

# Addressing your Big Data Challenges with IBM PureData Systems

**Mallarswami Nonvinkere** ([mnonvink@in.ibm.com](mailto:mnonvink@in.ibm.com))  
Technical Consultant, PureData Ecosystem Asia Pacific

# 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**

# Clients need to address critical imperatives in this environment

**Accelerate new applications, big data and analytics**



**34% of new IT Projects**  
deploy late

From a commissioned study conducted by Forrester Consulting on behalf of IBM

**Improve IT efficiency with expert integrated systems**



**• Only 1 in 5**  
Can allocate 50% or more of their IT budget to new projects<sup>1</sup>

IBM, *Data center operational efficiency best practices*, April 2012.

**Simplify cloud infrastructure and application platforms**



**90% plan to implement cloud**  
by 2015

\* IBM GBS 2011 IBV Study, "The power of cloud: driving business model innovation"

# Too many IT projects are over budget and deployed late

## Design/Deploy



## Manage/Maintain



Typical IT Project Time and Budget

Phase	Time (days)	Budget
Specify/design	73 - 96	14% - 16%
Procure	57 - 112	19% - 21%
Implement	74 - 93	12%
Configure/test	74 - 80	10% - 11%
Cluster & HA	66 - 104	11% - 12%
Backup	44 - 108	10%
Tune	89 - 98	9% - 10%
Management	67 - 110	9 - 10%

Top Causes of Project Delays

### Hardware

- Troubleshooting and tuning production environment 45%
- Integration, configuration and testing of the infrastructure 45%
- Installation, cabling and network access for the environment 29%

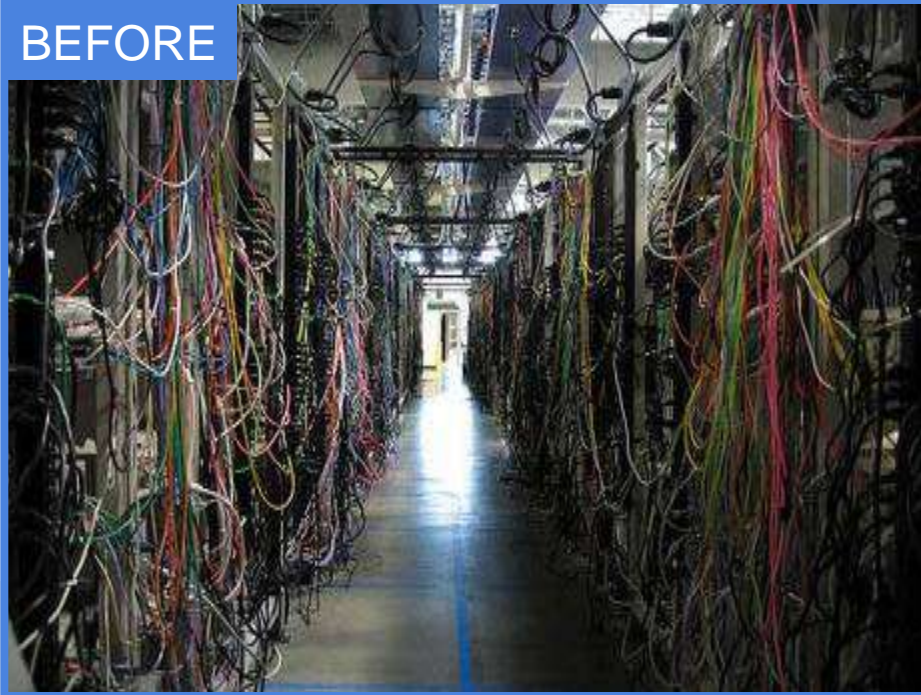
### Software

- Integration, configuration and testing of applications 41%
- Integration, configuration and testing of middleware 35%
- Configuration, build and deployment of applications 34%

