IBM PureData System for Analytics N3001

Powered by Netezza technology

Highlights

- Easy to deploy and manage; dramatically simplifies your data warehouse and analytic infrastructure

- Arrives ready to go with IBM Fluid Query, plus data integration, business intelligence and Hadoop starter kits

- Protection of all data from unauthorized access

- Integrated platform supporting thousands of users, unifying data warehouse, Hadoop, and business intelligence with advanced analytics

- Delivered with data integration, business intelligence and Hadoop starter kits

- Powered by Netezza technology

To gain competitive advantage, organizations must rely on sophisticated analytics mining large volumes of data. Yet many companies need faster time-to-value for new analytical capability, as well as maintaining service level agreements on existing analytics applications. IBM® PureData™ System for Analytics N3001, powered by Netezza technology, provides faster performance, is big data and business intelligence (BI) ready, and provides advanced security all in a wider range of appliance models. With IBM PureData System for Analytics N3001, IBM is again changing the game for data warehouse appliances.

IBM PureData System for Analytics is a high-performance, scalable, massively parallel system that enables clients to gain insight from their data and perform analytics on enormous data volumes. Realizing business value from today's volumes is made simpler and faster, because the data is more easily accessible. The N3001 model comes ready to deliver extra value with included business intelligence and Hadoop starter kits. These tools complement the IBM PureData System for Analytics N3001's ability to run complex analytics on very large data volumes, at much faster rates than competing solutions.

IBM PureData System for Analytics is a purpose-built, standards-based data warehouse and analytic appliance that integrates database, server, storage and advanced analytic capabilities into a single, easy-to-manage system. Designed for rapid and deep analysis of data volumes scaling into the petabytes, it delivers insight never before thought possible, at a low cost of ownership. IBM PureData System for Analytics N3001 family ranges widely from entry point to petabyte scale, covering a broad range of data capacity needs.

PureData System for Analytics delivers the proven performance, scalability, intelligence, and simplicity your business needs. It requires minimal administration and tuning both for the initial deployment as well as ongong maintenance, which translates into a lower total cost of ownership (TCO).

The analytics opportunity

Deep, sophisticated analytics on large data volumes are integral to enterprises in a competitive economy, giving them an edge over the competition. However, most organizations are challenged by both the time-to-market on new analytic capability, as well as maintaining service level agreements on existing analytics. IBM PureData System for Analytics shifts the focus to simplicity, instead of unproductively managing complexity.

Strategic analytics should not be complicated to deliver and difficult to manage.

IBM PureData System for Analytics is a scalable, hardware-accelerated, massively parallel system that enables clients to gain insight from enormous data volumes, 10–100 times faster than they can with traditional systems[1] without the need to copy the data into a separate analytics server.

Figure 1: PureData System for Analytics

Best practice: Routing queries to the data

IBM Fluid Query is the capability that unifies data access across the logical data warehouse and big data ecosystems. Users and analytic applications need access to data in a variety of data repositories and platforms without concern for the data's location or access method or the need to rewrite a query. IBM Fluid Query is the capability for a data store to route a query (or even part of a query) to the correct data store within the logical data warehouse so that the query can flow to the data, not the data flow to the query.

No matter where a user connects within the logical data warehouse, they can access all data through the same, standard API/SQL access. IBM Fluid Query is the foundation of the logical data warehouse, giving users the ability to combine their data, even if spread across various sources in a fast, agile manner to drive analytics and deeper insight, without understanding how to connect multiple data stores, use different syntaxes or APIs, or change their application.

IBM Fluid Query 1.0, included with IBM PureData System for Analytics, provides access to data in Hadoop from IBM PureData System for Analytics appliances. IBM Fluid Query 1.0 enables the fast movement of data between Hadoop and IBM PureData System for Analytics appliances. Enabling query and data movement, IBM Fluid Query 1.0 connects those appliances to common Hadoop systems: IBM BigInsights for Apache™ Hadoop®, Cloudera, and Hortonworks. IBM Fluid Query 1.0 allows queries against PureData System for Analytics, Hadoop or both by merging results from PureData System for Analytics database tables and Hadoop data sources thus creating powerful analytic combinations.

