

ALIBABA CLOUD

阿里云

E-MapReduce
产品简介

文档版本：20220713

 阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <code>Instance_ID</code>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1.什么是E-MapReduce	06
2.产品优势	09
3.产品架构	10
4.应用场景	11
5.基本概念	13
6.使用限制	14
7.技术支持	16
8.发行版本	18
8.1. 版本概述	18
8.2. 版本说明（EMR-5.x系列）	30
8.2.1. EMR-5.7.x版本说明	30
8.2.2. EMR-5.6.x版本说明	31
8.2.3. EMR-5.5.x版本说明	32
8.2.4. EMR-5.4.x版本说明	60
8.2.5. EMR-5.3.x版本说明	63
8.2.6. EMR-5.2.x版本说明	64
8.3. 版本说明（EMR-4.x系列）	66
8.3.1. EMR-4.10.x版本说明	66
8.3.2. EMR-4.9.x版本说明	68
8.3.3. EMR-4.8.x版本说明	69
8.3.4. EMR-4.6.x版本说明	71
8.3.5. EMR-4.5.x版本说明	72
8.3.6. EMR-4.4.x版本说明	73
8.3.7. EMR-4.3.x版本说明	74
8.4. 版本说明（EMR-3.x系列）	75
8.4.1. EMR-3.41.x版本说明	75

8.4.2. EMR-3.40.x版本说明	76
8.4.3. EMR-3.39.x版本说明	77
8.4.4. EMR-3.38.x版本说明	80
8.4.5. EMR-3.37.x版本说明	83
8.4.6. EMR-3.36.x版本说明	84
8.4.7. EMR-3.35.x版本说明	86
8.4.8. EMR-3.34.x版本说明	87
8.4.9. EMR-3.33.x版本说明	89
8.4.10. EMR-3.32.x版本说明	90
8.4.11. EMR-3.30.x版本说明	91
8.4.12. EMR-3.29.x版本说明	93
8.4.13. EMR-3.28.x版本说明	94
8.4.14. EMR-3.27.x版本说明	95
8.4.15. EMR-3.26.x版本说明	96
8.4.16. EMR-3.25.x版本说明	97
8.4.17. EMR-3.24.x版本说明	98
8.4.18. EMR-3.23.x版本说明	100
8.4.19. EMR-3.22.x版本说明	101
8.4.20. EMR-3.22.x之前版本说明	103

1.什么是E-MapReduce

开源大数据开发平台E-MapReduce（简称EMR），是运行在阿里云平台上的一种大数据处理的系统解决方案。

产品介绍

开源大数据开发平台EMR构建于云服务器ECS上，基于开源的Apache Hadoop和Apache Spark，让您方便地使用Hadoop和Spark生态系统中的其他周边系统分析和处理数据。EMR还可以与阿里云其他的云数据存储系统和数据库系统（例如，阿里云OSS和RDS等）进行数据传输。

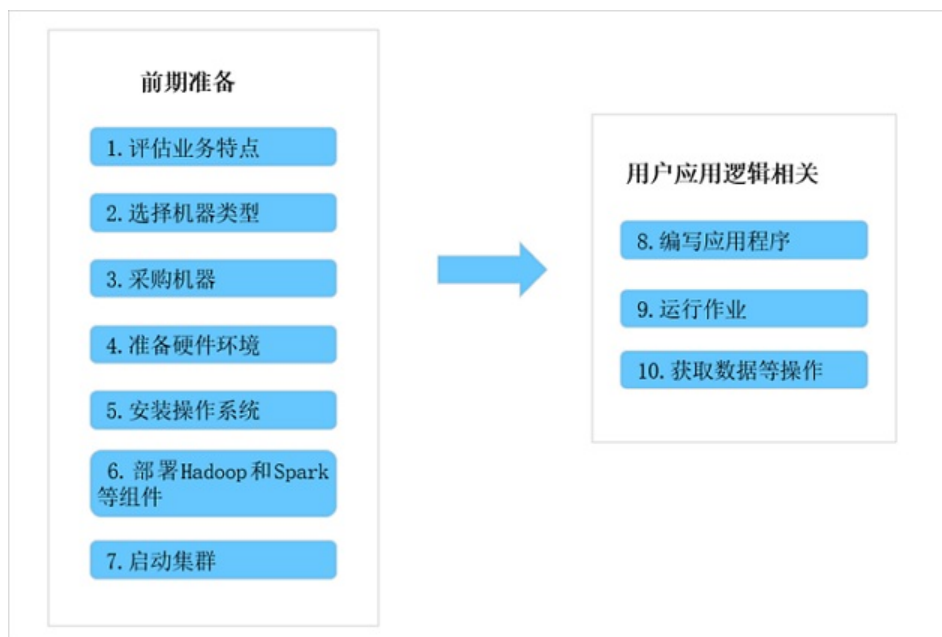
开源大数据开发平台EMR的Smart Data组件是EMR Jindo引擎的主要存储部分，为开源大数据开发平台EMR各个计算引擎提供统一的存储优化、缓存优化、计算缓存加速优化和多个存储功能扩展。

说明

- 关于Apache Hadoop的更多介绍，请参见[Apache Hadoop官网](#)。
- 关于Apache Spark的更多介绍，请参见[Apache Spark官网](#)。
- 关于Apache Hive的更多介绍，请参见[Apache Hive官网](#)。
- 关于Apache HBase的更多介绍，请参见[Apache HBase官网](#)。
- 关于Smart Data的更多介绍，请参见[Smart Data](#)。

E-MapReduce的用途

以往在使用Hadoop和Spark等分布式处理系统时，您通常需要执行如下步骤。



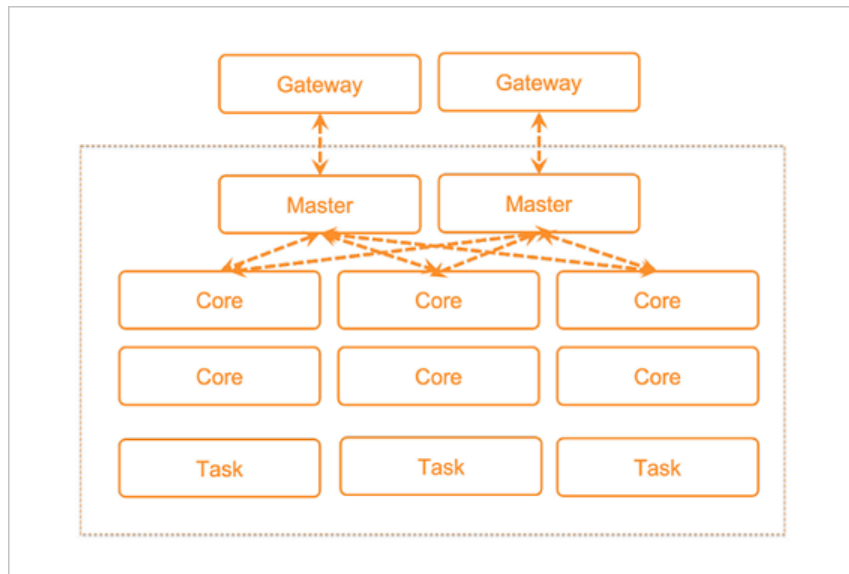
在上述使用流程中，真正跟用户的应用逻辑相关的是步骤8~10，而步骤1~7都是前期准备工作，但这些前期准备工作都非常冗长繁琐。E-MapReduce提供了集群管理工具的集成解决方案，例如，主机选型、环境部署、集群搭建、集群配置、集群运行、作业配置、作业运行、集群管理和性能监控等。通过E-MapReduce，您可以从繁琐的集群构建相关的采购、准备和运维等工作中解放出来，只关心自己应用程序的处理逻辑即可。

此外，E-MapReduce还为您提供了灵活的搭配组合方式，您可以根据自己的业务特点选择不同的集群服务。例如，如果您的需求是对数据进行日常统计和简单的批量运算，则可以只选择在E-MapReduce中运行Hadoop服务；如果您有流式计算和实时计算的需求，则可以在Hadoop服务基础上再加入Spark服务。

E-MapReduce的组成

E-MapReduce的核心是集群。E-MapReduce集群是由一个或多个阿里云ECS实例组成的Hadoop、Flink、Druid、ZooKeeper集群。以Hadoop为例，每个ECS实例上通常都运行了一些daemon进程（例如，NameNode、DataNode、ResourceManager和NodeManager），这些daemon进程共同组成了Hadoop集群。

例如，下图是一个包含Master节点、Core节点和Task节点的Hadoop集群和Gateway集群。



- Master节点，部署了Hadoop的主节点服务，包括HDFS NameNode、HDFS JournalNode、ZooKeeper、YARN ResourceManager和HBase HMaster等服务，可以根据集群的使用场景，选择高可用集群或非高可用集群。测试环境可以选择非高可用集群，生产环境建议选择高可用集群。高可用集群可以选择2个或3个Master节点，当选择2个Master节点时，HDFS JournalNode和ZooKeeper会部署在Core的emr-worker-1节点。生产环境建议创建高可用集群时选择3个Master节点。
- Core节点，部署了HDFS DataNode和YARN Nodemanager，用于HDFS数据的存储和YARN的计算，不可以弹性伸缩。
- Task节点，部署了YARN NodeManager，用于YARN计算，可以通过弹性伸缩的方式灵活扩容或缩容。
- Gateway集群，部署了Hadoop的客户端文件，您可以通过Gateway提交作业，避免直接登录集群产生的安全和客户端环境隔离问题。您需要先创建Hadoop集群，然后创建Gateway集群关联至Hadoop集群。