From a commissioned study conducted by Forrester Consulting on behalf of IBM

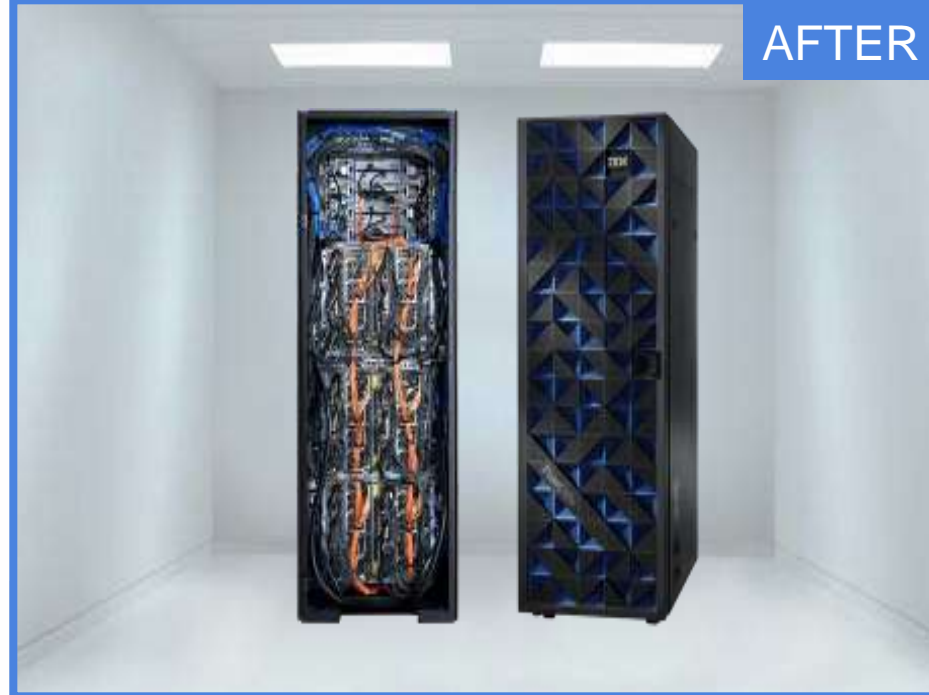
# Doing It The Better Way....

BEFORE



- Silo'd multi-rack infrastructure
- Under utilization
- Complex operations
- 1000s of points of maintenance
- Slow to provision, configure, and change

AFTER



- Pre-integrated infrastructure
- High density reduces space & energy
- Integrated & streamlined operations
- Simplified rack level maintenance
- Provisioned, changed in minutes

# Let's view IT from a new perspective

*Systems with integrated expertise and built for cloud*



## Integrated by Design

*Deeply integrate and tune hardware and software*



Get up and running faster: from 4 – 7 months to hours!

Streamline development and operations: months to minutes (with repeatability)

Optimize ongoing effort: single management console; no downtime upgrades!

## Built-in Expertise

*Capture and automate what experts do*

## Simplify the Experience

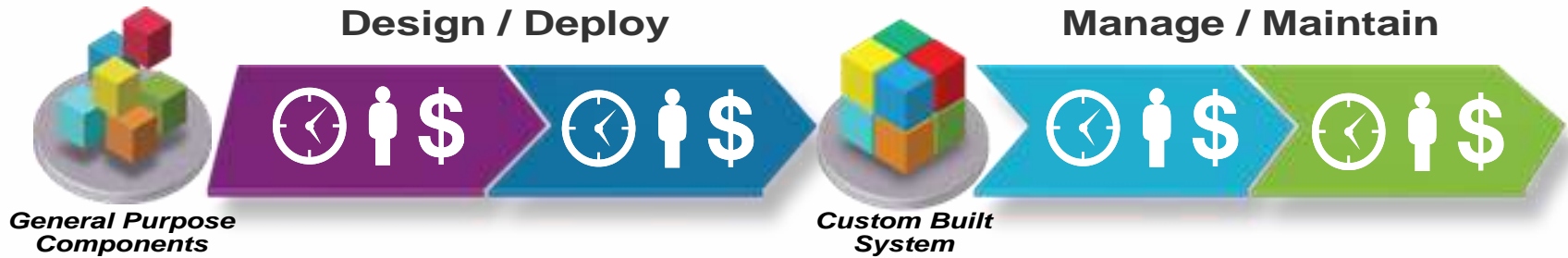
*Make every part of the IT lifecycle easier*

