



Highlights

- Analytic accelerators provide advanced analytics for various data types
 - Application accelerators address specific big data use cases
 - Included with IBM Big Data Platform components IBM InfoSphere BigInsights and IBM InfoSphere Streams, at no additional charge
-

IBM accelerators for big data

Speeding the development and implementation of big data solutions

With information streaming in from more sources than ever before—social media, outdoor sensors, cameras, videos, check-out transactions and so on—organizations face the daunting challenge of gaining insights from new data sources and types. Many business leaders know that big data is an important resource for creating new opportunities and becoming more agile. However, wanting to get started and knowing how to start are two different notions.

Building applications for handling big data typically requires a certain skill set: understanding numerous data sources and types, familiarity with particular programming languages, knowledge of advanced analytics and more. Unfortunately, companies often lack the time or money to invest in building these areas of expertise. Enter IBM® accelerators for big data.

IBM accelerators for big data are packaged software components that speed the development and implementation of specific big data solutions. They are included with two components of the IBM Big Data Platform—IBM InfoSphere® BigInsights™ and IBM InfoSphere Streams—at no additional cost. The accelerators offer organizations business logic, data processing and UI/visualization capabilities that can be tailored for a given use case or industry need. By leveraging IBM experience and information management best practices, they help eliminate the complexity of building big data applications and reduce the time-to-value for big data deployments.

IBM provides two types of accelerators for big data (see Figure 1):

1. **Analytic accelerators** address specific data types or operations with advanced analytics, such as text analytics and geospatial data.
2. **Application accelerators** address specific use cases, such as log analysis and social media insights, and include both industry-specific and cross-industry features.



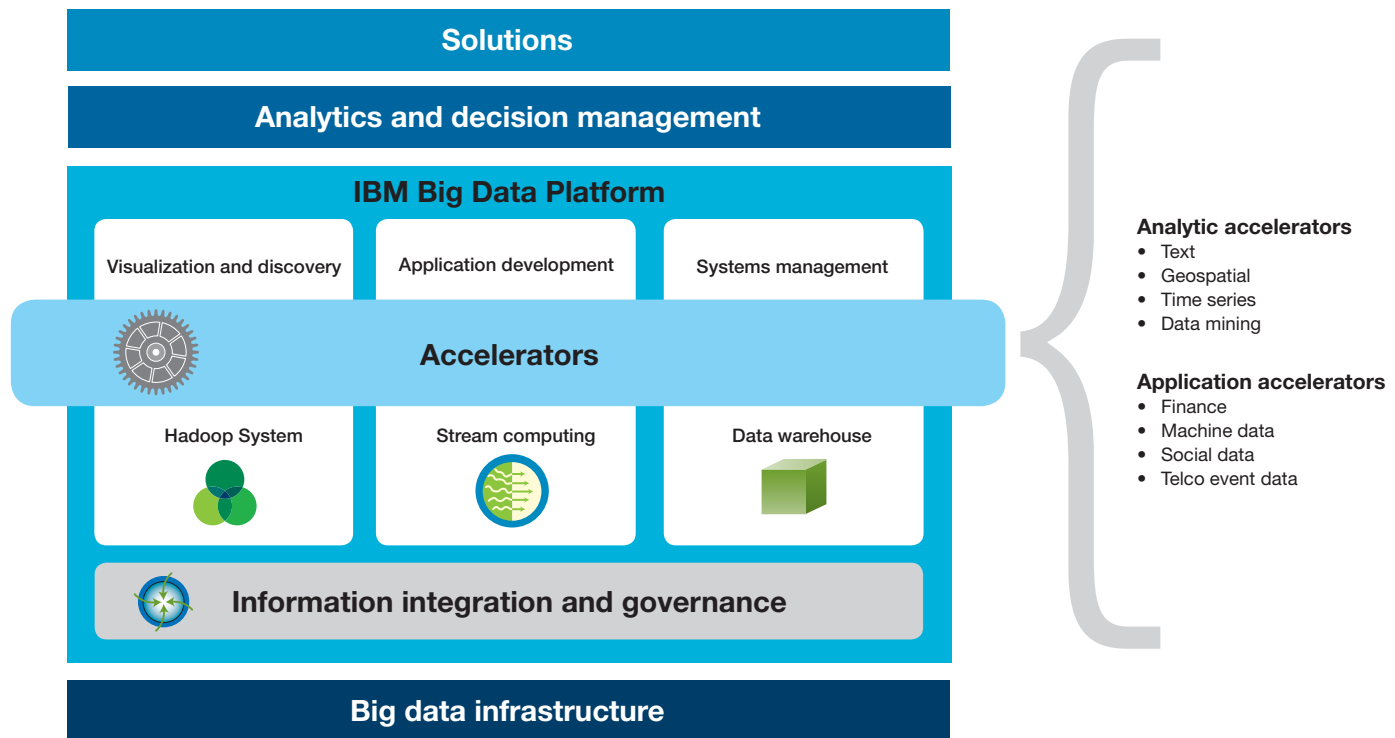


Figure 1: IBM accelerators for big data and the IBM Big Data Platform.

