KEYSTONE
strategy

White Paper

Data & Analytics Maturity Model & Business Impact

August 23, 2016

Keystone Strategy
Boston • New York • San Francisco • Seattle

www.keystonestrategy.com

KEYSTONE
strategy

Table of Contents

1. Executive Summary

The importance of "Big Data" to modern enterprises has drawn extensive coverage. Business leaders, technology analysts, the press, and the investment community have discussed how data is transforming business, work and society. Data has been heralded as nothing less than "the new raw material of business: an economic input almost on par with capital and labor."[1]

Amid the discussion of how data, business intelligence, and analytics are reshaping business, the concrete benefits of data and analytics have been disputed. Do large IT budgets translate into higher business performance? Can enterprises unlock the potential of data in the same way as hyper-scale internet companies whose very business models are often reliant on Big Data, millions of connected devices, and sophisticated software platforms and algorithms? What concrete business value does data have for enterprises in traditional industries like manufacturing, consumer packaged goods, financial services, and retail?

This white paper investigates the relationship between Data & Analytics technologies and business performance based on a large empirical study of major enterprises. To quantify the impact of data on business performance, Keystone Strategy developed a Data & Analytics maturity index to grade what companies can actually *do* with their data and their data platform. Companies were ranked based on the capabilities they have deployed in their business and then compared and contrasted in terms of business results they have achieved. This study evaluated whether companies who have sophisticated Data & Analytics capabilities also have better business performance.

The results are startling. The research found that Data & Analytics technologies are crucial for these companies: Enterprises who have realized advanced data capabilities are found to outperform their peers on measures of profitability and employee productivity. Companies who have developed the most sophisticated Data & Analytics platforms and apply these capabilities as a regular part of their business enjoy operating margins that are eight percentage points higher than lagging organizations. This translates to $100 million in operating profits on average for the more advanced companies in the sample controlling for factors such as company size and industry vertical.

In addition to having superior financial results, companies with top Data & Analytics capabilities also have business processes that are more sophisticated than their peers. Top performing enterprises have

[1] "Data, data everywhere." *The Economist* 25 February 2010. Web. 24 June 2016. http://www.economist.com/node/15557443

used data to transform how their business operates across sales and marketing, engineering, operations, finance, HR, and back-office. These business functions of top performers look dramatically different as a result of the way they store, process, and use data to make more effective and real-time decisions.

This paper profiles the capabilities and technical roadmaps applied by the Data & Analytics leaders and presents a framework to assist companies in planning their strategy. Using the Data & Analytics capabilities framework and online assessment tool (see: www.microsoft.com/datamaturity), organizations can assess their current capabilities and determine how they rate relative to industry peers. The framework also offers a pattern for how enterprises can advance their Data & Analytics capabilities across six product areas key to a modern data platform: Operational Databases, Enterprise Data Warehouse, Enterprise Data Lake, Business Intelligence, Advanced Analytics, and Cloud Computing Infrastructure.

2. Methodology

The findings in this paper are based on primary research Keystone Strategy conducted as to how major enterprises apply Data & Analytics within their business and use data to guide their business operations. Keystone conducted 344 one-hour long telephone interviews with senior business and technology decision makers to profile the technologies enterprises have deployed and assess the business and technical capabilities in place to manage, analyze, and generate insight from data. This research focused on upper midmarket and enterprise organizations, with a median company size of over 6000 employees and $3.4B in company revenue. Organizations represented include companies in the manufacturing, consumer packaged goods, financial services, and retail industry verticals.

Survey respondents answered approximately 150 closed-ended questions pertaining to their company's business, technologies deployed and Data & Analytics capabilities as well as their perceptions regarding data's strategic importance. To design the Data & Analytics capabilities framework, multiple inputs were used including analyst reports and white papers, case studies and marketing materials of technology companies providing Data & Analytics solutions, and pilot interviews with industry leading companies. Ultimately, seventy-four questions pertaining to Data & Analytics capabilities were used to grade the organization's level of data platform sophistication. These questions span six technology areas, which cover the most important elements of organization's data platform:

- Operational Databases
- Enterprise Data Warehousing (EDW)

- Enterprise Data Lake (EDL)
- Business Intelligence (BI)
- Advanced Analytics
- Cloud Computing Infrastructure

Respondents were grouped into quartiles based on the proportion of capabilities their organization has in place (i.e., the percentage of capabilities questions they answered affirmatively). Organizations that possess the highest number of the Data & Analytics capabilities rated in the top quartile, whereas the least sophisticated enterprises with the lowest number of capabilities rated in the lowest quartile.

Keystone complemented the survey data with company profile and business performance metrics from public information sources and regulatory filings, S&P CapitalIQ, and Dun & Bradstreet.