# PureSystems

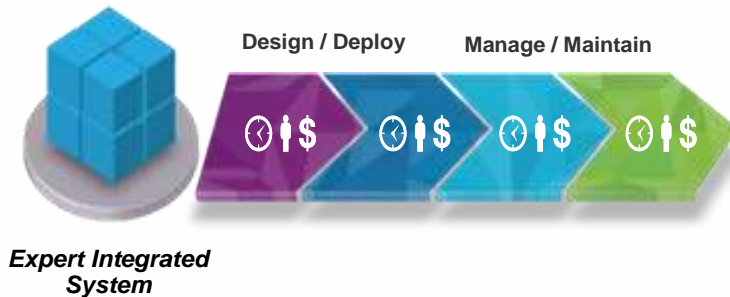


# PureSystems simplify the entire systems lifecycle

**Traditional:** *Time and effort is spent tuning general purpose components*



**PureSystems:** *Reduced Time, Cost and Risk*



**PureFlex**

**PureApplication**

**PureData**

# PureSystems Family: Introducing PureData!

## 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



# Data systems need to be optimized for different data workloads



Order Management

Sales Analysis

Real Time  
Fraud Detection

Customer Analysis

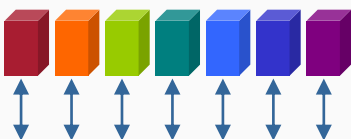


Transaction  
Processing

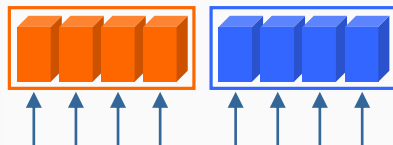
Reporting  
and Analytics

Operational  
Analytics

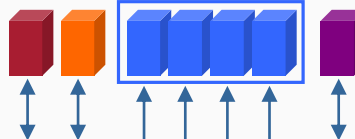
Any Structure  
Analytics



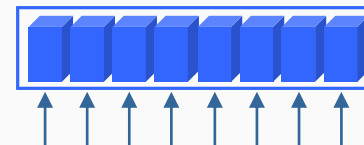
Scalable  
Transactional  
Database



Analytics  
Data Warehouse



Operational  
Data Warehouse



Map-Reduce  
Distributed File  
System

# 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

Simplicity & Optimization Out-of-the-box

# PureData



**Data Platform**

- *Select model*
  - *for Transactions*
  - *for Analytics*
  - *for Operational Analytics*
  - *for Hadoop (announced)*
- *Select size*
- *Load data*

---

*Delivering Data Services*

# Smart Consolidation of Application and Data Services

## PureApplication

## PureData



Next best offer  
Coupon

E-commerce  
Catalog & Cart

Order  
management

Credit card  
management

Customer  
Analysis

Real Time  
Fraud  
Detection

Analytics & Transactional  
Applications

Transactional  
Data Services

Analytic  
Data Services

PureApplication  
System

PureData System

for  
Transactions

for  
Analytics

for  
Operational Analytics

for  
Hadoop

# IBM PureData System for Transactions

*Optimized exclusively for transactional data workloads!*

## PureData System for Transactions

*Delivering data services  
for transactions*



### Speed

- Industry leading DB2 performance
- Database node recovery in seconds<sup>1</sup>

### Simplicity

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

### Scalability

- Scaling up to 30x<sup>3</sup>
- Designed to expand from small to medium to large configurations with no planned system downtime

### Smart

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

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.



# Three Configurations Available



Configurations	Upgrade		Upgrade
	T1500-96 Small ¼ Rack	T1500-192 Medium ½ Rack	T1500-384 Large Full Rack
Chassis	1	1	2
Compute Nodes (16 cores per node)	6	12	24
CPU Cores	96	192	384
# of clusters/ databases	3/ 30	6/ 60	12/ 120
Memory	1.5 TB	3.1 TB	6.1 TB
V7000 Storage Unit (each unit has: 18 x 900GB HDD, 6 x 400 GB SSD)	1	2	4
V7000 Storage Expansion (each unit has: 18 x 900GB HDD, 6 x 400 GB SSD)	1	2	4
User Capacity	<b>18.6 TB</b>	<b>37.2 TB</b>	<b>74.4 TB</b>
Raw SSD Storage	<b>4.8 TB</b>	<b>9.6 TB</b>	<b>19.2 TB</b>
Raw HDD Storage	<b>32.0 TB</b>	<b>64.0 TB</b>	<b>128.0 TB</b>



