Accelerate Big Data & Analytics with IBM PureData System

Dr Marko Milek

Big Data & Analytics Technical Leader , ASEAN

IBM Software

IBM

Today's increasing demand on data systems

Increasing
Volume of data
requires growing capacity

50x
35 ZB
by 2020

2010   2020

Increasing
Velocity of data
requires higher performance

Millions of transactions per second

Increasing
Variety of data
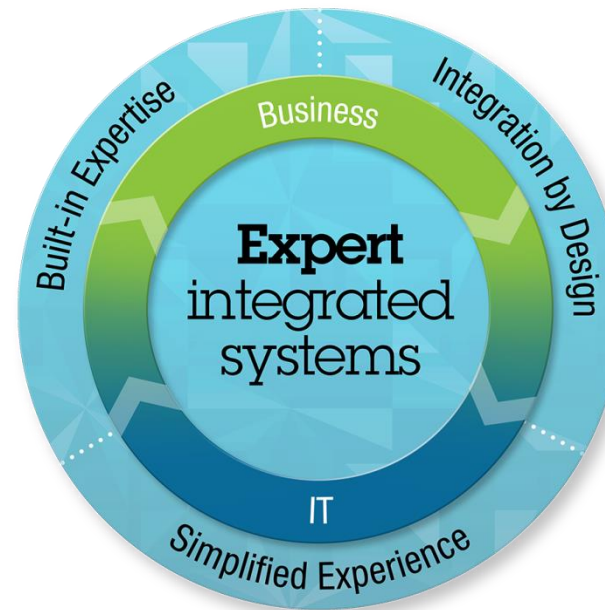*requires new techniques*

Billions of devices & sensors

A smarter approach to meeting data challenges is required to:
Reduce complexity ▪ Accelerate time to value ▪ Improve IT economics

It is 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

Delivering Big Data Platform Services

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

PureSystems

IBM

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

Data systems need to be optimized for different data workloads

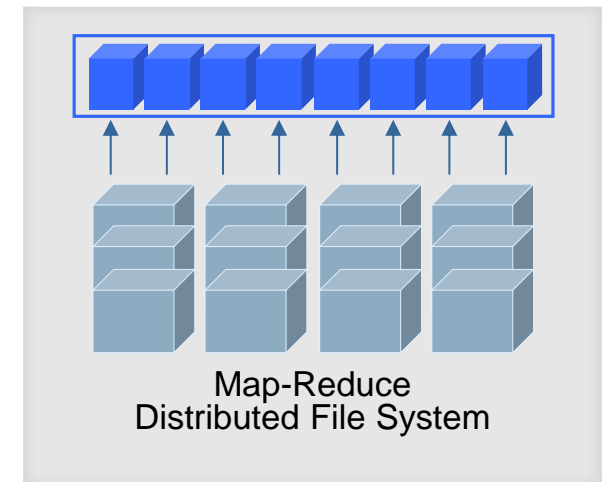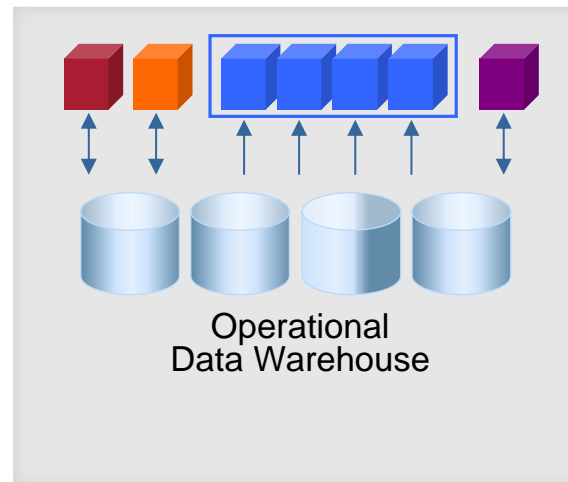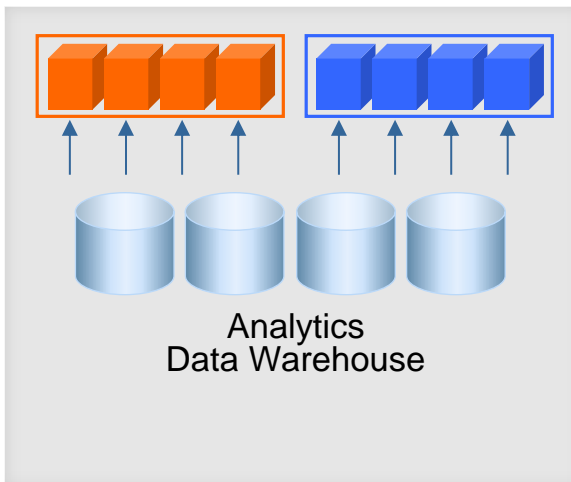| Order Management	Sales Analysis	Real Time Fraud Detection	Customer Analysis
Transaction Processing	**Reporting and Analytics**	**Operational Analytics**	**Any Structure Analytics**
Scalable Transactional Database	Analytics Data Warehouse	Operational Data Warehouse	Map-Reduce Distributed File System