与自建Hadoop集群对比

开源大数据开发平台EMR与自建Hadoop集群的优势对比如下表所示。

对比项	阿里云EMR	自建Hadoop集群
成本	资源按量付费，支持集群资源灵活调整，数据分层存储，资源使用率高。无额外软件License费用。	提前预估资源，且资源相对固定，资源使用率低。采用Hadoop发行版，需额外支付License费用。
性能	较开源版本性能大幅提升。	采用开源社区版本，性能需自行优化。
易用性	分钟级别启动Hadoop集群，敏捷响应业务需求。	采购服务器，部署Hadoop生态组件，周期长达数周。
弹性	可根据作业临时启动和销毁集群。集群资源可根据时间周期或集群负载动态自动调整。基于JindoFS计算存储分离架构，轻松分别扩展计算和存储资源。	计算和存储耦合，资源相对固定，无法弹性调整资源。

对比项	阿里云EMR	自建Hadoop集群
安全	支持企业级多租户资源管理能力，支持对表、列、行级别的权限控制和日志审计，支持数据加密。	多租户管理能力需自行配置，能力不完善，无法满足企业级需求。
可靠	大规模、企业级环境的检验，随开源版本升级，并经过专业的兼容性验证测试，提供优于社区版本的使用体验。	需自行更新和升级开源版本，验证各组件版本兼容性，自行修复社区bug。
服务	专业和资深大数据专家技术服务团队提供售后支持。	社区版本无服务支持，Hadoop发行版，需额外支付License和服务费用。

2. 产品优势

阿里云E-MapReduce（简称EMR）为您提供相对方便可控的企业级开源大数据服务。您可以快速搭建开源大数据服务，例如Hadoop、Spark、Flink、Kafka和HBase服务。

稳定可靠的开源组件

- 100%采用社区开源组件，随开源版本升级迭代，详情请参见[版本概述](#)。
- 适配开源组件，避免开源组件之间的版本兼容性问题。
- 基于开源组件，优化和增强阿里云部署环境，性能远高于开源版本。

节约成本

- 通过有效弹性伸缩和数据分层存储机制，相较于传统HDFS固定集群方式，可以节省高达50%以上的费用。
- 支持创建抢占式实例，相较于按量付费的购买方式，可以节省50%~80%的费用，详情请参见[抢占式实例概述](#)。

易用性

- 分钟级别创建和扩容集群，无需手动部署和启动服务。
- 完善集群的监控和告警体系，覆盖硬件和Hadoop服务，您可以配置告警模板。

弹性

- 计算存储分离：解耦了计算与存储之间的绑定关系，实现了资源的弹性利用。
- 自定义集群环境：您可以通过引导操作和集群脚本灵活配置集群环境，将第三方优化和集群管理工具部署到EMR环境，详情请参见[管理引导操作](#)和[集群脚本](#)。
- 自主运维：您可以登录Master节点，查看集群日志和部署环境，优化和部署配置，详情请参见[常用文件路径](#)。
- 弹性伸缩：可以通过弹性伸缩的方式灵活扩容或缩容。

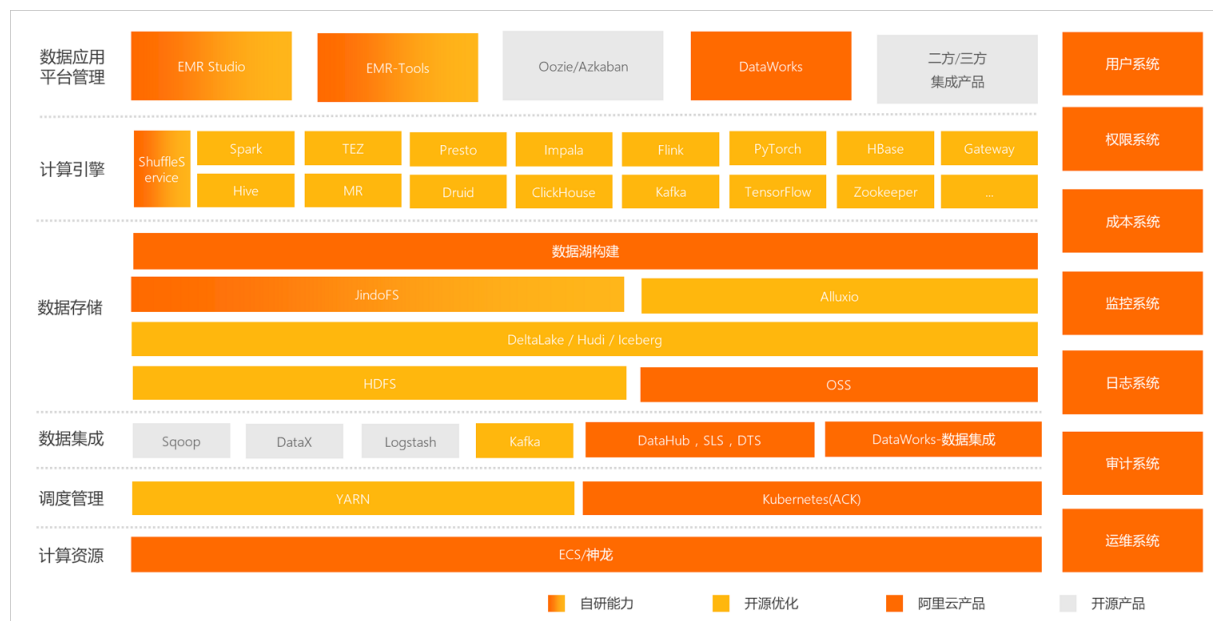
深度整合

- 支持基于阿里云ECS和ACK部署EMR集群，支持ECS多种实例规格，您可以根据使用场景灵活选择，详情请参见[ECS实例说明](#)。
- 集成在DataWorks，您可以在DataWorks上使用EMR作为作业计算和数据存储引擎。
- 集成了数据湖构建（Data Lake Formation），实现数据湖场景下多引擎的统一元数据管理。

3. 产品架构

本文介绍E-MapReduce（简称EMR）的产品架构，以便您直观的了解EMR的产品组成。

EMR的产品架构如下图所示。



从上图可以看出EMR由四部分组成：

- 社区开源产品

集成Apache社区开源大数据组件，例如Hadoop、Hive和HBase，随着EMR版本更新，开源软件也会相应的升级，详情请参见[版本概述](#)下各版本的版本说明。



注意 已经创建好的EMR集群不支持组件升级。

- EMR开源优化

EMR基于开源社区版本的组件，增强了其性能和功能，例如：

- Spark中增加了Spark Streaming SQL，性能较开源版本有大幅提升，详情请参见[简介](#)。
- Delta Lake相较开源版本增加了ZOrder和Data Skipping能力，详情请参见[Delta Lake概述](#)。

- EMR自研能力

为让开源大数据组件和服务更好的运行在阿里云技术设施上，EMR自研了如下组件：

- 数据应用平台，提供交互式开发、作业提交、作业调试和工作流一站式数据开发体验，详情请参见[EMR Studio概述](#)。
- Shuffle Service是EMR在优化计算引擎的Shuffle操作上，推出的扩展组件，详情请参见[ESS概述](#)。
- SmartData，为EMR各个计算引擎提供统一的存储优化、缓存优化、计算缓存加速优化和多个存储功能扩展，详情请参见[SmartData](#)。

- 阿里云产品

EMR衔接了开源大数据生态和阿里云生态。EMR可以部署在阿里云ECS（Elastic Compute Service）和Kubernetes（简称ACK）上；数据可以存储在阿里云OSS上；通过在EMR上创建Data Science集群可以使用及学习机器学习PAI；EMR集成在DataWorks，您可以在DataWorks上使用EMR作为作业计算和数据存储引擎。

4. 应用场景

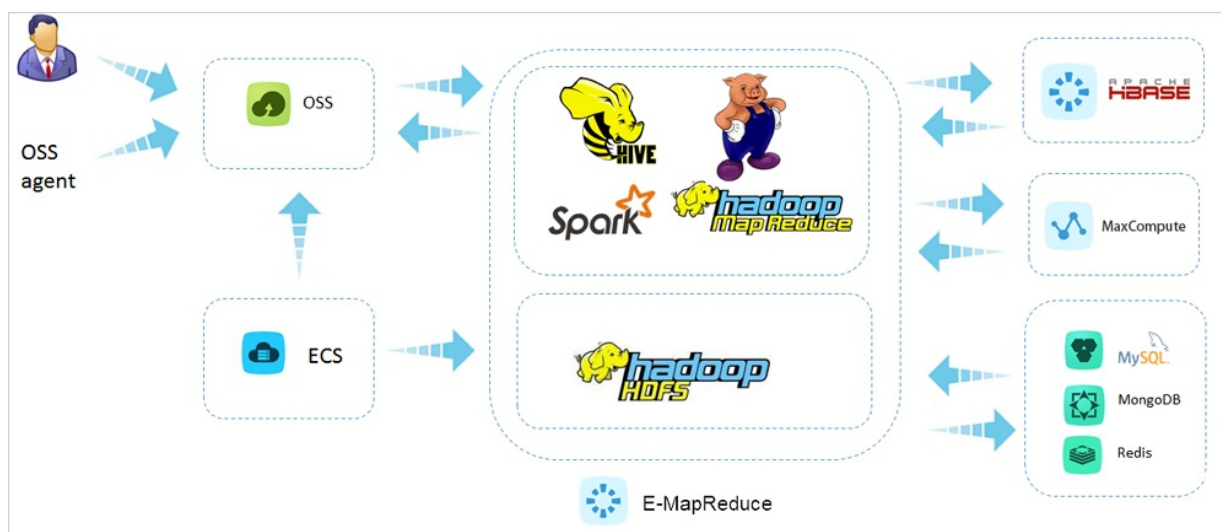
E-MapReduce集群适用多种使用场景，同时支持Hadoop EcoSystem和Spark能够支持的所有场景。