# Reduce time, skill and effort required to deploy reliable and highly scalable systems

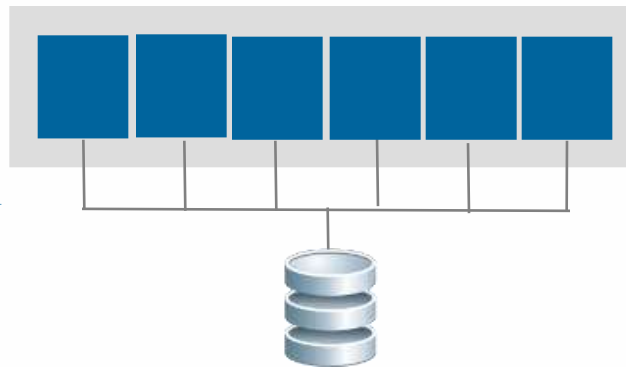
*Uninterrupted access to data with consistent performance*

## 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 - built-in expertise

### In minutes,

1. Just specify cluster name, description and topology pattern

# Simple deployment of database cluster instances and databases

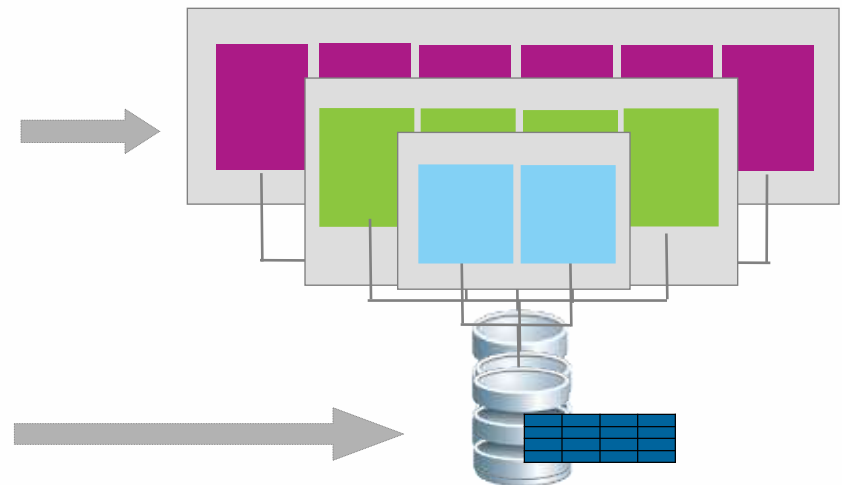
*Deploy topology and databases in minutes using patterns*

