

# EXPLORING THE BIG DATA REVOLUTION



## ABOUT THIS REPORT

Big data is a collection of data and technology that accesses, integrates and reports all available data by filtering, correlating and reporting insights not attainable with past data technologies. Big data describes data processing beyond the human scale. In the past, databases tended to be limited—they only had to meet the demands of human users entering and retrieving data. With the emergence of ecommerce and Internet search engines, database technology is evolving to manage humans and computers.

Today, with the amount of data growing by 50 percent each year, it is information technology that is capable of managing, processing and finding value.

Approximately 21,000 APICS members and customers were invited to complete this survey to discover existing and emerging insights and innovations in big data.

This report was developed by APICS Supply Chain Council, an organization that advances supply chain and operations management and innovation through research, education and publications. APICS SCC maintains the Supply Chain Operations Reference (SCOR) model, the supply chain management community's most widely accepted framework for evaluating and comparing supply chain activities and performance. **For more information, visit [apicsscc.org](http://apicsscc.org).**

APICS SCC research reports are based on practitioner surveys that explore trending topics in supply chain and operations management. They include survey results, analysis, tips and best practices to keep you and your organization informed of insights and innovations in supply chain and operations management.

# EXPLORING THE BIG DATA REVOLUTION

## TABLE OF CONTENTS

<b>Executive Summary</b>	<b>5</b>
Key Findings	6
Respondent profile	18
<b>Big Data Glossary</b>	<b>22</b>
<b>APICS Research</b>	<b>26</b>

### Additional Resources

These programs and publications provide resources and detailed information about the topic.

APICS Risk Management  
Certificate Program  
[apics.org/risk](https://apics.org/risk)

A full version of this report is available free to APICS Supply Chain Council affiliates and sponsors and APICS members. Log in to the website to access additional analysis and insights on this topic. If you aren't an APICS member, join APICS today. Nonmembers may also purchase the full report.

Visit [apicsscc.org](https://apicsscc.org) for information about becoming a corporate affiliate or sponsor.

# EXPLORING THE BIG DATA REVOLUTION

## Big Data:

The *APICS Dictionary*, 14th Edition, defines Big Data as a collection of data and technology that accesses, integrates and reports all available data by filtering, correlating and reporting insights not attainable with past data technologies. It describes data processing beyond the human scale. In the past, databases tended to be limited—they only had to meet the demands of human users entering and retrieving data. With the emergence of e-commerce and Internet search engines, database technology is evolving to manage humans and computers. Today, with the amount of data growing by 50 percent each year, it is information technology that is capable of managing, processing and finding value.

## EXECUTIVE SUMMARY

The supply chain of the future will depend on information. Lots of it. By increasing your understanding of big data, positioning your organization and career, and knowing what to expect, you can be ready to capitalize on the tremendous opportunities and advantages that big data presents.

What is big data? There are many definitions, but experts agree that big data encompasses creating, enabling and operating a data storage, processing and reporting structure that enables deep insight resulting from the analysis of all available data. Big data becomes a foundation that permits applications such as data mining or data analytics.

Big data exists today because of the rapid growth in data volumes, data locations, data types, as well as data currency. By today's standards, an enterprise-scale organization may need to routinely process, moment by moment, hundreds of terabytes of information. Supply chains are no exception. As the complexity and scale of supply chains grow, so does their data.

Big data seeks to address data challenges by creating tools that capture all data and find insights, correlations, patterns and other value-adding information that transforms previous uncertainty, or lack of information, to greater certainty and awareness. Professionals who use big data effectively seek to improve forecasting, planning, situational awareness and information flow by creating a comprehensive understanding of complex perspectives. Big data helps to create a fluid perspective—from high-level strategic views to tactical views—by showing the connections, relationships, dependencies and patterns that previously were too difficult to interpret.

## KEY FINDINGS

Big data is a revolution of information technology that is affecting industries around the globe. Big data is a collection of data and technology that accesses, integrates and reports all available data by filtering, correlating and reporting insights not attainable with past data technologies. It describes data processing beyond the human scale.

In the past, databases tended to be limited — they only had to meet the demands of human users entering and retrieving data. With the emergence of ecommerce and Internet search engines, database technology is evolving to manage humans and computers. Today, with the amount of data growing by 50 percent each year, it is information technology that is capable of managing, processing and finding value.

APICS SCC conducted a survey to collect the opinions and perspectives of industry practitioners on big data. The survey explores the needs and expectations related to big data in the fields of supply chain and operations management. This research reveals the expectations and innovations that will help shape the direction of big data and improve the capabilities of professionals globally.

The big data survey revealed the following trends:

- **Data overload and the abundance of trivial information are challenges many organizations face.** There is a lot of data in general, but also a lack of useful data. High-level data is available, but not the detailed data essential to many decisions, plans, tasks and functions.
- **Important data is not reaching practitioners in efficient timeframes.** Fifty percent of respondents report there is an undesirable delay in receiving information about actual sales, demand forecasts, customer changes and orders and materials or component shipment status.
- **Despite current systems, data is still not always easily accessible.** There may be access restrictions, support limitations, security requirements, or out-of-sight-out-of-mind challenges. Another example is that physical or paper records are used in many cases. Separate databases serving departments or partners in other nations may be used. Access to this data may depend on factors practitioners are not aware of, such as availability or security.