E-MapReduce本质是Hadoop和Spark的集群服务，您完全可以将其使用的阿里云ECS主机视为您专属的物理主机。以下示例列出了E-MapReduce使用的经典场景：

- 批量数据处理
- Ad hoc数据分析查询
- 海量数据在线服务
- 流式数据处理

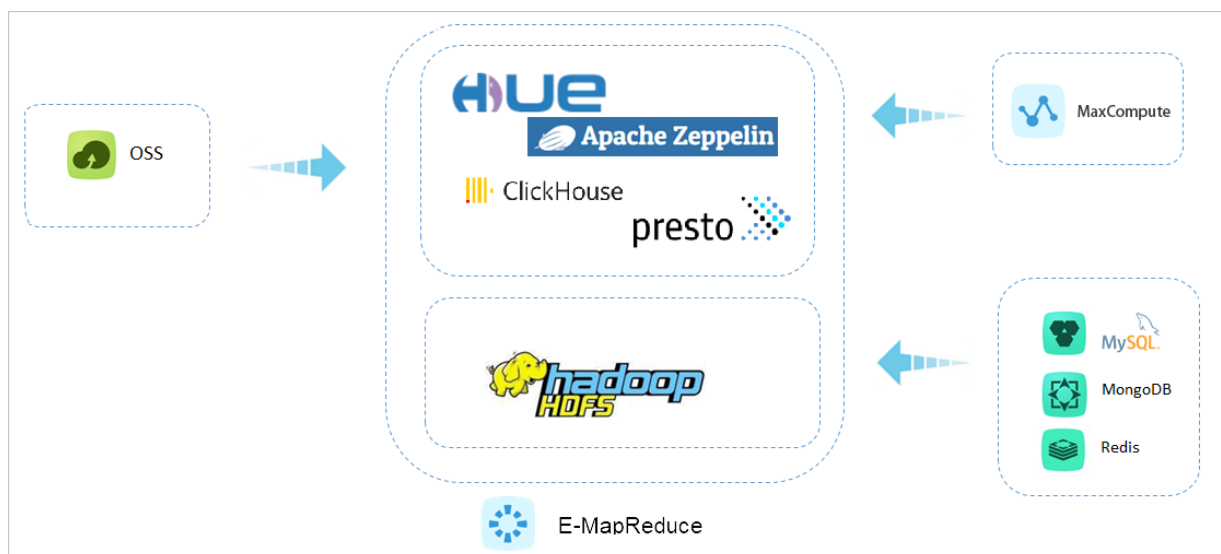
批量数据处理

将海量日志同步到E-MapReduce的数据节点后，您可以借助于Hue等工具使用Hive、Spark和Presto等主流计算框架快速获取数据洞察力。您还可以使用Sqoop等工具加载分散于各RDS或其他存储引擎的数据，并把分析后的数据同步到RDS，为数据可视化产品提供数据支撑。



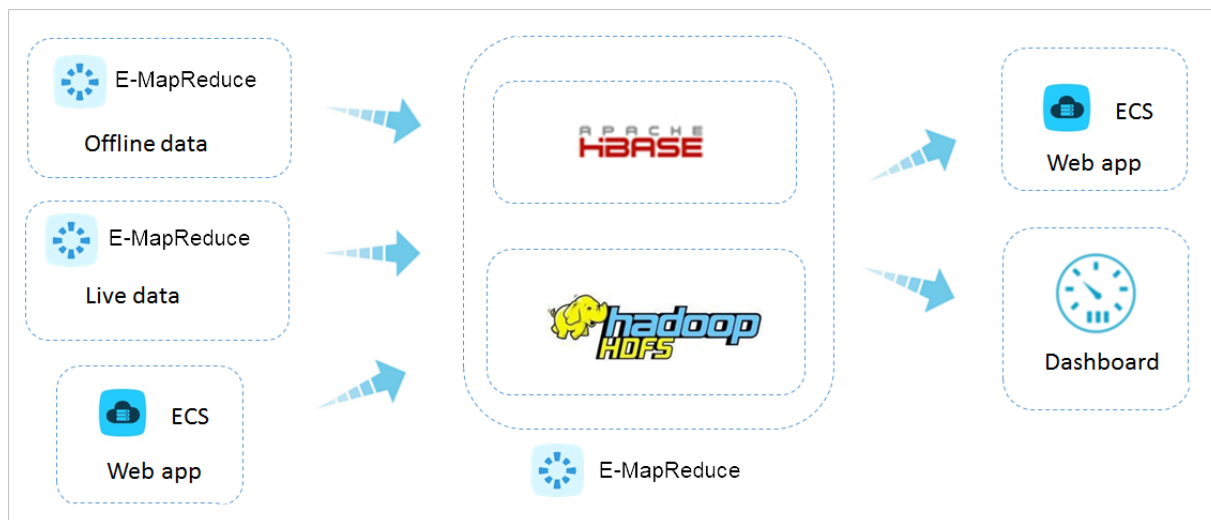
Ad hoc数据分析查询

E-MapReduce将海量数据通过导入或者外表等形式引入到OLAP分析引擎里，例如，Clickhouse、Presto和Impala，提供高效、实时和灵活的数据分析能力，满足用户画像、人群圈选、BI报表和业务分析等一系列的业务场景。



海量数据在线服务

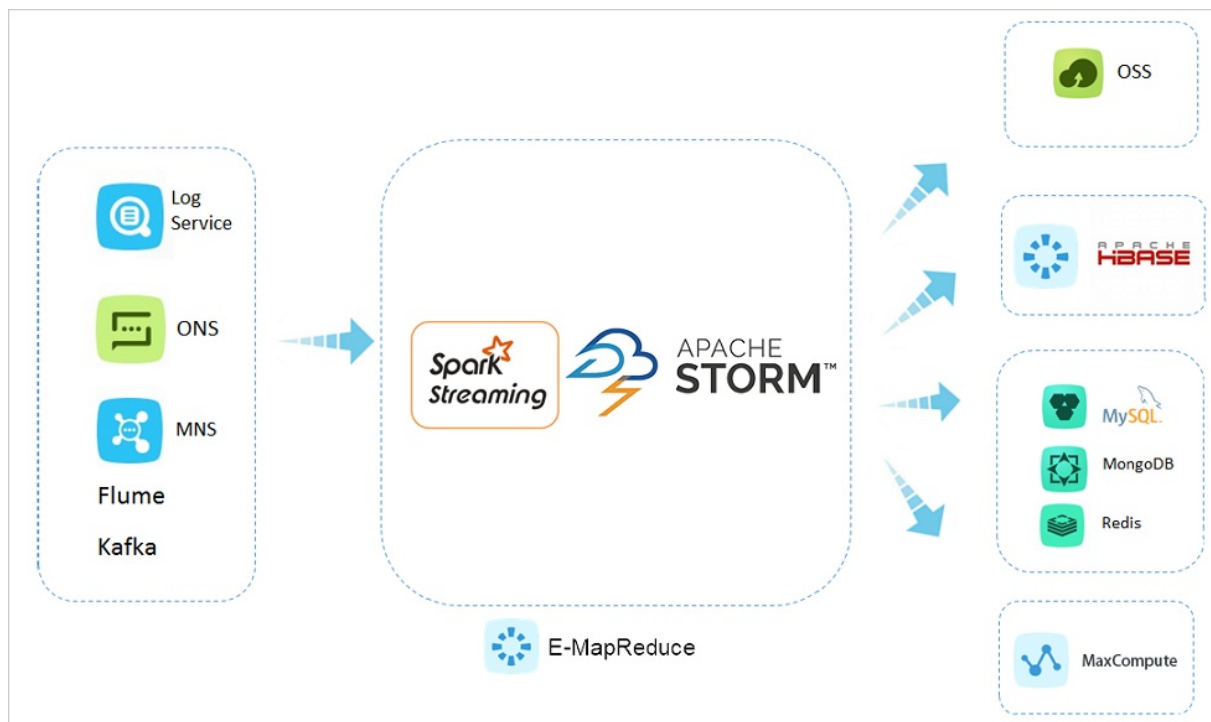
E-MapReduce基于Web和移动应用程序等生成的PB级别的结构化、半结构化或非结构化数据进行在线分析，以方便Web应用或者数据可视化产品获取分析结果进行实时展示。



流式数据处理

通过Spark Streaming和Storm，使用和处理来自阿里云日志服务Log Service、阿里云消息队列ONS（Message Queue）、阿里云消息服务MNS（Message Service）、Apache Kafka或其他数据流的实时数据。

采用容错方式执行流式数据分析，并将相应结果写入阿里云对象存储服务OSS（Object Storage Service）或HDFS中。



5. 基本概念

本文介绍开源大数据平台E-MapReduce使用过程中遇到的常用名词的基本概念和简要描述。

集群

E-MapReduce产品中的一个独立的机器组合，为上层业务提供一些基础的引擎服务。一个E-MapReduce集群是由一个或多个阿里云ECS实例组成的，包含Hadoop、Flink或ZooKeeper集群。

ECS实例

在E-MapReduce上创建的一台云服务器，等同于一台虚拟服务器，内含CPU、内存、操作系统、网络配置、磁盘等基础的组件。更多介绍，请参见[实例概述](#)。

E-MapReduce（简称EMR）支持的ECS实例类型，以及各实例类型适用的场景，请参见[ECS实例说明](#)。

节点

E-MapReduce集群由多个不同类型的实例节点组成，包括主实例节点（Master）、核心实例节点（Core）和计算实例节点（Task）。详细信息，请参见[实例类型](#)。

