Big data has been big news in recent years. Organizations recognize that they must now begin to focus on using big data technologies to solve business problems. The pressure is on for organizations to move past the discussion phase toward well-planned projects. A complete view of all the available information can help organizations gain deeper insights and make better decisions.

**Five high-value uses for big data**

IBM has conducted surveys, studied analysts’ findings, spoken with thousands of customers and prospects and implemented hundreds of big data solutions. As a result, it has identified five high-value use cases that enable organizations to gain new value from big data, as described below:

1. **Big data exploration**: find, visualize and understand big data to improve decision making.
2. **Enhanced 360-degree view of the customer**: enhance existing customer views by incorporating internal and external information sources.
3. **Security/intelligence extension**: reduce risk, detect fraud and monitor cybersecurity in real time.
4. **Operations analysis**: analyze a variety of machine data for better business results and operational efficiency.
5. **Data warehouse modernization**: integrate big data and traditional data warehouse capabilities to gain new business insights while optimizing the existing warehouse infrastructure.

These are not intended to be sequential or prioritized. It doesn’t matter where users start; it just matters that they start. The key is to identify which use cases make the most sense for the organization given the challenges it faces.
What is big data exploration?
Big data exploration addresses the needs of organizations at multiple points in their big data projects. Initially, it addresses the challenge that many organizations face – they may not have complete knowledge of all of the data available to them. All kinds of data may fall into this category, but unstructured data in particular (for example, documents, messages and other human-generated content) can be opaque, with little to indicate its potential value or hidden risks.

To address this challenge, organizations can index all their data using a product such as IBM Watson Explorer, enabling them to explore their data based on the topics in the content and metadata. This allows the organization to make better-informed decisions about what data to use in their big data projects, as well as to derive continuing value from improved access to the data by all parts of the organization.

Two key requirements are critical to this use case. The first is for the organization to gain visibility into all of its data, regardless of format or where it is managed – including data that is stored in Apache™ Hadoop® and other special-purpose repositories where large volumes of varied data are stored and processed. This use case can even extend into data stored outside the enterprise or in the cloud, as well as that managed by third parties. A second key requirement is to be able to explore data without having to move it from where it is stored and managed. This greatly reduces the time and effort required to execute big data exploration and enhances the organization’s ability to leverage the data for better decisions and day-to-day operations.

Viewing and exploring big data
Once an organization has created indices representing all of its data, the next step is to create ways to explore and navigate data. The indices provide a virtual integration that enables data from multiple sources to be used simultaneously. In addition to displaying and navigating information from multiple sources, Watson Explorer can dynamically generate virtual documents that contain data from multiple applications with a common element or theme.

In a typical interaction, a user can perform a single search that yields data from multiple data sources. The set of items – which could include documents, messages, database records, data stored in Hadoop or data stored externally – can then be explored in a variety of ways. For example, Watson Explorer can dynamically categorize items based on their content and group similar items together. This enables the user to detect common themes across multiple sources – and when represented as a tag cloud, clustering results can show the dominance of various themes.

Most unstructured documents also include metadata, such as dates, identifiers and tags, that can be used for classification and segmentation. Watson Explorer can categorize and assign tags to items based on taxonomies, ontologies and other semantic structures. All of this metadata can be used by Watson Explorer to enable faceted navigation – filtering data items and displaying the results visually. Like clustering, faceted navigation can enable users to detect themes and relationships that would be difficult, if not impossible, to detect using any other analysis method.
Security is always a concern when organizations provide access
to data across a large number of sources, whether for the general
user population or for specialized roles such as data scientists
and business analysts. To enable organizations to provide this
level of access with confidence, Watson Explorer maps the
access permissions of each repository or management system
and enforces them when providing access to users of any kind.
As a result, users are only able to view data through Watson
Explorer that they would be able to view if logged directly into
the target system.

With secure access to a broad range of data, authorized users can
explore data using a number of methods, including searching
and browsing through a number of user interaction models.
Figure 1 provides an example of a general purpose search
and exploration application created using Watson Explorer.