- **Current information technology has not yet delivered optimal satisfaction in terms of what is easily measurable, reportable or quantifiable data, such as scheduling, inventory levels and customer demand across the supply chain.** Survey results suggest possible reasons, such as different data formats and systems and timeliness and data access challenges. This may challenge processes such as sales and operations planning (S&OP) that seek a shared and integrated understanding of supply and demand.
- **Supply chain dataflow includes direct suppliers and customers, but there are gaps in a complete or true end-to-end supply chain dataflow model for most respondents.** For example, only about half of survey respondents report that logistics, distributors and minor direct suppliers are part of their supply chain dataflow. True end-to-end supply chain dataflow, including suppliers' suppliers and customers' customers, remains a challenge. Supply chain and operations management professionals have more work to do in terms of improving tools, technologies, strategies and relationships, and big data can play a major role in this progress.

## Big data in supply chain and operations management

Data volumes and complexity continue to grow. In addition, there is the increasing challenge of unstructured data. Unstructured data is information that does not fit into traditional databases. Digital photographs and emails are examples. A database does not automatically know how to classify or relate this information with other data. For example: is the photo showing a location? People? A device? Was the email related to a decision about a production plan?

Structured data, in contrast, is data with a clear or inherent meaning and relationship to other data, for example, a postal code or a bank balance. Unstructured data is growing rapidly and becoming more relevant to supply chain and operations management processes, but remain a challenge to many current data and information systems. Big data promises to improve management, processing and integration of unstructured data as the volume and complexity of data increase. Managing complexity, improvement and making informed decisions are the goals of supply chain and operations management professionals and big data technologies will become a powerful resource.

## Questions for discussion

Put big data to work at your organization. Consider the following questions:

**Compare your organization's responses to the survey results. Does your organization have relative strengths or weaknesses? If so, these may suggest competitive advantages or disadvantages in your data systems or processes.**

**Consider new systems, upgrades or dataflow strategies. Do these really improve data timeliness, accessibility, processes and more?**

Many tasks that present challenges have to do with retrieving and decoding information. Some of the most difficult tasks are

- collecting data from multiple people or sources (52 percent),
- working with different data formats, standards and files (49 percent)
- selecting useful information from a large amount of data (39 percent)

About **50 percent** of participants feel they have too little data to successfully forecast.

**The top benefit** that would come from having supply chain and operations management data more accessible is in the area of planning and decision making (78 percent).

**53 percent of respondents** either have too much trivial data but not enough useful data, or have high-level data but not enough information to make detailed decisions.

## **S&OP and big data**

Balancing supply and demand requires information sharing excellence. Data sharing must be accurate, timely, comprehensive and relevant to S&OP processes in order for good recommendations and decision making to occur. As complexity grows in the domains of supply and demand, there is potential for S&OP processes to struggle because useful or relevant information does not always reach S&OP participants quickly enough. Big data will help uncover relevant and useful information and give S&OP participants access to useful information faster. Data points previously obscured will become visible and this will help reduce uncertainty in the S&OP process. This leads to more objective, data-driven perspectives among S&OP participants. Data-driven decision making should become easier as big data makes correlations between the drivers of supply and demand.

## **Put big data to work at your organization**

Big data is an immensely powerful, continuously emerging skill and knowledge area. It may pay to make big data a professional development priority because the expertise will be rare and in demand. In addition, big data will probably be implemented in long-ranging projects in increments as technology and resources permit. If big data were to resemble past ERP implementations, practitioners can help develop project management skills and high-level implementation ideas now and share those with management teams. This interesting and thoughtful approach may position a professional to advance in his or her career, improve a team's performance, or gain strategic knowledge of an organization's supply chain and operations management environment.

Big data best practices

Follow these big data best practices at your organization:

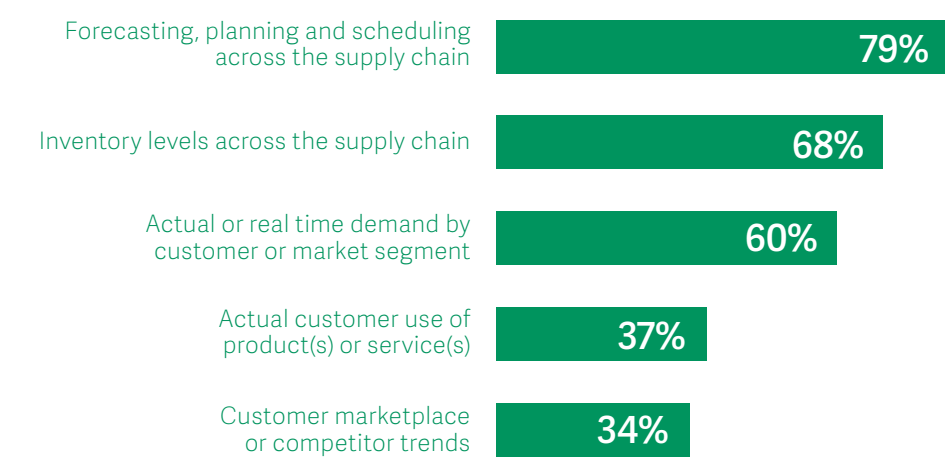
**Develop good relationships with supply chain partners.** Good relationships are necessary to facilitate shared data and insight. Build a foundation with these relationships in place, before you begin asking partners for advanced data sharing.