3. Data & Analytics Maturity Model & Business Impact

A. Data & Analytics Driving Business Performance

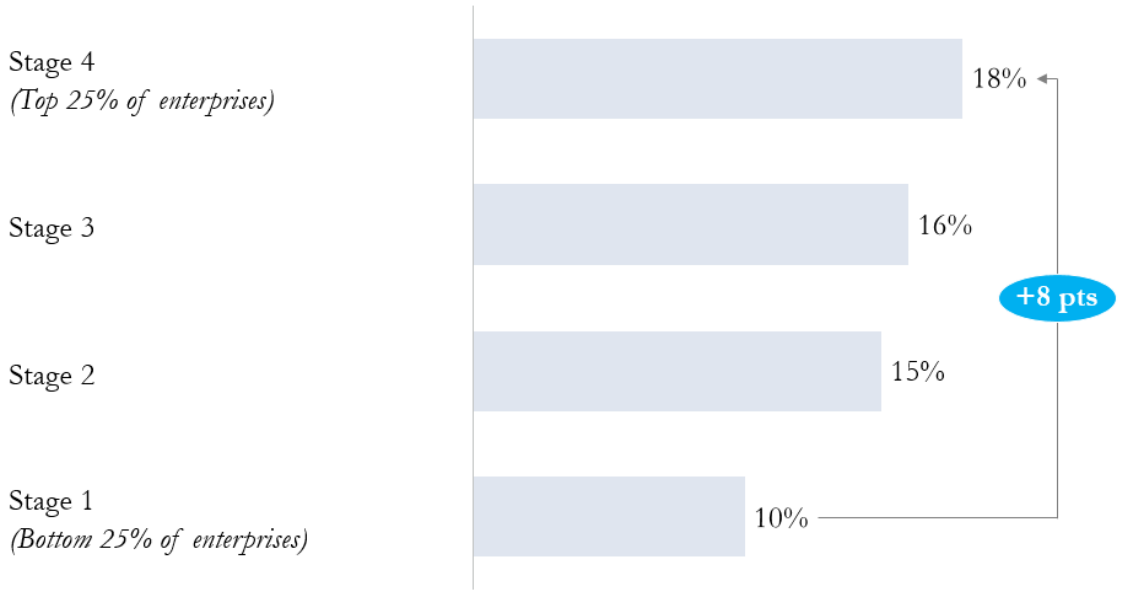
This study found that the enterprises with the most sophisticated Data & Analytics capabilities demonstrate higher levels of corporate business performance when holding constant factors such as industry vertical and company size. Enterprises within the top quartile have operating margins eight percentage points higher than enterprises in the bottom quartile. This difference in operating margins translates to a difference of $100 million per year in operating profit, controlling for company size and industry vertical.

"Nobody envisioned that data and analytics could provide this sort of value to our business. It's not a question of where we should apply data anymore, it's a question of where we can gain insights first."

– Lead Technical Architect, $1 Billion Consumer Electronics Firm

---

Data Platform Maturity Stage | **3-Year Operating Margin ('12-'14) by Maturity Stage**



This research finds that organizations with leading Data & Analytics capabilities also outperform the others on a variety of company productivity and profitability measures. The table below outlines the differences between top and bottom quartiles of companies:

	Laggards – Stage 1 *(Bottom 25% of enterprises)*	**Leaders – Stage 4** *(Top 25% of enterprises)*
Average revenue per employee[2]	$473K	$507K
Three Year Average Gross Margin[3]	37%	55%
Three Year Average Earnings before Taxes	11%	16%
Three Year Average Net Income	7%	11%

B. Data as a Strategic Asset

Organizations with the leading Data & Analytics capabilities recognize that data is a strategic asset which differentiates them in the market. Leading enterprises have pursued a strategy of aggressively and systematically collecting data and deploying systems to process and manage a large influx of data, develop

[2] Revenue per employee based on CY2014 revenue
[3] Three year averages based on CY2012 – CY2014

business insight, and take action based on analytical models, while simultaneously protecting sensitive or confidential information they maintain.

Leading organizations on average store and manage over 22 petabytes of data versus laggards who on average handle 0.5 petabytes. Leaders collect data from all manner of sources. Data is produced and captured from their operations through business applications and websites, collected from customers and partners, obtained from third-parties, gleaned from the internet and other public resources, and increasingly generated from sensors and connected devices embedded within their company's products and services.

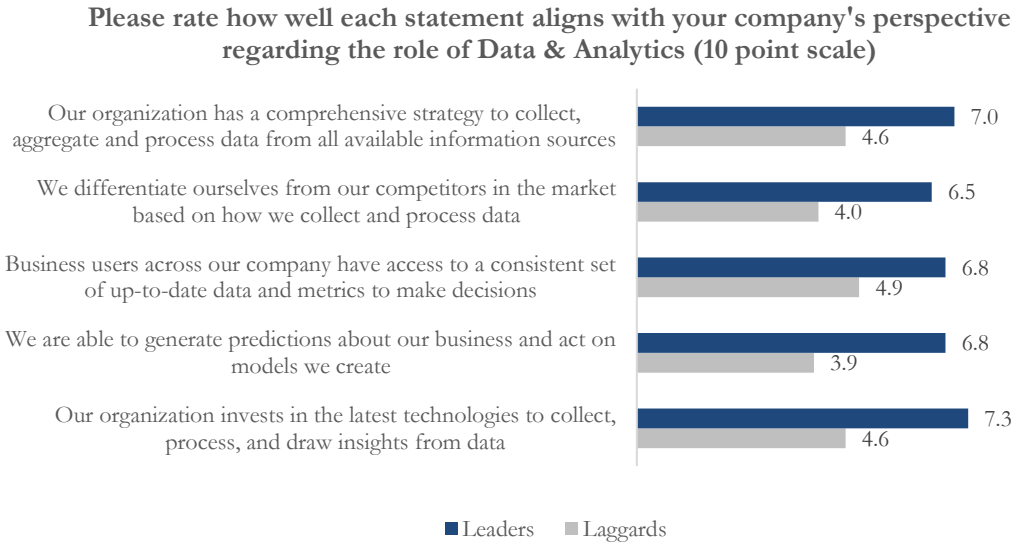
"You don't really know the value of data before you have it. Data is an investment, but once you have it, you can do amazing things. We're creating entirely new revenue streams from the information we've started collecting."

– VP, Data & Analytics, *Fortune 100* Industrial Goods Manufacturer

---

Enterprises with leading data platforms view this data and their technology investments to process, store and analyze data in starkly different terms. Leading organizations are more likely to state that they have a comprehensive data acquisition strategy, that their data platform is differentiated from those of competitors, that business users have access to a consistent set of up-to-date metrics for decision making, and that they are able to generate predictions about their business from data.

The following table characterizes some of the key attitudinal differences between leading and lagging organizations as to the importance of data to business strategy and operations:



These attitudes regarding the importance of data signal some differences as to how enterprises are using data to guide and optimize their business operations and the technical capabilities they have put into place to realize the full potential of their data. This research has found that the attitudes about the importance of data is reflected in the technological capabilities enterprises have put into place and in the business processes they have enabled.