Beyond the benefits of exploring the organization’s information
to support big data and analytics programs, implementing
a big data exploration solution enables an organization to deploy
360-degree information applications. 360-degree information
applications are an example of big data exploration that is often,
but not exclusively, focused on a specific business activity or role.

360-degree information applications leverage the same underlying
Watson Explorer index as the more general-purpose information
navigation example described above, but focus on presenting
all of the available information about key entity types, such as
a customer, product, account, physical asset or other identifiable
category. In addition to data about the entity, Watson Explorer
can deliver analytic artifacts – such as statistical graphs and
data visualizations – that relate to a specific instance of an
entity or a theme relevant to the current view.

Figure 1: Exploring big data using IBM Watson Explorer
Using a 360-degree information application, users browse data by navigating through related entity pages. Users typically begin from a personalized home page that contains data and links relevant to their role or task. A typical navigation path would progress from the user’s home page to an entity page, and then to a related entity. Each page dynamically displays all of the available information about the entity from multiple sources. In this navigation model, users would seldom search beyond retrieving an initial item, but would mainly navigate via a series of links.

360-degree information applications can improve the accessibility of data across all job functions while still observing security and privacy, and can boost productivity and innovation in many parts of an organization and enable better understanding of big data assets.

**IBM Watson Foundations**

IBM Watson Foundations provides all of the capabilities needed to gain maximum value from big data. The following products and services are part of Watson Foundations and are particularly important for supporting the big data exploration use case.

**IBM InfoSphere® BigInsights™** software builds on the open source Apache Hadoop framework to include analytics, and makes it easier for the enterprise customer to deploy and manage. IBM packages open-source Hadoop in its InfoSphere BigInsights software but also includes IBM technology that is not available with open-source Hadoop distributions. InfoSphere BigInsights can be used as a queryable data store and for a myriad of data- and compute-intensive tasks, including advanced text analytics, in support of big data exploration. Visit [http://ibm.biz/BdRyhU](http://ibm.biz/BdRyhU) for more information.
IBM’s information integration and governance (IIG) offerings support the quality, accuracy and security of data throughout its lifecycle, including all forms of big data. The IIG portfolio includes the following:

- **IBM InfoSphere Master Data Management** ensures an accurate and consistent view of key entities such as customers and products. InfoSphere MDM Probabilistic Matching for InfoSphere BigInsights, or Big Match, enables resolution of billions of records at big data scale and speed to provide more context, faster. Visit [http://ibm.biz/BdRSHx](http://ibm.biz/BdRSHx) for more information.

- **IBM InfoSphere Business Information Exchange** ensures alignment of terminology and metadata across the organization to enhance search, discovery and navigation. Visit [http://ibm.biz/BdRyVB](http://ibm.biz/BdRyVB) for more information.

**IBM consulting and implementation services** help organizations get started with big data initiatives. Our skilled professionals bring years of experience to each project and can help organizations get started quickly with the right foundation and strategy. Visit [http://ibm.biz/BdRyVR](http://ibm.biz/BdRyVR) for more information.

**IBM Watson Explorer** provides federated search, discovery and navigation across a wide range of applications and data sources in support of the enhanced 360-degree view of the customer. In addition, Watson Explorer provides an Application Builder framework for constructing and deploying 360-degree information applications. Visit [http://ibm.biz/BdRyVi](http://ibm.biz/BdRyVi) for more information.

---

**Do you need big data exploration?**

If you find yourself answering “no” to any of the following questions, you should consider implementing a solution based on the big data exploration use case:

- Do you know how to get started with big data initiatives, including enabling access to all of your data assets?
- Do you already have the capability to explore and survey all the content that could be important in your big data initiative?
- Are you able to leverage information in your enterprise systems (such as enterprise content management, customer relationship management, data warehouses and databases) and deliver it to users?
- Do your employees have a way to gain access to external information sources such as blogs, websites and social media?
- Can they access that content in a single user interface alongside enterprise information?
- Do you have a way to identify areas of data leakage or risk?

Get in touch today to learn more about how IBM can help.