云盘和本地盘

- 云盘：数据块级别的块存储产品，采用分布式三副本机制，为ECS实例提供99.9999999%的数据可靠性保证。
- 本地盘：ECS实例所在物理机上的本地硬盘设备，存储I/O性能、海量存储的性价比极高，但数据可靠性取决于物理机的可靠性，存在单点故障风险。

详细信息，请参见[存储说明](#)和[本地盘机型概述](#)。

Region（地域）

地域是指物理的数据中心。资源创建成功后不能更换地域。详细信息，请参见[地域和可用区](#)。

安全组

一种虚拟防火墙，您可以基于安全组控制实例的入流量和出流量。详细信息，请参见[管理安全组](#)。

SSH密钥对

一种安全便捷的登录认证方式，由公钥和私钥组成，仅支持Linux实例。详细信息，请参见[SSH密钥对概述](#)。

专有网络

您基于阿里云创建的自定义私有网络，不同专有网络之间通过隧道在逻辑上彻底隔离。您可以完全掌控自己的专有网络，例如选择IP地址范围、配置路由表和网关等。详细信息，请参见[什么是专有网络](#)。

标签

由一对键值（Key-Value）组成。使用标签标识具有相同特征的资源后，例如所属组织或用途相同的资源，您可以基于标签方便地检索和管理资源。详细信息，请参见[设置标签](#)。

资源组

供您从业务角度管理跨地域、跨产品的资源，并支持针对资源组的权限管理。为云资源指定资源组及相关的示例，请参见[使用资源组](#)。

6.使用限制

您在使用E-MapReduce（简称EMR）过程中，某些非产品预期的操作可能会导致集群的不稳定甚至集群的不可用，为了避免这种问题，请您认真阅读本文。本文为您介绍了EMR的一些使用限制。

 **注意** 在使用EMR时，所有操作都需要在EMR控制台上执行，通常不建议您在ECS控制台上进行操作，不合适的操作可能导致集群不稳定或异常。因此请您仔细阅读本文，如果遇到文档中的操作导致的问题或故障，需由您自行承担后果或提前购买[专家服务](#)。

高危操作（禁止）

操作	可能的结果	建议
禁止删除或者修改etc/hosts目录下的hosts文件	集群关联不到节点上的服务，导致服务异常。	增加hosts信息。
禁止直接在ECS控制台上修改组件配置文件的参数	服务重启后，导致修改的参数被覆盖。	在EMR控制台上修改参数配置。
禁止在ECS控制台重新部署ECS实例	影响EMR服务。	提交工单 处理。
禁止在ECS控制台对EMR节点挂载磁盘	因为EMR无法识别和初始化磁盘，所以会导致磁盘不可用。	如果您需要更换磁盘，请 提交工单 处理。
禁止在ECS控制台对EMR节点卸载磁盘	因为EMR无法感知到卸载磁盘操作，所以容易导致数据丢失。	
禁止在ECS控制台直接删除Core节点	导致数据丢失和在删除节点上的作业执行失败。	无
禁止在ECS控制台直接删除Master节点	- HA集群，删除Master节点，会导致集群HDFS NameNode HA、YARN ResourceManager或HBase HMaster无法正常切换，该情况只能重新购买EMR集群，迁移数据/任务。 - 非HA集群，删除Master节点，会导致集群不可用，无法迁移数据和任务。	无
禁止在ECS控制台直接删除Task节点	执行在删除节点上的作业执行失败。	无
禁止停止Master的MySQL服务（创建集群时，元数据选择集群内置MySQL）	emr-header-1上的MySQL服务，关联到Hive MetaStore、Oozie和Ranger，如果停止服务，会造成支持的服务无法访问到数据库。	无
禁止修改emr-header-1节点上预装的MySQL root密码（创建集群时，元数据选择集群内置MySQL）	导致EMR的Hue或Ranger等组件失败。	无
禁止在集群运行中修改ECS节点的安全组	- 导致集群节点间网络异常。 - 导致组件不可用。	无

操作	可能的结果	建议
禁止在ECS控制台转换计费方式	导致无法转回原来的计费方式。	在EMR控制台进行计费方式的转换，详情请参见按量转包月。  注意 EMR不支持从包年包月转换为按量付费的计费方式。如果您需要从包年包月转换为按量付费的计费方式，请购买专家服务。

常见问题处理

问题	建议
磁盘容量不足	因为EMR集群的Master节点不支持磁盘数量的增加，所以您可以在EMR控制台调大单块磁盘的容量，或扩容Core节点。
磁盘容量过剩	因为EMR集群不支持磁盘容量缩容，所以您可以重新购买集群，详情请参见 创建集群 。
计算能力不足	您可以在EMR控制台上扩容Task节点，详情请参见 扩容集群 。
计算能力过剩	根据集群区分如下： - 针对按量付费集群，您可以在EMR控制台上缩容Task节点。 - 针对包年包月集群，先在Task节点Decommission YARN Nodemanager服务，然后在ECS控制台将对应Task节点的ECS转为按量付费后再释放。
组件版本过低	因为EMR集群不支持单组件版本升级，所以您可以重新购买高版本的集群，详情请参见 创建集群 。
转化非HA集群为HA集群	EMR集群暂不支持非HA集群转为HA集群，建议您重新购买HA集群。
在EMR上部署第三方软件或服务	建议您在集群创建时通过引导操作安装第三方软件或服务。 如果集群创建后手工安装第三方软件或服务，在扩容时，新扩容节点需重新手工安装第三方软件或服务。

7. 技术支持

阿里云E-MapReduce（简称EMR）是运行在阿里云平台上的系统解决方案。EMR构建于ECS或ACK之上，通过半托管的形式供您使用，您可以对集群具有完全的管理操作权限。为了更好的支持EMR产品，现将EMR产品技术支持服务规范进行相应的说明。

- 支持范围
- 支持方式
- 服务SLA

支持范围

- 支持的服务
 - 负责EMR产品管控平台（不含平台上的具体组件）的使用、咨询服务、稳定性问题和产品缺陷问题处理。管控平台功能如下：
 - 集群购买流程
 - 集群创建流程
 - 集群扩容流程
 - 集群添加服务流程
 - 集群弹性伸缩
 - 集群告警监控功能
 - 数据开发功能
 - 负责EMR产品自研组件的使用、咨询服务、稳定性问题和产品缺陷问题处理。自研组件如下：
 - SmartData
 - Bigboot
 - ESS
 - DLF-Auth
 - 协助排查EMR产品开源组件的问题，针对开源组件缺陷或需求，将按照产品规划迭代排期解决。开源组件如下：
 - Hadoop，包括HDFS、YARN和MapReduce。
 - 数据集成组件，例如，Flume和Sqoop等。
 - 数据存储组件，例如，HDFS、Hbase、Kudu、Iceberg、DeltaLake、Hudi和Kafka等。
 - 计算引擎组件，例如，Spark、Hive、Tez、Druid、Flink和Phoenix等。
 - OLAP组件，例如，Presto、ClickHouse和Impala等。
 - 其他组件，例如，Zookeeper、Knox、Ganglia、OpenLDAP、Hue、Zeppelin、Livy、Superset、Ranger、Alluxio和Oozie等。
 - 协助排查EMR产品依赖的其他阿里云产品的问题，对应的缺陷和需求，由对应产品团队负责解决。依赖的其他阿里云产品如下：
 - ECS或ACK
 - OSS
 - 云监控
- 不支持的服务
 - EMR提供运维工具（例如，告警监控和参数配置等）并保证工具功能的可靠性，但不提供具体集群和组件管理的运维操作，包括集群容量规划、组件性能优化以及集群上具体的运维操作（例如，重启服务和修改配置等）。

- 不负责基于EMR组件之上的客户业务应用开发问题答疑和处理，例如，业务实现、代码开发、作业性能调优和业务迁移等。
- 在集群组件无明显异常或明确产品缺陷的情况下，不负责单个作业的问题排查。
- 不负责非产品标准能力的支持，例如，集群或组件升级、Core节点缩容和磁盘缩容等。
- 不负责处理用户自行安装的第三方组件的问题。
- 不负责处理用户进行非产品预期操作，导致集群不稳定甚至集群不可用的问题。非产品预期的操作详情，请参见[使用限制](#)。

支持方式

如果您有业务问题或者产品疑难问题，建议您在阿里云官网提交工单寻求支持服务。

1. 登录[工单系统](#)。
2. 在搜索栏中，输入E-MapReduce。
3. 选择对应的问题分类后，单击创建工单。

4. 工单信息填写完成后，单击提交。

您可以扫描下方二维码进入E-MapReduce交流群。在E-MapReduce交流群中，您可以和其他用户进行交流。



说明 此钉钉群仅用于用户之间的交流，如果您有业务问题或者产品疑难问题，请您[提交工单](#)。

服务SLA

支持的服务SLA的详情，请参见[服务内容对比](#)的内容。

8. 发行版本

8.1. 版本概述

本文介绍E-MapReduce产品的发行版本号格式、发行版本更新记录和发行版本说明。

产品发行版本号格式

E-MapReduce采用类似EMR-a.b.c格式的版本号，详细说明如下：

- a代表了版本有较大的变动。
- b代表了版本中一些组件的变动。
- c代表了版本中bug修复，可以向前兼容。

以版本升级为例，详细说明如下：

- 1.0.0~2.0.0：这是一个a位置的大版本的变动，在升级后请务必测试保证之前的作业都能正常运作。
- 1.0.0~1.1.0：这是一个b位置的变动，一般是组件版本进行了升级，大部分都相似，但也需要进行测试。
- 1.0.0~1.0.1：这是一个c位置发生了变化，这两个版本可以完全兼容。

