Informatica PowerCenter Big Data Edition

The Proven Path to Innovation That Lowers Data Management Costs

Huge numbers of new data types, such as social media data, Web logs, machine sensor data, and unstructured data, are created every day. Is your company using big data to develop innovative products and services and to improve business operations?

As data volumes continue to grow, they quickly consume the capacity of data warehouses and application databases. Is your IT organization forced into costly upgrades to expensive databases and data warehouse hardware appliances?

Are you putting off big data projects due to resource constraints and the perceived risk associated with rapidly changing and emerging technologies like Hadoop?

The Safe On-Ramp to Big Data

Informatica® PowerCenter® Big Data Edition is the safe on-ramp to big data that works with both emerging technologies and traditional data management infrastructures. With this edition, your IT organization can rapidly create innovative products and services by integrating and analyzing new types and sources of data. It provides a proven path of innovation while reducing big data management costs and minimizing risk.

Focus on Big Data Analysis—Not Integration

Your data scientists and analysts likely spend only 20 percent of their time on data analysis. The rest of their time is spent on the mechanics of data integration—accessing, parsing, standardizing, integrating, preparing, delivering, and managing data.

This is what PowerCenter Big Data Edition does best, and it does it at scale. No specialized coding is required to scale performance on distributed computing platforms like Hadoop. Simply use the same PowerCenter GUI that has been exercised by thousands of integration developers for years.

Productivity Goes Up, Costs Go Down

With PowerCenter Big Data Edition, you don’t need an army of developers to realize the benefits of big data. Productivity skyrockets when your development team uses this edition’s codeless development environment, reusable business rules and mapplets, efficient collaboration tools, and flexible deployment models across all projects. And because raw data can be staged and ETL processing can be executed on cheaper commodity hardware, PowerCenter Big Data Edition extends the capacity of your existing data warehouse investments.

Adopting New Technology Just Got Safer

With this edition, your company can minimize the risks associated with new technology. PowerCenter Big Data Edition is proven data integration software that works with emerging technologies like Hadoop and traditional data management infrastructures. So no matter what happens technologically, you’re safe.

Benefits

- Bring innovative products and services to market faster and improve business operations
- Reduce big data management costs while handling growing data volumes and complexity
- Realize performance and costs benefits by expanding adoption of Hadoop across projects
- Minimize risk by investing in proven data integration software that hides the complexity of emerging technologies
Universal Data Access
Your IT team can access all types of big transaction data, including RDBMS, OLTP, OLAP, ERP, CRM, mainframe, cloud, and others. You can also access all types of big interaction data, including social media data, log files, machine sensor data, Web sites, blogs, documents, emails, and other unstructured or multi-structured data.

High-Speed Data Ingestion and Extraction
You can access, load, replicate, transform, and extract big data between source and target systems or directly into Hadoop or your data warehouse. High-performance connectivity through native APIs to source and target systems with parallel processing ensures high-speed data ingestion and extraction.

“Electronic Arts unified big data platform depends on enterprise-ready data access, data ingestion, data integration, data processing, and data visualization to power the future of game development, marketing, sales, accounting, and customer experience.”

– Tom Kornegay, Director Analytics Worldwide Customer Experience
Electronic Arts

Unlimited Scalability
Your IT organization can process all types of data at any scale—from terabytes to petabytes—with no specialized coding on distributed computing platforms such as Hadoop.

Optimized Performance for Lowest Cost
Based on data volumes, data type, latency requirements, and available hardware, PowerCenter Big Data Edition deploys big data processing on the highest-performance and most cost-effective data processing platforms. You get the most out of your current investments and capacity whether you deploy data processing on SMP machines, traditional grid clusters, distributed computing platforms like Hadoop, or data warehouse appliances.

ETL on Hadoop
This edition provides an extensive library of prebuilt transformation capabilities on Hadoop, including data type conversions and string manipulations, high-performance cache-enabled lookups, joiners, sorters, routers, aggregations, and many more. Your IT team can rapidly develop data flows on Hadoop using a codeless graphical development environment that increases productivity and promotes reuse.

