Data-intensive computing for big data with IBM Spectrum Symphony

Improve the management of structured and unstructured data

Highlights

- Promotes low latency architecture
- Enables high resource availability
- Supports multiple applications and mixed workload types running in the same cluster
- Priority-based workload scheduler.

IBM® Spectrum Symphony Advanced Edition is an enterprise-class distributed runtime engine for MapReduce applications. It is designed to deliver production-ready capabilities such as high resource availability and predictability, multiple applications and file systems support, operational maturity, service level agreement (SLA) policy control and high resource utilisation for MapReduce applications. Drawing on the years of distributed workload scheduling and management experience developed by IBM Spectrum Computing professionals, IBM Spectrum Symphony offers modern distributed workload runtime services for your MapReduce applications.

With the explosion in data availability, organisations are struggling with the management and processing of structured and unstructured data. In fact, approximately 80 percent of data in some organisations is unstructured data. In order to extract value from this unstructured data, new systems and processes are needed. A popular mechanism to deal with this growing need is to employ the Hadoop File System (HDFS) and the Hadoop MapReduce framework. IBM provides such a system in the IBM InfoSphere BigInsights offering tailored for enterprise use. BigInsights delivers a number of additional management and analytics enhancements that enterprises find valuable as part of a Smarter Analytics strategy.

For some enterprises, the need for these systems goes beyond processing, running Hadoop applications in their distributed computing infrastructure. These enterprises require a system that that can process queries or tasks with extremely quick (sub-millisecond(ms) response times. They wish to leverage the distributed cluster for multiple types of applications, not just for Hadoop applications. In addition, many enterprises want to share the application amongst multiple subteams, each with their own SLAs and resource allocations within the shared environment. IBM Spectrum Symphony offers a distributed computing platform that provides heterogeneous application support within a single cluster. Applications executed in this environment are provisioned with very low latency and by providing a multitenant architecture, it allows better allocation of resources and SLAs for different subteams sharing the environment.

Low latency architecture

IBM Spectrum Symphony employs optimisation techniques that allow ms provisioning of workloads. This implies that short running jobs have a smaller percentage of time spent in the provisioning and deprovisioning steps, giving a higher useful work to overhead ratio. It also has a high job throughput rate whereby the system allows more than 5000 tasks per second to be submitted. Comparable open source systems are two orders of magnitude slower as of current publication.

Policy-driven workload scheduler

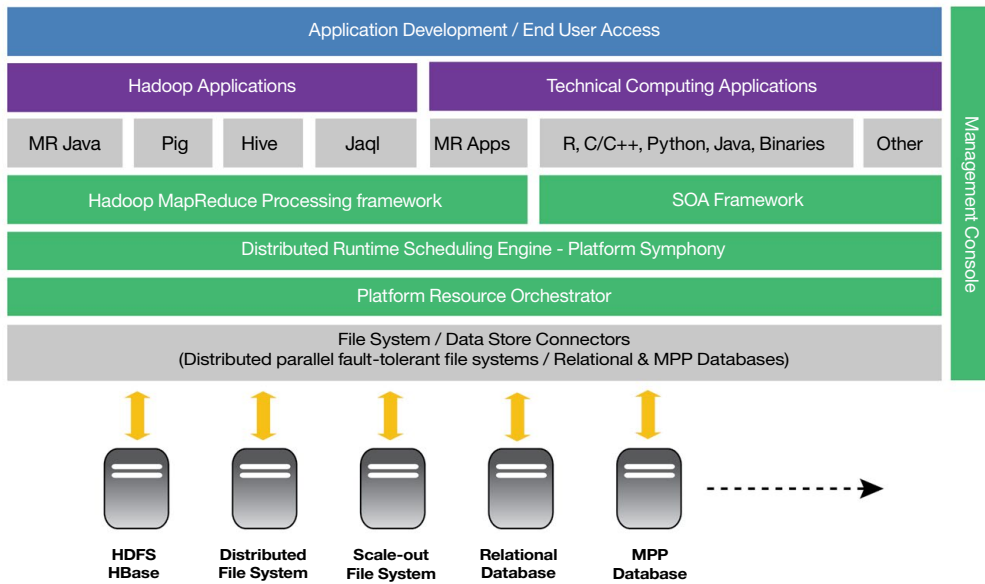The policy-driven workload scheduler in IBM Spectrum Symphony provides 10,000 levels of priority and supports multiple MapReduce jobs running in parallel. This policy-driven scheduler includes delivery of resource priority for preemptive jobs, as well as fair share scheduling of Mapper and Reducer jobs, all done at the job level to provide better granularity and control.