E-MapReduce每个版本的捆绑软件、集群创建以及集群升级说明如下：

- 捆绑软件：每个E-MapReduce发行版本上捆绑的软件和软件版本都是固定的。目前软件不支持多个不同版本的选择，也不推荐您自行更改软件的版本。
- 集群创建：创建某一个发行版本的E-MapReduce集群后，该集群版本不会自动升级。后续即使升级该版本对应的镜像，也不会影响已经创建的集群，但新建集群会使用新的镜像。
- 集群升级：集群升级完成后（例如，从1.0.x升级到1.1.x），请务必测试集群中的作业，以保证其在新软件环境中也能正常运行。

产品发行版本更新记录（EMR-5.x系列版本）

- [EMR-5.7.x版本说明](#)
- [EMR-5.6.x版本说明](#)
- [EMR-5.5.x版本说明](#)
- [EMR-5.4.x版本说明](#)
- [EMR-5.3.x版本说明](#)
- [EMR-5.2.x版本说明](#)

产品发行版本更新记录（EMR-4.x系列版本）

- [EMR-4.10.x版本说明](#)
- [EMR-4.9.x版本说明](#)
- [EMR-4.8.x版本说明](#)
- [EMR-4.6.x版本说明](#)
- [EMR-4.5.x版本说明](#)
- [EMR-4.4.x版本说明](#)
- [EMR-4.3.x版本说明](#)

产品发行版本更新记录（EMR-3.x系列版本）

- [EMR-3.41.x版本说明](#)
- [EMR-3.40.x版本说明](#)
- [EMR-3.39.x版本说明](#)
- [EMR-3.38.x版本说明](#)

- [EMR-3.37.x版本说明](#)
- [EMR-3.36.x版本说明](#)
- [EMR-3.34.0~EMR-3.35.0](#)
- [EMR-3.28.2~EMR-3.33.0](#)
- [EMR-3.25.1~EMR-3.28.1](#)
- [EMR-3.22.1~EMR-3.25.0](#)
- [产品发行版本更新记录（EMR-3.22.x之前版本）](#)

EMR-3.34.0~EMR-3.35.0

版本	EMR-3.35.0	EMR-3.34.0
发布时间	2021.4	2021.3
Hadoop	2.8.5	2.8.5
Knox	1.1.0	1.1.0
Spark	2.4.7	2.4.7
Hive	2.3.7	2.3.7
Tez	0.9.2	0.9.2
Sqoop	1.4.7	1.4.7
YARN	2.8.5	2.8.5
HDFS	2.8.5	2.8.5
Flink	1.11-vvr-2.0.1	1.11-vvr-2.0.1
Druid	0.20.1	0.20.0
HBase	1.4.9	1.4.9
Phoenix	4.14.1	4.14.1
Zookeeper	3.6.2	3.6.2
Livy	0.7.1	0.6.0
Presto	338	338
Storm	1.2.2	1.2.2
Impala	3.4.0	3.4.0
Flume	1.9.0	1.9.0
Hue	4.9.0	4.4.0
Oozie	5.1.0	5.1.0
Zeppelin	0.9.0	0.9.0

版本	EMR-3.35.0	EMR-3.34.0
Ranger	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2
SmartData	3.5.0	3.4.0
Kafka	1.1.1	1.1.1
Kafka-Manager	1.3.3.16	1.3.3.16
Superset	0.36.0	0.36.0
Bigboot	3.5.0	3.4.0
OpenLDAP	2.4.44	2.4.44
Kudu	1.10.0	1.10.0
Flink-Vvp	1.11-2.2.2	1.11-2.2.2
Alluxio	2.5.0	2.4.1
Hudi	0.6.0	0.6.0

EMR-3.28.2~EMR-3.33.0

版本	EMR-3.33.0	EMR-3.32.0	EMR-3.30.0	EMR-3.29.0	EMR-3.28.2
发布时间	2021.1	2020.11	2020.6	2020.6	2020.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
Knox	1.1.0	1.1.0	1.1.0	1.1.0	1.1.0
Spark	2.4.7	2.4.5	2.4.5	2.4.5	2.4.5
Hive	2.3.7	2.3.5	2.3.5	2.3.5	2.3.5
Tez	0.9.2	0.9.2	0.9.2	0.9.2	0.9.2
Pig	0.14.0	0.14.0	0.14.0	0.14.0	0.14.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7	1.4.7
YARN	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
HDFS	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
Flink	1.11-vvr-2.0.1	1.11-vvr-2.0.1	1.11-vvr-2.0.1	1.10-vvr-1.0.4	1.10-vvr-1.0.4
Druid	0.20.0	0.18.1	0.18.0	0.18.0	0.17.1
HBase	1.4.9	1.4.9	1.4.9	1.4.9	1.4.9
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1	4.14.1

版本	EMR-3.33.0	EMR-3.32.0	EMR-3.30.0	EMR-3.29.0	EMR-3.28.2
Zookeeper	3.5.6	3.5.6	3.5.6	3.5.6	3.5.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0	0.6.0
Presto	338	338	338	338	331
Storm	1.2.2	1.2.2	1.2.2	1.2.2	1.2.2
Impala	2.12.2	2.12.2	2.12.2	2.12.2	2.12.2
Flume	1.9.0	1.9.0	1.9.0	1.9.0	1.9.0
Hue	4.4.0	4.4.0	4.4.0	4.4.0	4.4.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0	5.1.0
Zeppelin	0.9.0	0.8.1	0.8.1	0.8.1	0.8.1
Ranger	1.2.0	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2	3.7.2
SmartData	3.2.0	3.1.0	3.0.0	2.7.301	2.7.2
Kafka	1.1.1	1.1.1	1.1.1	1.1.1	1.1.1
Kafka-Manager	1.3.3.16	1.3.3.16	1.3.3.16	1.3.3.16	1.3.3.16
Superset	0.36.0	0.36.0	0.36.0	0.35.2	0.35.2
Bigboot	3.2.0	3.1.0	3.0.0	2.7.301	2.7.2
OpenLDAP	2.4.44	2.4.44	2.4.44	2.4.44	2.4.44
Kudu	1.10.0	1.10.0	1.10.0	1.10.0	1.10.0
Flink-Vvp	1.11-2.2.2	1.11-2.2.2	1.11-2.2.2	1.11-2.2.2	1.11-2.2.2
PAI-Alink	1.1.0	1.1.0	1.1.0	1.1.0	1.1.0
Alluxio	2.4.0	2.4.0	不涉及	不涉及	不涉及
Hudi	0.6.0	0.6.0	不涉及	不涉及	不涉及

EMR-3.25.1~EMR-3.28.1

版本	EMR-3.28.1	EMR-3.28.0	EMR-3.27.2	EMR-3.26.3	EMR-3.25.1
发布时间	2020.6	2020.6	2020.5	2020.4	2020.1
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
Knox	1.1.0	1.1.0	1.1.0	1.1.0	1.1.0
Spark	2.4.5	2.4.5	2.4.3	2.4.3	2.4.3

版本	EMR-3.28.1	EMR-3.28.0	EMR-3.27.2	EMR-3.26.3	EMR-3.25.1
Hive	2.3.5	2.3.5	2.3.5	2.3.5	2.3.5
Tez	0.9.2	0.9.2	0.9.2	0.9.2	0.9.2
Pig	0.14.0	0.14.0	0.14.0	0.14.0	0.14.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7	1.4.7
YARN	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
HDFS	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
Flink	1.10-vvr-1.0.2	1.10-vvr-1.0.2	1.9.1	1.9.1	1.9.1
Druid	0.18.0	0.18.0	0.17.1	0.16.0	0.16.0
HBase	1.4.9	1.4.9	1.4.9	1.4.9	1.4.9
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1	4.14.1
Zookeeper	3.5.6	3.5.6	3.5.6	3.5.6	3.5.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0	0.6.0
Presto	331	331	331	310	310
Storm	1.2.2	1.2.2	1.2.2	1.2.2	1.2.2
Impala	2.12.2	2.12.2	2.12.2	2.12.2	2.12.2
Flume	1.9.0	1.9.0	1.9.0	1.9.0	1.9.0
Hue	4.4.0	4.4.0	4.4.0	4.4.0	4.4.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0	5.1.0
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1	0.8.1
Ranger	1.2.0	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2	3.7.2
SmartData	2.7.1	2.7.0	2.6.5	2.6.3	2.2.3
Kafka	1.1.1	1.1.1	1.1.1	1.1.1	1.1.1
Kafka-Manager	1.3.3.16	1.3.3.16	1.3.3.16	1.3.3.16	1.3.3.16
Superset	0.35.2	0.35.2	0.35.2	0.28.1	0.28.1
Bigboot	2.7.1	2.7.0	2.6.5	2.6.3	2.2.3
OpenLDAP	2.4.44	2.4.44	2.4.44	2.4.44	2.4.44
Kudu	1.10.0	1.10.0	1.10.0	1.10.0	1.10.0

版本	EMR-3.28.1	EMR-3.28.0	EMR-3.27.2	EMR-3.26.3	EMR-3.25.1
Analytics Zoo	不涉及	不涉及	不涉及	0.5.0	0.5.0
Tensorflow on Spark	不涉及	不涉及	不涉及	1.0.1	1.0.1