PowerCenter Big Data Edition triples your IT team’s productivity by moving away from hand coding to a codeless visual development environment that provides prebuilt transforms and promotes reuse.
With PowerCenter Big Data Edition, your IT team can easily coordinate, schedule, monitor, and manage all interrelated processes and workflows across your traditional and Hadoop environment. You can even drill down into individual Hadoop jobs.

**Profiling on Hadoop**
Data on Hadoop can be profiled through the Informatica developer tool and a browser-based analyst tool. This makes it easy for developers, analysts, and data scientists to understand the data, identify data quality issues earlier, collaborate on data flow specifications, and validate mapping transformation and rules logic.

**Design Once and Deploy Anywhere**
ETL developers can focus on data and transformation logic without having to worry where the ETL process is deployed—on Hadoop or traditional data processing platforms. Developers can design once, without any specialized knowledge of Hadoop concepts and languages, and easily deploy data flows on Hadoop or traditional systems.

**Complex Data Parsing on Hadoop**
This edition makes it easy to access and parse complex, multistructured, unstructured, and industry-standard data such as Web logs, JSON, XML, and machine device data. Prebuilt parsers for market data and industry standards like FIX, SWIFT, ACORD, HL7, HIPAA, and EDI are also available and licensed separately.

**Entity Extraction and Data Classification on Hadoop**
Using a list of keywords or phrases, entities related to your customers and products can be easily extracted and classified from unstructured data such as emails, social media data, and documents. You can enrich master data with insights into customer behavior or product information such as competitive pricing.

**Mixed Workflows**
Your IT team can easily coordinate, schedule, monitor, and manage all interrelated processes and workflows across your traditional and Hadoop environment to simplify operations and meet your SLAs. You can also drill down into individual Hadoop jobs.

**High Availability**
This edition provides 24x7 high availability with seamless failover, flexible recovery, and connection resilience. When it comes time to develop new products and services using big data insights, you can rest assured that they will scale and be available 24x7 for mission-critical operations.
Key Benefits

Bring Innovative Products and Services to Market Faster
With PowerCenter Big Data Edition, your data scientists and analysts can focus on new data insights—not on data integration—which your company can use to develop innovative products and services. This edition provides a data integration development environment into which existing PowerCenter mappings can be imported and through which Informatica resource skills can be reused. It provides an easy-to-use visual development environment and a repeatable development process paradigm. It extends data integration to social media data, Web logs, sensor device data, documents, email, and other multi-structured or unstructured data without hand coding. As a result, this edition dramatically increases your IT team’s productivity and accelerates time to deployment.

Lower Big Data Project Costs
PowerCenter Big Data Edition reduces big data costs. Your IT team can manage twice the data volume with your existing analytics environment. You can offload data from your warehouse and source systems and offload processing to low-cost commodity hardware. By moving away from hand coding to proven data integration productivity tools, you triple your productivity—you no longer need an army of developers. This edition provides unified administration for all data integration projects. You can build it once and deploy it anywhere, which keeps costs down by optimizing data processing utilization across both existing data platforms and emerging technologies like Hadoop.

Expand Hadoop Adoption Across the Enterprise
PowerCenter Big Data Edition enables wider adoption of Hadoop across multiple projects. You don’t need to hand code in Java MapReduce to realize the performance and cost benefits of Hadoop—you can readily employ 100,000+ Informatica-trained developers worldwide to staff your big data projects. This edition is easier to administer and supports multiple big data projects. It ensures optimal end-to-end performance and optimal deployment. Your IT organization can run on Hadoop or traditional grid computing environments.

Minimize Risk of Adopting New Technologies
PowerCenter Big Data Edition minimizes the risks associated with the latest technology like Hadoop. Your IT organization can make one investment that works with both emerging technologies and traditional data infrastructures. This edition of PowerCenter is based on the same proven Informatica data integration software used by thousands of customers worldwide. It provides a single, consistent data integration approach for all types of data. By profiling data up front to identify data quality issues and validate big data insights, you decrease project risk while increasing trust in the big data analysis your team generates.

About Informatica
Informatica Corporation (NASDAQ: INFA) is the world’s number one independent provider of data integration software. Organizations around the world rely on Informatica for maximizing return on data to drive their top business imperatives. Worldwide, over 4,630 enterprises depend on Informatica to fully leverage their information assets residing on-premise, in the Cloud and across social networks.