IBM big data for the automotive industry

Take advantage of new sources of data to reduce costs and increase competitive advantage
Businesses of all kinds are entering a new era of computing that is customer-centric and fueled by big data. The big data era is characterized by the growth of social media, an explosion of mobile devices and a physical world being outfitted with millions of networked sensors connected through the Internet. These factors have resulted in unprecedented growth of all types and volumes of data available to businesses.

The automotive industry is no exception. In fact, it is estimated that the auto industry will be the second-largest generator of data by 2015.¹ This estimate is not surprising, considering that some plug-in hybrid vehicles generate 25 GB of data in just one hour.²

While big data affects nearly every industry, specific market forces are driving change in the automotive industry (see Figure 1). For example, new alliances are being formed as consumer and environmental requirements create the need for collaborative partner ecosystems. Increased globalization is driving more integration within and between automotive OEMs and their suppliers. And rapidly integrating automotive enterprises with high numbers of joint ventures imply increasingly dynamic operations.

Additionally, sophisticated consumers are increasing the demand for innovative and sustainable vehicles. In response, new technologies and capabilities are making vehicles more intelligent. These dramatic forces require new approaches to maximize profitability and revenue, and are causing the automotive industry to focus on several business imperatives. Companies need ways to quickly launch sustainable vehicles, capitalize on services opportunities for intelligent connected vehicles, optimize the global value chain and transform the retail environment.

---

**Figure 1.** Market forces are focusing the automotive industry on imperatives supported by key big data use cases.
Automotive companies can use big data and analytics to help realize these business imperatives. IBM has built solutions based on key industry use cases, including the connected vehicle, predictive asset optimization, actionable customer insight and data warehouse optimization. This paper examines each of these use cases and shows how IBM big data and analytics solutions can help automotive manufacturers, suppliers and dealers use big data to increase competitive advantage and business success.

**Capitalizing on the connected vehicle**

A primary contributor to the dramatic increase in automotive industry data is the automobile itself, or devices integrated within the vehicle. Automakers are looking at ways to connect vehicle information with data about the environment in which the vehicle is operating—such as data about weather, traffic and even other vehicles. Analyzing vehicle data integrated with location and environmental data enables automakers or fleet operators to send drivers real-time alerts about impending external conditions.

The IBM big data platform is enabled by unique, streaming analytics for big data. With massive parallelism, advanced analytic models and associated development and management tools, the platform can capture, integrate and analyze real-time streaming data coming from the vehicle while in use. This capability provides automakers with real-time insight into how the various vehicle systems are performing under specific driving patterns and environmental conditions. The information gained can be analyzed to predict maintenance issues or transmitted to the driver using cloud-based services to enhance safety and the driving experience.

To provide an integrated view, the solution platform can utilize a variety of data-at-rest and data-in-motion sources. Data-at-rest for a specific vehicle identification number (VIN) includes option and feature codes, specifications, part costs, durability and stock information. Dealer data-at-rest includes repair history, aftermarket products, and services purchased and installed. Telematics can provide data-in-motion such as vehicle speed, data relating to the transmission control system, braking, air bags, tire pressure and wiper speed as well as geospatial and current environmental conditions data. Other data contributing to an integrated view of the vehicle includes social media content about customer sentiment and driving experiences.

Analysis of this data delivers business benefits by enabling the automotive company to strengthen customer relationships, better predict demand for replacement parts and service, and monetize telematic data. For example, a company might build customer relationships by offering an improved, interactive driving experience—even helping owners of hybrids maximize fuel economy by finding optimal speeds and timing of gear shifts. Telematic data can be monetized by providing integrated vehicle information to third-party providers of driver assistance services, or by providing raw data to rental and insurance companies.

**Leveraging predictive asset optimization**

Manufacturing companies frequently use predictive analytics to monitor and maintain in-plant assets for better availability, utilization and performance. For today's automotive industry, predictive analytics can be extended beyond in-plant equipment to include the vehicle itself, both pre-sale and post-sale.

Information from predictive analytics can be used to provide a feedback loop that enables carmakers and dealers to optimize quality and supply-chain processes while minimizing warranty costs. However, organizations struggle to take advantage of this big data opportunity because they lack the capability to rapidly analyze data from the myriad of sources available today. Predictive asset optimization (PAO), an IBM Signature Solution, integrates IBM industry, services, software and research expertise to augment traditional analytics with big data capabilities.
Combined with user-friendly industry dashboards, accelerators and methods, IBM PAO enables automotive companies to analyze data from available sources with low latency. By leveraging advanced analytics capabilities, companies can improve visibility into their operations and optimize asset health and yield. For example, OEMs can analyze years of vehicle operating data combined with publicly available data to identify potential leading indicators of vehicle wear, and then anticipate parts and service demand.

