

IBM Analytics Engine

Deploy analytics applications in minutes to unlock the value of big data



Highlights

- IBM® Analytics Engine offers a simple user experience when creating and managing Hadoop and Spark clusters, reducing the time required to configure or administer clusters so that data scientists can focus on exploring data
 - Data scientists can use it to spin up and deprovision clusters as and when required, and scale compute and storage separately to meet specific workload demands
 - IBM Cloud Object Storage acts as permanent data store, helping users cut their storage costs and increase resiliency.
-

In the search for competitive advantage in tomorrow's marketplace, it will be the businesses who turn their data into a strategic asset and uncover deep, actionable insights that surge ahead. That makes the role of data science increasingly important for your organization, and placing the right tools in the hands of your analysts could be the key to pinpoint the market opportunities that your competitors fail to spot.

Many businesses rely on Hadoop or Spark-based clusters to run big data analytics jobs and build predictive models. However, this approach can lead to a set of challenges that prevent you from realizing the true value of your data.

Firstly, managing the cluster is a specialist task that is often beyond the knowledge of data scientists. Maintaining each of the components within the Hadoop or Spark environment is a distraction for your data science team and takes up time that would be better spent focusing on core analytics tasks.

On top of this, finding tools to run management tasks on multiple clusters across a business has historically been difficult. Data science teams tend to use whichever solutions function best for specific jobs, leading to fragmentation rather than standardization, and limiting the potential for business-wide collaboration.



In many cases, data lakes also use Hadoop clusters as permanent data stores. In one respect, this makes sense: it is easier to run analytics jobs where data is stored rather than moving it to a separate location. But it also increases the risk of data loss as the cluster is responsible for compute and storage, makes maintenance even more complex, and is an inefficient way of utilizing processor power. What's more, since the number of nodes must be pre-defined, adapting the data lake as the workload grows can be costly.

IBM Analytics Engine is a cloud service that enables data scientists to provision Apache Hadoop and Apache Spark clusters in minutes. The solution, part of the IBM Watson® Data Platform, enables data scientists to concentrate on exploring data and makes it easy to unlock business value.

Scalability to suit your workflow demands

Where most Hadoop-based setups combine compute and long-term storage infrastructure, IBM Analytics Engine separates them. It utilizes IBM Cloud Object Storage, with data residing permanently in the object store, and the clusters connecting to the repository whenever they need access. The data is dispersed across multiple data centers, and can optionally be dispersed across multiple regions for security and protection. This unique approach combines encryption, erasure coding and information dispersal of data for protection without complex or expensive copies. The service is continuously available; it can tolerate even catastrophic regional outages without downtime or the need for customer intervention.

Separating compute and storage resources also allows users to scale independently; you can add capacity as data volumes expand, or extra compute nodes as your data science workload grows. The solution matches your individual project needs. Moving data off Hadoop clusters also cuts your storage costs, as additional compute power or memory is no longer required.

Provision clusters within minutes

Utilizing an object store also removes the need to have a fixed-size cluster permanently online. Instead, data scientists can spin up clusters whenever they require them, then deprovision once the job is completed. This is a simple process: users just log in to IBM Analytics Engine, select the number of nodes they need and link the cluster to the object store.

The cluster is ready within minutes and data scientists can begin running analytics jobs. Thanks to sophisticated virtualization technology, when the task is complete, users can save the results to the object store and then spin down the cluster.

Maintain clusters easily

As clusters exist only when in use for analytics tasks, maintenance and upgrades are much easier to complete. To move to updated versions of Spark or Hadoop, users just spin up a new cluster with the upgraded software. Configuration can be easily replicated using configuration scripts across clusters. With data held in the object store, there is no impact on data integrity due to the upgrade.

As IBM Analytics Engine utilizes Hortonworks Data Platform, an open-source Hadoop distribution, users can also leverage the expertise of the open source community to make the most of their investments.

Unleash the full potential of your data scientists

With a single point of control to create and manage Hadoop and Spark environments, IBM Analytics Engine removes the need for data scientists to dedicate time to configuring servers, troubleshooting and installing updates. Instead, they can focus on the more important task of mining data for valuable insights.

Data scientists can log in to IBM Analytics Engine in several ways, depending on user preference. They can interact through a web interface, a REST API, or via a command-line interface to customize Hadoop clusters, and install third-party libraries and configurations across the whole cluster.

Integrate fully into your data science workflow

As a standalone product, IBM Analytics Engine makes it simple to deploy big data analytics applications. The solution also integrates seamlessly with the other services in the IBM Watson Data Platform to form an end-to-end data science workflow, ensuring teams across your organization work from a standard set of tools.

IBM Analytics Engines integrates with:

- IBM Data Catalog, for metadata-rich data cataloging
- IBM Data Science Experience, for data analysis and exploration
- IBM Watson Machine Learning, for evaluating predictive models

Together, the solutions form an analytics ecosystem with the capabilities to simplify data governance, significantly reduce cut the cost of disaster recovery, and streamline your data science and machine learning workflows.

For more information

To learn more about IBM Analytics Engine, contact your IBM representative or IBM Business Partner, or visit:

ibm.com/cloud/analytics-engine



© Copyright IBM Corporation 2017

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
December 2017

IBM, the IBM logo, ibm.com and Watson are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.



Please Recycle