Information empowerment for your evolving data ecosystem

Highlights

- Enables better results for critical projects and key analytics initiatives

- Ensures the information that drives business and strategic initiatives is trusted, consistent and governed in real time

- Available on the cloud—move existing workloads and accelerate the time to value of new deployments

- Helps you better understand data and cleanse, monitor, transform and deliver it

Introduction

Scalable data platforms such as Apache Hadoop offer unparalleled cost benefits and analytical opportunities. IBM helps fully leverage the scale and promise of Hadoop, enabling better results for critical projects and key analytics initiatives. The end-to-end information capabilities of IBM® Information Server let you better understand data and cleanse, monitor, transform and deliver it. IBM also helps bridge the gap between business and IT with improved collaboration. By using Information Server "flexible integration" capabilities, the information that drives business and strategic initiatives—from big data and point-of-impact analytics to master data management and data warehousing—is trusted, consistent and governed in real time.

Since its inception, Information Server has been a massively parallel processing (MPP) platform able to support everything from small to very large data volumes to meet your requirements, regardless of complexity. Information Server can uniquely support the flexibility (through extract, transform and load [ETL] or extract, load and transform [ELT]), performance and scalability required to succeed with projects of any size.

Information Server capabilities are available in four packages to help your company overcome key information challenges (Figure 1):

- **Information Server for Data Integration:** Flexibly transform data in any style and deliver it to any system, supporting faster time to value and reduced IT risk.
- **Information Server for Data Quality:** Establish high-quality data and manage it, turning a deluge of data into trusted information.
- **Information Governance Catalog:** Better understand data and foster collaboration between IT and line-of-business teams to narrow the communication gap and create a framework for information governance.
- **Information Server Enterprise Edition:** Gain the capabilities of all three individual packages in one comprehensive offering; start information integration efforts in one area and then expand when needed to further optimize results.

Available on the cloud

Whether you want to move your entire Information Server Enterprise or Information Governance Catalog infrastructure to the cloud, or you want to start small by deploying development or test environments, IBM is ready to help with your data integration, data quality and governance needs through rapid deployment and flexible subscription pricing.

- Optimize resources and budgets to realize faster time to value and improve agility with flexible cloud deployment options. Whether you require two users, five users or more, flexible configuration and pricing options are available for your needs.
- Benefit from the same powerful and scalable ETL, data quality and information governance capabilities found in the on-premises deployment of Information Server Enterprise Edition.
- Integrate and transform large volumes of data, with data structures ranging from simple to complex.

Your organization can select the package that most closely aligns with your data needs or choose the comprehensive edition for the full range of capabilities
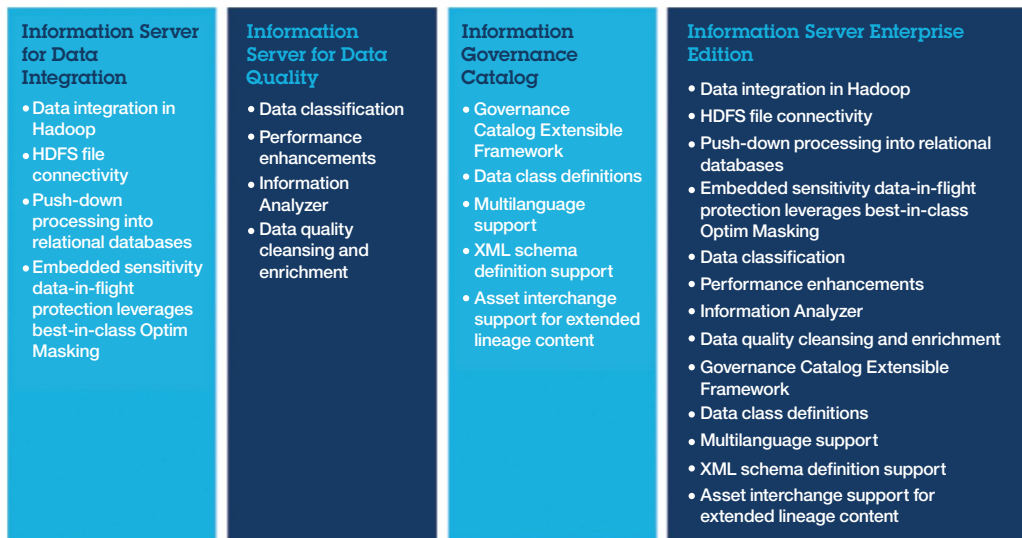
Information Server for Data Integration	Information Server for Data Quality	Information Governance Catalog	Information Server Enterprise Edition
• Data integration in Hadoop	• Data classification	• Governance Catalog Extensible Framework	• Data integration in Hadoop
• HDFS file connectivity	• Performance enhancements	• Data class definitions	• HDFS file connectivity
• Push-down processing into relational databases	• Information Analyzer	• Multilanguage support	• Push-down processing into relational databases
• Embedded sensitivity data-in-flight protection leverages best-in-class Optim Masking	• Data quality cleansing and enrichment	• XML schema definition support	• Embedded sensitivity data-in-flight protection leverages best-in-class Optim Masking
		• Asset interchange support for extended lineage content	• Data classification
			• Performance enhancements
			• Information Analyzer
			• Data quality cleansing and enrichment
			• Governance Catalog Extensible Framework
			• Data class definitions
			• Multilanguage support
			• XML schema definition support
			• Asset interchange support for extended lineage content

