



The value of OpenText™ Magellan™

Magellan is a flexible, AI-powered analytics platform that combines open source machine learning with advanced analytics, enterprise-grade business intelligence (BI), and capabilities to acquire, merge, manage, and analyze big data and big content stored in your Enterprise Information Management (EIM) systems. Magellan enables machine-assisted decision-making, automation, and business optimization.

Magellan cost-effectively supports innumerable use cases on a cohesive, highly scalable infrastructure perfectly equipped for handling massive amounts of structured and unstructured data.

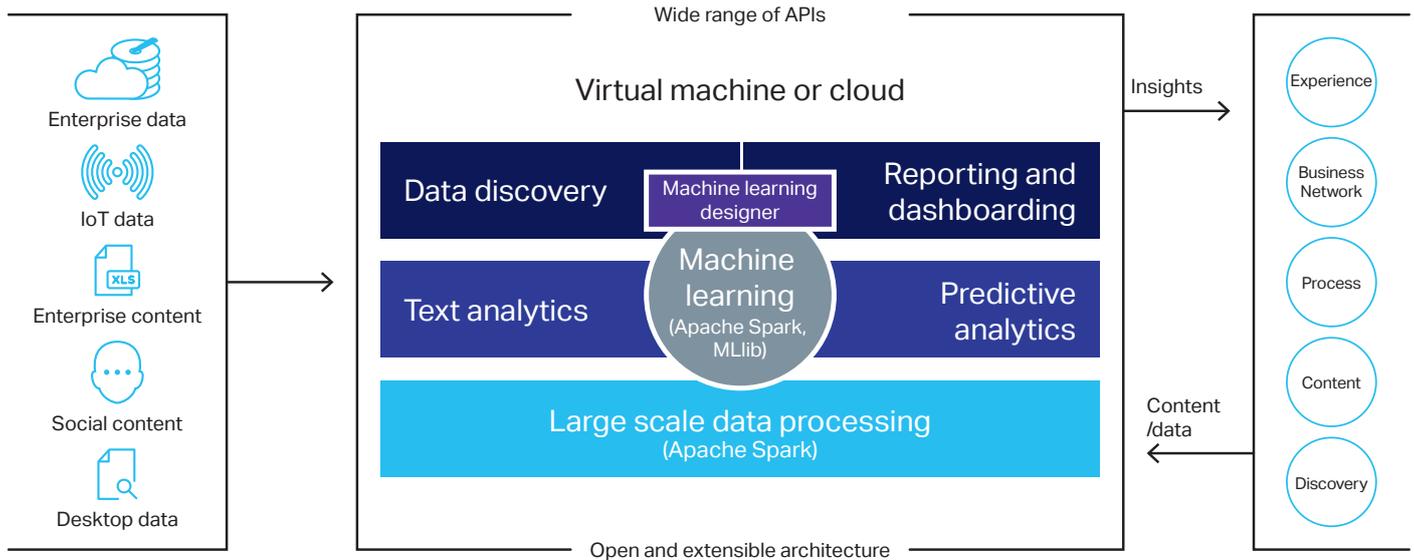
What exactly is Magellan?

Magellan is a platform comprised of pre-integrated technologies that provide the necessary features for a powerful Artificial Intelligence (AI) and analytics system, including business intelligence (BI), data discovery, predictive analysis, and text mining. The platform is anchored by Apache Spark™, an open source, extensible big data storage and processing technology from Apache. This allows the system to efficiently crunch through the immense amount of data that customers wish to analyze.

AI comes into the picture via Spark's machine learning library, where data scientists can create custom algorithms using the Magellan Notebook (based on Jupyter) and save them as drag-and-drop analytic assets for business users. The platform can source data from any stream or repository, especially data managed by OpenText EIM solutions, such as Experience Suite, Business Network, Content Suite, and Process Suite.

CONTACT US

magellan@opentext.com



What does Magellan deliver?

Magellan, a prebuilt and integrated platform, combines existing OpenText business intelligence and analytics features with custom machine learning modules in a framework that leverages open standards for easier integration of components and with external applications and data sources.

The Magellan BI package delivers:

- Data discovery
- Reporting and dashboarding
- Text analytics (including sentiment analysis)
- Sophisticated predictive analytics
- Machine learning (via the MLlib library)
- Machine learning designer (Magellan Notebook)
- Large-scale data processing (Apache Spark)
- A wide range of APIs for easy customization
- Open and extensible architecture

Data discovery

The data discovery component of Magellan makes it easy for anyone to quickly access, blend, store, and analyze data, identifying business trends, gaps, and relationships they may never have known existed. It provides a drag-and-drop experience for exploring data including set analysis with Venn diagrams, Pareto and distribution charts, profiles and correlations, and more.