6
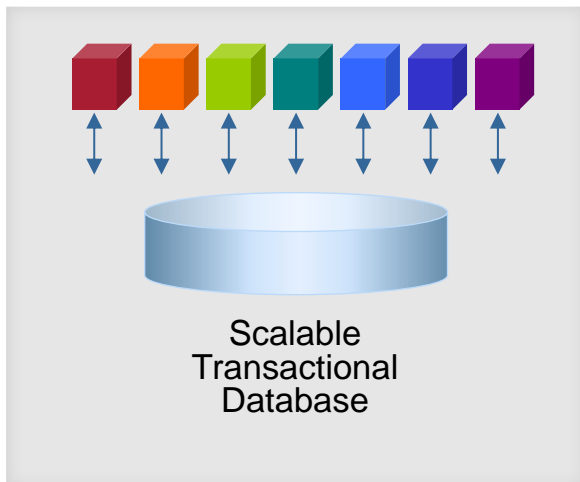
Delivering workload optimized performance

Meeting Big Data Challenges – Fast and Easy!

PureData
System for Transactions

For apps like
Order Management

Database cluster services optimized 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

New

For apps like
Big Data
Exploration

Hadoop services optimized for exploration of large volumes of data with any type of structure; and as a queryable archive to augment traditional data warehousing
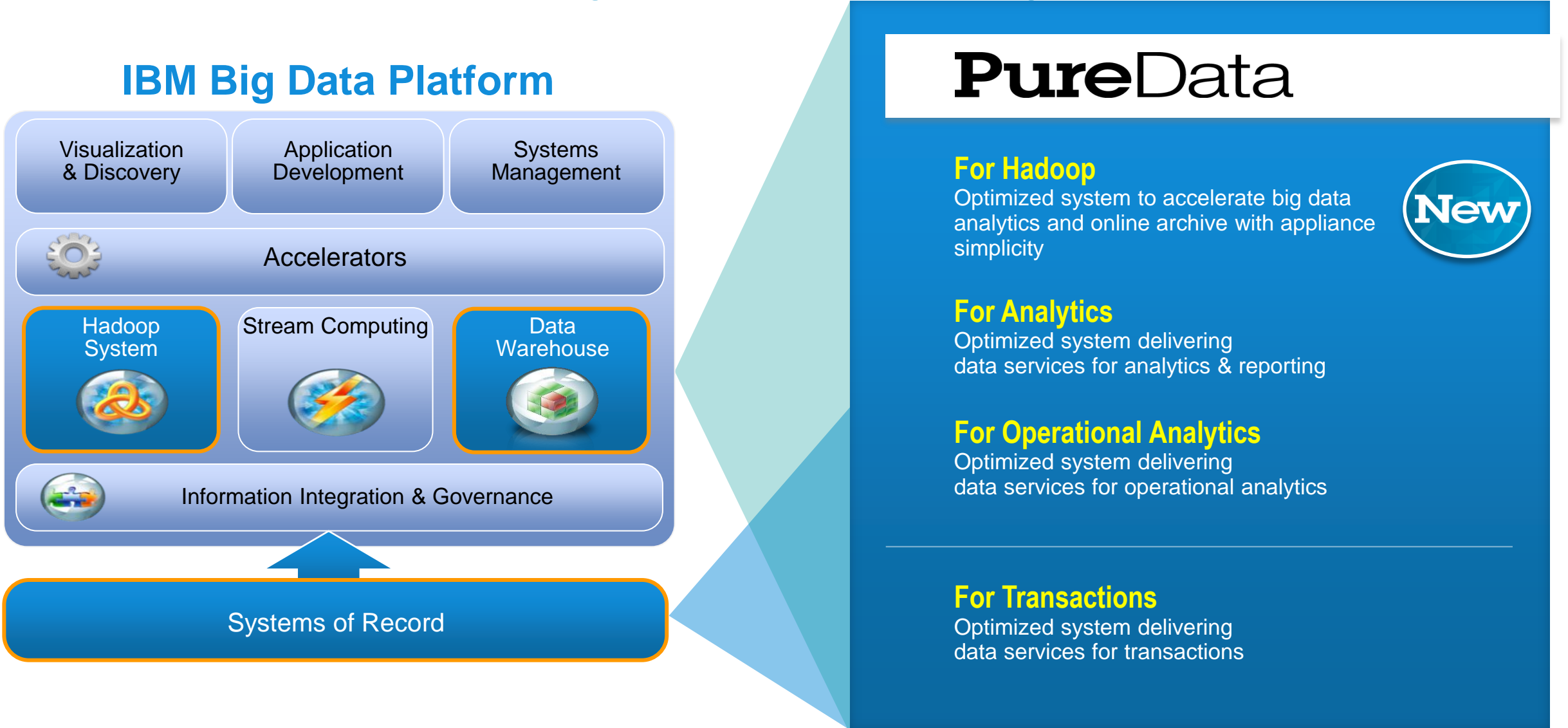
Simplified, integrated system management
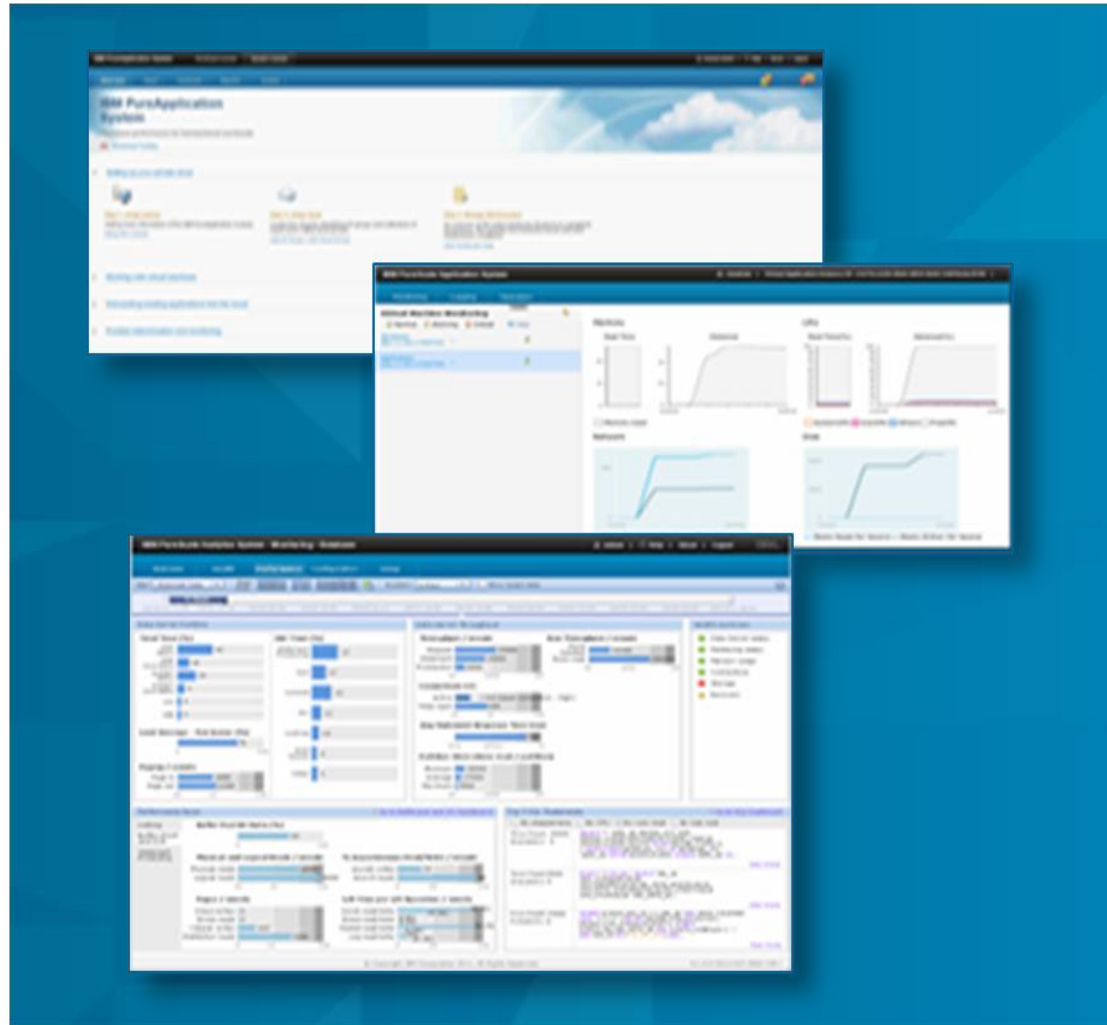
- Consistent IBM PureSystem console to manage all resources and workloads

