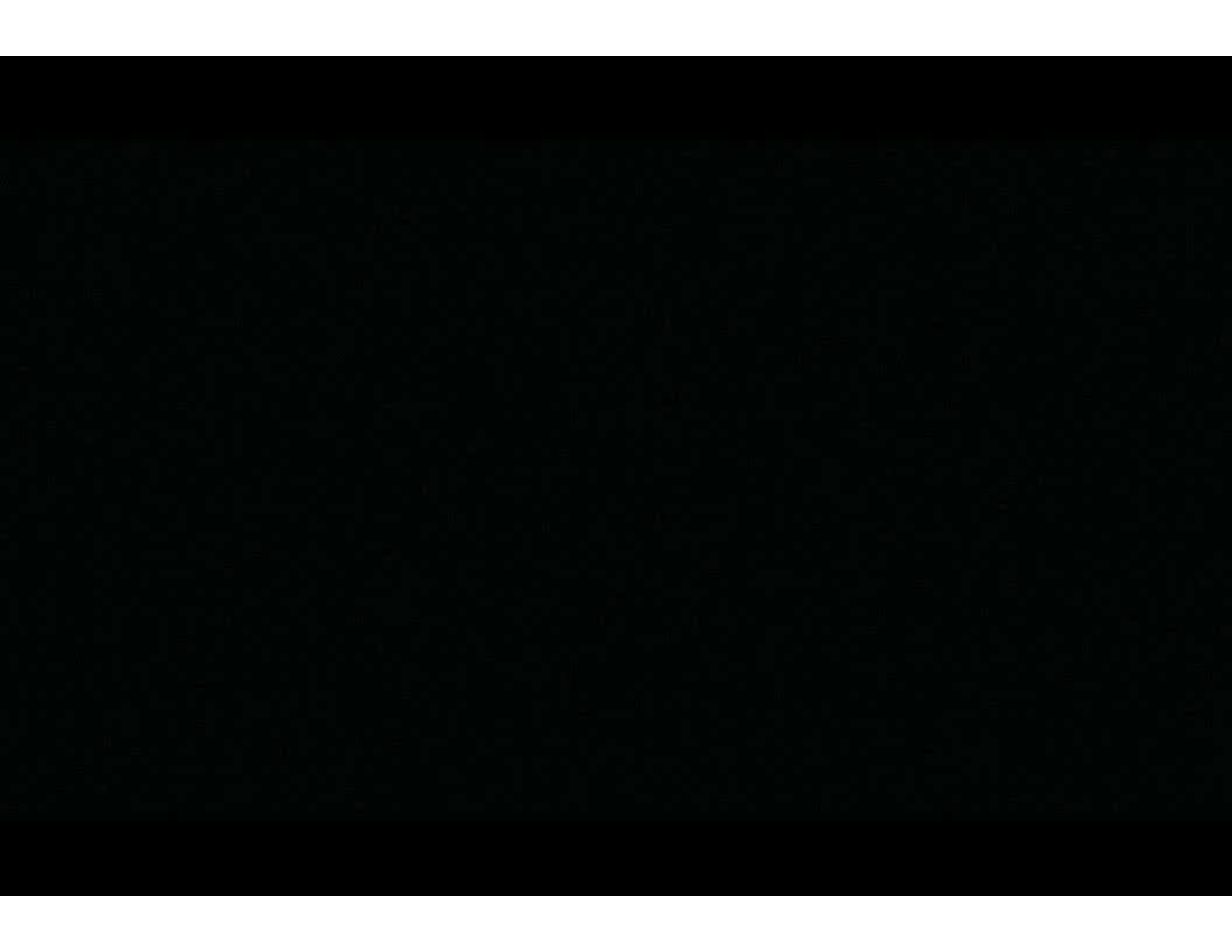
IBM Innovation by Design

**Implement Enterprise Class  
Big Data**

- **Only integrated Hadoop system**  
*with built-in archiving tools*<sup>2</sup>
- **Delivered with more robust security**  
*than open source software*
- **Architected for high availability**

<sup>1</sup>Based on IBM internal testing and customer feedback. "Custom built clusters" refer to clusters that are not professionally pre-built, pre-tested and optimized. Individual results may vary.

<sup>2</sup>Based on current commercially available Big Data appliance product data sheets from large vendors. US ONLY CLAIM.





# Experience Program: Let us prove it at no charge

- Guided analysis of business value
- PureSystems Technology Demonstration
- 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



[ibm.com/puresystems](http://ibm.com/puresystems)



# PureData



- A platform for today's analytic challenges  
**IBM Big Data Platform**
- A new class of Expert Integrated Systems  
**IBM PureSystems**
- A workload optimized data system  
**PureData System**
- A way to experience the value  
**PureExperience**

**IBM®**