## Topology patterns

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

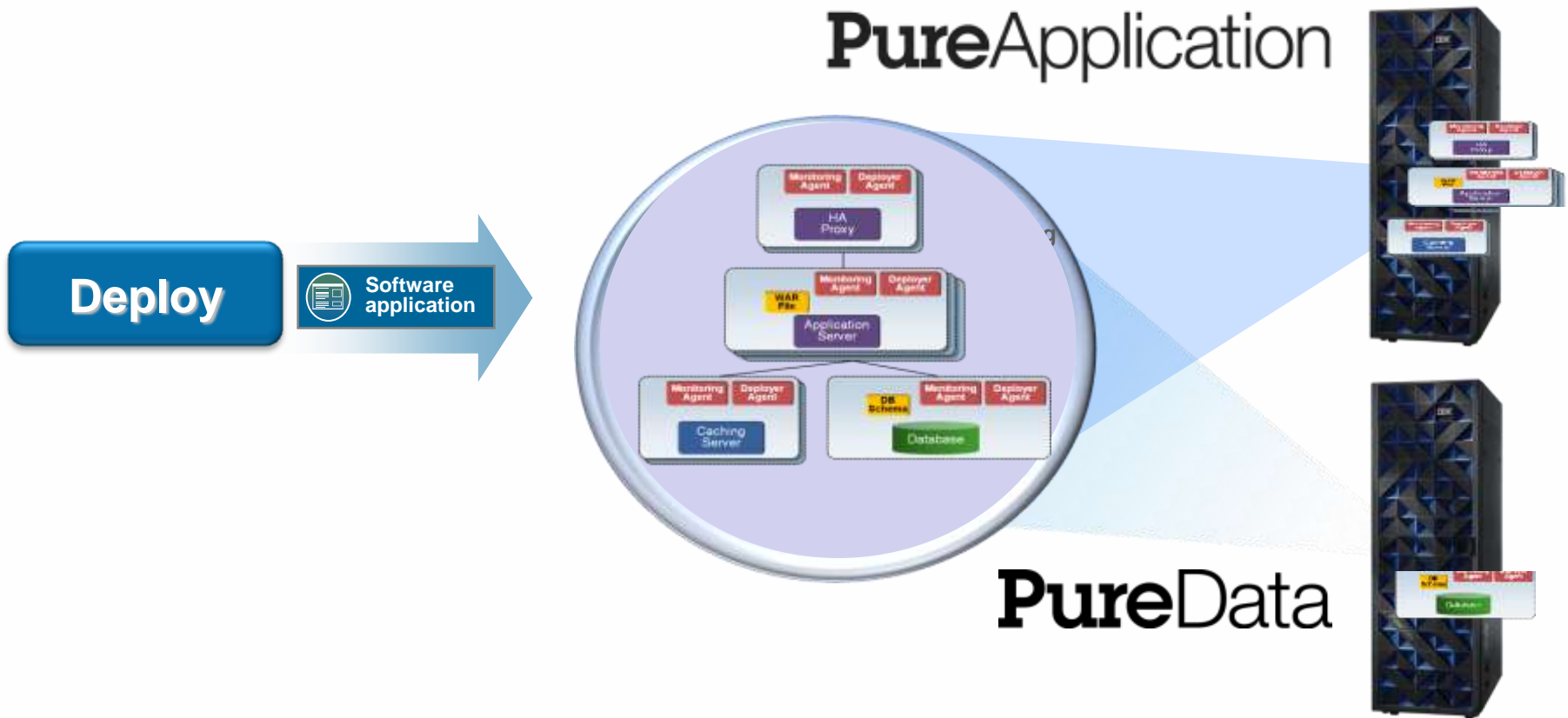
## Database patterns

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



# Pattern based deployment across systems

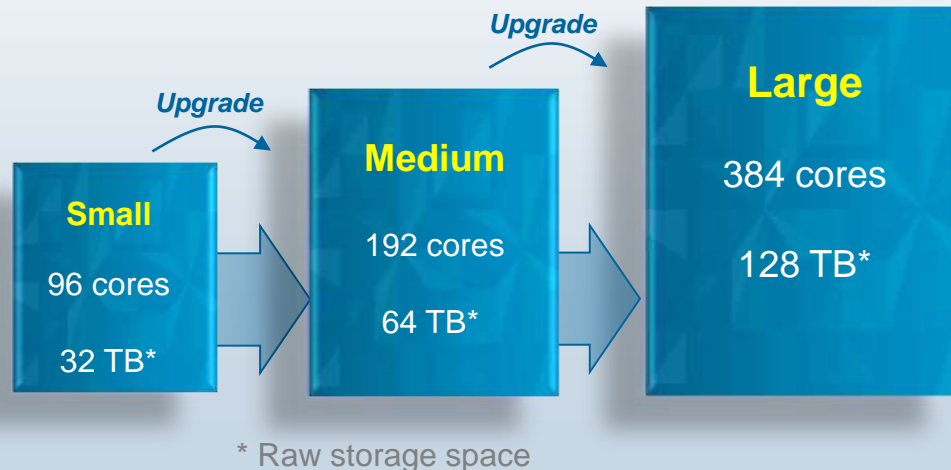
An application pattern can automatically deploy a new database to the PureData System



# Three configurations from which to choose

- High density, high scale computing
- Integrated solid state and disk storage
- High speed networking
- Scalable database management
- Integrated systems management

## PureData System for Transactions



<sup>1</sup> Based on internal DB2 10 tests and reported client experience from 28 Sep 2011 to 07 Mar 2012

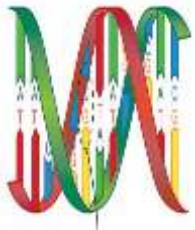
\* Raw storage capacity

# PureData Systems for Transactions in Action!



## 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

# Ready for DB2 applications work with PureData TX Unchanged





# 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 workload management
- Integrated backup on the system
- 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 PureData System for Operational Analytics

## Built-in Expertise

- Real-time operational analytics
- Adaptive compression
- Integrated backup
- Built-in temporal analytics
- Fully parallel, optimized In Database Analytics