IBM Fluid Query 1.0 enables your existing PureData System for Analytics applications to gain insight from even more data. You can now run your existing queries, reports, and analytics against data on Hadoop, in addition to the data in your appliance.

Fast. Scalable. Smart. Simple. Completely integrated.

IBM PureData System for Analytics is designed specifically for running complex analytics on very large data volumes with faster execution times than competing solutions. It delivers the proven performance: scalability, intelligence, and simplicity that organizations need to leverage their data.

Fast

IBM PureData System for Analytics N3001 delivers a performance advantage over other analytic options. This comes from its unique asymmetric massively parallel processing (AMPP)™ architecture that combines open IBM blade servers and disk storage with IBM's patented, hardware-accelerated data filtering, using field programmable gate arrays (FPGAs). This combination delivers fast query performance on analytic workloads supporting thousands of business intelligence and data warehouse users, providing sophisticated analytics for satisfying business requirements.

Scalable

With the IBM PureData System for Analytics solution, organizations can deploy the right-sized environments for their data volumes and workloads, and be confident that as data volumes grow, larger systems can be deployed quickly and easily. The IBM PureData System for Analytics N3001 family of seven different configurations, starts with a data capacity of 16 TB (new N3001-001) and can grow to well over a petabyte for an eight-rack system (new N3001-080), assuming a 4X compression rate.

IBM PureData System for Analytics provides near linear performance scalability as the size of the appliance grows, which means that organizations can pick the appropriate sized appliance to meet both their data volume and performance requirements. This is accomplished with predictable, scalable performance with no need to add significant resources to manage and maintain the appliance as data volumes grow.

Smart

IBM PureData System for Analytics dramatically simplifies analytics by consolidating all analytic activity to one place, where the data resides. Moving analytics to the IBM PureData System is straightforward with IBM's embedded analytic platform. With support for PMML 4.0 models, data modelers and quantitative teams can operate on the data directly inside the appliance instead of having to off load massive data volumes to a separate infrastructure, and then have to deal with the associated data preprocessing, transformation, and movement.

Data scientists can build their models using all the enterprise data, and then iterate through different models much faster to arrive at the best solution. Once the model is developed, it can be seamlessly executed against the relevant data in the appliance. Prediction and scoring can be done where the data resides. Users can get their predictive scores in near real-time, helping operationalize advanced analytics and making it available throughout the enterprise.

Included with every PureData System for Analytics system is IBM Netezza Analytics software. IBM Netezza Analytics offers a built-in analytical infrastructure and extensive library of statistical and mathematical functions, supporting a breadth of analytic tools and programming languages, including Open Source R. It is delivered with a library of more than 200 prebuilt, scalable, in-database analytic functions that execute analytics in parallel while abstracting away the complexity of parallel programming from the developers, users and DBAs.

The Netezza Analytics functionality also includes in-database geospatial analytics that are compatible with the industry-standard ESRI GIS formats. This enables easy integration with existing geospatial analytic environments. In addition, if models are developed using SPSS Modeler or SAS, IBM Netezza Analytics will accelerate the development and scoring of these models.

The IBM PureData System for Analytics N3001 brings advanced security to your data in this insecure world. Building on the appliance simplicity model, all data is stored on self-encrypting disk (SED) drives, providing security while not impacting performance. The protection provided by the SED implementation supports the leading industries in security compliance — health care, government, and the financial sectors. This system utilizes strong authentication that prevents threats due to unauthorized access, based on the industry-standard Kerberos protocol.

Simple and completely integrated

The IBM PureData System for Analytics N3001 also offers a great value bundle as complementary software licenses to use in conjunction with the appliance. Data movement, reporting, analytic tools, and Hadoop licenses make for a full service offering.

Included software entitlements:

- IBM Cognos® Business Intelligence—five Analytics User licenses, one Analytics Administrator license.
- IBM DataStage (280 PVUs)—2 concurrent Designer Client licenses and IBM InfoSphere Data Click (with PureData System for Analytics as a source or target).
- IBM BigInsights for Apache Hadoop, software licenses to manage around 100 TB of Hadoop data.
- Two non-production user licenses for the IBM InfoSphere Streams Developer Edition.

All of these new features are delivered with the same simplicity and ease-of-use that distinguish all IBM PureSystems® family offerings and what sets the IBM PureData System for Analytics apart.