The discovery feature provides a range of data engineering and enrichment methods enabling users to aggregate and decode data, build expressions to create calculated fields, create numeric and quantile ranges, build parametric columns consisting of query-based values, and rank records.

Business intelligence, reporting, and dashboarding

The business intelligence reporting feature of Magellan lets users create their own dashboards, reports, and visualizations from insights developed in other components, then easily share or embed them.

Text analytics

The text analytics component of Magellan provides sophisticated natural language processing (NLP) capabilities, including sentiment analysis, against unstructured content from a vast number of sources, including the web, social media, documents, and more.

The feature provides transformation, enrichment, and persistence of processed unstructured data into the Magellan data store. Metadata processed by text analysis is then leveraged for predictive analytics, dashboards, and visualizations, providing an easy way to understand and share insights.

Predictive analytics

Magellan’s predictive analytics component leverages custom data models created by data scientists in Apache Spark, presenting them as drag-and-drop models. This allows operational users to apply advanced algorithms to their data to learn likely behaviors, forecast outcomes, anticipate risk, and make recommendations. The predictive component includes built-in analytic techniques, such as profiling, mapping, clustering, forecasting, creating decision trees, and more—all without requiring statistical expertise or coding.

Large-scale data processing and machine learning

The core technology of Magellan is Apache Spark, the underlying general execution engine that provides in-memory computing for maximizing hardware capabilities. Data scientists leverage Spark to create and process custom machine learning algorithms using programming languages such as Scala, Python, SQL, and R, familiar environments that help make it easier for developers and data scientist to get to work.

Spark provides the MLlib machine learning library that is ready to go out of the box, so developers can immediately leverage predefined processing pipelines and statistical routines on their data and content. Spark enables the flexibility and extensibility of an open stack while ensuring that enterprises maintain full ownership of their data and custom models.

Machine learning designer

The Magellan platform includes the Magellan Notebook (based on the Jupyter Notebook), a web application used to create and share code, equations, visualizations, and explanatory text encapsulated in a single document. It is used by data scientists to design custom machine learning models and data processing routines leveraging Spark MLlib. These Notes can be written in several popular languages, such as Python, Scala, or R, which ensures that data scientists can get to work quickly in a familiar environment.

Within the Magellan ecosystem, the machine learning designer becomes a central opportunity to create new machine learning routines and share them with operational users, who can leverage them to make smarter decisions without needing to understand the science or logic underneath.

APIs for easy customization

Magellan can be quickly extended and customized by drawing on a wide range of built-in APIs, including REST and Javascript, which allow it to be integrated with any other standards-based technology. These APIs enable very detailed control over analytic features and functionality at any level to meet any functional requirement, from an interactive chart on a webpage to an embedded predictive process.

Open, standards-based foundation

Magellan leverages open standards in components, including the Apache Spark computing platform and Jupyter Notebook, to create a flexible, easily extensible solution. Open technology standards mean the technology is continually updated and kept compatible with a wide range of complementary technologies by a large community of trained experts, on whom Magellan customers can draw for consulting expertise.

The technology can be customized and extended to meet any customer requirement. Plus, any customizations remain the intellectual property of the customer.

What are Magellan's main benefits?

At a high level, Magellan provides four main benefits:

- Injects AI and analytics-based decision-making capabilities into our customers' applications—Magellan can be plugged in to any application to provide AI-based analysis that enhances and automates decision-making
- Specializes in acquiring, federating, and managing complex and highly diverse data
- Provides massive capacity and horsepower to process huge amounts of information
- Puts advanced analytics into the hands of every user, thanks to its roots in OpenText Analytics, allowing the organization to deliver advanced analysis capabilities that everyone can use, regardless of their technical skill

What is the "AI" in Magellan?

AI, or Artificial Intelligence, refers to computer-based decision-making. As defined by Alan Turing, "Artificial Intelligence (AI) is defined as the science of making computers do things that require intelligence when done by humans." While the field of AI is large and growing, it is important to make the distinction that Magellan

harnesses AI for the analysis of data. It does so through machine learning, a branch of AI ideally suited for analysis. As defined by MathWorks.com, "Machine learning teaches computers to do what comes naturally to humans and animals: learn from experience."

Magellan leverages AI in two ways:

1. Machine learning—Magellan includes a machine learning designer, which allows a data scientist to create sophisticated, predictive algorithms leveraging MLlib, the built-in machine learning library in Apache Spark. These routines are called "machine learning" because they use algorithms that learn from repetition.