**Address supply chain dataflow gaps where possible as soon as possible.** Your organization may have developed procedures or practices that make such gaps manageable, but are not as good as they could be. As powerful as the promise of big data is, it probably can't correct processes and procedures built around existing gaps.

**Look for areas of correlation that don't seem obvious now, or seem too complex to study at this time.** For example, if demand decreases at times when it is expected to increase and there isn't a clear explanation, this is a good assignment for big data systems. A new big data implementation may not otherwise know what has high priority to the organization in terms of correlation and investigation.

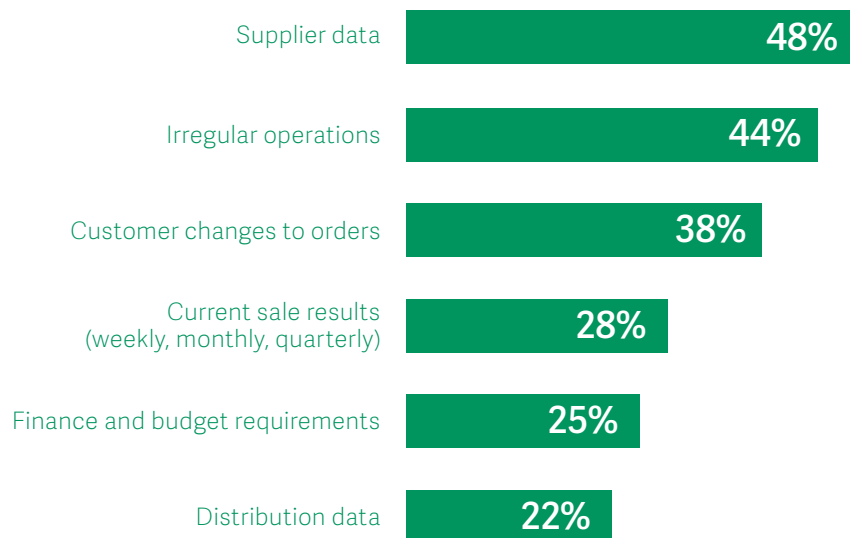
Delivery of actionable information

Respondents were asked what areas of supply chain information should deliver more actionable information and insight.



## Timeliness of data

Respondents were asked which data are not reaching them in a timely manner.



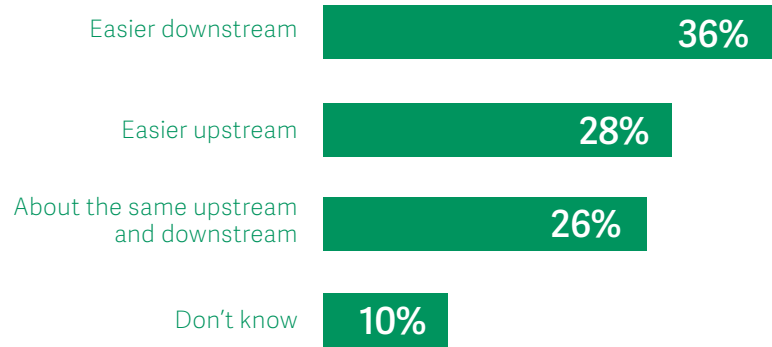
## Supply chain dataflow

Respondents were asked which of the following are part of their supply chain dataflow.



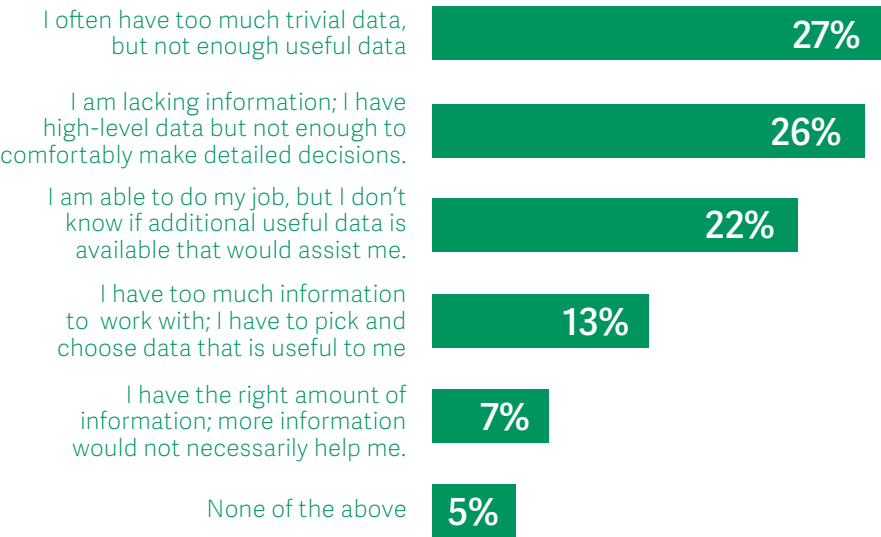
Upstream and downstream dataflow

Respondents were asked if their supply chain dataflow and data access is easier in the upstream or downstream direction.