C. Data & Analytics Enabling Business Operations

Leading organizations have realized the benefits from their data strategy in how they carry out their business processes. Data is being used to make decisions more rapidly with a more complete understanding of the market and customer preferences, optimize business operations, develop differentiated products and services, and augment workforce productivity. Top organizations have not only consolidated information across the organization to develop a "single version of the truth" about their business, but they are using real-time data to anticipate changes in their business and take corrective action. Top enterprises are using Business Intelligence tools and analytical models within their systems to develop tailored customer experiences, mitigate the risk of customer churn, identify proactively customer support issues, preempt an equipment failure, and make real-time decisions to run their business more efficiently. Through this research, we have observed striking differences between leading and lagging enterprises across Sales & Marketing, Engineering & Operations, Risk & Fraud, and Finance, HR & Back-office functions.

1. Sales & Marketing

Leading enterprises are using data to better understand the market, more efficiently acquire new customers and optimize their advertising effectiveness, better engage customers, anticipate their needs, and limit customer churn. Data gathered from across the customer lifecycle helps these enterprises make more informed decisions, present customers with tailored offers and experiences, and mitigate support issues using a 360-degree view of their customers based on an understanding of customer behavior online and offline.

Key Sales & Marketing capabilities that set apart leading enterprises include:

- Customer Acquisition
  - 2.2 times more likely to personalize their website experience in real-time based customer segment or customer behavior
  - 1.7 times more likely to identify high value customers based on their behavior using predictive models (e.g., identify customers that are currently "in-market")
  - 1.6 times more likely to be able to attribute value to marketing touchpoints that influence customer behavior
- Customer Retention & Monetization
  - 2.7 times more likely to programmatically identify at risk customers and suggest appropriate corrective action
  - 2.5 times more likely to develop perceptual intelligence about their customers, using data to infer characteristics about their customers that cannot be observed directly such as mood, gender, etc.
  - 1.4 times more likely to develop models to forecast demand and set prices

2. Engineering & Operations

Within Engineering, Manufacturing, and Operations functions, leading enterprises are using data to develop an integral understanding of their business's functioning by consolidating information across the product development lifecycle and supply chain. This aggregated data is being used to understand drivers of operational efficiency and product quality, anticipate equipment or operational downtime, and ensure process excellence across distributed production and operations facilities. Increasingly, top-performing enterprises are using Internet of Things technologies to instrument their products and services with connected sensors and gather data about product use and performance. This data in turn allows

enterprises to optimize their service operations and transform how they deliver and capture value from their customers.

Distinguishing Engineering & Operations capabilities of top enterprises include:

- Operations & Process Optimization
  - 1.7 times more likely to be able to predict equipment downtime using advanced analytics
  - 1.5 times more likely to optimize production runs based on demand forecast
  - 2.3 times more likely to use predictive modeling to anticipate customer support requests
- Data-Driven Products & Services
  - 2.3 times more likely to inform product design by capturing data on product use and performance
  - 1.8 times more likely to monitor products remotely and drive customer support based on insights
  - 1.9 times more likely to use data they collect to benchmark their customers and advise them on how to realize more value from products and services

3. Risk & Fraud

Data is often applied to minimize and limit fraud, and this is particularly prevalent among financial services institutions. Data can be used to anticipate operational risks to the business associated with market factors and credit risk. Increasingly top-performing companies are using unstructured data gathered from the internet, social media, and websites to identify patterns of customer behavior that may impact the business. Similarly, data can be used to detect and avert fraud in real time, identify security breaches and suspicions activity at it happens, and detect employee actions that might be linked to fraudulent activities within minutes.

Distinguishing Risk & Fraud capabilities of top enterprises include:

- Risk Estimation & Management
  - 2.2 times more likely to programmatically analyze unstructured data to assess risk posed to business performance
  - 1.6 times more likely to analyze multiple sources of data such as financial data, credit history, emails or social media to estimate credit or operational risk
  - 1.6 times more likely to collect and analyze information from the internet to estimate and manage risk within organization

- Fraud & Threat Prevention
    - 1.4 times more likely to monitor and analyze employee actions in real-time to identify fraudulent activities
    - 1.3 times more likely to identify and track suspicious activities as it happens
    - 1.2 times more likely to use pattern recognition models to identify security breaches or hacks in real time ad flag for corrective action

4. Finance, HR & Back-Office

Leading companies across multiple verticals employ data to increase agility of their business by analyzing their performance quicker and in centralized fashion. In addition, these companies use data to develop forecasts and recommendations across the whole spectrum of supporting functions, from optimizing business strategy to automating creation of individual development plans for employees.

Distinguishing Finance, HR & Back Office capabilities of top enterprises include:

- Business Operations Agility
    - 1.5 times more likely to use financial reporting tools to assess financial performance across the enterprise in near real time
    - 2.7 times more likely to automatically analyze employees' performance and actions and suggest actions for improvement
- Forecasting & Budgeting
    - 1.2 times more likely to view financials and conduct financial forecasting and budgeting from an enterprise-wide dashboard
    - 1.3 times more likely to run what-if analyses within budgeting and forecasting systems to project financial results

4. Data & Analytics Evolution

A. Comparison of the Data Platforms of Leading & Lagging Enterprises

Leading enterprises have very different technical capabilities than organizations who have not invested in their Data & Analytics platform. Enterprises whose data platforms were profiled as part of this research were graded across seventy-four technical capabilities spanning six major product areas.

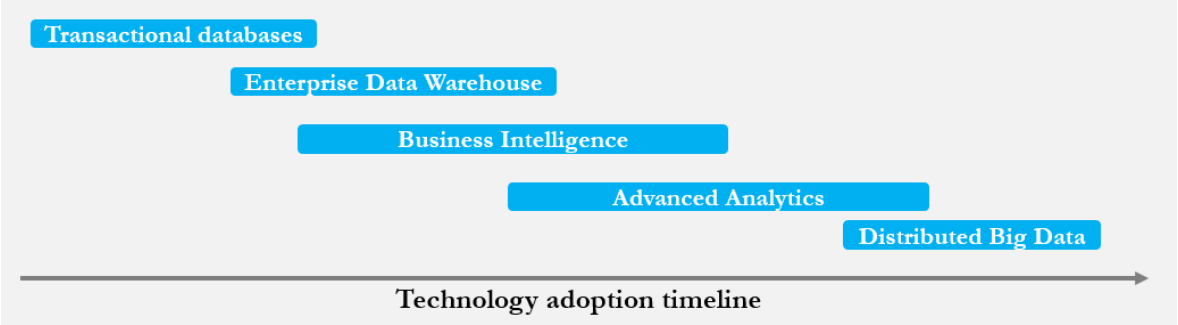
The following table describes the differences in platform capabilities of Leading and Lagging organizations:

