IBM Integrated Analytics System

Do data science faster

Highlights

- Support users with a high-performance, easy-to-deploy, cloud-ready integrated hardware and software platform.
- Create advanced analytics models with the included IBM® Data Science Experience Local.
- Experience increased data load speed and reduced data querying response time for analytics workloads.
- Achieve seamless interoperability across a hybrid data management infrastructure using the common SQL engine from IBM and built-in data virtualization.

Analytics accelerates decision-making and can make the difference between being an industry leader and being left behind by the competition. However, efforts to provide analytics to an organization are complicated by many factors, including:

- The importance of including your data scientists in the process with support for their complex analytics requirements, including machine learning
- The existence of multiple data types, which must be supported, including unstructured and semi-structured data
- The rise of new use cases like the Internet of Things (IoT), which demand real-time and streaming analytics
- The exponential growth of data that continues to create data silos, which need to be managed both across your on-premises environment and in the cloud
- The need for performance and scalability within the data warehouse infrastructure to meet your service level agreements

To control some of these factors, while still maintaining your existing investments, an analytics and data warehouse solution for your organization must:

- Deliver a solution to your data scientists that minimizes their difficulties in developing and deploying even the most advanced analytics workloads.
- Support multiple data types, locations, and both traditional data warehouses and near-real-time operational or mixed workloads.
- Fit a hybrid data strategy, with a path to the cloud.
- Deliver performance and scalability for any size or type of workload.
- Provide the flexibility and elasticity of a cloud option, but be housed in your data center.

As your organization's analytics needs expand, investing in the right solution is vital. Making the right choice could be the difference between meeting service level agreements and being the bottleneck to success.

The IBM Integrated Analytics System drives the insights needed to increase your competitiveness by matching accelerated development and deployment times for your data scientists with a high-performance, optimized and cloud-ready data platform.

As a unified data science solution, the built-in IBM Data Science Experience can be used by your data scientists to connect with your organization's data in place. This connection helps data scientists develop machine learning analytics that benefit from a performance-optimized common SQL engine with embedded Apache Spark processing.

From the start, the IBM Integrated Analytics System requires little or no tuning and maintenance to deploy and manage even demanding workloads that require high performance and petabyte-level scalability. The IBM Integrated Analytics System enables machine learning with the Apache Spark processing engine embedded on the system for higher performance analytics. At the same time, this feature can help reduce the complexity of moving analytics and data to separate environments. A common SQL engine shared across the IBM hybrid data management offering family lets you work with your existing on-premises and cloud applications. This flexibility allows you to pick the right environment for the right tasks.

The integrated architecture combines software enhancements, such as asymmetric massively parallel processing (AMPP), with IBM Power® technology and flash memory storage hardware. The IBM Integrated Analytics System handles traditional data warehouse workloads and operational mixed workloads. These workloads often require processing queries against large data volumes, quick point queries on small data sets and multiple concurrent operational accesses. As a result, the IBM Integrated Analytics System supports a wide variety of analytics use cases across broad data types and locations on a single solution. This flexibility provides your data scientists with almost endless possibilities.

Embedded analytics and machine learning with Spark

The IBM Integrated Analytics System helps simplify data scientists' efforts to train and evaluate predictive models with embedded Apache Spark processing. This feature helps eliminate the need for time-consuming movement and transformation of data to other systems. Once the models are developed using the tools of the data scientists' choice, the testing, deployment and training can be done where the data resides. With each node containing its own Spark executor process, latency is minimized, which helps speed data access and calculations compared to a stand-alone Spark cluster. In those cases where data scientists need to take the workloads off the system, industry-standard tools and the common SQL engine provide the option to seamlessly move models to a Spark cluster.

In addition to streamlining processes, this ability can also provide advanced performance and flexibility for analytics, including machine learning capabilities. Your data scientists can immediately connect to data in the system and begin building models with the five authorized user licenses included with the IBM Data Science Experience Local. This interactive, collaborative, cloud-based environment allows data scientists to use multiple tools to activate their insights. Data scientists also have the option of using Python, R or Scala using Jupyter Notebook with a Jupyter Notebook container included on the system. Jupyter can be used to execute interactive code with one-click deployment that transforms the code into a compiled and deployed Spark application.