- 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

Visualization & Discovery	Application Development	Systems Management

Accelerators

Hadoop System	Stream Computing	Data Warehouse

Information Integration & Governance

Systems of Record

PureData

New

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

Client 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
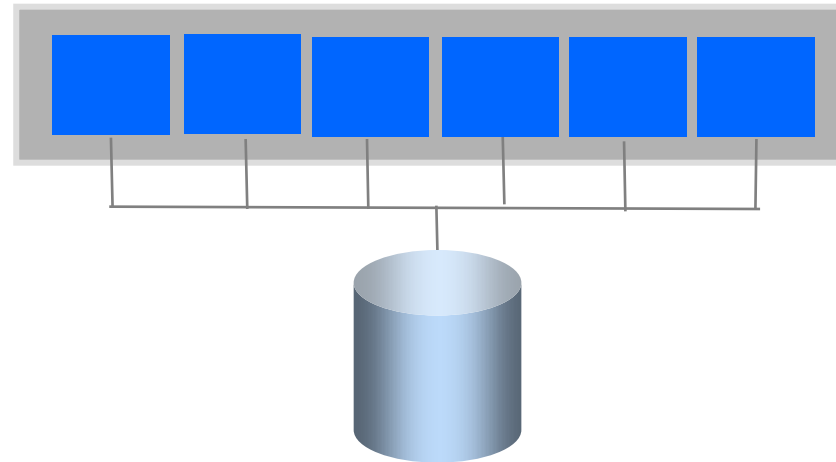
Simplified deployment of high availability

Uninterrupted access to data with consistent performance

Traditional systems – build it yourself

PureData System with built-in expertise

6-node database cluster

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

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 a database system topology with built-in redundancy and high performance