Analytic accelerators

The IBM portfolio of analytic accelerators for big data includes capabilities for text, geospatial and time series analysis, as well as data mining.

Text analytics: Gain broad insights with highly accurate analysis of textual context

The IBM text analytics engine is a powerful information extraction system, designed for use in big data and MapReduce parallel processing environments with exceptional performance. Developed by IBM Research, this analytics engine can identify meaning within text using technology similar to that demonstrated in IBM Watson™ for natural language processing. It includes more than 100 pre-built rules (called annotators) that understand textual

meaning. For example, the text analytics accelerator can differentiate between a first and last name, or indicate whether an address is for a street or apartment. The annotators are context-sensitive and can determine the relationship between terms even if they are separated by other text.

Case in point: Text analytics

In the healthcare industry, text analytics can pull information out of patient notes taken during routine clinical rounds, understand the context and detect meaning. This leads to more accurate records and analysis, better diagnoses and improved patient care.

Geospatial analytics: Do more with location data

The geospatial accelerator calculates distances and directions between points on the globe using flat earth, spherical and spheroid models. Key functions of this accelerator include:

- Real-time processing of location data
 - Efficient search for entities in or around an area of interest
 - GPS location data mapped to line—for example, road network segments
 - Calculators to determine the spatial relationship between different features on the Earth
-

Case in point: Geospatial analytics

In an urban environment, geospatial analytics can improve the efficiency of public transportation by enabling city bus systems to detect traffic patterns, calculate the fastest routes and optimize directions in real time.

Time series analytics: Predict the future

The time series accelerator facilitates real-time predictive analysis of regularly generated data, including:

- Physiology/nature data: Audio, ECG, weather data and seismic data
- Systems data: Response times and performance measures
- Market/finance data: Stocks, sales and other metrics
- Sensor data: Smart meters, temperature sensors, pressure sensors and water salinity sensors

The time series accelerator also includes digital filtering capabilities to remove noise or unwanted artifacts, smooth out data and decompose in low-varying and high-varying patterns. It can perform pattern and correlation analysis to identify patterns in the data and expose correlations in various data streams, as well as decomposition to estimate spectral components of the data stream and transform and project them into new data representations. To help “predict the future,” the accelerator includes forecasting capabilities to detect general trends and estimate the value of future data streams before they arrive (also known as look-ahead stream processing).

Case in point: Time series analytics

Energy utilities depend on data to measure demand for resources, manage distribution and react to outages or other disruptions. Utilities can use the time series accelerator to analyze huge volumes of environmental measurements to detect anomalies in weather patterns, or tap into constant streams of data from smart meters to help predict energy usage based on price and temperature.

Data mining: Score data against predictive models

The IBM accelerator for data mining provides a set of IBM Streams Processing Language (SPL) operators for scoring real-time data against predictive models represented in Predictive Model Markup Language (PMML). The PMML standard is supported by multiple analytics platforms such as IBM SPSS®, SAS and open-source R. Analytics operators provided by this accelerator include:

- Classification: Decision trees, Naïve Bayes and logistic regression
- Clustering: Demographic clustering and Kohonen clustering
- Regression: Linear regression, polynomial regression and transform regression
- Associations: Association rules

With this accelerator, operators can take a PMML file describing a predictive model as an input stream and update the model dynamically for rapid insight. For seamless integration, this accelerator embeds libraries directly from IBM InfoSphere Warehouse.

Case in point: Data mining

A bank can use the data mining accelerator to detect fraud in real time, based on analysis of historic records that include fraudulent purchases. With the insights gleaned, banks can greatly improve their decision making and analysis of customer purchase patterns over time for better retention and relationships.

Application accelerators

The IBM portfolio of application accelerators for big data includes capabilities for finance, machine data, social data and telecommunications (telco) event data analytics.

Finance analytics: Incorporate critical real-time insights for financial markets

The finance analytics accelerator provides automated options and equity trading analytics including:

- Real-time market data ingestion and management
- Real-time decision support for equities, derivatives, commodity and foreign exchange (Forex) trading
- Ability to incorporate additional contextual awareness (news, weather and so on) into trading decisions
- Real-time, cross-asset pricing and real-time, continuous enterprise risk-level monitoring and liquidity management across trading desks and geographies
- Continuous real-time trade monitoring to identify fraudulent trading
- SPL building blocks to develop InfoSphere Streams-based financial applications
- Out-of-the-box adapters and operators that save organizations time by implementing functions that don't need to be built from scratch:
 - Out-of-the-box adapters:
 - Financial Information Exchange (FIX)
 - IBM WebSphere® Front Office (WFO)
 - Low-latency messaging for InfiniBand support
 - Operators:
 - European style for options (11 methods)
 - American style for options (11 methods)
 - Greeks (for example, Delta and Theta)

Case in point: Finance analytics

A financial services company can take the out-of-the box capabilities of the finance analytics accelerator, including sample trading options algorithms, and plug them into its own decision-making algorithms. This helps facilitate automated options trading.