EMR-3.22.1~EMR-3.25.0

版本	EMR-3.25.0	EMR-3.24.3	EMR-3.24.0	EMR-3.23.0	EMR-3.22.1
发布时间	2020.1	2019.11	2019.11	2019.9	2019.7
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
Knox	1.1.0	1.1.0	1.1.0	1.1.0	1.1.0
Spark	2.4.3	2.4.3	2.3.5	2.4.3	2.4.3
Hive	2.3.5	2.3.5	2.3.5	2.3.5	2.3.5
Tez	0.9.2	0.9.1	0.9.1	0.9.1	0.9.1
Pig	0.14.0	0.14.0	0.14.0	0.14.0	0.14.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7	1.4.7
YARN	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
HDFS	2.8.5	2.8.5	2.8.5	2.8.5	2.8.5
Flink	1.9.1	1.9.1	1.9.1	1.8.2	1.7.2
Druid	0.16.0	0.16.0	0.16.0	0.15.1	0.14.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9	1.4.9
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1	4.14.1
Zookeeper	3.5.6	3.5.5	3.5.5	3.5.5	3.5.5
Livy	0.6.0	0.6.0	0.6.0	0.6.0	0.6.0
Presto	310	0.228	0.228	0.221	0.221
Storm	1.2.2	1.2.2	1.2.2	1.2.2	1.2.2
Impala	2.12.2	2.12.2	2.12.2	2.12.2	2.12.2
Flume	1.9.0	1.9.0	1.9.0	1.9.0	1.9.0
Hue	4.4.0	4.4.0	4.4.0	4.4.0	4.4.0
Oozie	5.1.1	5.1.0	5.1.0	5.1.0	5.1.0
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1	0.8.1

版本	EMR-3.25.0	EMR-3.24.3	EMR-3.24.0	EMR-3.23.0	EMR-3.22.1
Ranger	1.2.0	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2	3.7.2
Tensorflow on Spark	1.0.1	1.0.1	1.0.1	1.0.0	1.0.0
Analytics Zoo	0.5.0	0.5.0	不涉及	0.5.0	0.5.0
Kafka	1.1.1	1.1.1	1.1.1	1.1.1	1.1.1
Kafka-Manager	1.3.3.16	1.3.3.16	1.3.3.16	1.3.3.16	1.3.3.16
Superset	0.28.1	0.28.1	0.28.1	0.28.1	0.28.1
Bigboot	2.2.3	2.2.2	2.2.1	2.0.2	2.0.1
SmartData	2.2.3	2.2.2	2.2.1	2.0.2	2.0.1
OpenLDAP	2.4.44	2.4.44	2.4.44	2.4.44	2.4.44
Kudu	1.10.0	1.10.0	1.10.0	1.10.0	1.10.0

产品发行版本更新记录（EMR-3.22.x之前版本）

● EMR-3.19.0~EMR-3.21.0

版本	EMR-3.19.0	EMR-3.19.1	EMR-3.20.0	EMR-3.21.0
发布时间	2019.3	2019.4	2019.5	2019.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Knox	1.1.0	1.1.0	1.1.0	1.1.0
ApacheDS	2.0.0	2.0.0	2.0.0	2.0.0
Spark	2.4.1	2.4.1	2.4.2	2.4.3
Hive	3.1.1	3.1.1	3.1.1	3.1.1
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
YARN	2.8.5	2.8.5	2.8.5	2.8.5
HDFS	2.8.5	2.8.5	2.8.5	2.8.5
Flink	1.7.2	1.7.2	1.7.2	1.7.2
Druid	0.13.0	0.13.0	0.13.0	0.14.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9

版本	EMR-3.19.0	EMR-3.19.1	EMR-3.20.0	EMR-3.21.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Zookeeper	3.4.13	3.4.13	3.4.13	3.4.13
Livy	0.6.0	0.6.0	0.6.0	0.6.0
Presto	0.213	0.213	0.213	0.213
Storm	1.2.2	1.2.2	1.2.2	1.2.2
Impala	2.12.2	2.12.2	2.12.2	2.12.2
Flume	1.8.0	1.8.0	1.8.0	1.8.0
Hue	4.1.0	4.1.0	4.1.0	4.4.0
Oozie	4.2.0	4.2.0	5.1.0	5.1.0
Zeppelin	0.8.0	0.8.0	0.8.1	0.8.1
Ranger	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
OS	CentOS 7.4	CentOS 7.4	CentOS 7.4	CentOS 7.4
Tensorflow	不涉及	不涉及	1.8.0	1.8.0
Kafka	2.11-1.1.1	1.1.1	2.11	1.1.1
Superset	0.28.1	0.28.1	0.28.1	0.28.1
Jupyter	不涉及	不涉及	4.4.0	4.4.0
Analytics Zoo	不涉及	不涉及	0.2.0	0.5.0
Bigboot	不涉及	不涉及	1.0.0	1.0.0
OpenLDAP	不涉及	不涉及	不涉及	不涉及
Kudu	不涉及	不涉及	不涉及	不涉及

● EMR-3.15.0~EMR-3.18.1

版本	EMR-3.15.0	EMR-3.16.0	EMR-3.17.0	EMR-3.18.1
发布时间	2018.11	2018.12	2019.1	2019.2
Hadoop	2.7.2	2.7.2-1.3.2	2.7.2	2.8.5
Knox	0.13.0	0.13.0	1.1.0	1.1.0
ApacheDS	2.0.0	2.0.0	2.0.0	2.0.0
Spark	2.3.2	2.3.2-1.0.1	2.3.2	2.3.2

版本	EMR-3.15.0	EMR-3.16.0	EMR-3.17.0	EMR-3.18.1
Hive	2.3.3	2.3.3-1.0.3	2.3.3	3.1.1
Tez	0.9.1	0.9.1-1.0.2	0.9.1	0.9.1
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Sqoop	1.4.7	1.4.7-1.0.0	1.4.7	1.4.7
YARN	2.7.2	2.7.2	2.7.2	2.8.5
HDFS	2.7.2	2.7.2	2.7.2	2.8.5
Flink	1.4.0	1.6.2-1.0.0	1.6.2	1.6.2
Druid	0.12.3	0.12.3-1.0.1	0.12.3	0.13.0
HBase	1.1.1	1.1.1-1.0.2	1.1.1	1.4.9
Phoenix	4.10.0	4.10.0-1.0.0	4.10.0	4.14.1
Zookeeper	3.4.13	3.4.13	3.4.13	3.4.13
Livy	0.5.0	0.5.0	0.5.0	0.5.0
Presto	0.208	0.208	0.213	0.213
Storm	1.1.2	1.2.2	1.2.2	1.2.2
Impala	2.10.0	2.10.0-1.0.0	2.12.2	2.12.2
Flume	不涉及	1.8.0	1.8.0	1.8.0
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
Ranger	1.0.0	1.0.0	1.0.0	1.0.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
OS	CentOS 7.4	CentOS 7.4	CentOS 7.4	CentOS 7.4
Tensorflow	1.8.0	1.8.0	1.8.0	不涉及
Kafka	2.11-1.0.1	2.11-1.1.0	2.11-1.1.1	2.11-1.1.1
Superset	0.27.0	0.28.1	0.28.1	0.28.1
Jupyter	4.4.0	4.4.0	4.4.0	不涉及
Analytics Zoo	0.2.0	0.2.0	0.2.0	不涉及

● EMR3.11.0~EMR-3.14.0

版本	EMR-3.11.0	EMR-3.12.0	EMR-3.13.0	EMR-3.14.0
发布时间	2018.6	2018.7	2018.8	2018.10
Hadoop	2.7.2-emr-1.2.14	2.7.2-emr-1.2.14	2.7.2	2.7.2
Knox	0.13.0	0.13.0	0.13.0	0.13.0
ApacheDS	2.0.0	2.0.0	2.0.0	2.0.0
Spark	2.2.1	2.3.1	2.3.1	2.3.1
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Sqoop	1.4.6	1.4.7	1.4.7	1.4.7
YARN	2.7.2	2.7.2	2.7.2	2.7.2
HDFS	2.7.2	2.7.2	2.7.2	2.7.2
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Druid	0.11.0	0.12.0	0.12.2	0.12.3
HBase	1.1.1	1.1.1	1.1.1	1.1.1
Phoenix	4.10.0	4.10.0	4.10.0	4.10.0
Zookeeper	3.4.11	3.4.12	3.4.12	3.4.13
Presto	0.188	0.188	0.208	0.208
Storm	1.1.2	1.1.2	1.1.2	1.1.2
Impala	2.10.0	2.10.0	2.10.0	2.10.0
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Zeppelin	0.7.3	0.7.3	0.8.0	0.8.0
Ranger	0.7.3	1.0.0	1.0.0	1.0.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
OS	CentOS 7.4	CentOS 7.4	CentOS 7.4	CentOS 7.4
Tensorflow	不涉及	不涉及	1.8.0	1.8.0
Kafka	不涉及	不涉及	2.11-1.0.1	2.11-1.0.1
Superset	不涉及	不涉及	0.25.6	0.25.6

版本	EMR-3.11.0	EMR-3.12.0	EMR-3.13.0	EMR-3.14.0
Jupyter	不涉及	不涉及	不涉及	不涉及
Analytics Zoo	不涉及	不涉及	不涉及	不涉及

● EMR-3.7.1~EMR-3.10.1

版本	EMR-3.7.1	EMR-3.8.1	EMR-3.9.1	EMR-3.10.1
发布时间	2018.1	2018.1	2018.2	2018.4
Hadoop	2.7.2-emr-1.2.10	2.7.2-emr-1.2.12	2.7.2-emr-1.2.13	2.7.2-emr-1.2.14
Knox	0.13.0	0.13.0	0.13.0	0.13.0
ApacheDS	2.0.0	2.0.0	2.0.0	2.0.0
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Tez	0.8.4	0.8.4	0.8.4	0.9.1
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
YARN	2.7.2	2.7.2	2.7.2	2.7.2
HDFS	2.7.2	2.7.2	2.7.2	2.7.2
Flink	不涉及	1.4.0	1.4.0	1.4.0
Druid	不涉及	不涉及	0.11.0	0.11.0
HBase	1.1.1	1.1.1	1.1.1	1.1.1
Phoenix	4.10.0	4.10.0	4.10.0	4.10.0
Zookeeper	3.4.11	3.4.11	3.4.11	3.4.11
Presto	0.188	0.188	0.188	0.188
Storm	1.0.1	1.0.1	1.0.1	1.1.2
Impala	2.10.0	2.10.0	2.10.0	2.10.0
Hue	3.12.0	3.12.0	3.12.0	4.1.0
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
OS	CentOS 7.4	CentOS 7.4	CentOS 7.4	CentOS 7.4