In addition to prebuilt functions for data mining, prediction, transformations, statistics, geospatial data and data preparation, the Spark capability embedded in the IBM Integrated Analytics System supports open source R and other programming languages like Python, Java, C, C++ and Lua.

The IBM Integrated Analytics System also includes embedded IBM Netezza® Analytics technology with multiple algorithms, including linear regression, decision tree clustering, k-means clustering and Esri-compatible geospatial extensions. The system is designed to work with business analytics and visualization tools, including IBM Cognos®, SAP BusinessObjects, Kognitio, Microsoft Excel, QlikView, SAS, Microsoft SQL Server Reporting Services (SSRS) and Tableau. The system also handles model-building and scoring tools such as IBM SPSS®, Fuzzy Logix, open source R and SAS.

Integration and performance

IBM helps simplify the deployment and management of the analytics system using a design based on more than 20 years of experience with thousands of clients across multiple industries and regions. The software and hardware arrive at your data center configured to work together as a single performance-optimized solution. Within hours, you can load data without creating database indexes or struggling to tune and retune the data warehouse once it's operational.

Clients using the IBM Integrated Analytics System and the included IBM Db2® Warehouse technology should immediately recognize the common SQL engine used across the entire IBM hybrid data management solution portfolio. The IBM Db2 Warehouse is designed for data warehouse and analytics workloads. The common SQL engine uses dynamic in-memory columnar technologies for multi-workloads based on IBM Db2 and IBM BLU Acceleration® technology. BLU Acceleration massively parallel processing (MPP) architecture is designed for rapid and deep analysis of data that can scale into the petabytes. With query response times up to 100 times faster than earlier systems,[1] BLU Acceleration columnar tables can coexist with traditional row tables in the same schema, storage and memory so you can query both row and BLU Acceleration columnar tables at the same time. Adding BLU acceleration technology to traditional in-memory capabilities can accelerate performance even when data sets exceed the size of the memory. The dynamic in-memory columnar technologies of BLU Acceleration with data skipping offer an efficient method to scan and find relevant data even when the data is compressed.

The IBM Integrated Analytics System leverages IBM Power Systems™ and IBM FlashSystem® technology to improve reliability and performance at the hardware level. Today's IBM Power architecture enables denser systems that can achieve similar performance with less nodes than previous offerings. As the default storage for the system, IBM FlashSystem offers ultra-low latency and high near-in-memory I/O speeds with outstanding reliability.

While the analytics applications run at peak performance, the IBM Integrated Analytics System also brings new levels of reliability to help you meet or exceed your service level agreements. Power Systems and IBM FlashSystem storage is rated with increased uptime thanks to the fault tolerant design that helps eliminate a single point of failure. According to a 2017 Information Technology Intelligence Consulting (ITIC) survey, IBM Power Systems has the least amount of unplanned downtime—with 2.5 minutes per server per year—of any mainstream Linux server platform.[2]

Redundancies are built into components throughout the system, helping ensure continued operation in case of a hardware failure. The system also includes additional built-in, high-availability features to provide automated failovers for performance continuity. Monitoring and management for all components—hardware and software—is provided by a built-in console powered by IBM Data Server Manager that's used across the Db2 family.

A hybrid approach to the cloud and your data

When it comes to your data, a one-size-fits-all approach rarely works. The IBM Integrated Analytics System is built on the common SQL engine, a set of shared components and capabilities across the IBM hybrid data management offering family that helps deliver seamless interoperability throughout your infrastructure.

For example, a data warehouse that your team has been using might need to be moved to the cloud to meet seasonal capacity demands. Migrating this workload to IBM Db2 Warehouse on Cloud can be done seamlessly with tools like IBM Bluemix® Lift. The common SQL engine helps ensure no application rewrites are required on your part.

The common SQL engine provides a view of your data regardless of where it physically sits or whether it's unstructured or semi-structured data. The system's built-in data virtualization service in the common SQL engine helps unify data access across the logical data warehouse allowing you to federate across Db2, Hadoop and even third-party data sources.