Figure 1. Four IBM Information Server packages address a spectrum of challenges.

Information Server for Data Integration

The end-to-end information capabilities of Information Server allow you to understand your data; cleanse, monitor, transform and deliver data; and collaborate to bridge the gap between business and IT.

New capabilities include:

- **Enhanced scalability and high performance, enabling faster access to clean data:** An MPP engine that runs natively in Hadoop—where the data resides—supports a rich set of integration and governance features. The engine helps your organization achieve high levels of performance and speed for data integration and governance activities such as data connectivity, transformation, cleansing, enrichment and delivery features.
- **All connectivity, transformation and data delivery features are now able to natively run in Hadoop:** You also get expanded and simplified access to Hadoop Distributed File System (HDFS) files in various formats and character sets, including critical security features such as Kerberos and secure gateways.
- **Low cost of integration due to increased productivity:** By leveraging the Information Server engine, your developers can use its built-in graphical design tool capabilities to significantly reduce the cost of integration compared to manual hand-coding.
- **Additional features at no extra cost:** Your organization can use additional features such as access to popular enterprise applications, including:
  – Salesforce.com
  – Hyperion Essbase
  – Siebel
  – JD Edwards Enterprise One
  – PeopleSoft Enterprise
  – Oracle applications

You can also take advantage of push-down optimization through IBM DataStage® Balanced Optimization features. These features allow you to fully harness available capacity and computing power in your relational databases and in DataStage.

A new way to work—flexibility for hybrid environments integrating and governing all types of data.

Information Server for Data Quality

Data quality is a key part of information governance and is a core discipline within the Information Server platform, helping to enable the delivery of consistent, accurate, trusted information. Information Server delivers a wide range of data quality capabilities from data profiling, standardization, matching and enrichment to active data quality monitoring.

Discover where sensitive data is stored

Organizations must be responsible and diligent when protecting personally identifiable information (PII). A key starting point is to identify where such data is stored. Information Server for Data Quality supports data privacy, data masking and test data management initiatives by identifying where sensitive information, PII and other classes of data are stored.

- Realize quick time to value by identifying the type of data contained within a table or file using three dozen predefined data classes including credit card, taxpayer ID, US phone number, email address and others.
- Add more data classes that are unique to your organization. Create and customize them with three types of extensible data classes: valid values list, regular expression (regex) and Java.
- Put your analyst in full control. Column analysis suggests data classifications based on data values. Users may review these classifications and when necessary, select a different data class to better represent the data stored in each column.

Improved performance to meet ever increasing data demands

Take advantage of enhanced scalability and reduced resource consumption during data profiling and analysis. Information Server for Data Quality features a new ability to limit the number of distinct values stored, which can dramatically reduce the size of the results database and speed overall processing of data profiling. The platform is also designed to help improve

scalability for cross-domain and foreign key analysis. Its enhanced methods no longer require a full column analysis with capture of all distinct values.

Flexibility to choose where you run

All features to assess and monitor the quality of your data are now supported executing directly inside a Hadoop cluster. You can run the same profiling and data rules inside or outside your Hadoop environment, including:

- Data investigation
- Standardization
- Matching
- Survivorship
- IBM QualityStage® Module for US Address Certification
- IBM QualityStage Module for Address Verification Interface

Information Governance Catalog

Automated data compliance ensures you can find, understand, trust and work with the best data possible, faster than ever before. Information Governance Catalog provides industry-leading data lineage capabilities that automatically capture data sources and data movements across the information landscape. Empower all of your organization's users with business-level, natural-language understanding of information and data quality metrics that let users know which data can be trusted. Real-time detection of private and sensitive data facilitates fast compliance with corporate and regulatory requirements.

Provide a comprehensive view of data sources and data movements across your organization's information landscape, including legacy and proprietary data sources
- Define the display, structural and containment properties for registering new information assets in the catalog.
- Automate the import and cataloging of new information assets through the IGC REST application programming interface (API).
- Search, display and govern the new information assets.