● EMR-2.9.2~EMR-2.11.0

版本	EMR-2.9.2	EMR-2.10.0	EMR-2.11.0
发布时间	2018.2	2018.4	2018.7
Hadoop	2.7.2-emr-1.2.12	2.7.2-emr-1.2.12	2.7.2-emr-1.2.12
Spark	1.6.3	1.6.3	1.6.3
Hive	2.3.2	2.3.2	2.3.3
Tez	0.8.4	0.9.1	0.9.1
Pig	0.14.0	0.14.0	0.14.0
Sqoop	1.4.6	1.4.6	1.4.6
Hue	3.12.0	4.1.0	4.1.0
Zeppelin	0.7.1	0.7.1	0.7.3
HBase	1.1.1	1.1.1	1.1.1
Phoenix	4.10.0	4.10.0	4.10.0
Storm	1.0.1	1.1.2	1.1.2
Presto	0.188	0.188	0.188.0
Impala	2.10.0	2.10.0	2.10.0
Zookeeper	3.4.6	3.4.11	3.4.11
Oozie	4.2.0	4.2.0	4.2.0
Ranger	0.7.1	0.7.1	0.7.3
Ganglia	3.7.2	3.7.2	3.7.2
OS	CentOS 7.4	CentOS 7.4	CentOS 7.4

● EMR-1.0.0~EMR-1.3.0

版本	EMR-1.0.0	EMR-1.1.0	EMR-1.2.0	EMR-1.3.0
发布时间	2015.11	2016.3	2016.4	2016.5
Hadoop	2.6.0	2.6.0	2.6.0	2.6.0-emr-1.1.1
Spark	1.4.1	1.6.0	1.6.1	1.6.1
Hive	1.0.1	1.0.1	2.0.0	2.0.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Sqoop	不涉及	不涉及	不涉及	1.4.6

版本	EMR-1.0.0	EMR-1.1.0	EMR-1.2.0	EMR-1.3.0
Hue	不涉及	不涉及	不涉及	3.9.0
Zeppelin	不涉及	不涉及	不涉及	0.5.6
HBase	不涉及	不涉及	1.1.1	1.1.1
Phoenix	不涉及	不涉及	不涉及	不涉及
Zookeeper	不涉及	不涉及	3.4.6	3.4.6
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

 说明 Hadoop版本说明：

基于开源Hadoop的版本，在完全不改变原有接口的基础上，加入了emr-core组件，深度支持阿里云的OSS。这个组件的版本会跟在Hadoop版本后。

8.2. 版本说明（EMR-5.x系列）

8.2.1. EMR-5.7.x版本说明

本文介绍EMR-5.7.x版本的发布日期、更新内容和发行版本信息。

发布日期

EMR-5.7.0 2022年6月15日

更新内容

服务	变更点
JindoData	升级至4.3.1版本。
JindoSDK	升级至4.3.1版本。
StarRocks	- 升级至2.2.0版本。 - 支持资源隔离，详情请参见资源隔离。
Spark	- 支持选择Spark2或Spark3版本。 - 支持Spark ThriftServer和History监控指标。
Presto	- 修复了Presto集群无法查询OSS的问题。 - 支持HA。
ZooKeeper	- 修复了服务启动时未能正确加载环境变量的问题。 - 修复了Heap size无法配置的问题。
Kyuubi	- 新增组件，版本为1.5.1。 - 支持Spark3 SQL。

发行版本信息

Data Lake集群

OLAP集群

8.2.2. EMR-5.6.x版本说明

本文介绍EMR-5.6.x版本的发布日期、更新内容和发行版本信息。

发布日期

EMR-5.6.0 2022年4月21日

更新内容

服务	变更点
JindoData	新增组件，版本为4.3.0，详情请参见 JindoData 4.3.0版本简介 。
JindoSDK	升级至4.3.0版本。
Spark	升级至3.2.1版本。
Hive	修复了TEZ开启Speculation后重复Commit的缺陷。
Presto	修复了Hadoop集群初始化完成后，添加Presto服务后，Presto服务无法启动的缺陷。
DeltaLake	DML支持SubQuery。
Hudi	升级至0.10.1版本。
Iceberg	升级至0.13.1版本。
YARN	增加了限制AM只运行在CORE组节点的功能配置。
HBase	修复了HBase2.3.4内核的缺陷。
Zookeeper	优化了JVM的参数配置。
Impala	适配JindoSDK 4.3.0。
Sqoop	升级PostgreSQL版本。
Zeppelin	解决了JDBC Interpreter启动失败的问题。
Ranger	Ranger 1.2.0版本Spark Plugin支持Delta和Hudi。
Flume	适配JindoSDK 4.3.0。
Oozie	升级Log4j至2.17.2版本。
DLF-Auth	升级至2.0.0版本。

发行版本信息

② 说明

- OLAP集群需要在[新版控制台](#)查看。
- Shuffle Service集群需要在[旧版控制台](#)查看。

Hadoop集群

Shuffle Service集群

OLAP集群

8.2.3. EMR-5.5.x版本说明

本文介绍EMR-5.5.x版本的发布日期、更新内容和发行版本信息。

发布日期

版本	日期
EMR-5.5.1	2022年3月25日
EMR-5.5.0	2022年2月15日

更新内容

EMR-5.5.1

② 说明 仅[新版控制台](#)的OLAP集群支持该版本。

服务	变更点
Clickhouse	修改了部分参数的默认值。
StarRocks	升级至2.1.1版本，详情请参见 StarRocks概述 。

EMR-5.5.0

服务
SmartData
BIGBOOT

服务
RSS

服务	
JindoSDK	

服务

服务

服务

服务
Spark

服务

服务

服务

服务	
Hive	

服务

服务

服务

服务

服务
Delta Lake

服务

服务

服务

服务
Hudi

服务
HDFS
YARN
Flume

服务
Impala
Ranger

服务
HBase

服务
Clickhouse

服务

服务
Iceberg

服务
DLF-Auth

发行版本信息

 说明

OLAP集群需要在[新版控制台](#)查看。

- Hadoop集群
- ClickHouse集群
- Shuffle Service集群
- OLAP集群

8.2.4. EMR-5.4.x版本说明

本文介绍EMR-5.4.x版本的发布日期、更新内容和发行版本信息。

发布日期

版本	日期
EMR-5.4.3	2021年12月
EMR-5.4.2	2021年12月

版本	日期
EMR-5.4.1	2021年11月
EMR-5.4.0	2021年10月

更新内容

• EMR-5.4.3

修复了所有相关组件的Log4j安全漏洞，详细信息请参见[漏洞公告 | Apache Log4j2远程代码执行漏洞](#)。

服务	变更点
Presto	修复了Elasticsearch连接器的Log4j漏洞。
DLF Metastore	- 将Metastore log默认开启改为关闭。 - 修复了Metastore gettablestats URI超长出错的问题。
Delta Lake	修复了变更Schema同步到Metastore的问题。
Sqoop	修复了Sqoop导入HCatalog表时，Decimal类型精度丢失的问题。

• EMR-5.4.2

服务	变更点
SmartData	- 升级SmartData至3.8.0版本，详情请参见SmartData 3.8.x版本简介。 - 支持OSS基于Kerberos和Ranger的认证和授权管理。

• EMR-5.4.1

服务	变更点
SmartData	升级SmartData至3.7.3版本，详情请参见 SmartData 3.7.x版本简介 。
Oozie	修复了HA情况下Oozie的Jetty Server依赖JAR包冲突，导致Jetty Server无法启动的问题。
Impala	修复了查询DLF元数据表时提示 <code>no such method error</code> 的问题。
DLF-Auth	升级DLF-Auth至1.0.1版本。

• EMR-5.4.0

服务	变更点
SmartData	升级SmartData至3.7.2版本，详情请参见 SmartData 3.7.x版本简介 。

服务	变更点
Spark	- 升级Spark至3.1.2版本。 - Spark 3.x系列，SparkSQL优化了Distinct计算性能，即当聚合算子中包含多个count(distinct case ... when ...)时会触发优化功能。 - 修复了AQE在Stats缺失情况下数组越界的问题。 - 修复了AQE和Cache在特定场景下报错的问题。
Hive	Hive on JindoFS (Block) 支持批量元数据优化功能，默认未开启。
Presto	Delta表支持StorageHandler查询。
DeltaLake	- 升级DeltaLake至1.0.0版本。 - 统一Hive 2和Hive 3的delta-connectors。 - 修复了delta-connectors查询多级分区表时的报错问题。 - 支持多个功能的SQL语法。例如，DataSkipping、Optimize和Zorder。 - 支持同步Meta到MetaStore。
Hudi	- 升级Hudi至0.9.0版本。 - 修复了DeltaLake和Hudi的sql.extension的兼容性问题。  说明 支持Spark 3.1.2版本。
HDFS	NameNode预留空间默认参数自适应增加，保证磁盘空间不足时，NameNode及时进入SafeMode。
Storm	组件下线。
Zeppelin	升级Zeppelin至社区0.10.0版本。
Hue	- 修复YARN Job Browser在部分情况下无法正常展示和终止作业的问题。 - 默认配置中放开YARN Job Browser。 - 默认配置中支持Presto协议。
Druid	修复了服务器掉电时Pid文件残留，导致节点重启失败的问题。
ClickHouse	- 更新了默认配置。 - 支持集群扩容，详情请参见扩容ClickHouse集群。 - 支持MetaChecker功能。 - 支持OSS表引擎和OSS表函数读数据。
Iceberg	- 升级Iceberg至0.12.0-1.0.1版本。 - 修改了Hive Runtime依赖报错问题。
Knox	修复了Spark Task第一次访问时失败的问题。

