IBM PureData System for Analytics N2002

Powered by Netezza technology

Highlights

- 128 GB/sec effective scan rate per rack[1] helps tackle big data faster! (3x faster than the N1001[2])

- Fifty percent greater data capacity per rack[3] with no increase in power or cooling

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

- Arrives ready to go with expert integration

- Powerful platform supporting thousands of users, unifying business intelligence and advanced analytics

- Powered by Netezza technology

To gain an edge over the competition, organizations must rely on deep, sophisticated analytics on large volumes of historical data. Yet many companies are challenged by time-to-value for enabling new analytical capability, as well as maintaining service level agreements on existing analytics.

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. Big data volumes are made simpler, **faster and more accessible**. This system, powered by Netezza® technology, is designed specifically for running complex analytics on very large data volumes, orders of magnitude faster than competing solutions.

The PureData System for Analytics delivers the proven performance, scalability, intelligence, and simplicity your business needs. It is a low cost option; requiring minimal ongoing administration or tuning for a low total cost of ownership (TCO).

The IBM approach to data analysis is patented and proven. Minimize data movement, while processing the data at physics speed. Do this in parallel, on a massive scale, inside an easy-to-use data warehouse appliance. Extremely fast and at a low cost. And run business intelligence (BI) and advanced analytics that were previously impossible or impractical.
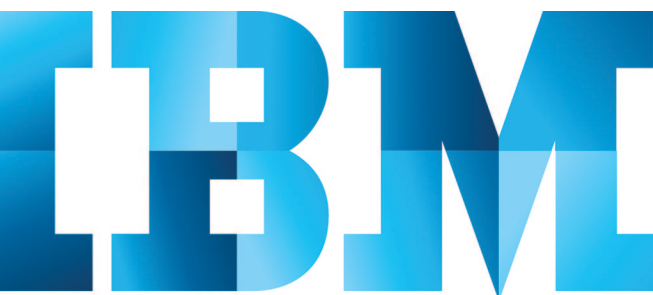
IBM PureData System for Analytics is a purpose-built, standards-based data warehouse and analytic appliance that architecturally integrates database, server, storage and advanced analytic capabilities into a single, easy-to-manage system. Designed for rapid and deep analysis of data volumes scaling into the petabytes, it can help deliver insight never before possible, at a low cost of ownership.

PureData System for Analytics takes the success of Netezza's proven platform technology one step further. The new N2002 model was designed not only for performance, but for data center efficiency. Why? Because today's world depends on more advanced analytics to gain a competitive edge. This increase in capability should come in a simple, agile, powerful system that is also highly efficient. The new PureData System for Analytics offers increased performance with no increase in power requirements. All this comes with less floor space and less resources in the data center. That is innovation—where it matters.

The analytics opportunity

Deep, sophisticated analytics on volumes of historical data are integral to enterprises in an intelligent economy, giving them an edge over the competition. However, most organizations are challenged by both the time-to-market on new analytical capability as well as maintaining service level agreements on existing analytics. The time has come to shift the focus from managing complexity to simplicity and agility.

Strategic analytics should not be complex to deliver and 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 onto a separate analytics server.

Figure 1. PureData System for Analytics

Fast. Scalable. Smart. Simple. Completely integrated.

IBM PureData System for Analytics is designed specifically for running complex analytics on very large data volumes, orders of magnitude faster than competing solutions. It delivers the proven performance, scalability, intelligence, and simplicity that organizations need to dive deep into their data.

Fast

The IBM PureData System for Analytics N2002 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 that supports tens of thousands of business intelligence (BI) and data warehouse users, sophisticated analytics at the speed of thought, and extreme capability.

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 if needed. The IBM PureData System for Analytics half-rack system starts with a data capacity of 96 TB and can grow to well over 700 TB for a four-rack system with built-in hardware compression of four times. This provides 50 percent greater capacity per rack than the N1001 series—without increasing the power or cooling requirements.

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 and no need to add significant additional 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 in one place, right 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 offload it to a separate infrastructure and 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 find the best fit. Once the model is developed, it can be seamlessly executed against the relevant data in the appliance. Prediction and scoring can be done right where the data resides, inline with other processing, on an as-needed basis. Users can get their prediction scores in near real-time, helping operationalize advanced analytics and making it available throughout the enterprise.