Automate compliance over private or sensitive data by automatically scanning and detecting data classes and profiling statistics
- View the detected data classifications as published by IBM Information Analyzer and select the standard or expected data class.
- Identify data elements that contain sensitive or restrictive information by evaluating their data class.
- Govern data classes, associating the class with the governance rules that need to control or restrict the data element.
- Create and manage data classes, such as the regular expression, valid value list or range.

Deploy the catalog across all of your organization's users anywhere in any language, with multilingual and multi-geography user experiences
- Users can select the display and business definition search language for the catalog.
- Users can view the defined language for business definitions and the related business definitions and shared asset assignments.
- Administrators can manage and define the set of the supported languages for business definitions.

Enable your organization to include your most important data sources, such as XSD models, in your governance landscape
- Display, explore and govern XML schema definitions (XSD models).
- Administrators may import XSD models through IBM Metadata Asset Manager and capture the XSD elements, attributes and types.
- Users may search, view and govern the details of an XSD model.

Information Server Enterprise Edition

This offering combines the capabilities of Information Server for Data Integration, Information Server for Data Quality and Information Governance Catalog for broad flexibility and scalability.

Embedded in-flight sensitive data protection leverages best-in class Optim Masking

Information Server Enterprise Edition licensing now includes Optim™ data masking libraries for protecting sensitive information in a parallel job integrating or cleansing data.

Native Hadoop execution: Scalability without limits

Information Server now offers revolutionary capabilities that enable your organization to operationalize Hadoop in ways that were previously not possible. In this release, Information Server can execute directly inside a Hadoop cluster. This means all of the data connectivity, transformation, cleansing, enhancement and delivery features that thousands of enterprises have relied on for years can be immediately available to run within the Hadoop platform. This supports a rich set of integration and governance features, used to solve some of the industry's most complex data challenges and designed to operate where your organization stores its big data. These features minimize data movements, thereby decreasing time to deliver business value.

- Scale your workloads to all data nodes of your Hadoop environments
- Automatically distribute runtime libraries to all Hadoop data nodes
- Leverage features such as Hadoop Node Labels and Yarn schedule queues to pinpoint resources
- Utilize HDFS data locality for the most efficient access to HDFS data
- Automatically exploit in-memory features , such as data pipelining, data partitioning and dynamic repartitioning, to minimize workload runtimes
- Capture design and runtime metadata and built-in data lineage support for Hadoop workloads
- Support Kerberos-enabled clusters

Hadoop version currency

Information Server supports Hadoop distributions that conform to the Open Data Platform (OPD) requirements such as IBM BigInsights® and HortonWorks Data Platform. It also supports Cloudera Distribution for Hadoop.

IBM supports these Hadoop distributions for data reading and writing as well as deploying the Information Server engine tier for running data integration and governance operations natively in Hadoop.

Updated repository and application server options

The repository tier now supports Oracle 12g (including RAC) and Microsoft SQL Server 2014. The services tier now supports IBM WebSphere® Application Server 8.5.5. For more details on specific system requirements, visit: **ibm.com**/support/docview.wss?uid=swg27046843

Updated connectivity support

Information Server expands its rich connectivity to enterprise data sources. It includes enhancements to the file connector for connectivity and metadata management for HDFS files. It also now features support for various file types, extended data type support, automatic field-level data lineage and importing of file metadata.

Additional updates to connectors and stages to support the latest versions of those data sources include:

- SQL Server stage for Microsoft SQL Server 2014
- Sybase stage for Sybase ASE and Sybase IQ, Version 16
- Teradata Connector for Teradata Database 15.10
- IBM Netezza® Connector for Netezza Performance Server 7.1
- IBM WebSphere ILOG® JRules connector for WODM 8.7

Why IBM?

IBM Information Integration and Governance (IIG) provides market-leading functionality to handle the challenges of big data. IIG provides optimal scalability and performance for massive data volumes; agile, right-sized integration and governance for the increasing velocity of data; and support and protection for a wide variety of data types and big data systems whether running on premises or in the cloud. IBM helps make big data and analytics projects successful by giving business users the confidence to act on insight.

For more information

To find the version of Information Server that's right for you, visit:

- Information Server:
  ibm.com/software/products/en/infosphere-information-server
- Information Server on Cloud:
  ibm.com/marketplace/cloud/information-server/us/en-us
- Information Server Enterprise Edition:
  ibm.com/software/products/en/infoinfoserventeedit

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: **ibm.com**/financing

IMS14481-USEN-02