Database patterns

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

PureScale™ Instances
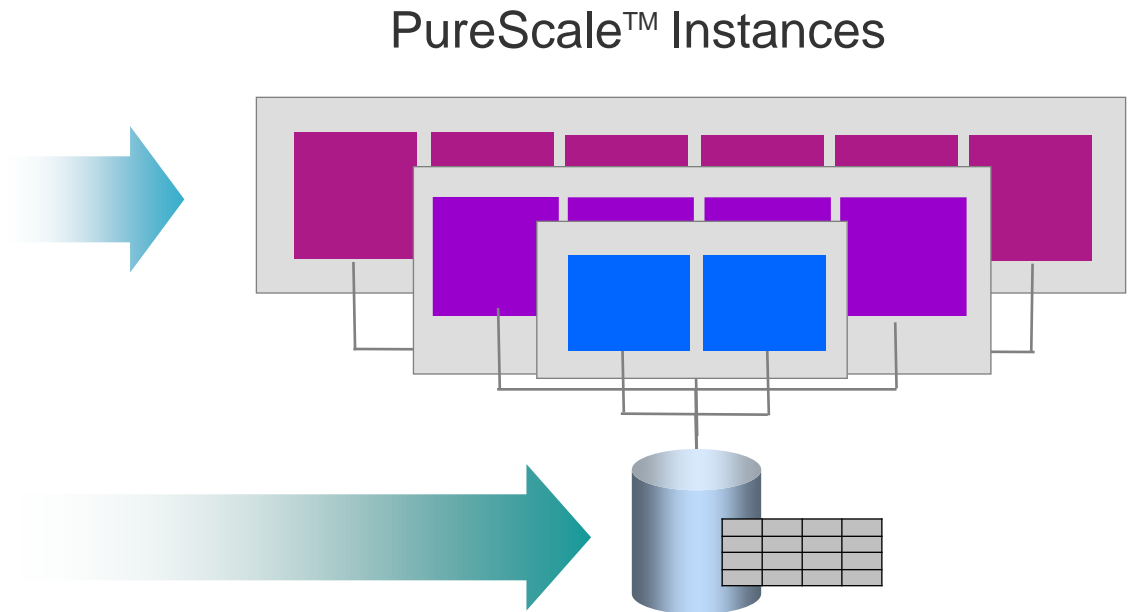
PureData
System for Analytics

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
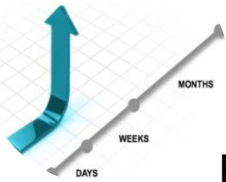
Powered by Netezza Technology

IBM

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

ice

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

XO Communications

---

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

NYSE Euronext

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

University at Buffalo
The State University of New York

IBM

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

Accelerate Performance of Analytic Queries

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

Increase Efficiency of your Data Center

- **50% greater data capacity per rack[3]**
  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

1 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 query times were an average of 3x faster than those of the N1001. Individual results may vary.
2 128 GB/sec scan rate assuming an average of 4x compression across the system. Individual results may vary.
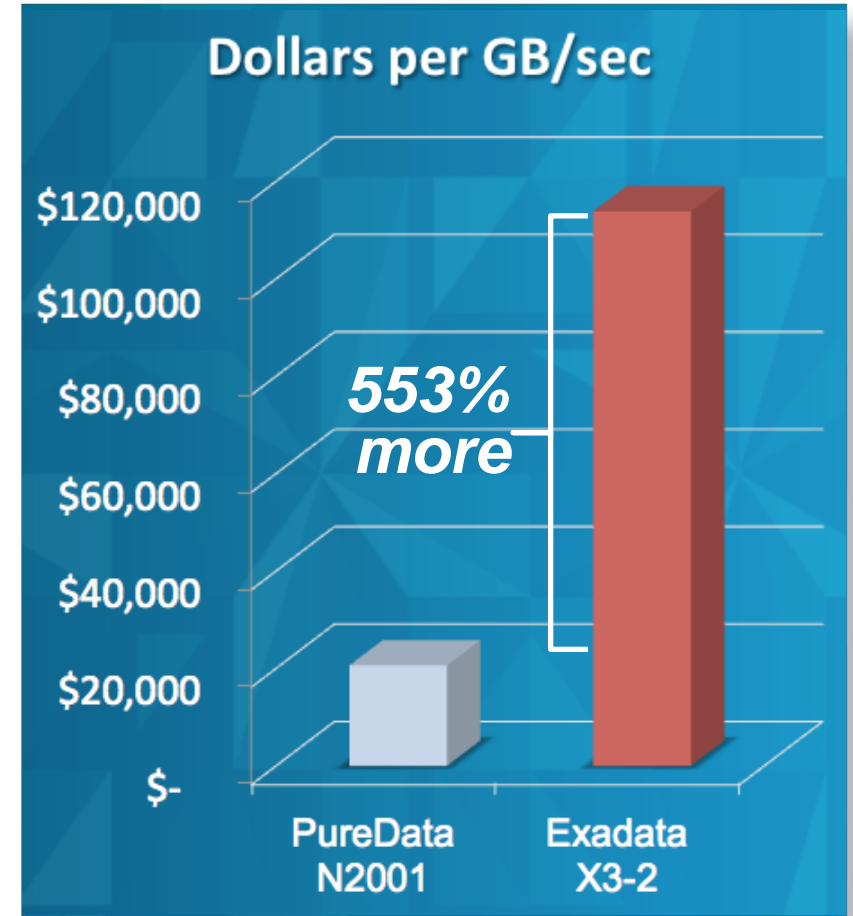3 Capacity of IBM PureData System for Analytics N2001 compared to previous generation IBM PureData System for Analytics N1001.