Machine data analytics: Create new possibilities for operational data applications

The IBM accelerator for machine data analytics ingests and processes large volumes of machine data to provide in-depth business insights. Machine data comprises information that was automatically created by a computer process, application or other device; sources include machines ranging from IT equipment and sensors to meters and network devices. The accelerator provides a range of data-intensive capabilities, including:

- Data ingestion, metadata validation and writing to Apache Hadoop Distributed File System (HDFS)
- Data parsing and extraction, available out of the box and as extendable rules
- Data indexing for complete re-indexing (ingest new batches) or batch-incremental indexing updates (update already indexed batches) using InfoSphere BigIndex, a component of InfoSphere BigInsights that delivers low-latency, full-text search capabilities for big data, with user-configurable fields and fact hierarchies
- Configurable faceted search leveraging Lucene wildcard syntax support for text search
- Data transformation to link records and create sessions based on time windows, a set of “seed” records and transitive linkage
- Statistical modeling to model patterns and perform correlation analysis

Case in point: Machine data analytics

In the energy and utilities industry, data streams in from the field 24x7—distribution grid monitors, sensors on drilling platforms, pipeline gauges, local and regional transmission stations and so on. The IBM accelerator for machine data analytics can ingest data from multiple equipment sources; analyze, index and parse it; and turn it into accessible insights that help companies manage peak demand, address constrained resources and improve system reliability and efficiency.

Social data analytics: Get closer to your customers

The IBM accelerator for social data analytics is designed to process large volumes of social media data to provide additional views into customer-facing activities such as retention efforts, lead generation, brand management and marketing campaigns. Highly extendable and customizable, this accelerator supports:

- Data ingestion from sources such as Gnip and Boardreader
 - Processing of both streaming data and data at rest
 - Micro-segmentation attributes, such as personal information (gender, location, parental status, marital status, employment and so on) as well as interests and products owned or previously purchased
 - Entity resolution across different social media sources
 - Monitoring of outputs and measures such as buzz, sentiment, intent to buy or start service and intent to attend or see, so organizations can then take specific and necessary actions
 - Visualization using IBM BigSheets, a browser-based, spreadsheet-like tool that comes with InfoSphere BigInsights and allows users to explore data stored in InfoSphere BigInsights applications and create analytic queries without writing any code
-

Case in point: Social data analytics

If a movie studio wants to study the effectiveness of trailers for the latest film release, the social data analytics accelerator can help process real-time feedback culled from millions of social media profiles as the trailer airs.

Telco event data analytics: Enhance customer profiles with streaming data

The telco event data analytics accelerator provides the ability to ingest, process and analyze large volumes of streaming data from telecommunications systems in real time. It can take in different types and volumes of streaming data, and then analyze subscriber information and call data records (CDRs) to support billing mediation. CDRs can be in ASN.1 (Abstract Syntax Notation.1), ASCII or binary format. Once the data is ingested, the accelerator can perform rule-based data transformations, including lookups against tables, computation of key fields and generation of unique CDR IDs. Registration and error-handling capabilities extend to file-level duplicate detection and error-checking of individual CDRs. Information can be output to different locations, including filesystems, databases and parallel databases. CDR deduplication is performed using a Bloom filter with a variable window of up to 15 days. The accelerator also features KPI computation, which aggregates user and cell information, offers summary statistics and facilitates end-to-end consistency, as well as dynamic table and rule updates.

This level of insight enables telecommunications companies to offer targeted, differentiated services for a high-quality customer experience, which helps to strengthen customer loyalty and reduce churn. Features such as personalized billing and real-time fraud detection also help increase operational efficiency.

Case in point: Telco event data analytics

One telecommunications company uses the IBM accelerator to access real-time analyses of 7 billion CDRs a day, reducing its data processing time from 12 hours to 1 minute and cutting hardware costs to one-eighth of the original amount.

Getting started with accelerators

IBM accelerators for big data are designed to help organizations address big data challenges and leverage industry-leading expertise without having to redesign existing systems and processes. As part of the IBM Big Data Platform, the accelerators help organizations integrate and manage the full variety, velocity and volume of data; apply advanced analytics to information in its native form; visualize all available data for ad hoc analysis; and build new analytic applications based on workload optimization and scheduling.

The IBM Big Data Platform

The IBM Big Data Platform is a comprehensive collection of best-of-breed technologies and services that help organizations integrate data from disparate sources, analyze big data in real time, anticipate future outcomes and rapidly generate insights for capitalizing on new opportunities. In addition to the accelerators, the enterprise-class platform includes stream computing, Hadoop-based analytics, visualization and discovery, data warehousing and information integration and governance capabilities.

To learn more about the platform, please visit: ibm.com/software/data/bigdata/enterprise.html

For more information

To learn more, please contact your IBM representative or IBM Business Partner about how the IBM accelerators included with InfoSphere Streams and InfoSphere BigInsights can help you solve your big data business challenges.

- ibm.com/software/data/infosphere/streams
- ibm.com/software/data/infosphere/biginsights



© Copyright IBM Corporation 2013

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States
January 2013

IBM, the IBM logo, ibm.com, BigInsights, InfoSphere, SPSS and Watson are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at ibm.com/legal/copytrade.shtml

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 and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.



Please Recycle