Scalability and expansion options

Both the software and hardware architecture have been designed to grow and scale as you bring more workloads to support your business onto the system. Compute and storage capacity can be expanded independently, providing almost cloud-like levels of flexibility and elasticity. Hardware expansion is non-disruptive to your business and can be done in place on the system.

The IBM Integrated Analytics System also supports multi-temperature tiered storage to help ensure the highest levels of performance, even with large volumes of data. The system manages the most recently used and active hot data directly on the system storage nodes, while older, less active cooling data resides on more cost-efficient, high-density IBM Storwize® storage devices.

Use cases

The common SQL engine used in the IBM Integrated Analytics System lets you match the right workload with the right deployment platform, while helping ensure that data is accessible regardless of type, location or size. The following are a few use cases to inspire you in getting started:

- Make operational excellence the new normal by creating a logical data warehouse using Hadoop, data marts or other associated deployments that all interact and offer a unified view of your data whether they sit on premises or in the cloud.
- Create personalized customer experiences in real time by using your internal data with analytics that requires high performance and simplified scalability.
- Improve time to market for new product innovations by using embedded machine learning to accelerate performance on complex analytics and deliver insight to your users.
- Deliver the requested workloads to your business users more quickly with a system that loads data faster and requires little or no tuning, nor extensive levels of configuration.
- Expand the workload options available to business users with the addition of operational mixed workloads.
- Help ensure operational compliance with more accurate views of your data.
- Offer new levels of insights to grow and expand your business by building analytics across different data types and data sets using data virtualization to unify data access across the logical data warehouse and other technologies.

Specifications

The IBM Integrated Analytics System integrates and optimizes all compute, storage and networking resources with analytics and data warehouse software. It's available in rack configurations as shown in the table.

Specifications—single rack systems			
IBM Integrated Analytics System model number	M4001-003	M4001-006	M4001-010
Active nodes	3	5	7
CPU cores	72	120	168
Memory (TB)	1.5	2.5	3.5
User data (TB) (assumes 4x compression)	64	128	192
Maximum power per rack (Watts)	4,400	6,900	9,400
Cooling (BTU/hour)	15,000	23,500	32,000
Rack weight (Kg)	479.9	563.8	647.7
Height (cm)	202		
Depth (cm)	110		
Width (cm)	64.8		
Power cords	For North America and Japan—200–240 V, Single Phase, 30 A For EMEA—220–240 V, single phase, 32 A		
Drops/rack	4		
Safety	UL/CSA 60950-1 and, also, have EN 60950-1 and IEC 60950-1		
Emissions	FCC part 15, ICES-003, AUS/NZ CISPR 22, VCCI, EN55022/55032 Class A; European immunity: EN 55024, EN 61000-3-2 and EN 61000-3-3		
Note	Multirack systems are also available. Please contact your sales representative or IBM Business Partner for more details.		

About IBM hybrid data management

IBM provides one of the most comprehensive portfolios of hybrid data management solutions to help you maximize the value of your information assets and discover new insights to make better and faster decisions, and optimize business outcomes.

For more information

To learn more about IBM Integrated Analytics System, contact your IBM representative or IBM Business Partner, or visit the following website: **ibm.biz**/BdjfvP.

Additionally, IBM Global Financing can help you more quickly acquire the software capabilities that your business needs in a cost-effective and strategic way. IBM customizes payment solutions to suit the business and development goals of credit-qualified clients, and helps them achieve effective cash-flow management, balance costs and benefits, and improve return on investment. Fund all your critical IT investments and propel your business forward with IBM Global Financing. For more information, visit: **ibm.com**/financing.

IBM Corporation
IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
September 2017

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary. It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

1 http://www.redbooks.ibm.com/technotes/tips1204.pdf

2 http://itic-corp.com/blog/2017/06/ibm-lenovo-servers-deliver-top-reliability-cisco-ucs-hpe-integrity-gain/

Please Recycle

IMS14523-USEN-00