IBM Software
Big Data

The IBM big data platform

IBM

Highlights

- Helps users explore large, complex data sets
- Streamlines the process of developing big data applications
- Monitors and manages big data systems for secure and optimized performance
- Speeds time-to-value with analytical and industry-specific modules

What is big data?

Every day we create 2.5 quintillion bytes of data—in fact, 90 percent of the data in the world today has been created in the last two years alone. This data comes from a wide variety of sources: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records and cell phone GPS signals, to name a few.

Big data spans four dimensions:

- **Volume:** Enterprises are awash with ever-growing data of all types, easily amassing terabytes and even petabytes of information.
- **Velocity:** For time-sensitive processes such as catching fraud, big data must be analyzed as it streams into the enterprise to maximize its business value.
- **Variety:** Big data extends beyond structured data to include unstructured data of all varieties: text, sensor data, audio, video, click streams, log files and more.
- **Veracity:** As the complexity of information grows, organizations must improve the level of trust users have in information, ensure consistency across the organization and safeguard the information. Establishing confidence is essential for driving better business results.

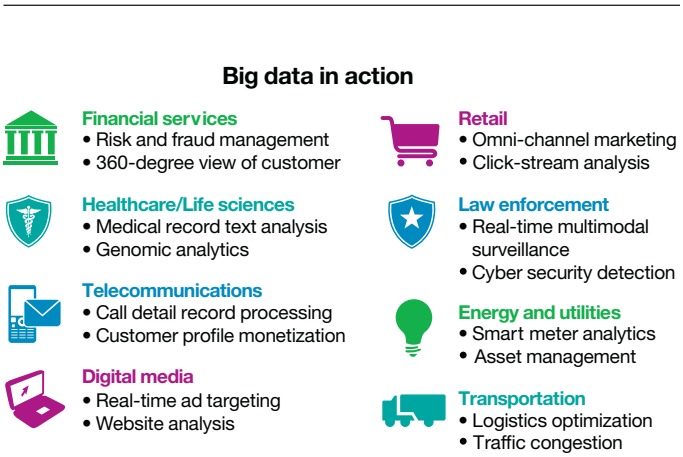

Figure 1. Big data plays a role in information-driven processes across all industries, from fraud detection and management to logistics optimization.

The IBM big data platform

IBM has developed a comprehensive, integrated and industrial-strength big data platform that lets you address the full spectrum of big data business challenges. The core capabilities of the platform include:

- **Hadoop-based analytics:** Enables distributed processing of large data sets across commodity server clusters
- **Stream computing:** Enables continuous analysis of massive volumes of streaming data with sub-millisecond response times
- **Data warehousing:** Delivers deep operational insight with advanced in-database analytics
- **Information integration and governance:** Allows you to understand, cleanse, transform, govern and deliver trusted information for critical business initiatives

IBM also offers a variety of supporting platform services:

- **Visualization and discovery:** Helps users explore large, complex data sets
- **Application development:** Streamlines the process of developing big data applications
- **Systems management:** Monitors and manages big data systems for secure and optimized performance
- **Accelerators:** Speed time-to-value with analytical and industry-specific modules