High resource availability

IBM Spectrum Symphony helps ensure uptime within the distributed runtime engine – there are no single points of failure. It provides job tracker and task tracker automatic failover and job recovery, without the need to restart jobs. For the Hadoop file system, IBM Spectrum Symphony offers automatic failover of the NameNode within the Hadoop Distributed file system and provides file system recovery and dependent job recovery.

Open architecture for application development and choice

IBM Spectrum Symphony is built on an open architecture to support multiple MapReduce applications, including 100 percent Hadoop application compatibility for Java™-based MapReduce jobs. The application adapter technology built into the product delivers seamless application integration with IBM Spectrum Symphony so that jobs built with Hadoop MapReduce technology (Java, Pig, Hive and others) require no changes to the programming logic for execution on IBM Spectrum Symphony. This open architecture also provides a method for leveraging multiple file system types, as well as database architectures. IBM Spectrum Symphony fully supports HDFS, IBM Spectrum Scale and other distributed file system types and data types. In addition, for MapReduce processes, the input data source file system type can be different from the output data source file system. This provides support for many uses, including extract, transformation and load (ETL) workflow logic.

Application Development / End User Access								
Hadoop Applications				Technical Computing Applications				
MR Java	Pig	Hive	Jaql	MR Apps	R, C/C++, Python, Java, Binaries	Other		
Hadoop MapReduce Processing framework				SOA Framework				
Distributed Runtime Scheduling Engine - Platform Symphony								
Platform Resource Orchestrator								
File System / Data Store Connectors (Distributed parallel fault-tolerant file systems / Relational & MPP Databases)								

HDFS HBase · Distributed File System · Scale-out File System · Relational Database · MPP Database

Management Console

IBM Spectrum Symphony Advanced Edition Architecture

Support for multiple MapReduce applications

IBM Spectrum Symphony Advanced Edition includes an Apache Hadoop-compatible MapReduce implementation which supports up to 300 separate applications (Job Trackers) for MapReduce workloads, as well as other types of distributed applications, simultaneously. This allows customers to leverage both existing and new resources and maximise their IT infrastructure while maintaining a single management interface.

Support for rolling upgrades

IBM Spectrum Symphony MapReduce supports multiple versions of MapReduce applications running on the same clusters; there is no need to take down the entire cluster for software upgrade. The servers running upgraded applications can co-exist with the previous version of the product on other nodes and thus allow upgrades to be done incrementally over a set of servers without taking down the entire cluster.

Greater monitoring and troubleshooting capabilities

IBM Spectrum Symphony MapReduce monitors central processing unit (CPU) and memory utilisation level and allocates resources accordingly. It provides the ability to pull log data from individual servers and manage them from a single interface.

IBM Spectrum Symphony MapReduce data affinity

IBM Spectrum Symphony MapReduce includes powerful data affinity capabilities to significantly improve application performance and resource utilisation by taking into account data locality when scheduling MapReduce workloads. Its data affinity solution virtually eliminates the time it takes to access large data volumes required by MapReduce applications. It significantly increases overall application performance through faster file access.

Why IBM?

IBM Spectrum Computing offers a comprehensive portfolio of software defined infrastructure solutions designed to help your organisation deliver IT services in the most efficient way possible, optimising resource utilisation to speed time to results and reduce costs. These offerings help maximise 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 centre (DC) 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, organisations 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 data-intensive computing for big data with IBM Spectrum Symphony, contact your IBM representative or IBM Business Partner (BP), or visit:
ibm.com/systems/spectrum-computing/products/symphony/

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