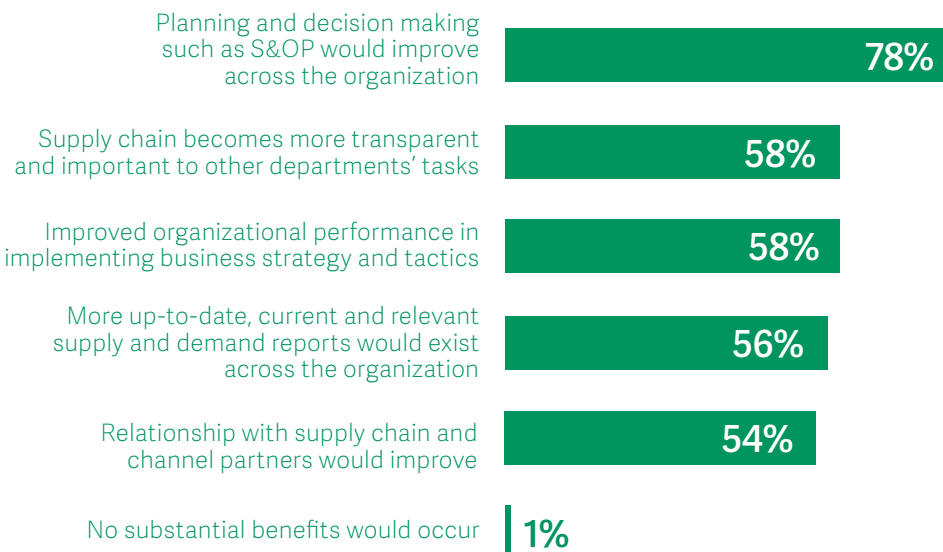
Data sufficiency

Respondents were asked which statement best matches their opinions about actual data sufficiency on the job:



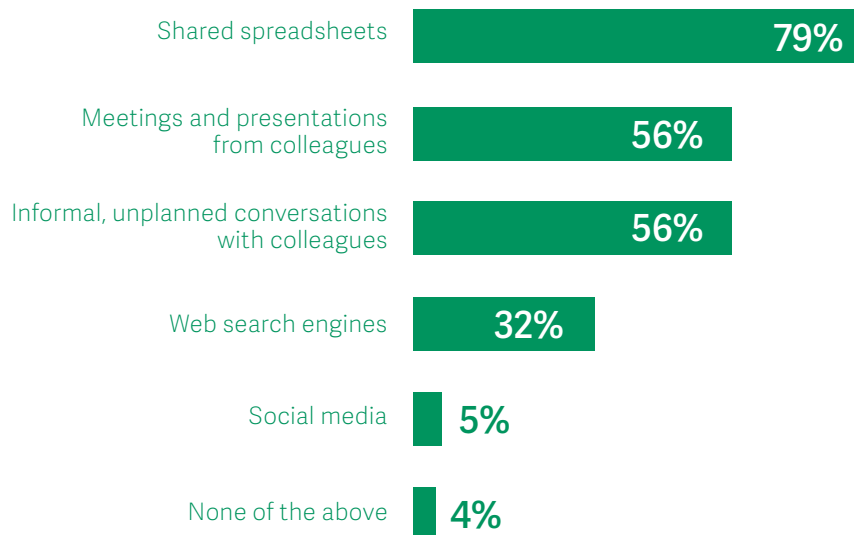
## Benefits of accessible data

Respondents were asked which benefits they would expect if supply chain and operations management data became more accessible across their organizations.



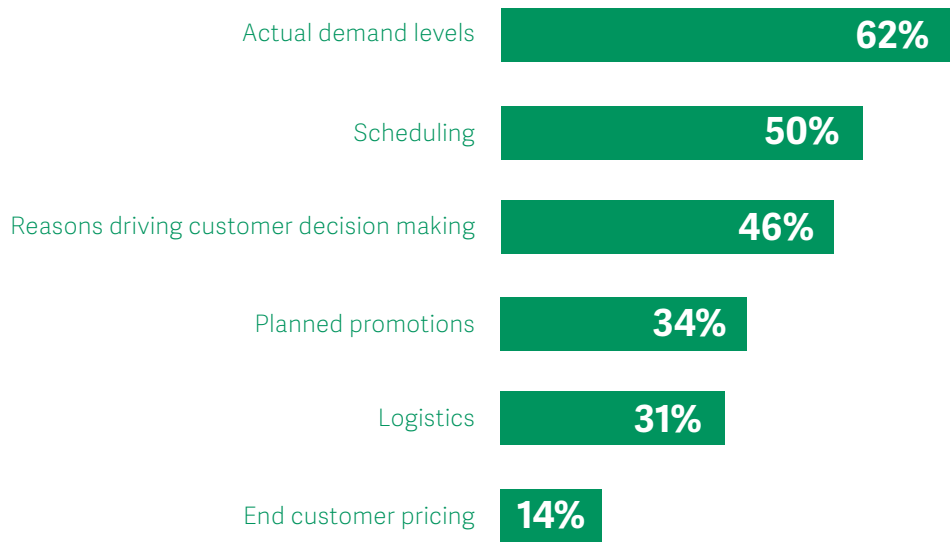
## Other methods of data management

Respondents were asked which technologies or methods of data management they rely on beyond ERP and MRP.



Information from third-party partners

Respondents were asked in which areas they wish they had more information from third-party partners.