	Laggards – Stage 1 *(Bottom 25% of enterprises)*	**Leaders – Stage 4** *(Top 25% of enterprises)*
Operational Databases	▪ Older versions of traditional relational databases	▪ Newer versions of relational databases ▪ In-memory technologies ▪ Implementation of real-time analytics in key transactional systems ▪ Non-relational databases ▪ Public cloud for customer-facing, and mission critical applications
Enterprise Data Warehouse (EDW)	▪ Multiple data marts and data warehouses ▪ No single version of truth ▪ Data often not current	▪ Comprehensive enterprise data warehouse ▪ Single version of truth ▪ Real-time refresh of data
Business Intelligence (BI)	▪ Multiple business intelligence tools ▪ Limited "self-serve" BI capabilities for ad hoc analysis	▪ Centralized dashboards for key business metrics ▪ SaaS-based BI packages ▪ Access to reports and alerts on mobile devices ▪ Embedded BI in key systems
Advanced Analytics	▪ Limited ability for statistical modeling, typically performed one-off or in batch	▪ Real-time predictive and prescriptive analytical models integrated in key systems ▪ Use of perceptual intelligence ▪ Use of public cloud for Big Data analysis and real-time stream analytics
Enterprise Data Lake (EDL)	▪ N/A	▪ Centralized repository for non-relational and unstructured data ▪ Ability to process and analyze unstructured data
Cloud Computing Infrastructure	▪ Virtualization of infrastructure but no use of public cloud	▪ Automated provisioning of virtual machines in private cloud ▪ Cloud-based back-up and recovery ▪ Hybrid cloud

B. Data & Analytics Journey

Enterprises who have attained a high-level of business and technical capabilities with their data platform have done so over time with a clear vision of maximizing the value of their data. Enterprises face different business priorities and develop distinct technical roadmaps to evolve their capabilities and

realize business impact from their Data & Analytics investments. However, certain commonalities can be observed across our sample.



Operational databases are the core of the Data & Analytics platform, and top-performing companies have focused on steadily evolving their database capabilities through software upgrades and the deployment of the latest technologies. Operational databases are becoming more scalable, available and secure. Leading enterprises are applying in-memory database technologies and are beginning to implement real-time analytics within transactional databases. This allows organizations to take business actions immediately without having to extract and analyze transactional data in another system or after the fact. Further, leaders are deploying non-relational databases to complement their relational database investments. NoSQL is often used to enable high-scale, high performance web applications or to power Internet of Things-related use cases. The public cloud extends the capabilities of many organizations' transactional databases through back-up and recovery and hosted applications.

"We store data and run analytics extensively in relational databases, we use SQL Server on Azure for that. It provides us with better flexibility and control of our data."

– VP, Data Analytics, $10 Billion Industrial Tools Manufacturer

---

Establishing a modern Enterprise Data Warehouse is a second major step in building an advanced data platform. Major enterprises are often faced with a profusion of isolated data stores which have grown over time through mergers and acquisitions, the launch of different business applications, and uncoordinated technical projects within company divisions or departments. This problem is compounded by the lack of unified customer, supplier and product identifiers. The result is that data crucial to the business is stored in multiple, non-integrated systems and decision makers do not have a single, up-to-date, and reliable version of the truth. Hence, business decision making is impeded by

discussions as to which numbers are correct and how best to obtain current data rather instead of being focused on substantive discussions as to which actions to take based on data insights.

Enterprise Data Warehouses are often at the foundation of Data & Analytics roadmaps. Identifying key data elements, defining common data dictionaries and interpretations of business metrics, reducing duplication of data across multiple data stores, and aggregating the information most crucial to reporting, analytics, and decision making is essential to the development of a robust Enterprise Data Warehouse. Over time, enterprises with successful Data & Analytics strategies develop data warehouses and archiving strategies that support the needs of business stakeholders across the enterprise with information that is both comprehensive and current. Some of the most mature enterprises are beginning to put parts of their data warehousing infrastructure in the cloud for higher scalability, greater flexibility in collecting data from geographically distributed locations, and accelerated time to market.

"We didn't want to spend too much money or time building up a server farm with 24/7 availability. We decided to use Azure SQL to aggregate all the data from multiple channels."

– Senior Director, Information Technology, Fashion Retailer

The next step for most enterprises is to standardize their Business Intelligence tools for reporting and self-service analysis. Most enterprises are faced with numerous overlapping BI tools which have been added over time to support different business applications or the needs of different divisions. Top performing companies have standardized on one enterprise BI tool for standard report generation and syndication and one "self-serve" BI system to allow data analysts and a broad cross-section of business users to run ad-hoc analyses and delve deeper into data to uncover new insights and adapt to changing business circumstances. SaaS-based BI is also becoming popular among top-performing enterprises who are starting to provide BI system access to a wider range of internal and external users.

"My team is using self-serve BI because it allows us to cut the data very quickly in different ways. We also use it as a customer-facing tool, to provide clients with interactive dashboards."

– VP, Data Analytics, $10 Billion Industrial Tools Manufacturer

---

With the deployment of an enterprise data warehouse and BI tools, most enterprises next add advanced analytics systems for in-depth statistical analysis, modeling, and decision support. Advanced analytics tools leverage data contained in the Enterprise Data Warehouse and can be used to forecast demand from distinct customer segments, model a customer's propensity to purchase a product or churn, anticipate equipment failure and take corrective action, optimize pricing models, minimize financial risk or fraud, or identify high performing employees. Often, advanced analytics tools are used in isolation to explore data and uncover new insights. However, leaders frequently deploy sophisticated models within their business operations to enable real-time decision making and optimize business operations. For instance, top retailers might use advanced analytics to identify "in-market" customers based on their online or offline behavior and present him or her with customized offers or a tailored user experience to maximize the chances of appealing to the customer. Similarly, machine learning technologies might be used to identify patterns in website traffic and optimize advertising copy based on which text or images on the website optimizes customer conversion rates. Companies might utilize the benefits of public cloud to offload compute-intensive tasks to on-demand virtual machines and use powerful cloud-based analytics services to derive insight from very large scale data sets. Companies pursuing IoT strategies often employ the cloud to analyze in real-time massive amounts of machine data using stream analytics.

