IBM PureData System for Operational Analytics

An integrated, high-performance data system for operational analytics

Highlights

- Provides an integrated, optimized, ready-to-use system with built-in expertise for operational analytics
- Delivers outstanding performance and throughput for in-database analysis of large data sets that include both historic and operational data
- Continuously ingests data to support near-real-time responsiveness to dynamic business environments
- Designed as a modular, scalable system that can grow with your business
- Designed to handle more than 1,000 concurrent operational queries[1]
- Integrated and simplified monitoring and maintenance
- Compatible with market-leading analytic and BI tools, applications and infrastructure
- Powered by IBM DB2–based IBM InfoSphere Warehouse software and IBM POWER7 processor-based IBM Power Systems servers

Businesses across industries need actionable insight into their operations to gain a competitive advantage. To gain this insight, they often must analyze large data sets comprising both historic and current (or operational) data. However, a high percentage of queries for operational analytics systems—often up to 80 percent—are interactive lookups focused on data about a specific customer, account or patient. These operational queries can originate from call centers, mobile sales apps, real-time fraud detection systems or other applications that support real-time decision making.

To deliver the correct information as rapidly as possible, the data warehouse supporting these systems must be optimized for the right balance of analytics performance and operational query throughput.

The IBM® PureData™ System for Operational Analytics—a member of the IBM PureSystems™ family—helps organizations meet these complex requirements with an expert integrated data system that is designed and optimized specifically for the demands of an operational analytics workload. Built on IBM Power Systems™ servers with IBM System Storage® and powered by IBM DB2®–based InfoSphere® Warehouse software, the system is a complete, out-of-the-box solution for operational analytics that provides both the simplicity of an appliance and the flexibility of a custom solution. Designed to handle more than 1,000 concurrent operational queries,[2] it delivers mission-critical reliability and scalability with outstanding performance.

The IBM PureData System for Operational Analytics

The PureData System for Operational Analytics enables IT departments to easily deploy, optimize and manage data-intensive workloads for operational analytics. It delivers exceptional value in three ways:

- **Built-in operational analytics expertise,** based on years of IBM experience and best practices from thousands of client engagements, is embedded into the system to provide a complete solution that can rapidly deliver value.
- **Integration by design** of software, server, storage and networking results in factory-optimized systems designed for fast time-to-value, efficiencies and high performance.
- **A simplified experience** from design to purchase to maintenance helps reduce total cost of operations.

Turn data into insight with in-database analytics

With in-database analytics, you run analytics on your data where it resides—in the warehouse. This eliminates the time, cost and risk associated with copying data out of the warehouse to analyze it.

Using multidimensional cubing services, this PureData System delivers rapid insight into high volumes of fast-moving data. Users can create, edit, import, export and deploy cube models over the relational warehouse schema to analyze multiple business variables. Cubing services help optimize performance for online analytical processing (OLAP) queries, providing more power for users to analyze data and generate business insight that can enhance both profitability and customer satisfaction. Overall, performance for complex queries is up to 3.3 times faster than the previous release of InfoSphere Warehouse software.[3]

Powerful data mining capabilities also enable integrated analytics of both structured and unstructured data in the system. Standard data mining models (including clustering, associations, classification and prediction) are supported and can be developed via drag and drop in an intuitive design environment. The data mining models are executed in the production environment to provide real-time scoring of data records. Additionally, rich presentation components are provided to enable visual analysis of data mining results.

Get real-time analytical insights with continuous data ingest

Continuous data ingest capabilities enable organizations to transparently load data from external sources into the PureData System without downtime—supporting real-time business analysis and decision making during the loading process. The continuous data ingest feature allows IT departments to load data across multiple threads at the same time to get the data in very quickly, while also dynamically switching between the various external load sources to help maximize resource utilization. It helps eliminate the latency created by batch-loading data on infrequent schedules, making it extremely valuable for business users who need current operational data in the warehouse.