## Reporting delays

Respondents were asked what reporting delays they experience with their information systems including ERP and production systems.

### Reporting day length

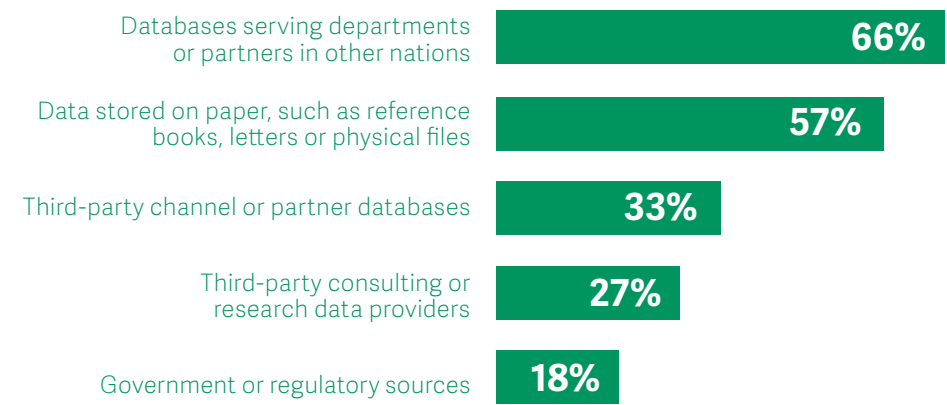
	No delay (live)	Short delay (few minutes to a few hours)	Between a few hours to a few days	Between a few days to a few weeks	Regular monthly, quarterly or annual reporting
Materials or components shipment status	21%	23%	37%	18%	1%
Finished goods shipment status	32%	36%	24%	7%	0.3%
Work in process inventory	26%	32%	24%	11%	6%
Finished goods inventory	46%	31%	16%	5%	3%
Customer changes to orders	12%	31%	34%	20%	3%
Actual sales-to-demand forecast	8%	12%	22%	26%	33%
Performance -to-production plan	12%	16%	35%	17%	19%
Performance-to-current budget	4%	8%	18%	22%	48%

### Your satisfaction

	Faster than I need	Acceptable delay	More delay than I prefer	Unacceptable delay
Materials or components shipment status	8%	43%	38%	11%
Finished goods shipment status	10%	64%	20%	6%
Work in process inventory	8%	55%	28%	10%
Finished goods inventory	12%	65%	18%	5%
Customer changes to orders	4%	38%	42%	16%
Actual sales-to-demand forecast	3%	34%	42%	22%
Performance -to-production plan	5%	45%	38%	13%
Performance-to-current budget	2%	51%	32%	14%

Data sources involved in job responsibilities

Respondents were asked to select all data sources used for their job roles or responsibilities.



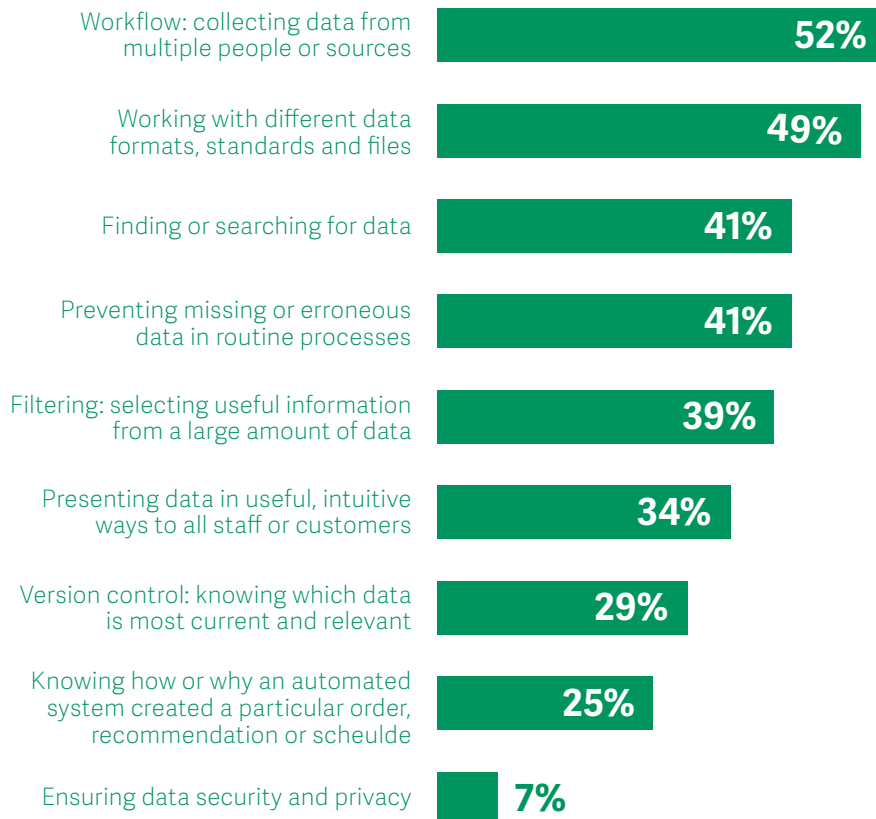
Areas benefitting from big data

Respondents were asked the top three areas that would benefit the most from big data (big data is defined as an innovative technology that accesses, integrates and reports all available data).