PAO allows organizations to collect and integrate structured, unstructured and streaming data, including:

- Data from sensors and actuators on plant equipment and vehicle systems
- Image, video and graphic data across various types and formats
- Maintenance and engineering information such as enterprise asset management data
- Call-center information logs, voice and audio records, email and text
- In-service usage data from service providers and third-party partners
- Geospatial and temporal information, including GPS, weather and environmental data

Automotive organizations can use the IBM solution to analyze the root causes of failure, determine the leading indicators of failure and assess the risk of failure and outage versus preventative maintenance. By detecting anomalies and patterns in equipment and parts performance, companies can optimize the supply chain and ensure that parts suppliers can meet anticipated demand, helping to increase customer satisfaction.

Potential business benefits provided by the PAO solution include reduced capital and operating expenditures, improved production yield and the ability to curb warranty costs by detecting and correcting problems early. Companies can also maximize asset value by reducing equipment downtime and outages.

---

The IBM big data platform

The IBM big data platform is designed to help you gain insight from big data by delivering enterprise-class data management and advanced analytics. The platform supports ad hoc data exploration, discovery and unstructured analysis as well as structured, repeatable tasks to improve business insight regardless of the volume, variety, velocity or veracity of data.

The IBM big data platform effectively manages and analyzes data in its native format—unstructured, structured, at rest or in motion. To accelerate time-to-value, all platform components are pre-integrated. By leveraging this extensible set of capabilities, you can start with a single project using one capability and add others as needed. Components of the IBM big data platform include:

- **IBM® InfoSphere® BigInsights™** provides an integrated solution for analyzing hundreds of terabytes, petabytes or more of raw data derived from an ever-growing variety of sources.
- **IBM InfoSphere Streams** helps organizations turn burgeoning, fast-moving volumes and varieties of data into actionable information and business insights.
- **IBM InfoSphere Data Explorer** provides federated discovery, search and navigation over a broad range of data sources to help organizations get started quickly with big data initiatives and gain more value from their information.
- **Robust data warehouse software and integrated systems** developed by IBM to simplify and accelerate the delivery of insights derived from your data.
- **IBM PureData™ System for Analytics** is a high-performance, scalable, massively parallel system that enables clients to gain deep insight from their data and perform analytics on enormous data volumes.
- **IBM PureData System for Operational Analytics**—part of the IBM PureSystems™ family—is an expert integrated data system designed and optimized specifically for the demands of an operational analytics workload.
Uncovering actionable customer insight

Big data gives marketing professionals an opportunity to understand customers with greater accuracy and depth. Taking advantage of this opportunity requires the ability to analyze many varieties and volumes of data, which can then be used to reduce marketing cost, improve marketing results and decrease churn.

The strength of the actionable customer insight solution lies in combining multiple data sources (see Figure 2). Valuable insight can be gained by combining external data such as social media content with internal data such as part and product information, customer information, emails and call-center interactions. Deep analysis of customer demographics, transactions and clickstream data can allow companies to create new segmentation that enables marketing to make more personalized offers that yield more sales. Combining social sentiment, emails and competitors’ buzz can provide new insight into which products are trending up, helping to anticipate a potential increase in demand. And knowing where the customer is located at a given time enables the delivery of offers when the customer is most receptive.

Organizations can also improve customer satisfaction and sales by creating a single view of the customer, relating customers to the products they own and providing information consistency across business entities, including dealers, suppliers and product/part record management. Benefits range from better use of marketing budgets and more effective local marketing tactics to more market-relevant products and increased customer loyalty.

Figure 2. Customer insight can be delivered through multiple data sources.
IBM big data for the automotive industry

In the era of big data, the ability to query a data warehouse for business intelligence and decision support can be more powerful than ever, but also challenging for existing systems to support. There is simply far more data to manage. Businesses need to leverage a variety of data from operational systems and channels of engagement, including structured, unstructured and streaming data sources, to enable deep analysis. They must be able to interactively and iteratively query growing data in real time or near-real time because answers are needed in minutes and hours, rather than in weeks or months.