"Investing in unstructured data analytics and Hadoop has allowed us to upsell and cross-sell our customers in real time when they visit our website leading to greater customer lifetime value"

– Director, Data Science, $2 Billion Online Furniture Retailer

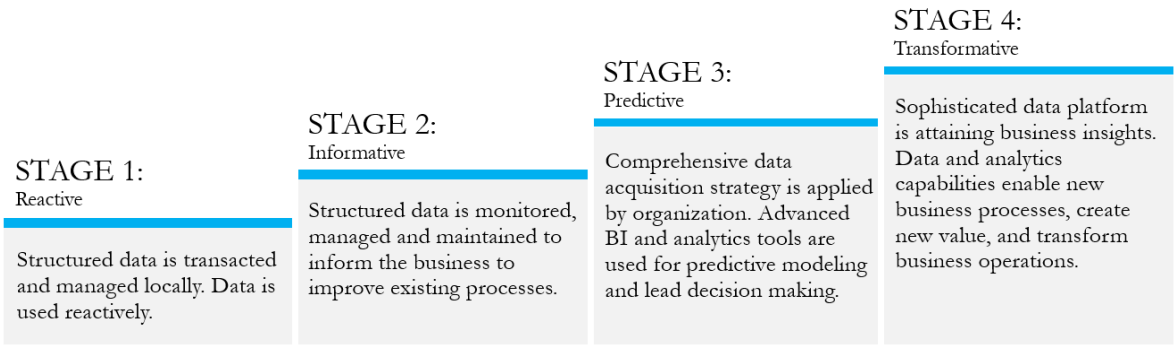
---

With a solid Data & Analytics foundation, companies tend to focus next on deriving value from large volumes of historical data and unstructured data like web, machine log, geospatial, and video and image data using Distributed Big Data technologies, such as Hadoop. 31% of enterprises in our study have a material Distributed Big Data deployment today. Of those who are using technologies such as Hadoop, just over 10% have designated an Enterprise Data Lake to consolidate, store, and process unstructured data as traditional Distributed Big Data technologies have required specialized data science and programming training. Hadoop and other Distributed Big Data technologies can be deployed on premises or in the cloud. Frequently, customers will opt to deploy Hadoop in the cloud, especially for use cases involving a large number of distributed connected devices or in cases where very large amounts of data that must be processed by a large cluster.

5. Data & Analytics Maturity Model as a Guide for Roadmap Planning

A. Data & Analytics Maturity Model Overview

Enterprises in the study were split into four stages based on overall Data Platform maturity. In the most basic stage, companies are only starting to generate benefits from data and use it *reactively*. Going forward, organizations begin to use data more in *informative* way to support the business and improve processes, and then utilize *predictive* capabilities to lead decision making. In the final stage data becomes a *transformative* force, driving the business and creating new value.



B. Maturity Levels by Data Platform Product Area

Enterprises at different stages in the development of their data platform demonstrate differing sets of capabilities. A comparison of companies at different points in their data platform evolution outlines key stages in how enterprises can progressively develop their data platform.

The following table outlines how enterprises add capabilities within each of the product areas. This framework can guide roadmap planning for how companies might enhance their Data & Analytics capabilities over time.

	Reactive	Informative	Predictive	Transformative
Operational Databases	Organizations primarily use relational databases for key business applications. Databases can be scaled up by adding or upgrading existing hardware. Transactional data is backed up to a secondary data center for failover and recovery.	Operational databases can be scaled up through massively parallel processing and elastically scaled out through pools. Uptime availability is "five nines" or better. Role-based security is defined at the database level and data can be encrypted at rest and in motion.	Performance tuning is achieved through solid-state drives and flash-based storage. Database schemas can be dynamically updated. Some organizations deploy customer-facing apps on public cloud.	In-memory database technologies and non-relational databases are in use. Analytical models are often implemented within transactional systems to enable real-time operational decision making. Public cloud is often used for data back-up, recovery, failover, and mission critical applications.
Enterprise Data Warehouse (EDW)	Organization maintains multiple data marts for individual business functions and systems. Some organizations have begun consolidating silos of data into an Enterprise Data Warehouse, but refresh it intermittently.	Enterprise Data Warehouse aggregates data from multiple business applications and data sources. The enterprise data warehouse can scale to handle growth with automated scheduled frequent refreshes.	Enterprise Data Warehouse contains comprehensive data for most reporting and analytics needs. Data warehouse contains six or more years of data and allows users to query historical data without pulling from cold storage. Redundant data marts have been retired in favor of a centralized enterprise data warehouse.	Some or all data stored in the Enterprise Data Warehouse is refreshed in real time or is streaming. Predictive modeling is often an extension of data warehouse. Data from public web and social media sites augments internal data sets. Public cloud is used for some data warehousing needs.
Business Intelligence (BI)	Business intelligence tools are in place for enterprise reporting and self-serve analytics. Users can customize workspace in business intelligence tools and reports through limited service software packages.	Enterprise-wide definitions are in place for key business metrics. Power users create data mashups from multiple sources. BI tools often deployed on the cloud or as a SaaS solution.	Organizations have broad user access to BI using embedded dashboards within key applications and mobile BI. Users can perform ad hoc analyses and visualizations. Well-defined data dictionaries and data governance policies are established.	Organizations have established interactive and streaming BI dashboards for enterprise-wide KPIs. Business users are enabled to use BI tools to customize their view of the state of business and identify opportunities and risks through user-defined, custom alerts.