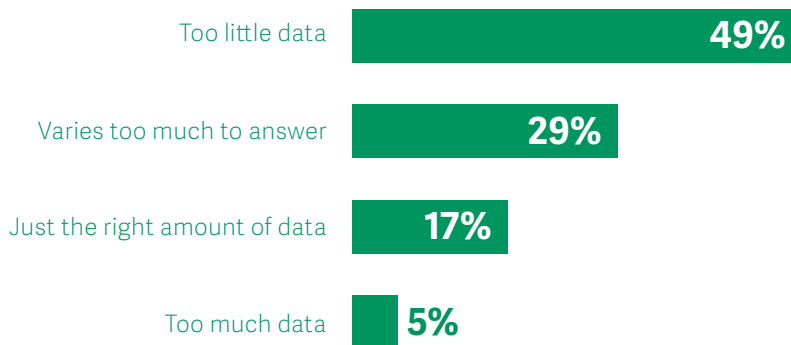
## Difficult data tasks

Respondents were asked which data tasks were particularly difficult or dissatisfying.



## Forecasting

Respondents were asked how much data they have to work with when it comes to forecasting.



RESPONDENT PROFILE

Respondents were asked which technologies or methods of data management they rely on beyond enterprise resources planning (ERP) and material requirements planning (MRP). Seventy-nine percent of respondents indicated shared spreadsheets. More than 50 percent of respondents also said they relied on meetings and presentations, as well as informal conversations with colleagues.

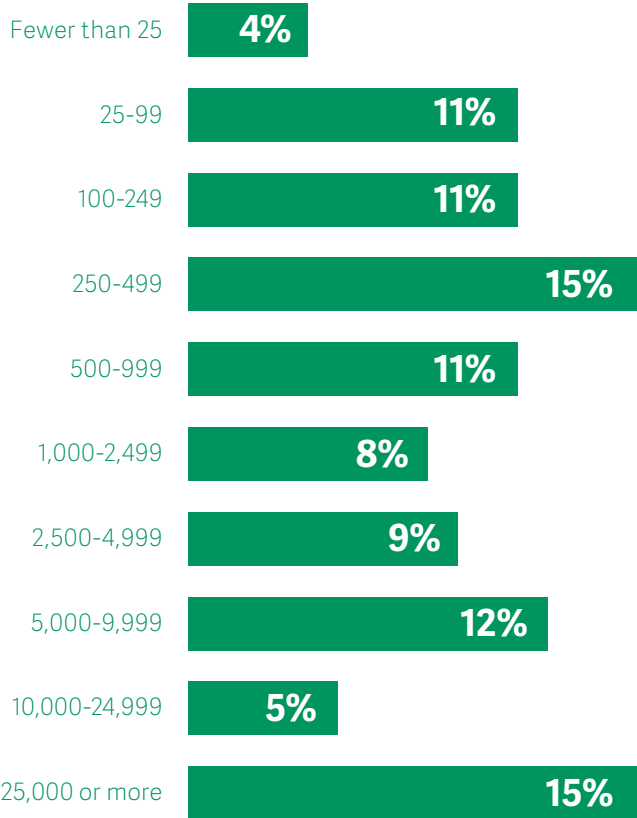
Years of experience

Respondents were asked how many years of experience they have implementing or practicing supply chain or operations management.

No experience	0-1 years	2-3 years	4-5 years	6-8 years	9-11 years	12-15 years	More than 16 years
1%	4%	9%	10%	12%	14%	15%	36%