## Integration by Design

- Server, Storage, Database in one easy to use package
- Preset and configured for top performance, throughput, and efficient resource utilization
- Enterprise-class security, availability and platform management

## Simplified Experience

- Fast time-to-value
- Solution-level management with integrated console
- Automated updates for faster maintenance
- Standard interfaces to best of breed Analytics, BI, and integration tools
- Built-in analytics capabilities allow users to derive insight from data quickly



# PureData System for Operational Analytics hardware and capacity sizes



- 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

# PureData System for Operational Analytics

## Data Integration

- Ab Initio
- Cloudera
- Composite Software
- IBM Big Insights
- IBM Information Server
- IBM InfoSphere Streams
- IBM InfoSphere Change Data Capture
- Informatica
- Oracle Data Integrator
- Oracle GoldenGate
- SAP Business Objects



## Reporting & Analysis

- IBM Cognos
- IBM SPSS
- IBM Unica
- Information Builders
- Kalido
- KXEN
- Microsoft Excel
- MicroStrategy
- Oracle OBIEE
- SAP Business Objects
- SAS
- Actuate



# Part of the IBM Big Data Platform



*Workload Optimized Solutions for all your analytic needs*

## Analytic Applications

BI / Reporting | Exploration / Visualization | Functional App | Industry App | Predictive Analytics | Content Analytics

## IBM Big Data Platform

Visualization & Discovery

Application Development

Systems Management



Accelerators

Hadoop System

Stream Computing

Data Warehouse



Information Integration & Governance

## PureData System for Operational Analytics

New



# IBM PureData System for Analytics



*Optimized exclusively for analytic data workloads*

## PureData System for Analytics

*Delivering data services  
for analytics*



### Speed

- 10-100x faster than traditional custom systems\*
- Patented MPP hardware acceleration (Massively Parallel Processing)

### Simplicity

- Data load ready in hours
- No database indexes
- No tuning
- No storage administration

### Scalability

- Peta-scale data capacity

### Smart

- Designed to runs complex analytics in minutes, not hours
- Richest set of in-database analytics



# 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



# PureData System for Analytics Hardware Overview



- 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 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

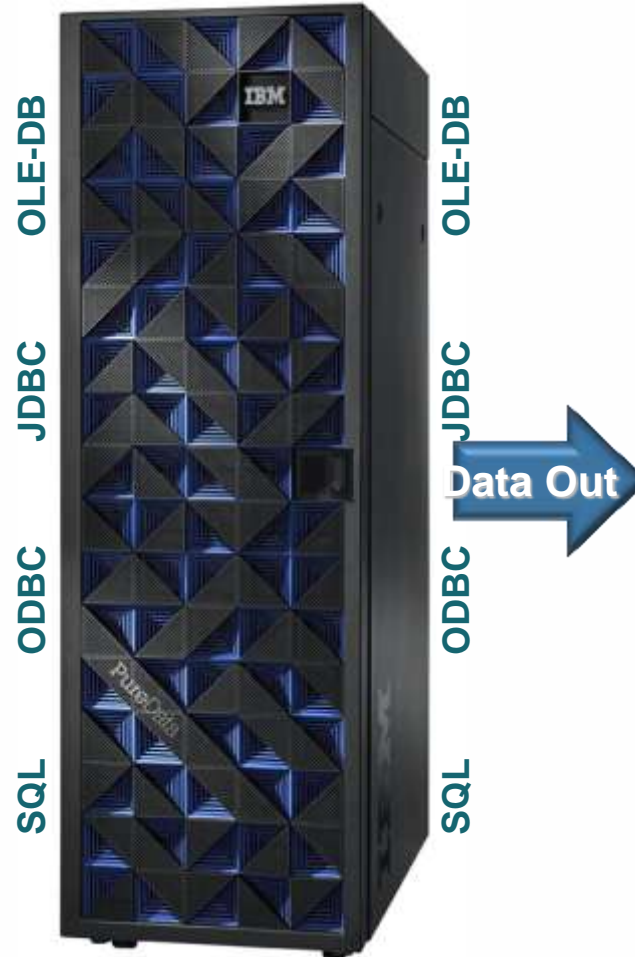
- Power Requirements: 7.6 kW
- Cooling Requirements: 7.8 kW