Grow as needed with a flexible and efficient system design

A modular, flexible system design enables organizations to acquire the PureData System for Operational Analytics at the correct size for their current needs and scale incrementally up to a petabyte of capacity as their data grows.[4]

The system includes many features that are preset to optimize performance, throughput and resource utilization of operational analytics workloads. It is also designed to significantly reduce disk space requirements and improve query performance. Using Adaptive Compression features, the PureData System for Operational Analytics can automatically compress indexes and temporary tables to help reduce storage costs. Data row compression contributes to storage space savings and helps reduce I/O overhead—and the stored pages are also compressed, which further enhances the compression on disk. Because data is compressed, it significantly reduces the I/O requirements and helps improve query response time without the need to frequently reorganize the data. Adaptive Compression can also adapt to changing patterns in the data. Clients have experienced cases of 10x storage space savings via Adaptive Compression.[5]

Rely on highly available, high-performance operational analytics

The fault-tolerant design of the PureData System for Operational Analytics virtually eliminates single points of failure and includes standby server capacity to support continued operations in the event of a hardware failure. With built-in automated workload management features, IT departments can establish and enforce service levels for end users by prioritizing queries from different users and applications and then controlling the number of underlying resources dedicated to those processes.

Key capabilities

Database management

IBM DB2 9.7 or 10 Enterprise Edition

Continuous data ingest

Storage compression with Adaptive Compression

Label- and Row-Based Access Control

IBM Workload Manager

Data movement

IBM SQL Warehousing Tool

IBM InfoSphere Federation Server

Operating system

IBM AIX® 7.1

Analytics

Cubing services

IBM Cognos® Business Intelligence (5 user entitlements)

Text analytics

Intelligent miner

Tooling

IBM PureData System Console

IBM Design Studio

IBM Optim™ Development Studio

Table 1: IBM PureData for Operational Analytics configurations

	Extra Small Foundation rack 1 foundation module	**Small** Foundation rack + 1/3 rack 1 foundation node 1 data node	**Medium** Foundation + 2/3 rack 1 foundation node 2 data nodes	**Large** Foundation + full rack 1 foundation node 3 data nodes
Cores	32	64	80	96
Memory	256 GB	512 GB	640 GB	768 GB
SSD storage	4.8 TB	9.6 TB	12 TB	14.4 TB
HDD storage				
HDD unformatted raw capacity	64.8 TB	151.2 TB	237.6 TB	324 TB
HDD RAID capacity	54 TB	126 TB	198 TB	270 TB
User data capacity - uncompressed	29.7 TB	69.3 TB	108.9 TB	148.5 TB
Primary servers	1	2	3	4
Standby servers	1	2	2	2
Uncompressed disk bandwidth Specification: 3.2 GB/s for foundation node, 6.4 GB/s for data node	3.2 GB/s	9.6 GB/s	16 GB/s	22.4 GB/s
Database disk IOPS	34 KB	57 KB	148 KB	205 KB
Data load rate	Uncompressed: 1,161 GB/h Compressed: 890 GB/h	Uncompressed: 3,484 GB/h Compressed: 2,670 GB/h	Uncompressed: 5,807 GB/h Compressed: 4,450 GB/h	Uncompressed: 8,130 GB/h Compressed: 6,230 GB/h
Database software and tools	IBM data warehousing and analytics software entitlements included			
Processors and operating system	IBM POWER7® with AIX			
Power (watts maximum)	Foundation rack: 6,196	Foundation rack: 6,196 Data rack: 4,647	Foundation rack: 6,196 Data rack: 7,551	Foundation rack: 6,196 Data rack: 10,454
Typical cooling (BTU/hour)	Foundation rack: 14,160	Foundation rack: 14,160 Data rack: 11,601	Foundation rack: 14,160 Data rack: 19,534	Foundation rack: 14,160 Data rack: 27,467
Weight	Foundation rack: 1,450 lbs (658.3 kg)	Foundation rack: 1,450 lbs (658 kg); data rack: 1,250 lbs (567.5 kg)	Foundation rack: 1,450 lbs (658 kg); data rack: 1,650 lbs (749.1 kg)	Foundation rack: 1,450 lbs (658 kg); data rack: 2,150 lbs (976.1 kg)
Rack dimensions (W x D x H)	644 mm (25.4 in) x 1,465 mm (57.7 in) x 2,015 mm (79.3 in), including doors			
Voltage drops/rack	200-240 V ac; frequency: 47-63 Hz			
Drops/rack	4 x 30A	4 x 30A and 4 x 60A	4 x 30A and 4 x60A	4 x 30A and 4 x 60A
Safety	IEC 60950-1; UL 60950-1; CSA 60950-1			
Emissions	CISPR 22; CISPR 24; FCC, CFR 47, Part 15 (US); VCCI (Japan); Directive 2004/108/EC (EEA); ICES-003, Issue 4 (Canada); ACMA radio communications standard (Australia, New Zealand); CNS 13438 (Taiwan); Radio Waves Act, MIC Rule No. 210 (Korea); Commodity Inspection Law (China); TCVN 7189 (Vietnam); MoCI (Saudi Arabia); SI 961 (Israel); GOST R 51318.22, 51318.24 (Russia).			