Number of employees in company

Respondents were asked how many employees are at their organization



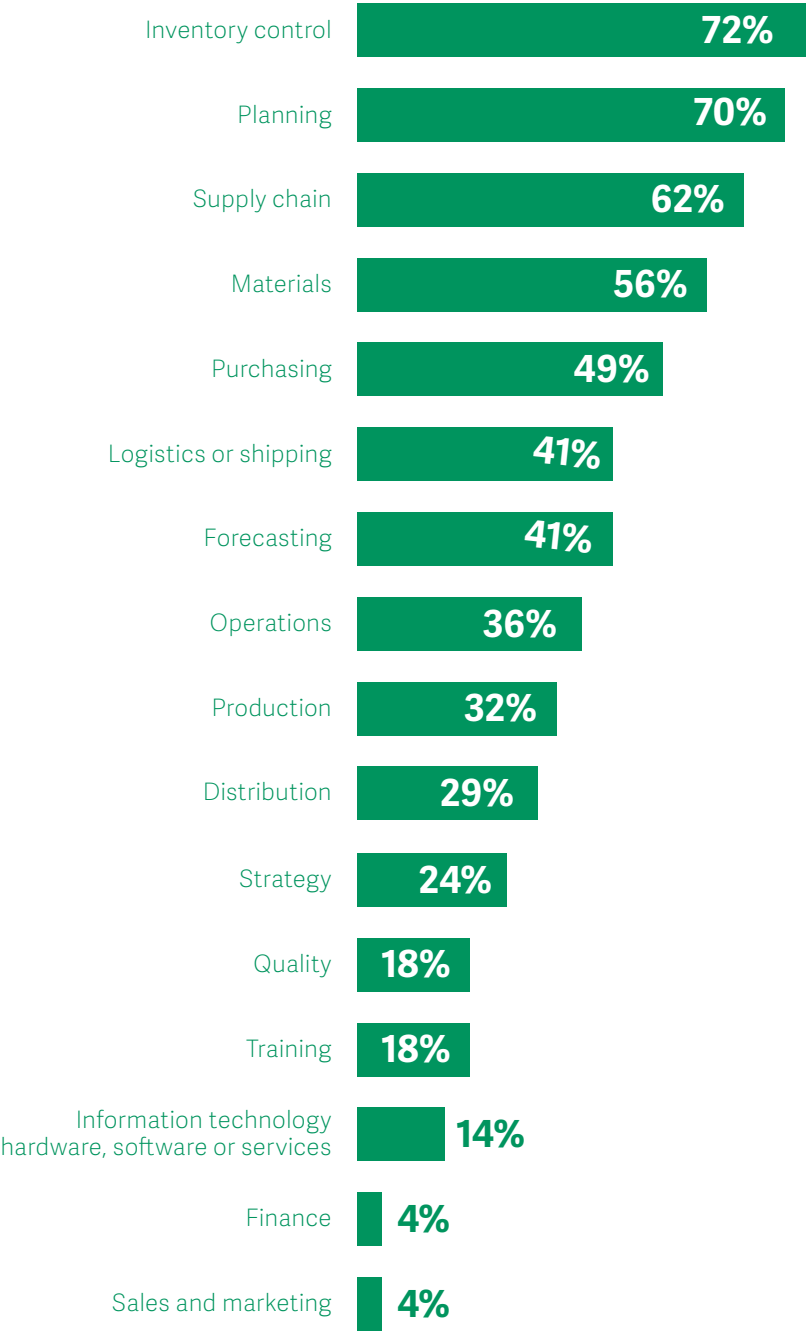
## Areas of responsibility

Respondents were asked which areas fall under their responsibility.



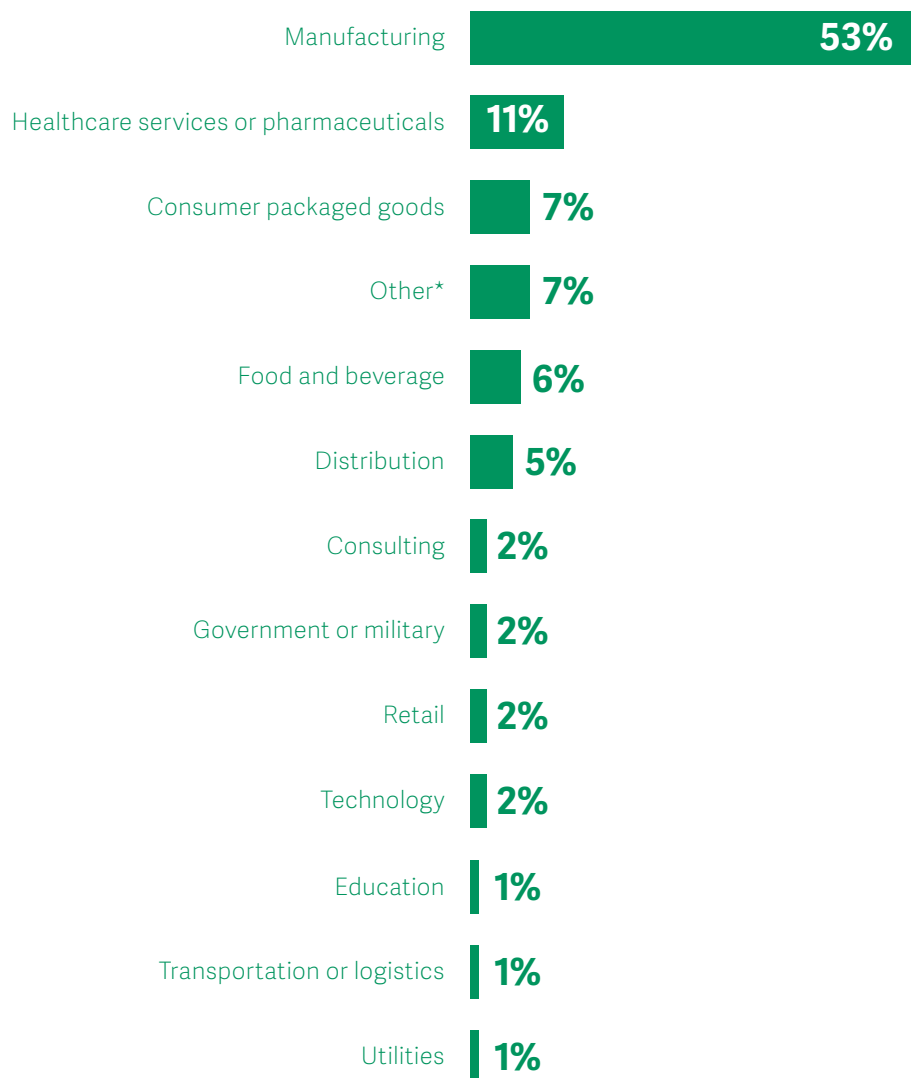
Key responsibilities

Respondents were asked which of the following are key responsibilities for their team or group.



## Industry type

Respondents were asked which term best describes their industry.