As an appliance, the integration of hardware, software and storage is done for you, leading to shorter deployment cycles and industry leading time-to-value for business intelligence and analytic initiatives. The appliance is delivered ready-to-go for immediate data loading and query execution. The appliance integrates with leading ETL, BI and analytic applications through standard ODBC, JDBC and OLE DB interfaces.

Included with every system, the PureData System for Analytics Performance Portal provides a web-based GUI that helps administrators monitor and manage hardware, administer database objects, configure workload management, view active sessions and monitor system resource utilization for capacity planning. The portal provides a consolidated administrative interface supporting PureData Systems for Analytics from one, easy-to-use access point.

IBM PureData System for Analytics is architected for high availability. All components are internally redundant, and the failure of a processing node (S-Blade) causes no significant performance degradation for a robust, production-ready environment from the moment the appliance is installed in your data center.

IBM eliminates complexity at every step so you can redirect valuable resources to initiatives that will positively impact the bottom line.

The best value

IBM PureData System for Analytics is a cost-effective analytics option. It requires minimal ongoing administration or tuning, minimizing internal resources as well as implementation costs, for an extremely low total cost of ownership. The performance and scalability of IBM appliances is available immediately, without requiring tuning, indexing, or aggregated tables.

IBM offers your company fast time-to-value for important BI and analytic initiatives. Your organization is armed with more accurate intelligence to react quickly and accurately to opportunities and risks as they may present themselves.

At a time when companies need flexibility to react to changing market conditions and growing analytic demands, an uncomplicated, easy-to-maintain system that runs fast and analyzes your growing data volumes makes sense.

The original data warehouse appliance

Patented hardware acceleration

IBM PureData System for Analytics N3001 adheres to IBM's basic principle of moving processing to the data, not moving the data to the processors. Each IBM PureData System for Analytics contains multiple snippet blades or S-Blades, where SQL query code segments (or "snippets") and complex analytic processes are executed. The S-Blades are intelligent processing nodes that make up the massively parallel processing appliance engine. Each S-Blade is an independent server that contains multi-core Intel CPUs, IBM's unique FPGAs, and gigabytes of RAM Additionally, dedicated storage devices work concurrently with the blades to deliver peak performance.

Software

Database
IBM Netezza Platform Software (NPS) v7.2 or greater

Operating system
Red Hat Linux Advanced Server 6.5

IBM Fluid Query 1.0
Run PureData System for Analytics queries against Hadoop data; Move PureData System for Analytics data quickly to Hadoop file systems, and quickly move Hadoop data to PureData System for Analytics

Supported APIs
SQL, OLE DB, ODBC 3.5, JDBC 3.0 Type 4

SQL standards
SQL-92 compliant, with SQL-99 extensions

Programming languages
Java, Python, Open Source R, R[3], Fortran, C/C++, Perl, Lua

Netezza Analytics foundation
In-Database Analytics, Open Source R, R[2], Matrix, MapReduce, Geospatial technology with ESRI support

High-speed load/unload
Interoperable with ETL and EAI tools at rates of 10 TB/hour

Backup and restore
Interoperable with IBM® Tivoli®, EMC Legato and Symantec Netbackup

Database portability
From IBM DB2®, IBM Informix®, Microsoft SQL Server, MySQL, Oracle Database, Red Brick, Sybase IQ, Teradata

Included software entitlements

Business intelligence
IBM Cognos Business Intelligence[4] (five analytics users licenses plus one analytics administrator license; the IBM PureData System for Analytics N3001 must be the data source)

Data integration
IBM InfoSphere DataStage[4] (280 PVUs), Designer Client (two concurrent users) and InfoSphere Data Click (all must work with the IBM PureData System for Analytics N3001 as the data source or target.)

Hadoop data services
IBM BigInsights for Apache Hadoop[4] (five virtual servers to manage ~100 TB of Hadoop data; IBM PureData System for Analytics appliances must be a source or target[2])

Real-time analytics
IBM InfoSphere Streams Developer Edition[4] (developer license: two users, not for production use, must work with IBM PureData System for Analytics N3001 appliances)

Additional tools
Windows and web-based DB Admin GUI; CLI and high-speed loading/unloading for IBM AIX®, HP-UX, Linux, Solaris and Windows

The IBM PureData System for Analytics 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 contact your IBM representative.