Spend less on PureData than Oracle Exadata for faster performance!

Accelerate Performance of Analytic Queries

Comparing dollars per GB per second…

Oracle will **cost you more than 500% more** to scan the same gigabyte of data![1]

**Dollars per GB/sec**

553% more

PureData N2001   Exadata X3-2

[1] 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 X3-2 = $11,384,280 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 pricing documents at 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 based on out of box configurations for both systems

IBM

Increase data center efficiency
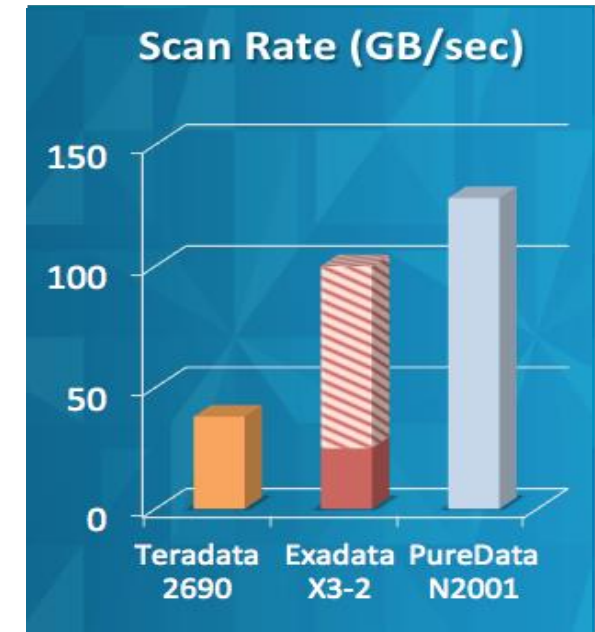## *with faster, more efficient systems*