\*\*: 4X compression assumed

# PureData System for Analytics

## *Data Integration*

- IBM Big Insights
- IBM Information Server
- IBM InfoSphere Streams
- Ab Initio
- Cloudera
- Composite Software
- Informatica
- Oracle Data Integrator
- Oracle GoldenGate
- SAP Business Objects



## *Reporting & Analysis*

- IBM Cognos
- IBM SPSS
- IBM Unica
- Information Builders
- Kalido
- KXEN
- Microsoft Excel
- MicroStrategy
- Oracle OBIEE
- SAP Business Objects
- SAS
- Actuate

# If this were easy, everyone would already be leveraging big data

“Big Data offers big business gains but hidden costs and complexity present barriers that most organizations will struggle with”

- The Cost of Big Data, Eric Savitz, Forbes 5/2012

- Open source Apache Hadoop for enterprise usage is incomplete
- Hadoop skills are in short supply
- Custom built solutions lack integrated cluster management
- Requires integration effort within the existing analytic ecosystem
- Most integrated solutions do not help with archival

# 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





# Benefits of IBM PureData System for Hadoop

Built-in 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*

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*<sup>2</sup>
- **Delivered with more robust security**  
*than open source software*
- **Architected for high availability**



# LiveAssist for 30 days at your site at no charge

## Rapid deployment of a PureData at your site

Day one: On-site installation of PureData

Day 2 – 30:

- Guided demonstration of business value
- Execution of on-site **PureStart** service:
  - **Education** on PureData capabilities
  - **Validation** of first workload to be deployed
  - **Assistance** with data migration to PureData
  - **Readiness** review for best performance at your site
- Access to a lab advocate for usage questions & advice
- Single line of IBM support and maintenance

Note: The service begins when the system arrives at your site.



# No upfront costs and defer payments for 90 days<sup>1</sup>

## IBM Global Financing makes it even simpler with flexible financing options

- Turn upfront costs into low monthly payments and conserve your cash for other strategic needs
- Accelerate PureSystems' cash flow breakeven point and better align cash outlays to anticipated benefits
- Drive your total cost of ownership up to 16% lower through fair market value leasing<sup>2</sup>
- Reduce risk of technology obsolescence and increase flexibility through our mid and end of lease options



IBM Global Financing offerings are provided through IBM Credit LLC in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates and availability are based on a client's credit rating, financing terms, offering type, equipment and product type and options, and may vary by country. Non-hardware items must be one-time, non-recurring charges and are financed by means of loans. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice and may not be available in all countries.

1. 90 day deferral subject to client credit approval. Interest accrues during deferral period. Restrictions may apply. Contact IBM for more details.

2. Represents Net Present Value savings over a 36 month period for an IT infrastructure Fair Market Value lease, "best credit" customer. Current IBM Global Financing monthly rates for IBM hardware used to calculate PV savings.

[ibm.com/financing/us/eis](http://ibm.com/financing/us/eis)

# 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**

# Questions?

Thank you!!!



© International Business Machines Corporation 2013

International Business Machines Corporation New Orchard Road Armonk, NY 10504

IBM, the IBM logo, PureSystems, PureFlex, PureApplication, PureData and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide.

A current list of IBM trademarks is available on the Web at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)

All rights reserved.



IBM, the IBM logo, ibm.com, DB2, POWER, Power7, and Storwize are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Other company, product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

All information contained in this document is subject to change without notice.

The client examples described are presented as illustrations of how those clients have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by client.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices are taken from publicly available information, including vendor announcements and vendor worldwide webpages. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The actual performance that any user will experience will vary depending upon 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 throughput or performance improvements equivalent to the ratios stated here.

Starting price may not include a hard drive, operating system or other features. Prices are subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

The products described in this document are NOT intended for use in implantation, life support, space, nuclear, or military applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change IBM product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of IBM or third parties.

THE INFORMATION CONTAINED IN THIS DOCUMENT IS PROVIDED ON AN "AS IS" BASIS WITHOUT WARRANTY OR CONDITION OF ANY KIND, WHETHER EXPRESS OR IMPLIED. In no event will IBM be liable for damages arising directly or indirectly from any use of the information contained in this document.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

© IBM Corporation 2013. All rights reserved.