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. 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 into existing geospatial analytic environments, and analytics across location and corporate data at the same time and speed. Users can get their prediction scores in near real-time, helping operationalize advanced analytics and making it available throughout the enterprise. In addition, if models are developed using SPSS Modeler or SAS, IBM Netezza Analytics has built-in integration to accelerate the development and scoring of these models as well.

Simple and completely integrated

All of these new features are delivered with the same simplicity and ease-of-use that distinguish all IBM PureSystems®. Based on proven Netezza technology, this simplicity, and ease of development and deployment is what sets the IBM PureData System for Analytics apart. IBM PureData System for Analytics delivers high performance automatically, with no indexing or tuning required.

As an appliance, all of the integration of hardware, software and storage is done for you, leading to shorter deployment cycles and industry leading time-to-value for BI and analytic initiatives. The appliance is delivered ready-to-go for immediate data loading and query execution and 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 to 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 one to many PureData Systems for Analytics from one, easy-to-use access point.

IBM PureData System for Analytics is architected for high availability from the ground up. 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 plugged into your data center.

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

The best value

IBM PureData System for Analytics is a cost effective analytic 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, aggregations, etc.

IBM offers your company fast time-to-value for important BI and analytic initiatives, which will have a positive impact on your bottom line. With IBM on board, your organization is armed with more accurate intelligence to react quickly and accurately to any opportunities or threats the market may present.

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

How is this possible?
Patented hardware acceleration

IBM PureData System for Analytics adheres to IBM's basic principle of moving processing to the data, not moving the data to the processing. 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 engine of the appliance. Each S-Blade is an independent server that contains powerful multi-core Intel CPUs, IBM's unique multi-engine FPGAs and gigabytes of RAM as well as dedicated storage devices—all balanced and working concurrently to deliver peak performance.

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, right where the data resides, for industry-leading performance. Visit: **ibm.com**/PureSystems to see how our family of expert integrated systems eliminates complexity at every step and helps you drive true business value for your organization

About IBM Data Warehousing and Analytics Solutions

IBM provides the broadest and 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.
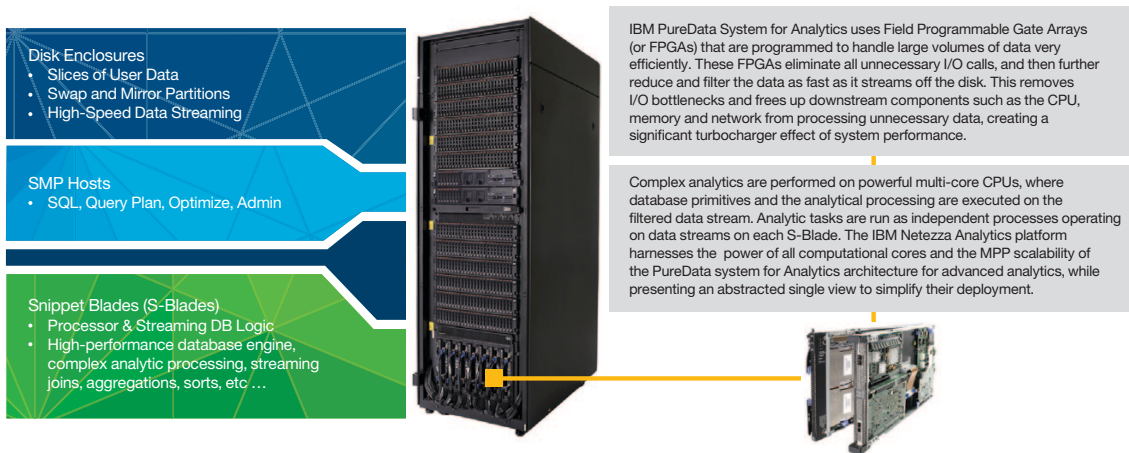
Disk Enclosures
- Slices of User Data
- Swap and Mirror Partitions
- High-Speed Data Streaming

SMP Hosts
- SQL, Query Plan, Optimize, Admin

Snippet Blades (S-Blades)
- Processor & Streaming DB Logic
- High-performance database engine, complex analytic processing, streaming joins, aggregations, sorts, etc …

IBM PureData System for Analytics uses Field Programmable Gate Arrays (or FPGAs) that are programmed to handle large volumes of data very efficiently. These FPGAs eliminate all unnecessary I/O calls, and then further reduce and filter the data as fast as it streams off the disk. This removes I/O bottlenecks and frees up downstream components such as the CPU, memory and network from processing unnecessary data, creating a significant turbocharger effect of system performance.