For example, an algorithm is created to detect fraudulent activity in banking transaction data. The algorithm is fed a sample of data to "train" against, which it does multiple times, all the while "learning" the signs of a fraudulent transaction based on the sample data. When the algorithm is sufficiently trained to recognize fraud, it is saved as a model and can be used with any banking transaction data to detect signs of fraud.

Once custom algorithms/models are created, they can be saved as drag-and-drop objects for other Magellan users to leverage with their own data in dashboards and reports, expanding the scope of AI across the organization.

2. Text analytics—Magellan provides sophisticated NLP (natural language processing) capabilities against unstructured content from a vast number of sources, including web, social media, documents, and more. As it processes the information ("reads the text"), it performs concept identification, categorization, entity extraction, and sentiment analysis. The machine learning is in concept identification, where the NLP engine learns to match terms and topics with complex concepts. The metadata from unstructured sources can then be merged with structured data for more contextual analysis, and fed through Magellan's predictive analytics, finally producing dashboards and visualizations for users to easily understand and share the insights with other users.

What is unstructured data and how does Magellan handle it?

Structured data refers to data that is organized into a common structure, such as numbers in rows and columns in a Microsoft® Excel® spreadsheet, which can be easily measured and quantified. Unstructured data is any data that lacks a common, repeatable structure, is not primarily made up of numbers, and is generally presented in "natural language." Unstructured data sources include documents such as contracts, letters, or manuals, as well as web-based content such as tweets, blogs, or online news articles. These sources are complex to analyze because of their disparate nature and varying formats (How does one link a Tweet to weather data, for example?).

Magellan simplifies the complexity by providing sophisticated NLP capabilities against unstructured content from a vast number of sources. The feature provides transformation, enrichment, and persistence of processed unstructured data into the Magellan data store where it can be easily merged with other sources of data, including structured data. Metadata processed by the text analysis feature can be used for predictive analytics, dashboards, and visualizations, providing an easy way for stakeholders to understand and share insights.

What parts of Magellan are “open?”

Apache Spark and Eclipse™ BIRT are the underlying open source technologies that power Magellan: Spark for data processing and BIRT for data visualization. Open source software leaves its source code exposed so that it can be customized and extended by customers for their own use. This is a huge benefit, as customers don't have to wait for a commercial product release to incorporate new or custom features.

Both Apache Spark and Eclipse BIRT are hugely popular and have a community of open source users and contributors, which lets our customers draw upon a large pool of trained experts who can work with Magellan. The technology can be customized and extended to meet any customer requirement. And any customizations remain the property and IP of the customer, unlike competitive offerings such as IBM® Watson™.

How Is Magellan implemented?

Unlike competitive offerings that force a customer to use a vendor-managed environment, Magellan can be deployed anywhere the customer wishes: on-premises, in their own third-party cloud, or on OpenText Cloud as a managed service.

The flexible deployment options allow customers to easily transition from one deployment model to another. For example, they could start with Magellan on-premises. Later, if they require more capacity or horsepower, they could move the application to the cloud—no rebuilding or redesign required. OpenText Professional Services will provide implementation packages to help our customers deploy Magellan.

What are the main differentiators for Magellan?

Unlike alternative options, Magellan provides a pre-built platform, so customers do not have to install or integrate the various technologies required for an AI and analytics system. This allows them to focus immediately on what's important: analyzing the data.

And we don't dictate where Magellan needs to live; customers can implement on-premises or in the cloud as a virtual appliance.

The stack is based on open source Apache Spark and Eclipse BIRT, so it can be customized and extended, and customers can leverage a vast pool of technical expertise in the open source community.

And most importantly, custom algorithms created by customers, which represent a proprietary competitive advantage for them, remain entirely their own IP.

What does all of this mean?

Magellan takes all your customers' content and data, both structured and unstructured, merges and processes it by leveraging the power of Apache Spark, and gets to work mining, learning from, and understanding the data. From here, Magellan can begin to uncover trends, anomalies and correlations, as well as make predictions about future outcomes or likely behavior.

Finally, Magellan provides visual depictions of the insights it uncovers, which users consume as dashboards and reports that they can edit and share. And through machine learning, Magellan continues to improve its understanding of the data, becoming more accurate and efficient as it learns.

Magellan can automate parts of our customers' business that were previously manual, uncover operational inefficiencies, optimize cost control, contain risk, detect fraud, and deliver impactful customer experiences.

www.opentext.com/contact