



---

## Highlights

- Simplify Apache Spark deployment and management with a multitenant framework
  - Improve time to results through efficient resource scheduling and shared infrastructure
  - Cut costs and increase resource utilization with dynamic resource allocation
  - Integrate Spark with other applications such as Hadoop, Cassandra and MongoDB
  - Maintain and simplify management of multiple data sources
  - Eliminate resource silos tied to multiple instances and different versions of Spark
  - Enhance security with role-based access control, Kerberos authentication, and SASL and SSL encryption
  - Automated and dynamic cloud bursting
- 

# IBM Spectrum Conductor with Spark

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

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.

IBM® Spectrum Conductor™ with Spark is designed to address those issues, helping users overcome the challenges of Spark deployment and management. Unlike competitive open-source offerings that require piecemeal assembly of components, IBM Spectrum Conductor with Spark is an integrated solution that is backed by IBM services and support. It incorporates a Spark distribution and supports multitenancy for Spark 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 performance and efficiency.

IBM Spectrum Conductor with Spark allows organizations to deploy Spark applications efficiently and effectively. The enterprise-grade, multi-tenant 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 with Spark supports integration of Spark with other application frameworks such as Hadoop, Cassandra and MongoDB.



### **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 with Spark enables applications to take full advantage of available resources. A proven, efficient resource scheduler provides fine-grained 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 with Spark allocates resources so service levels are met while preserving security isolation between application instances.

IBM Spectrum Conductor with Spark offers up to 58 percent higher throughput for Spark jobs than competitive open-source resource managers.<sup>1</sup> 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 resource definition data sets (RDDs) can be shared across users 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 with Spark 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 with Spark also supports multi-tenancy, 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 without the need for them to be upgraded in lockstep.

When additional resources are temporarily needed, Spectrum Conductor with Spark 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 cloud resources. 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 with Spark 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 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 with Spark is an integrated solution that includes a Spark distribution for data analytics, workload management, monitoring, reporting and enterprise-grade security. For storage management, IBM Spectrum Conductor with Spark can be combined with IBM Spectrum Scale™, which provides significant storage efficiencies compared to the Hadoop Distributed File System (HDFS). IBM Spectrum Conductor with Spark also supports HDFS for users who prefer that option. The included Spark distribution makes the framework simple to deploy both for exploratory projects and in production environments.

### Deploy Spark with confidence

IBM Spectrum Conductor with Spark offers an efficient and highly effective solution for organizations that need to:

- Share resources among Spark instances to speed time to results and improve utilization
  - Manage fast-moving Spark lifecycles
  - Provide fine-grained control of resource allocation for multiple Spark instances
  - Consolidate multiple Spark environments and integrate Spark with other frameworks
  - Take advantage of GPU processing power for compute-intensive tasks
  - Deploy Spark in a hybrid cloud environment
- 

### 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 Hadoop, Spark and cloud-native applications at any scale, extract insight from your data and get higher-quality products to market faster.

Whether deployed in a data center or on the cloud, IBM Spectrum Computing solutions fuel product development, 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 [software-defined infrastructure solutions](#) for big data, analytics, HPC and cloud to improve business results.

### For more information

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

- [ibm.com/us-en/marketplace/spark-workload-management](https://ibm.com/us-en/marketplace/spark-workload-management)
- [ibm.com/storage/software-defined-storage](https://ibm.com/storage/software-defined-storage)

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



---

© Copyright IBM Corporation 2017

IBM Systems  
New Orchard Rd  
Armonk, NY 10504

Produced in the United States of America  
November 2017

IBM, the IBM logo, ibm.com, IBM Spectrum, and IBM Spectrum Scale 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](http://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.

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.

<sup>1</sup> "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.  
<https://stacresearch.com/news/2016/03/29/IBM160229>



Please Recycle

---