	Reactive	Informative	Predictive	Transformative
Advanced Analytics	Limited or only ad-hoc use of data analysis and modeling capabilities through manual entry and calculation.	Limited use of predictive and statistical modeling capabilities mostly through manual calculation and forecasting. Basic batch predictive models have begun to emerge.	Predictive models have been developed for multiple purposes including customer churn analysis and predictive maintenance for equipment. Organizations beginning to develop prescriptive models that generate recommended business actions for users based on analysis of data.	Predictive models are based on real-time data streams and update dynamically. Models are deployed within key business applications to support real-time operational decision making and personalized recommendations. Data scientists have ability to build, refine, and select best model after having run multiple in parallel.
Enterprise Data Lake (EDL)	Organization makes little or no use of unstructured data. There is no use of Hadoop or other distributed big data technologies.	Unstructured data is stored in a centralized repository, but limited analysis is performed on unstructured or non-relational data.	Organizations use Hadoop and other distributed big data technologies on a pilot basis. Key unstructured data are transformed in batch into structured for analysis.	Organization has established an Enterprise Data Lake on premises or in the cloud to house structured and unstructured data. Data often arrives in real time and analytics can be run within data lake.
Cloud Computing Infrastructure	The majority of enterprise infrastructure has been virtualized, but most organizations are not using public Infrastructure- or Platform-as-a-Service (not using IaaS or PaaS).	Organizations auto-provision infrastructure in a private cloud. Some organizations are experimenting with public cloud, often for development and test environments.	Organizations use cloud infrastructure for specific production applications. For instance, for temporary tasks where it is necessary to spin up a large number of instances rapidly.	Organizations deploy "hybrid" applications where certain elements of the system are deployed using a public IaaS or PaaS provider and other aspects (e.g., sensitive corporate data) are maintained behind the corporate firewall.

6. Appendix: Data & Analytics Maturity Model Case Examples

A. Online Furniture Retailer

1. Company Overview

The company is a US-based e-commerce home furnishing retailer, generating $2 billion in annual revenue, with a year-over-year growth rate of 70% from 2014 to 2015. This retailer has a catalog of tens of millions of products and houses hundreds of terabytes of data, most of which are unstructured clickstream data and server logs.

2. Business Use Cases and Data Pipeline

The company has identified several business goals for its data strategy. Business use cases include sales and marketing effectiveness and business operations:

- *Optimization of customer acquisition*:
  - Personalizing website experience in real-time to increase sales by providing tailored recommendations to customers and optimizing online merchandising based on user behavior
  - Optimizing email marketing effectiveness by adjusting messaging and content based on products viewed during customer's website visitation
- *Operations optimization using analytical models*:
  - Creating demand forecasting models based on historical inventory holding levels and supplier performance data to minimize back orders
  - Using analytical models to predict damage to goods in shipment. Fulfillment data is used to inform supply chain operations (e.g., goods found especially prone to damage are over packed)

"Initially we focused on sales and marketing scenarios and personalization. We tailored our website and digital marketing to each customer's specific needs. Now, we are turning our attention to how we can impact operations and fulfillment."

– Director, Data Science, $2 Billion Online Furniture Retailer

---

The company created an advanced data pipeline to solve these business needs. To support marketing, unstructured data (such as website clickstream activity) is accumulated in Distributed Big Data environment. There, it is cleansed, processed and piped into the company's analytics package. Analytical models are developed, proven and ultimately implemented in transactional systems. Machine learning models have for instance been developed to identify in-market customers or customers who have just moved. These models are then used to customize the customer's website experience and special offers to maximize the likelihood of purchase.

3. Data & Analytics Maturity Assessment

The following figure indicates the maturity of the company in each of the six data platform workloads. This benchmarking analysis illustrates that the company has advanced data platform capabilities in all areas of their data platform, except in its use of cloud where it lags relative to other companies. To achieve this level of data platform maturity, the company has deployed modern transactional databases which use in-memory technologies for operational analytics. The retailer has provisioned their own Distributed Big Data cluster and created a data science team which develops analytical models using programming languages with machine learning libraries and assists in the implementation of these models. The company has made a conscious decision not to use the public cloud in order to have greater control of its technology stack, minimize costs, and reduce a potential competitive threat from Amazon.com.

	Maturity stage	Current State	Potential Gaps
Operational Databases	✓ **Stage 4** *Transformative*	▪ In-memory technologies ▪ Real-time operational analytics	▪ Store unknown data type
Enterprise Data Warehouse	✓ **Stage 4** *Transformative*	▪ EDW comprehensive for BI, reporting, and analytics	▪ Real-time / near real-time refresh in EDW ▪ Real-time analytics
Enterprise Data Lake	✓ **Stage 4** *Transformative*	▪ Enterprise Data Lake (EDL) ▪ Real-time refresh in EDL	▪ Real-time EDL analytics ▪ Cloud-based EDL
Business Intelligence	✓ **Stage 4** *Transformative*	▪ KPIs in central dashboard ▪ Real-time alerts for business users	▪ Mobile BI ▪ SaaS-based BI
Advanced Analytics	✓ **Stage 4** *Transformative*	▪ Real-time predictions ▪ Models implemented in key software systems	▪ Cloud-based analytics
Cloud Infrastructure	✓ **Stage 2** *Informative*	▪ Majority of systems virtualized	▪ Auto-provisioning for private cloud ▪ Dev/test of data solutions in cloud
OVERALL	✓ **Stage 4:** *Transformative*		

4. Future Plans

The company is planning to continue developing its data platform to address new business needs. This includes implementing mobile BI technologies, which will allow employees to gain access to business dashboards remotely and make better decisions in the field. Mobile BI is viewed as especially important for buyers who are evaluating suppliers. Additionally, this company seeks to enhance its ability to programmatically identify and tag product attributes in its online catalog based on a product's image and text description. To enable this, company's data scientists are applying open-source machine learning technologies to the repository of tagged product descriptions and images that company accumulated. This capability will ultimately make it easier for customers discover products in the catalog.

B. Industrial Conglomerate

1. Company Overview

The company is a multinational innovative technology and manufacturing conglomerate — a *Fortune 100* company that has about 130,000 employees worldwide and $40 billion in sales. The manufacturer provides aerospace equipment, develops automation & control solutions, advanced materials, and process control technologies.

