IBM Spectrum Conductor

Simplify Apache Spark deployments, speed time to results and maximize resource utilization

Highlights

- Simplify management of Spark, Anaconda, Dask and other analytics frameworks.
- Eliminate resource silos with a shared, multitenant platform.
- Cut costs and increase server utilization with dynamic resource allocation.
- Integrate Spark with other applications such as IBM Watson Studio and H2O.
- Access additional resources with automated and dynamic cloud bursting.

Apache Spark offers compelling performance advantages as an open source, big data analytics framework. However, implementing Spark poses significant challenges including investment in new expertise, tools, and workflows; and integration with other frameworks. Setting up ad-hoc Spark clusters can lead to inefficient use of resources, as well as management and security challenges.

The IBM Spectrum Conductor® platform is designed to address those issues, helping users overcome the challenges of Spark deployment and management at scale. Unlike competitive open source offerings that require piecemeal assembly of components, IBM Spectrum Conductor is an integrated solution that is backed by IBM® services and support. It incorporates a Spark distribution and supports multitenancy for Spark, Anaconda/Python, Dask and other frameworks, augmented by technologies for granular and dynamic resource allocation. These technologies have been widely and successfully implemented in many demanding customer environments to improve infrastructure performance and efficiency.

IBM Spectrum Conductor allows organizations to deploy Spark, Anaconda/Python and Dask applications efficiently and effectively. The enterprise-grade, multitenant management solution can support multiple instances of Spark, maximizing resource utilization, increasing performance and scale, and eliminating silos of resources that would otherwise be tied to separate Spark implementations. IBM Spectrum Conductor supports integration of Spark with other application frameworks such as Anaconda, Dask, IBM Watson® Studio and H2O.

Accelerate time to results

By supporting the simultaneous running of multiple instances of Spark and other frameworks on a single shared infrastructure, IBM Spectrum Conductor enables applications to take full advantage of available resources. A proven, efficient resource scheduler provides fine-grain resource allocation, helping to deliver superior application performance, improved utilization and a faster response to business-critical demands. In environments running multiple application workloads, IBM Spectrum Conductor allocates resources, so service levels are met while preserving security isolation between application instances.

IBM Spectrum Conductor offers up to 58%[1] higher throughput for Spark jobs than competitive open source resource managers. It also provides advanced graphic processing unit (GPU) support to take advantage of its full power with automated scheduling and monitoring for improved utilization and management. In addition, cached or persisted resilient distributed data sets (RDDs) can be shared across applications to avoid reloading or recomputing previous results. All of these elements combine to provide the fastest possible time to results while minimizing expenditure on computing infrastructure.

Increase resource utilization

IBM Spectrum Conductor helps organizations avoid cluster sprawl and inefficient use of resources. By running workloads on a single shared platform, the solution enables individual applications to use resources that would normally be dedicated to other application instances and might otherwise be idle. IBM Spectrum Conductor also supports multitenancy, which allows users to run multiple instances and different versions of Spark simultaneously in a shared environment. This capability helps organizations manage fast-moving Spark lifecycles by allowing various groups to run different versions of Spark, Anaconda, Dask and Jupyter notebooks without the need for them to be upgraded simultaneously.

When additional resources are needed temporarily, IBM Spectrum Conductor can automatically and dynamically take advantage of cloud resources using its Resource Connectors. Cloud bursting dynamically grows and shrinks your cluster while supporting a hybrid mix of on-premises and cloud hosts. When your workload crosses a configured threshold, additional hosts are created on the most cost-effective cloud resources including AWS Spot instances. When the workload running on the cloud resources falls below a configured threshold the hosts are shut down, managing and minimizing the use of cloud resources.

Reduce administration costs

By providing advanced service orchestration and workload management, IBM Spectrum Conductor helps contain infrastructure and management costs. A sophisticated policy-based resource manager offers dynamic resource allocation, allowing organizations to optimize existing hardware usage and defer the need for incremental capital investment. A unified interface lets administrators manage multiple Spark, Anaconda and Dask frameworks, eliminating the need to collect and aggregate metrics from each framework individually.

Easily implement a complete solution

Organizations are looking to move to solutions that optimize storage, analysis and protection of their information assets. IBM Spectrum Conductor is an integrated solution that includes a Spark distribution for data analytics, workload management, monitoring, reporting, IT chargeback and enterprise-grade security. For storage management, IBM Spectrum Conductor can be combined with IBM Spectrum® Scale, which provides significant storage efficiencies compared to the Hadoop Distributed File System (HDFS). IBM Spectrum Conductor also supports NFS, HDFS, object storage and relational databases. The included Spark distribution makes the framework simple to deploy for both exploratory projects and production environments.

Why IBM?

IBM Spectrum Computing offers a comprehensive portfolio of software-defined infrastructure solutions designed to help your organization deliver IT services in the most efficient way possible, optimizing resource utilization to speed time to results and reduce costs. These offerings help maximize the potential of your infrastructure to accelerate your analytics, high-performance computing (HPC), Apache Spark, Anaconda, AI and cloud-native applications at any scale and extract insight from your data faster. Whether deployed in a data center or on the cloud, IBM Spectrum Computing solutions fuel data engineering, critical business decisions and breakthrough insights in financial services, manufacturing, digital media, oil and gas, life sciences, government, research, and education. From designing Formula One race cars to credit risk analysis, organizations in a wide variety of industries are using IBM Spectrum Computing as a foundation for big data, analytics, HPC and cloud to improve business results.

For more information

To learn more about IBM Spectrum Conductor, contact your IBM representative or IBM Business Partner, or visit:

- ibm.com/products/spark-workload-management
- ibm.com/analytics/spectrum-computing

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business, with full lifecycle management of IT products and services from acquisition to disposition. For more information, visit ibm.com/financing.

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
March 2021

XWA5DPBL

1 STAC Report: Spark Resource Manager Comparison of IBM Platform Conductor for Spark, Apache YARN and Apache Mesos–Phase 1, Securities Technology Analysis Center, March 28, 2016.