Another challenge is that storage costs may escalate as the volume of data increases. To capitalize on big data while controlling costs, automotive companies can optimize storage, maintenance and licensing by using Apache Hadoop as a lower-cost component of the data warehouse for landing, preprocessing and analyzing large amounts of structured and unstructured data. Hadoop can enable data analysis in a lower-cost environment and offload processing requirements in the more expensive data warehouse component of the system.

The IBM data warehouse optimization solution includes:

- A preprocessing hub to help determine what data should be moved to the warehouse
- A query-able archive where infrequently accessed or aged data can be offloaded from warehouse and application databases
- Exploratory analysis to enable analytics that might have previously been performed in the warehouse, thereby optimizing the warehouse and enabling new types of analysis
- Real-time analytics and filtering of streaming data

With the solution in place, organizations can process and act on information in real time to reduce storage in the warehouse by creating a staging area or landing zone for all automotive enterprise data. This zone is used to determine if data should be retained for deeper analytics, moved to more cost-effective storage or not saved at all. The solution also cleanses and transforms retained data before loading it into the data warehouse.

With these capabilities, companies gain faster, more accurate information, better alignment of cost with business objectives and insight from all available data to drive improved business performance. IT costs can be dramatically reduced by creating a landing zone and an analytics zone of enterprise information that features analytics and combines enterprise data with external data across multiple data platforms.

The IBM Big Data Hub is a source for information, content and conversation regarding big data analytics for the enterprise, filled with whitepapers, books, videos and customer case studies. IBM Briefings and Solution Centers provide face-to-face opportunities to learn about big data, while the Big Data University and IBM YouTube Big Data channel offer downloads, a test environment and a sandbox designed for learning at the user's own pace.
Exploration: Defining the business case and road map
The focus of the exploration stage is on developing an organization’s road map for big data development. Key objectives include developing a quantifiable business case and creating a big data blueprint. This strategy takes into consideration existing technology and skills, and then outlines where to start and how to develop a plan aligned with the business strategy of the organization.

IBM can perform a big data workshop briefing for businesses to help them apply big data capabilities based on strategic considerations, business value and total cost of ownership (TCO). IBM can also recommend the big data platform that will best support a company’s use case.

Engagement: Embracing big data
In the engagement stage, organizations begin to prove the business value of big data, as well as perform an assessment of their technologies and skills. A proof of concept (POC) can help validate the requirements associated with implementing big data initiatives, as well as articulate the expected returns. At this stage, IBM helps clients validate a big data platform supporting their business environment, using IBM value-added technologies on top of Hadoop, before clients deploy a solution enterprise-wide.

---

**Big data adoption**

1. **Educate**
   - Join the business community
     - Big data case studies, white papers and IBM Institute for Business Value reports
   - IBM Readiness Assessment for big data
     - Prioritize business use cases
     - Recommend big data platform
   - Self-paced learning and exploration with downloads and test environments
   - Join the technical community

2. **Explore**
   - Validate and realize business value
   - Solution design and proof of concept
     - Validate business value of the big data use case
     - Demonstrate big data capabilities to execute use case

3. **Engage**
   - Enterprise-wide big data initiatives
     - Incremental value across multiple use cases
     - Leverage investment from reusing the same big data platform
     - Enterprise data platform to support analytics

4. **Execute**
   - Learn the technology and gain expertise

*Figure 3. Stages of the big data journey.*
Big data and analytics capabilities are more widely implemented within the organization during the execution phase. Organizations should be prepared to implement big data solutions at scale as the threshold for advancing to this stage. IBM offers a full breadth of big data platform solutions to support analytics needs across the enterprise and across multiple business use cases.

Realizing the benefits of big data
Automotive companies are entering a new era in which they can capitalize on new sources of data to achieve their business objectives quickly, efficiently and at minimum cost—if they have the necessary knowledge, experience and solutions. With IBM big data and analytics, companies can meet these needs at each stage of the big data adoption process.

Organizations can realize concrete benefits of big data with key auto industry use cases ranging from optimizing the data warehouse to leveraging the connected vehicle. The IBM big data platform is designed to scale as the volume and variety of data grows, provide real-time processing capability, and transparently deliver timely, accurate data that businesses can trust.

For more information
To learn more about IBM big data solutions, see your IBM representative or IBM Business Partner, or visit: ibm.com/bigdata