2. Business Use Cases and Data Platform Roadmap

The company views data as a strategic resource to create new products and provide new services for its customers. For instance, recently the company announced the launch of an IoT division to provide industrial customers with software and analytics services for connected devices.

Company's strategy with regards to data includes both creating better data-enabled products for their customers, and building an ecosystem of various industrial and service partners who receive value from the data company aggregates. Some business cases where Data & Analytics technologies are applied include:

- *Failure prediction service for power management*: Detecting patterns in the data collected from devices and predicting energy use across the country to respond to demand and react to predicted equipment failures.
- *Smart home products for customers*: Developing innovative connected "smart home" appliances that allow customers to use their phones for remote control and over time learn from consumer behavior to improve user experience.

In order to enable these strategic applications of data, the company needed to significantly enhance its data platform. On top of the relational databases, data warehouses and legacy reporting, the company has built several Hadoop clusters on Azure Public Cloud to collect data from connected devices, added programming languages for predictive models, and NoSQL databases to support real-time analytics.

3. Data & Analytics Maturity Assessment

The following graph indicates how the company's abilities in each of the six data platform workloads compare with other enterprises included in the study. The company significantly outperforms the average organization in almost all areas. For instance, the company makes extensive use of public cloud infrastructure (deployed to overcome performance limitations with its legacy on-premises infrastructure) and is using SQL Server extensively for custom applications. For Advanced Analytics, the company develops predictive models using programming languages and has deployed these analytics in transactional databases. This company lags other organizations in Business Intelligence which has been deprioritized in favor of Operational Databases, Enterprise Data Warehousing, Enterprise Data Lakes, Advanced Analytics, and Cloud Infrastructure.

	Maturity stage	Current State	Potential Gaps
Operational Databases	✓ **Stage 4** *Transformative*	▪ In-memory technologies ▪ Real-time operational analytics ▪ Store unknown data type	▪ "5 nines" availability ▪ Dynamic data schema
Enterprise Data Warehouse	✓ **Stage 4** *Transformative*	▪ Real-time/near real-time data refresh and analytics in EDW ▪ Cloud-based EDW	▪ EDW comprehensive for BI, reporting, and analytics ▪ Query data going back 6 years+ ▪ Unique customer identifier
Enterprise Data Lake	✓ **Stage 4** *Transformative*	▪ Enterprise Data Lake (cloud-based) ▪ Real-time EDL refresh ▪ Real-time EDL analytics	▪ N/A
Business Intelligence	✓ **Stage 1** *Reactive*	▪ BI tools query data directly and users can create data mashups ▪ SaaS-based BI	▪ Users create interactive visualizations of data ▪ Mobile BI ▪ Users have well-defined dictionaries
Advanced Analytics	✓ **Stage 4** *Transformative*	▪ Real-time and stream analytics ▪ Predictive models implemented in key software systems ▪ Dynamic updates with new data	▪ Prescriptive models ▪ Quickly build, iterate, and test models in parallels
Cloud Infrastructure	✓ **Stage 4** *Transformative*	▪ Cloud for dev/test and compute-intensive tasks ▪ Hybrid cloud applications	▪ Auto-provision infrastructure in private cloud
OVERALL	✓ **Stage 4** *Transformative*		

4. Future Plans

The company is currently migrating many systems to the cloud and is actively assessing which workloads are best suited to the public cloud.

"Our goal is to get out of business of running data centers."

– Director, Information Technologies, *Fortune 100* Industrial Goods Manufacturer

---

Additionally, the company continues prioritize investments in its Data & Analytics technologies which can generate new revenues for the company to advance its data platform and expand data-based services for their customers. For instance, this company seeks to monetize data it collects by analyzing the performance of home appliances. When devices are performing poorly or likely to require maintenance, the company delivers alerts to utilities and local contractors who can assist customers with service and equipment repairs. Data serves to differentiate the company's products and can create a better customer experience while also providing value to the company's ecosystem and partners.

C. Property & Causality Insurer

1. Company Overview

The company is one of the leaders in auto, home and life insurance in the US, serving more than 15 million households nationwide. It is a public S&P 500 company, with annual revenues of more than $30 billion and 40,000 employees.

2. Business Use Cases and Data Platform Roadmap

The company applies Data & Analytics technologies both to their core business problems and to enable novel business scenarios.

- *Risk and Fraud Prediction*: Company has built "self-learning services" for predictive models and fraud detection by collecting data in Hadoop and applying machine learning technologies. New technologies have significantly improved the accuracy of predictions for business challenges such as assessing insurance risk and a customer's propensity to file claims.
- *New personalized pricing through telematics*: Company has built a custom IoT platform to collect telematics data about driver behavior. Telematics data allows the company to gauge the level of risk a customer presents and more accurately price insurance policies based on driver habits.

"We are very proactive with our business stakeholders and constantly look for ways we can provide value with data."

– VP, Big Data Analytics, *S&P 500* Insurance Company

The company's data platform has evolved significantly since 2012, when a CDO (Chief Data Officer) was appointed to build a Data & Analytics team. The new department has implemented an Enterprise Data Lake on Hadoop and introduced several open-source analytics technologies.

3. Data & Analytics Maturity Assessment

Relative to other companies in the research, this company has mature EDL, BI and Advanced Analytics capabilities. The data lake and analytics are managed by a technology team which reports to the CDO and resides within a function separate from corporate IT. IT supports the operational database and data warehousing systems. Operational database capabilities have not evolved as quickly as other areas of this

company's data platforms. Many systems still run older versions of database software because of the complexities associated with upgrading systems for legacy business applications. Nevertheless, the company has started using in-memory technologies in operational databases with some systems and is using NoSQL for master data management in its Enterprise Data Warehouse.

"One of the key things in our company is having an analytics team outside of IT. Our data scientists sit in R&D and report to the business. This gives them more freedom to innovate."

– VP, Big Data Analytics, *S&P 500* Insurance Company

The company has a sophisticated private cloud with auto-provisioning, but has opted not to use the public cloud because of regulatory concerns.