Complex analytics are performed on powerful multi-core CPUs, where database primitives and the analytical processing are executed on the filtered data stream. Analytic tasks are run as independent processes operating on data streams on each S-Blade. The IBM Netezza Analytics platform harnesses the power of all computational cores and the MPP scalability of the PureData system for Analytics architecture for advanced analytics, while presenting an abstracted single view to simplify their deployment.

Figure 2. Illustration of IBM PureData System for Analytics

Software:

Database:
IBM Netezza Performance Server v7.0.4 or greater

Operating system:
Red Hat Linux Advanced Server 6

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, R, Fortran, C/C++, Perl, Lua

Netezza Analytics foundation
In-Database Analytics, R, Matrix, MapReduce, Geospatial technology with ESRI support

High-speed load/unload:
Interoperable with ETL and EAI tools at rates up to 5 TB/hour

Backup and restore:
Interoperable with IBM® Tivoli®, EMC Legato and Symantec Netbackup, at rates higher than 4 TB/hour

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

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

IBM InfoSphere® BigInsights™ 2 – starter edition – 5 users

IBM InfoSphere Streams 3 – developer license – 2 users

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 N2002-002** data sheet: http://ibm.co/1e4NNDu

Specifications	Single rack systems		Multiple rack systems	
IBM PureData System for Analytics	IBM PureData System for Analytics N2002-005	IBM PureData System for Analytics N2002-010	IBM PureData System for Analytics N2002-020	IBM PureData System for Analytics N2002-040
Racks	1	1	2	4
Active S-Blades	4	7	14	28
CPU cores	64	112	224	448
FPGA cores	64	112	224	448
User data in TB (assumes 4X compression)	96	192	384	768
Power/rack (Watts maximum/rack)	4,200	7,500	7,500	7,500
Cooling - BTU/hour	15,500	27,000	54,000	108,000
Weight/rack kg	771.1	907.2	907.2	907.2
Height/rack cm	202	202	202	202
Depth/rack cm	101.6	101.6	101.6	101.6
Width/rack cm	64.8	64.8	64.8	64.8
Power	220-240 VAC, 50Hz/60 Hz (Single Phase), 24A	220-240/380-415 VAC, 50 Hz (EU models — 3-Phase Wye), 16A 200-208 VAC, 60 Hz (NA models — 3-Phase Delta), 24A 200-240 VAC, 50Hz/60 Hz (Single Phase), 48A	220-240/380-415 VAC, 50 Hz (EU models — 3-Phase Wye), 16A 200-208 VAC, 60 Hz (NA models — 3-Phase Delta), 24A 200-240 VAC, 50Hz/60 Hz (Single Phase), 48A	220-240/380-415 VAC, 50 Hz (EU models — 3-Phase Wye), 16A 200-208 VAC, 60 Hz (NA models — 3-Phase Delta), 24A 200-240 VAC, 50Hz/60 Hz (Single Phase), 48A
Drops/rack	2	2	2	2
Safety	US/CSA/EN60950			
Emissions	FCC Part 15, ICES-003, AUS/NZ C-Tick, VCCI and EN55022 Class A; European Immunity: EN55024			

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

[1] 128 GB/sec scan rate assuming an average of 4X compression across the system. Individual results may vary.

[2] 3x faster performance refers to the query times on both macro-analytic and mixed workload tests as conducted in IBM engineering lab benchmarks where the IBM PureData System for Analytics N200x was shown to be an average of 3x faster than N1001. Individual results may vary.

[3] 50 percent greater capacity when compared to previous model PureData System for Analytics N1001. Power and cooling specifications within 97 percent of previous model PureData System for Analytics N1001.

WAD12363-USEN-01