



要点

- 推进低延迟架构
 - 实现高资源可用性
 - 在相同集群内支持多个应用程序和混合工作负载类型
 - 基于优先级的工作负载调度器
-

利用 IBM Spectrum Symphony 实现面向大数据的数据密集型计算

改善结构化和非结构化数据的管理

IBM® Spectrum Symphony Advanced Edition 是面向 MapReduce 应用程序的企业级分布式运行时引擎。它旨在提供面向 MapReduce 应用程序的生产就绪功能，例如高资源可用性和可预测性、多种应用程序和文件系统支持、运营成熟度、SLA 策略控制和高资源利用率。借鉴 IBM Spectrum Computing 专业人员数年来积累的分布式工作负载调度和管理经验，IBM Spectrum Symphony 为您的 MapReduce 应用程序提供现代分布式工作负载运行时服务。

面对数据爆炸式增长，组织艰难应对结构化和非结构化数据的管理和处理。事实上，一些组织中近 80% 的数据是非结构化数据。为了从非结构化数据获取价值，需要新的系统和流程。为了应对这种日益增长的需求，一种流行的机制是应用 Hadoop 文件系统（HDFS）和 Hadoop MapReduce 框架。IBM 在专为企业用途而定制的 IBM InfoSphere® BigInsights® 解决方案中提供了这种系统。作为“智慧的分析”战略的组成部分，BigInsights 提供了一些企业认为有价值的额外管理和分析增强功能。



对于一些企业而言，对这些系统的需求超越了在他们的分布式计算基础架构中处理、运行 Hadoop 应用程序的范畴。这些企业需要一个能够在极短（亚毫秒级）响应时间内处理查询或任务的系统。他们希望利用面向多种应用程序类型而非仅仅是 Hadoop 应用程序的分布式集群。此外，许多企业期望在多个子团队之间共享应用程序，其中每个子团队在共享环境中都有自己的 SLA 和资源分配。IBM Spectrum Symphony 提供分布式计算平台，该平台在单个集群内提供异构应用程序支持。在此环境中执行的应用程序被配置极低的延迟，而且，通过提供多租户架构，它支持更好地为共享该环境的不同子团队分配资源和 SLA。

低延迟架构

IBM Spectrum Symphony 应用可实现亚毫秒级工作负载配置的优化技巧。这意味着短时间运行的作业将有更小比例的时间用于配置和取消配置步骤，从而实现更高的有效工作与开销比例。它还具备高作业吞吐率，使得系统支持每秒提交超过 5000 项任务。截至本文发布时，同等开源系统的速度慢两个数量级。

策略驱动的工作负载调度器

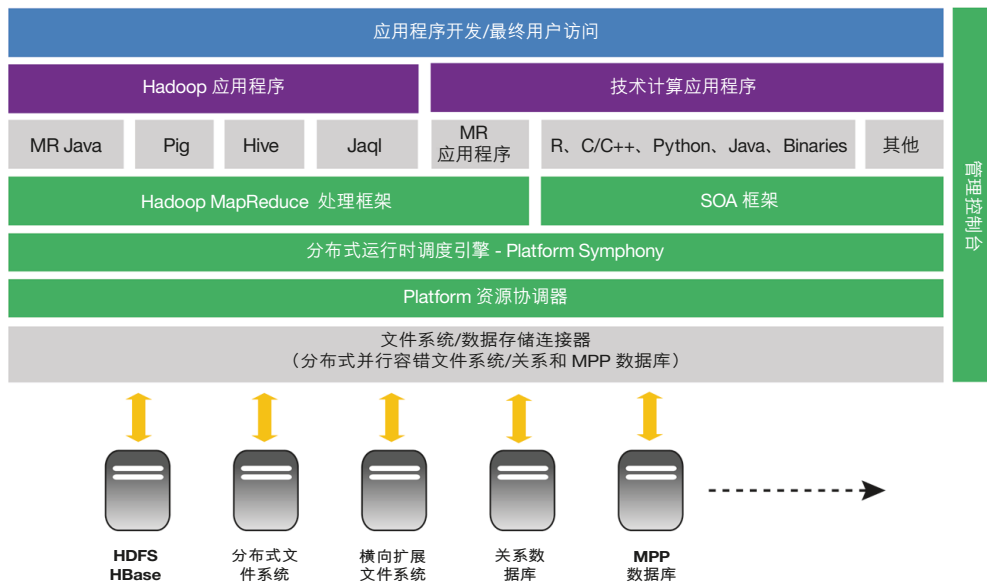
IBM Spectrum Symphony 中策略驱动的工作负载调度器提供 10,000 个优先级级别，支持多个 MapReduce 作业并行运行。这种策略驱动的调度器包括抢占式作业资源优先级交付以及 Mapper 和 Reducer 作业公平共享调度，全部在作业级别完成，以提供更好的粒度和控制。

高资源可用性

IBM Spectrum Symphony 有助于确保分布式运行时引擎中的正常运行时间 — 不存在单点故障。它提供 Job Tracker 和 Task Tracker 自动故障转移和作业恢复，无需重启作业。对于 Hadoop 文件系统，IBM Spectrum Symphony 在 Hadoop 分布式文件系统中提供 NameNode 自动故障转移，并提供文件系统恢复和依赖作业恢复。