For clients interested in a smaller entry-point model, please review the IBM PureData System for Analytics N3001-001 data sheet.

Specifications	Single Rack Systems			Multiple Rack systems		
IBM PureData System for Analytics	IBM PureData System for Analytics N3001-002	IBM PureData System for Analytics N3001-005	IBM PureData System for Analytics N3001-010	IBM PureData System for Analytics N3001-020	IBM PureData System for Analytics N3001-040	IBM PureData System for Analytics N3001-080
Racks	1	1	1	2	4	8
Active S-Blades	2	4	7	14	28	56
CPU cores	40	80	140	280	560	1,120
FPGA cores	32	64	112	224	448	896
User data in TB (assumes 4X compression)	32	96	192	384	768	1,536
Power (Watts maximum)/ rack	3,200	4,200	7,600	7,600	7,600	7,600
Cooling - BTU/hour	11,000	14,400	27,000	54,000	108,000	216,000
Rack Weight Kg	620	771	907	907	907	907
Height/rack cm	202	202	202	202	202	202
Depth/rack cm	110	110	110	110	110	110
Width/rack cm	64.8	64.8	64.8	64.8	64.8	64.8
Power	200-240 V, 50Hz/60 Hz (Single phase), 24A	200-240 V, 50Hz/60 Hz (Single phase), 24A	200-240 V, 50Hz/60 Hz (Single phase), 2X 24A	200-240 V, 50Hz/60 Hz (Single phase), 2X 24A, per rack	200-240 V, 50Hz/60 Hz (Single phase), 2X 24A, per rack	200-240 V, 50Hz/60 Hz (Single phase), 2X 24A, per rack
Drops/rack	2	2	4	4	4	4
Safety	US/CSA/EN60950					
Emissions	FCC Part 15, ICES-003, AUS/NZ CISPR 22, VCCI and EN55022 Class A; European immunity: EN55024					

About IBM PureData System for Analytics

The IBM PureData System for Analytics, powered by Netezza technology, integrates database, server and storage into a single, easy-to-manage appliance that requires minimal setup and ongoing administration while producing faster and more consistent analytic performance. The IBM PureData System for Analytics simplifies business analytics dramatically by consolidating all analytic activity in the appliance, where the data resides. Visit: **ibm.com**/PureSystems to see how our family of expert integrated systems eliminates complexity and helps you drive true business value for your organization.

About IBM Data Warehousing and Analytics Solutions

IBM provides the most comprehensive portfolio of data warehousing, information management and business analytic software, hardware and solutions to help clients maximize the value of their information assets and discover new insights to make better and faster decisions and optimize their business outcomes.

Why IBM?

IBM PureSystems offerings combine the flexibility of a general purpose system, the elasticity of cloud and the simplicity of an appliance. They are integrated by design and come with built-in expertise gained from decades of experience to deliver a simplified IT experience. Members of the PureSystems family include: IBM PureFlex® System, IBM PureApplication® System, IBM PureData System for Transactions, IBM PureData System for Operational Analytics and IBM PureData System for Analytics.

The IBM PureData System for Analytics, powered by Netezza technology, integrates database, server and storage into a single, easy-to-manage appliance that requires minimal setup and ongoing administration while producing faster and more consistent analytic performance. The IBM PureData System for Analytics simplifies business analytics dramatically by consolidating all analytic activity in the appliance, right where the data resides, for industry-leading performance. Visit: **ibm.com**/PureData to see how our family of expert integrated systems eliminates complexity at every step and helps you drive true business value for your organization.

For more information

Help IT make the shift to the strategic center of your business. Leverage proven expertise to take the lead. To learn more about IBM PureSystems and the PureData System for Analytics, contact your IBM representative or IBM Business Partner, or visit the following website: **ibm.com**/PureSystems/PureData

Additionally, IBM Global Financing can help you acquire the software capabilities that your business needs in the most cost-effective and strategic way possible. We'll partner with credit-qualified clients to customize a financing solution to suit your business and development goals, enable effective cash management, and improve your total cost of ownership. Fund your critical IT investment and propel your business forward with IBM Global Financing. For more information, visit: **ibm.com**/financing