*PureData uses **Less Power** than other systems[1]*
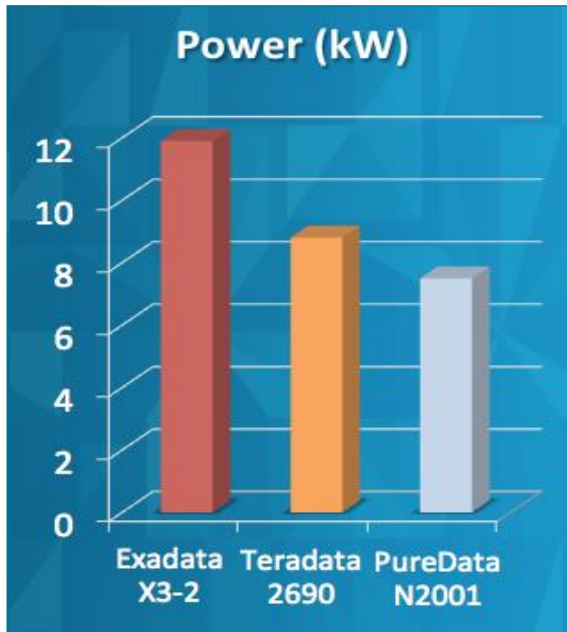
*PureData has **More Capacity** than other systems [2,3]*
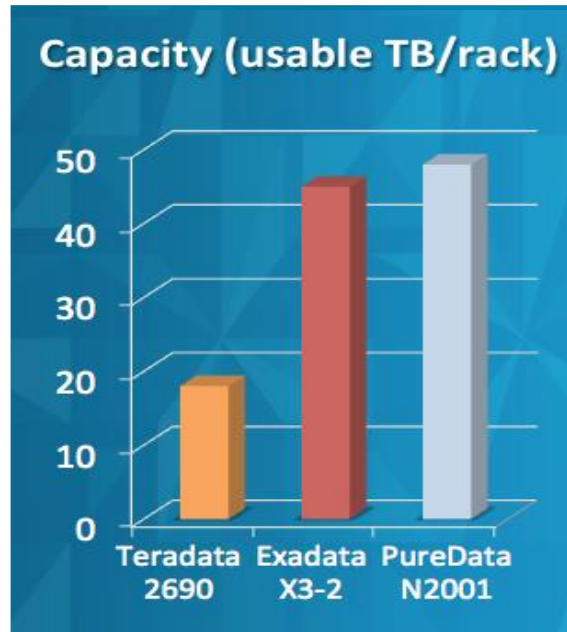
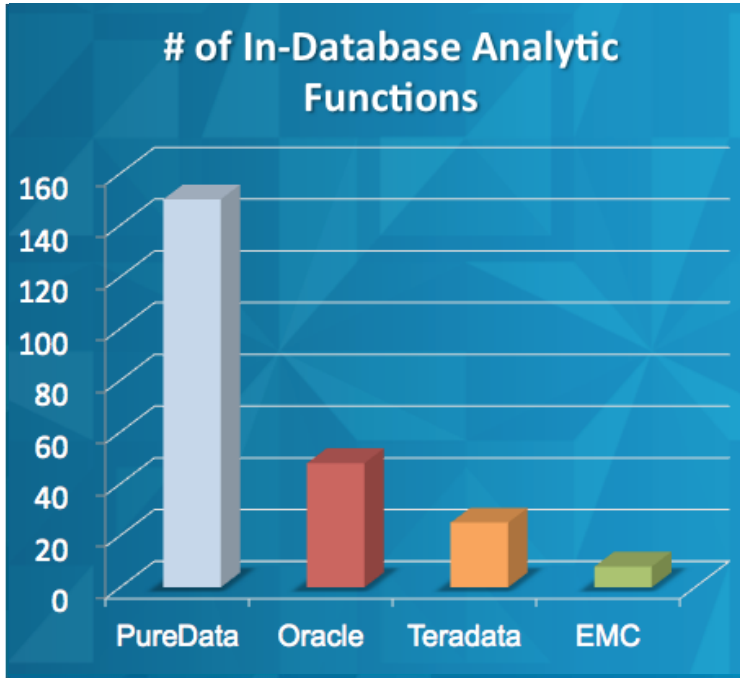*PureData has "Out of the box" **Faster Scan Rates** than other systems*

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 tile

Enjoy the richest set of in-database analytics

# of In-Database Analytic Functions

(Bar chart showing values on y-axis: 160, 140, 120, 100, 80, 60, 40, 20, 0; x-axis categories: PureData, Oracle, Teradata, EMC)

PureData in-database analytics

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

- *SDK for developing your own analytics*
- *3rd 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 workloads

- **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

Client 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

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

- Up to 10x storage space savings via adaptive compression

PureData
System for Hadoop

***Optimized exclusively for
any structure analytic data workloads***

- **Accelerate Hadoop analytics with appliance simplicity**
- **CoIntegrated system management**
- **Application development accelerators**

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

Let's 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

DEPLOY 8x FASTER![1]

[1] 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.

IBM

Benefits of IBM PureData System for Hadoop

Built-in Expertise

Accelerate Big Data Time to Value

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

Simplified Experience

Simplify Big Data Adoption & Consumption

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

Integration by Design

Implement Enterprise Class Big Data

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

[1] 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.
[2] Based on current commercially available Big Data appliance product data sheets from large vendors. US ONLY CLAIM.

AmBank's Journey
to Greater Insights with *IBM PureData System*

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

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