面向应用程序开发和选择的开放架构

IBM Spectrum Symphony 基于开放架构而构建，支持多个 MapReduce 应用程序，包括面向基于 Java 的 MapReduce 作业的 100% Hadoop 应用程序兼容性。该产品内置的应用程序适配器技术提供与 IBM Spectrum Symphony 的无缝集成，因此采用 Hadoop MapReduce 技术 (Java、Pig、Hive 等) 构建的作业无需对编程逻辑进行调整便可在 IBM Spectrum Symphony 上执行。这种开放架构还提供一种利用多种文件系统类型和数据库架构的方法。IBM Spectrum Symphony 全面支持 HDFS、IBM Spectrum Scale™ 和其他分布式文件系统类型和数据类型。此外，对于 MapReduce 流程，输入数据源文件系统类型可以不同于输出数据源文件系统。这提供对许多用途的支持，包括提取、转换和加载 (ETL) 工作流逻辑。



IBM Spectrum Symphony Advanced Edition 架构

支持多个 MapReduce 应用程序

IBM Spectrum Symphony Advanced Edition 包含兼容 Apache Hadoop 的 MapReduce 实施，该实施同时支持多达 300 个面向 MapReduce 工作负载的单独应用程序 (Job Trackers) 以及其他类型的分布式应用程序。这让客户可以利用现有和新的资源，并充分利用他们的 IT 基础架构，同时维持单一管理界面。

支持滚动升级

IBM Spectrum Symphony MapReduce 支持在相同集群中运行多个版本的 MapReduce 应用程序；进行软件升级无需让整个集群停机。运行升级版应用程序的服务器可与其他节点上的前一版本产品并存，因此，可以对一组服务器进行渐进升级，无需让整个集群停机。

更出色的监控和故障排除功能

IBM Spectrum Symphony MapReduce 监控 CPU 和内存利用率水平并相应分配资源。它能够从各个服务器拉取日志数据，并从单个界面管理这些数据。

IBM Spectrum Symphony MapReduce 数据亲和性

IBM Spectrum Symphony MapReduce 包含强大的数据亲和性功能，通过在调度 MapReduce 工作负载时考虑数据局部性，可大幅改善应用程序性能和资源利用率。它的数据亲和性解决方案几乎可以消除访问 MapReduce 应用程序所需大量数据所花费的时间。它通过更快的文件访问速度大幅提高整体应用程序性能。

为何选择 IBM?

IBM Spectrum Computing 推出了全面的软件定义基础架构解决方案产品组合，有助于组织以最高效的方式交付 IT 服务，优化资源利用率以缩短实现成果的时间并降低成本。这些产品有助于最大程度发挥基础架构的潜能，无论规模大小，都能加快分析、高性能运算 (HPC)、Apache Hadoop、Spark 以及原生云应用程序，从数据中获取洞察，以更快的速度将更优质的产品推向市场。

无论是部署在数据中心还是云端，IBM Spectrum Computing 解决方案都能以更快的速度完成产品开发，制定关键业务决策并获得突破性洞察，该方案适用于金融服务、制造、数字媒体、油气、生命科学、政府、研究与教育等领域。从设计一级方程式赛车到信用风险分析，各行各业的组织都在使用 IBM Spectrum Computing，将它作为软件定义基础架构解决方案的基础，应对在大数据、分析、HPC 和云方面的挑战，从而获得更优异的业务成果。

如需更多信息

如需了解有关利用 IBM Spectrum Symphony 实现面向大数据的数据密集型计算的更多信息，请联系您的 IBM 代表或 IBM 业务合作伙伴，或者访问：

ibm.com/systems/spectrum-computing/products/symphony/

此外，IBM 全球融资部提供多种付款选项，可以帮助您获得发展业务所需的技术。从采购到处理，我们提供 IT 产品和服务全生命周期管理。有关更多信息，请访问：

ibm.com/financing



© Copyright IBM Corporation 2016

IBM Systems
Route 100
Somers, NY 10589

2016 年 6 月

IBM、IBM 徽标、ibm.com、BigInsights、InfoSphere 和 IBM Spectrum Scale 是 International Business Machines Corp. 在全球许多司法辖区的注册商标。其他产品或服务名称可能是 IBM 或其他公司的商标。有关 IBM 商标的最新列表，请访问以下网站的“版权与商标信息”部分：
ibm.com/legal/copytrade.shtml

Java 和所有基于 Java 的商标与徽标均为 Oracle 和/或其关联公司的商标或注册商标。

本文档为初始发布日时的最新文档，IBM 可能随时对其进行更改。IBM 并未在每个开展业务的国家/地区提供所有产品/服务。

本文档中的信息“按原样”提供，不带任何明示或暗示的保证，包括不带任何适用性、对特定用途的适用性的保证以及任何不侵权的保证或条件。IBM 根据提供产品时的协议条款与条件提供产品担保。



请回收再利用