Table 2: Expansion options

	1/3 rack 1 data module	**2/3 rack** 2 data modules	**Full rack** 3 data modules
Cores	32	48	64
Memory	256 GB	384 GB	512 GB
SSD storage	4.8 TB	7.2 TB	9.6 TB
HDD storage			
HDD unformatted raw capacity	86.4 TB	172.8 TB	259.2 TB
HDD RAID capacity	72 TB	144 TB	216 TB
User data capacity - uncompressed	39.6 TB	79.2 TB	118.8 TB
Power (watts maximum)	4,647 KW	7,551 KW	10,454 KW
Cooling (BTU/hour)	11,601	19,534	27,467
Weight	1,250 lbs (567.5 kg)	1,650 lbs (749.1 kg)	2,150 lbs (976.1 kg)
Rack dimensions (W x D x H)	644 mm (25.4 in) x 1,465 mm (57.7 in) x 2,015 mm (79.3 in), including doors		

The system is also designed to provide very high throughput and concurrency. With the underlying strengths of IBM DB2 software, the system is designed to handle more than 1,000 concurrent operational queries.[6] Integrated local backup capabilities support rapid backup and recovery without having to move data on or off the system.

The system provides advanced capabilities for data partitioning, giving IT users multiple ways to distribute data across servers for large-scale parallelism and linear scalability. The shared-nothing architecture helps ensure that performance will not degrade as the warehouse grows. Database partitioning for massively parallel processing architecture splits the database across multiple partitions and uses the processing power of multiple servers to satisfy requests for large amounts of information. SQL statements are automatically decomposed into subrequests that are executed in parallel across the partitions. Results of the subrequests are joined to provide final results at extremely fast speeds.

The PureData System for Operational Analytics offers significant performance enhancements for business intelligence (BI) queries. The new zigzag join feature means DBAs can significantly reduce the time for complex multidimensional business queries compared to the previous release of InfoSphere Warehouse software. Enhanced query joins and optimizer enhancements help to further increase performance of other analytic queries, reducing the need for additional indexes. Materialized query tables (MQTs) can provide enhancement for queries by precomputing and storing results of a query. The query optimizer of the system transparently redirects queries from base tables to matching MQTs, thereby improving the performance of complex aggregate queries.

Simplify deployment, maintenance and operation

Intuitive user interfaces support solution-level management of the PureData System with service packs, automated patch maintenance and firmware updates. Installation services, one-call support and hardware and software maintenance are also available.

The PureData System for Operational Analytics comes in multiple configurations, which can be sized to an organization's specific needs (see Tables 1 and 2). The IBM solutions portfolio for the PureData System is supported by a wide range of market-leading business partners including complementary technology partners, resellers, systems integrators and service providers. For a complete list or to find out if a particular company or solution is part of our program, please visit **ibm.com**/partnerworld or contact your IBM representative.

Meet the needs of business and IT leaders by design

The IBM PureData System for Operational Analytics is designed, built and tuned to help organizations drive smarter business outcomes at the speed of business. From IT professionals struggling to meet changing business requirements for high-performance analytics capabilities to business executives who need information to produce fast, accurate answers to critical business questions, the PureData System for Operational Analytics provides a single, trusted version of truth—whenever it is needed.

See for yourself: Take a test drive at no charge

Organizations can try out the PureData System for Operational Analytics through the IBM PureExperience™ program. This program is available at no charge and allows you to test drive the system with your own data. The program offers on-site installation and demonstration of business value, education and data migration services, use of the system for a specified period and a single line of support. For details on this program and to see what is available in your area, please visit **ibm.com**/PureExperience or contact your IBM representative.

For more information

To learn more about the IBM PureData System for Operational Analytics, contact your IBM representative or IBM Business Partner or visit **ibm.com**/puredata

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
October 2012

WAD12351-USEN-01