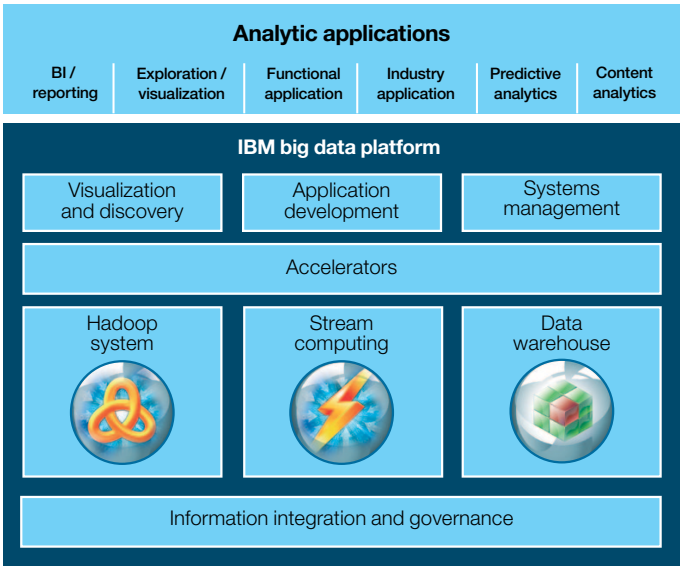

Figure 2. The platform blends traditional technologies that are well suited for structured, repeatable tasks together with complementary new technologies ideal for ad hoc data exploration, discovery and unstructured analysis.

The following products are the foundation of the IBM big data offering:

- **IBM® InfoSphere® BigInsights™:** Brings the power of Hadoop to the enterprise and enhances it with performance, reliability, security and administrative features. It helps firms discover and analyze new business insights hidden in large volumes of structured and unstructured data.
- **IBM PureData™ System for Hadoop:** A purpose-built, standards-based expert integrated system that architecturally integrates InfoSphere BigInsights Hadoop-based software, server and storage into a single, easy-to-manage system.
- **IBM InfoSphere Streams:** Supports ultra-low latency analytics on diverse data types, enhancing your organization's insights and decision making, and providing an opportunity to respond to events as they happen.
- **IBM PureData System for Analytics:** Powered by IBM Netezza® technology, IBM PureData for Analytics is a simple data appliance for serious analytics. It simplifies and optimizes performance of data services for analytic applications, enabling very complex algorithms to run in minutes not hours, delivering speed, simplicity and fast time-to-value.
- **IBM PureData System for Operational Analytics:** This expert integrated data system is designed and optimized specifically to handle the demands of operational analytic workloads. It provides high levels of consistent performance and reliability while handling thousands of concurrent operational queries. The continuous data ingest capability keeps the most current data is available for analysis while the adaptive storage compression helps keep storage costs down.
- **IBM Smart Analytics System:** Provides a modular and flexible integrated system with data warehousing and analytics software, along with an IBM server and storage optimized for operational analytics.

- **IBM DB2® with BLU Acceleration:** Speeds analytics and reporting using dynamic in-memory columnar technologies, providing an extremely efficient way to scan and find relevant data. Coupled with innovations such as parallel vector processing, data skipping and actionable compression, it helps analytic queries run faster.
- **IBM Informix®:** In addition to providing support for both online transaction processing (OLTP) and online application processing (OLAP) workloads, Informix has built-in capabilities to manage spatial and interval data used in time-based and space-based applications and analysis.
- **IBM InfoSphere Information and Integration Governance:** This unified set of capabilities brings together data from different sources for diverse targets, manages its quality and maintains master data for multiple domains. Plus, it helps secure and protect data, manages it across its lifecycle and facilitates information-based collaboration across business and technical teams.

Why IBM?

IBM offers an enterprise-class platform that supports the full breadth of big data use cases. The IBM big data platform has five distinct advantages:

- **Comprehensive:** Designed to be a complete platform for managing and analyzing the volume, variety and velocity of big data, while increasing veracity and confidence in your decisions
- **Enterprise-class capabilities:** Delivers the management, security, reliability and usability features necessary for large-scale deployments
- **Analytic accelerators:** Integrates analytic engines optimized for big data, as well as pre-built accelerators for industry-specific and cross-industry applications
- **Visualization tools:** Gives users tools to explore all available data for ad hoc analysis
- **Integration and governance:** Helps organizations understand their data, improve it and then allow business users and others to act on it, bringing confidence to big data

For more information

To learn more about IBM big data platform and big data solutions, please contact your IBM representative or IBM Business Partner, or visit: **ibm.com**/bigdata

To get started with our big data platform, visit:

- **ibm.com**/infosphere/quickstart
- **ibm.com**/infosphere/streams-quickstart

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
September 2013

IMB14135-USEN-02