	Maturity stage	Current State	Potential Gaps
Operational Databases	✓ **Stage 1** *Reactive*	▪ Failover to secondary data center ▪ In-memory technologies	▪ Vertical and horizontal scaling ▪ "5 nines" availability
Enterprise Data Warehouse	✓ **Stage 3** *Predictive*	▪ EDW in place with daily refresh ▪ Append 3rd-party and web data to customer records	▪ EDW comprehensive for BI, reporting, and analytics
Enterprise Data Lake	✓ **Stage 4** *Transformative*	▪ Enterprise Data Lake established ▪ Real-time refresh of data in EDL ▪ Real-time analytics on EDL	▪ Data Lake on cloud for streams from sensors deployed through telematics initiative
Business Intelligence	✓ **Stage 4** *Transformative*	▪ Self-serve BI ▪ KPIs in central dashboard ▪ Mobile BI	▪ Real-time alerts for business users
Advanced Analytics	✓ **Stage 4** *Transformative*	▪ Models implemented in key software systems ▪ Real-time predictions	▪ Models update dynamically as new data comes in
Cloud Infrastructure	✓ **Stage 2** *Informative*	▪ Auto-provision infrastructure in private cloud	▪ Dev/test of new data solutions in cloud
OVERALL	✓ **Stage 4:** *Transformative*		

4. Future Plans

The company continues to develop its data platform, in particular its Enterprise Data Lake and Advanced Analytics capabilities. For its Data Lake, the company is experimenting with new technologies to gain more flexibility in how it manages data and uses machine learning algorithms on Hadoop. Specifically, the team is seeking to secure funding to implement open-source ETL tools for Hadoop. The company

is also evaluating technologies that would decrease the deployment time of analytical models. Currently, it takes between 6 to 9 months to implement a model in production systems. The company hopes to make this process substantially faster. Finally, the company is considering using public cloud infrastructure more broadly, starting with use cases that fall outside of government and industry regulations. For example, the company is considering using the cloud to capture data streamed from sensors deployed through its telematics program.

D. Toy manufacturer

1. Company Overview

The company is a 70-year old toy manufacturer with $6 billion revenue. The company was historically selling through a wholesale channel, but recently started to increasingly reach end customers directly through e-commerce. Additionally, it is transforming the products it develops by instrumenting them with sensors and gathering data on how children and parents interact with the toys they manufacture.

2. Business Use Cases and Data Platform Roadmap

The company started to invest in new technologies and data-driven products to support shifts in their business:

- *New data sources for better, direct engagement with customers*: The company started to collect unstructured web data from social networks in addition to point-of-sale and e-commerce data to enrich its understanding of its customers such as their perceptions of the brand and their preferences in how they find out about and purchase toys. This company also uses data to refine its demand forecasting models using new metrics and data sources it has not had traditionally.
- *Products transformed through data*: The company introduced new, differentiated digital products for a technology-savvy generation of customers. Data is used to understand how consumers interact with their products and to enrich the user's experience. Examples include:
    - Connected toys that recognize the user's voice and respond accordingly.
    - Baby monitors that collect data about a child and his or her environment in order to provide feedback to parents.

"We want to consolidate all the data – from social networks, point of sale, ecommerce, etc. to better understand customers and support our marketing and product development"

– Senior Director, BI Platform, $6 Billion Toy Manufacturer

Many aspects of this company's data platform have been in place for many years and have not been updated. It continues to use traditional relational databases and enterprise BI packages extensively. Within the past two years, the company started to do significant investments in its core data infrastructure to support the strategy of better customer engagement. It moved its traditional data warehouse to a modern data appliance and has added self-serve and SaaS-based BI packages to empower business users and get insights from new sources of data like social networks, and is beginning to experiment with cloud-based Big Data and analytics technologies.

3. Data & Analytics Maturity Assessment

This company has mature Enterprise Data Warehouse and Advanced Analytics capabilities. For the Enterprise Data Warehouse, where the company implemented a modern appliance solution, it has integrated the data from recently acquired companies. In analytics, modern cloud-based technologies have enabled advanced capabilities such as predictive and prescriptive models, and data stream processing. In other product areas, such as operational databases and Data Lake, efforts to modernize are under way.

	Maturity stage	Current State	Potential Gaps
Operational Databases	**Stage 2** *Informative*	▪ 5 nines availability ▪ Scale up ▪ SSDs / Flash tuning	▪ Elastically scaling out ▪ In-memory technologies ▪ Dynamic data schema
Enterprise Data Warehouse	**Stage 3** *Predictive*	▪ EDW comprehensive for BI, reporting, and analytics	▪ Real-time / near real-time refresh in EDW ▪ Real-time analytics
Enterprise Data Lake	**Stage 1** *Reactive*	▪ N/A	▪ Centralized repository for unstructured data ▪ Establishment of EDL
Business Intelligence	**Stage 2** *Informative*	▪ Self-serve BI – data mashups by power users ▪ SaaS-based BI	▪ Uniform set of definitions for KPIs ▪ Centralized reporting dashboard ▪ Mobile BI
Advanced Analytics	**Stage 3** *Predictive*	▪ Prescriptive models ▪ Real-time data stream processing ▪ Cloud-based analytics tool	▪ Real-time predictions ▪ Models implemented in key software systems
Cloud Infrastructure	**Stage 2** *Informative*	▪ Majority of systems are virtualized	▪ Dev/test of new data solutions in the cloud ▪ Compute-intensive tasks
OVERALL	**Stage 3:** *Predictive*		

4. Future Plans

The company just started the transformation of its data platform and sees several opportunities for improvement. For instance, the company is rationalizing business applications and operational databases, and considers expanding its experiments with cloud-based Hadoop which it plans to develop into a full-scale Enterprise Data Lake over time.

The company also sees value in enhancing capabilities in workloads where it is relatively mature. It is expanding its analytics technologies, such as machine learning, to analyze consumer interaction with products. For its Enterprise Data Warehouse, it is focusing on data preparation and virtualization software for faster integration of new data from recently acquired companies, POS and business applications, and on implementing a Master Data Management (MDM) platform.