\*Respondents who selected "Other" indicated aerospace, oilfield service, mining service, maintenance and repair, engineering, animal nutrition, telecommunication, grocery, airline carrier, corporate wellness and aviation.

# GLOSSARY



## GLOSSARY

### **Big data storage**

Big data storage is the recording of all types of data using storage that permits live or near-live access. Big data storage has no single location. It uses all networked storage available. Data warehouses and enterprise data warehouses, cloud computing or local area network servers, are some locations that might house big data storage. The amount of data storage that qualifies as big data storage has no fixed definition. It is large enough, however, to store vast arrays of data, typically in petabyte or exabyte ranges.

### **Big data vendors**

Large and small information technology vendors serve the big data market. Smaller companies tend to specialize in specific areas, such as cloud storage, programming or statistical analytics. Larger vendors usually offer integrated product platforms and services. A recent trend is the sale of big data appliances that combine big data hardware and software platforms, easing implementation and scalability. Many big data implementations rely on low-cost commodity computer hardware, freeing up customer budget for greater focus on software, services and customization.

### **Business analytics (BA)**

The capability of business systems and processes to use algorithms and statistics to derive meaning and insight from data, such as for decision making, planning and analysis.

### **Business intelligence (BI)**

The capability to gather, sort, classify and maintain data and knowledge for the purpose of improving competitive positioning and business management.

### **Cloud storage**

Cloud storage is remotely hosted online data storage accessible through the internet. A cloud storage data provider usually operates one or more large data centers and provides, sells or leases a specific amount of data storage capacity to its customers. Cloud storage may be public, offered to anyone, or private (available only to specific people or organizations).

## Components summary

Big data analytics accesses big data storage and the organization created by Hadoop. In classic big data architecture, big data analytics process data found in big data storage in efficient amounts, using massively parallel processing in order to find useful insight, relationships, dependencies for decision support systems, competitive analysis and improvement, and increasing productivity across a strategic scale. The analytics can be designed for human use, such as reports, charts and tables. Or analytics can serve other programs and data processes.

## Hadoop

Hadoop is emerging as a primary system for organizing big data. It is an open source software framework designed to allow many independent computers working in parallel to access and organize vast amounts of data. It was originally derived from the Google File System. Many organizations contribute to it, including Yahoo. The history of Hadoop began with Google's need to gather vast amounts of textual and structural information from websites all over the world. Google then needed a way to present that data to individuals in useful ways. Given the limited capabilities of products at the time, Google developed its own innovative platform. This effort spun off into open source projects that became Hadoop. Cloudera CEO Mike Olson says, "The Hadoop platform was designed to solve problems where you have a lot of data—perhaps a mixture of complex and structured data-and it doesn't fit nicely into tables. It's for situations where you want to run analytics that are deep and computationally extensive, like clustering and targeting. That's exactly what Google was doing when it was indexing the web and examining user behavior to improve performance algorithms." According to an Oracle white paper, Hadoop performs "brute force scans" through massive sets of data "to produce results that are output into other files."

## MapReduce

MapReduce is a programming model created by Google in 2004, and implemented in Hadoop, designed to provide distributed computer processing of very large amounts of data. Such data processing can occur on unstructured file system data or data found in structured databases. MapReduce reflects a two-step process. A "map" step begins data processing by dividing processing tasks into sub-tasks and distributing the subtasks to other resources for completion. A "reduce" step combines all completed subtasks and reports the results as output.

## **NoSQL databases**

NoSQL databases are designed for large amounts of dynamic data where a relational model is not required. Common uses include enabling real-time access for data elements such as Twitter posts, Internet server logs and security keys.

## **Relational databases**

Relational databases expand their reach into less structured datasets to analyze big data.

## **R programming language**

The R programming language is an open source programming language that serves statistical computing and graphics needs. The R programming language provides a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification and clustering.

## **Structured data**

Structured data is information that exists in fixed fields of a computer record, file or database. Structured data also includes data that can be easily looked up, processed, analyzed and reported with little uncertainty. Examples of structured data include records of product prices, customer names and postal codes.

## **Unstructured data**

Unstructured data are data that do not exist in fixed fields within a record or file, or are difficult to label. Examples of unstructured data include audio and video files, photographs and text-based data (documents, journals, emails and reports).

# APICS SUPPLY CHAIN COUNCIL RESEARCH

For more insider information explore  
APICS research topics at [apicsscc.org](http://apicsscc.org).

**Reports are available on:**

- Chronic disruption
- Project management
- Remanufacturing
- Risk management
- Supply chain strategy
- Sustainability
- And other topics

If you have comments or questions,  
contact [askapics@apics.org](mailto:askapics@apics.org).

## ABOUT APICS SUPPLY CHAIN COUNCIL

APICS SCC is a nonprofit organization that advances supply chains through unbiased research, benchmarking and publications. APICS SCC maintains the Supply Chain Operations Reference (SCOR) model, the supply chain management community's most widely accepted framework for evaluating and comparing supply chain activities and performance. APICS SCC enables corporations, academic institutions and public sector organizations to address the ever-changing challenges of managing a global supply chain to elevate supply chain performance. APICS SCC is part of APICS, the premier professional association for supply chain and operations management.

**Visit [apicsscc.org](http://apicsscc.org) to learn more.**