服务	变更点
DLF-Auth	新增组件。 支持DLF权限，可以控制Hive和Spark权限。版本为1.0.0，详情请参见 DLF-Auth 。

发行版本信息

Hadoop集群

Kafka集群

ClickHouse集群

8.2.5. EMR-5.3.x版本说明

本文介绍EMR-5.3.x版本的发布日期、更新内容和发行版本信息。

发布日期

版本	日期
EMR-5.3.1	2021年9月
EMR-5.3.0	2021年8月

更新内容

• EMR-5.3.1

服务	变更点
SmartData	升级SmartData至3.7.1版本。
Hue	修复了高安全集群无法使用Impala的问题。
Kudu	支持Kerberos。
HBase	- 修复了高安全集群重启HBase耗时过长的问题。 - 修复了Spark 3.1.1版本与HBase集成使用时失败的问题。 - 优化了Graceful Stop流程。

• EMR-5.3.0

服务	变更点
SmartData	升级SmartData至3.7.0版本。
Spark	修复了与Delta Lake兼容性的问题。
Hive	Hive on JindoFS（Block模式）支持批量元数据优化功能。 默认未开启此功能。
DeltaLake	- 支持DeltaLake的分区特性。 - 修复了 <code>desc detail</code> 命令和Spark 3.1.1版本不兼容的问题。

服务	变更点
YARN	- 节点Containers REST API增加了appld, CPU和Memory资源使用信息。 - 修复了弹性伸缩释放节点上AM日志无法查看的问题。 - 修复了State Store历史数据造成集群不可用的问题。 - 支持弹性伸缩Decommission后清理释放的节点。 - 完善了弹性伸缩Graceful Decommission的操作逻辑, 待NM进程结束后再标记下线完成。
Zookeeper	升级至社区3.6.3版本。
Flink	- 新增了SmartData组件。 - 修复了SSH到DataFlow-Flink集群中提交作业时, 无法免密访问OSS的问题。
Impala	修复了直接删除OSS分区目录出现list目录循环的问题。
Hue	修复了Hue与Oozie集成使用时界面显示的问题。
Kudu	升级至社区1.14.0版本。
Clickhouse	组件版本为21.3.13.9。
Iceberg	新增Iceberg组件, 组件版本为0.12.0。

发行版本信息

Hadoop集群

ClickHouse集群

Shuffle Service集群

Kafka集群

8.2.6. EMR-5.2.x版本说明

EMR-5.x系列的第一个稳定版本。本文介绍EMR-5.2.x版本的发布日期、更新内容和发行版本信息。

发布日期

EMR-5.2.1 2021年7月16日

更新内容

服务	变更点
SmartData	升级SmartData至3.6.1版本。 版本详情, 请参见 SmartData 3.6.x版本简介 。
Hive	- 修复使用DLF元数据执行 <code>show create table</code> 命令, 结果显示不正确的问题。 - 优化Hive默认参数, 以提升作业性能。 - 修改E-MapReduce控制台上, Hive服务配置页面的hive-env页签的配置项名称为大写, 便于用户使用。 - 修复UDF (User Define Function) 导致HiveServer2内存泄露的问题。 - 优化文件系统与MetaStore不一致时写Hive表的报错信息。

服务	变更点
HDFS	支持ZSTD（Zstandard）压缩格式。
Delta Lake	- 升级Delta Lake至0.8.0版本。 - 支持Spark 3。
Flink	升级Flink至1.12-vvr-3.0.2版本。
Hudi	- 升级Hudi至0.8.0版本。 - 支持和Spark SQL集成。
Spark	 注意 EMR-5.2.1版本的Spark（3.1.1）与Kudu（1.11.1）不兼容。 - 支持数据湖格式Delta Lake和Hudi。 - 支持Remote Shuffle Service。 - 支持Livy。 - 优化E-MapReduce控制台上，Spark服务配置页面的spark-defaults页签的配置项名称。 - 优化CBO（Cost-Based Optimization）、DPP（DynamicPartitionPruning）以及Z-Order等功能，性能比开源Spark 3版本提升50%。 - 支持阿里云Log Service、DataHub和消息队列RocketMQ版（简称ONS）等数据源。
Tez	优化Tez默认参数，以提升作业性能。
Ranger	- 修复Ranger启用Spark日志中的Warning报错。 - 修复对接LDAP后，自动同步用户失败的问题。
Knox	- 适配Kudu组件。 - 适配Hbase组件。
Kafka	- 支持Cruise Control组件，提供Kafka集群Balance功能。 - Kafka上线热换盘功能，在不启停Broker前提下，可以更换坏盘。 - 修改部分参数的默认值。
Hue	修复HUE查询历史记录时，中文乱码的问题。
Phoenix	修复Hive和Spark SQL访问Phoenix表时，报未找到JDBC Driver的问题。
ESS（EMR Remote Shuffle Service）	支持Spark 3。

发行版本信息

Hadoop集群

Shuffle Service集群

Kafka集群

8.3. 版本说明（EMR-4.x系列）

8.3.1. EMR-4.10.x版本说明

本文介绍EMR-4.10.x发行版本的发布日期和更新内容信息。

发布日期

EMR-4.10.0 2022年3月23日

更新内容

服务	变更点
SmartData	组件下线。
BIGBOOT	
JindoSDK	- 升级架构为JindoData。 - EMR首次集成JindoSDK 4.0，支持OSS和OSS-HDFS服务等，详情请参见JindoData 4.0.0版本简介。
Spark	- 升级至2.4.8版本。 - 修复了Adaptive Execution部分场景无法生效的问题。 - 修复了统计聚合函数行为和Hive不一致的问题。 - 修复了读取Hive ORC表 <code>char</code> 类型数据正确性问题。 - 优化了Thriftserver的默认配置。 - 优化E-MapReduce控制台上，Spark服务配置页面的<code>spark-defaults</code>页签的配置项名称。 - 优化了Hive on Spark。 - 修复了AQE在Stats缺失情况下数组越界的问题。 - 修复了AQE和Cache在特定场景下报错的问题。 - 移除了无效配置Log4j MetricsAppender。 - 修复了SparkContext启动过程中空指针异常的问题。 - 支持ZSTD（Zstandard）压缩格式。
Hive	- 修复了UDF导致HiveServer2内存泄露的问题。 - 修复了使用DLF（DataLakeFormation）元数据执行show create table命令时，结果显示不正确的问题。 - 优化Hive默认参数，以提升作业性能。 - 修改了E-MapReduce控制台上，Hive服务配置页面的<code>hive-env</code>页签的配置项名称为大写，便于用户使用。 - 优化文件系统与MetaStore不一致时写Hive表的报错信息。 - Hive on JindoFS（Block）支持批量元数据优化功能，默认未开启。
Ranger	- 修复了Ranger启用Spark日志中的Warning报错。 - 修复了对接LDAP后，自动同步用户失败的问题。

服务	变更点
HDFS	- 支持ZSTD（Zstandard）压缩格式。 - NameNode预留空间默认参数自适应增加，保证磁盘空间不足时，NameNode及时进入SafeMode。
YARN	- 节点Containers REST API增加了appld，CPU和Memory资源使用信息。 - 修复了弹性伸缩释放节点上AM日志无法查看的问题。 - 修复了State Store历史数据造成集群不可用的问题。 - 支持弹性伸缩Decommission后清理释放的节点。 - 完善了弹性伸缩Graceful Decommission的操作逻辑，待NM（NodeManager）进程结束后再标记下线完成。
Knox	- 适配Kudu组件。 - 适配HBase组件。 - 修复了Spark Task第一次访问时失败的问题。
Tez	优化了Tez默认参数，以提升作业性能。
Sqoop	修复了Sqoop导入HCatalog表时，Decimal类型精度丢失的问题。
Delta Lake	- 元数据管理 - 使用Spark内置Catalog替代Hive CLI API同步元数据及分区信息。 - 自动上报表的统计信息（dataProfiling）到MetaStore。 - SQL - 支持Time Travel语法。 - 支持DropPartition SQL语法。 - 支持指定位置（FIRST和AFTER）的ADD COLUMN操作。 - 表管理能力增强 - 支持并默认开启可以根据表大小动态调整filesize。 - 支持并默认开启自动Vacuum，支持并发Vacuum。 - 优化了自动Compaction的逻辑，默认关闭。 - 新增Zorder语法，并加速了Zorder的处理过程。
Hudi	- 升级至0.10.0版本。 - 修复了DeltaLake和Hudi的sql.extension的兼容性问题。
Iceberg	新增组件。 版本为0.13.0，文档详情请参见Iceberg概述。
Hue	- 修复了Hue查询历史记录时，中文乱码的问题。 - 修复了Hue与Oozie集成使用时界面显示的问题。 - 修复YARN Job Browser在部分情况下无法正常展示和终止作业的问题。 - 默认配置中放开YARN Job Browser。 - 默认配置中支持Presto协议。

服务	变更点
DLF-Auth	新增组件。 版本为1.0.4，文档详情请参见 DLF-Auth 。
HBase	- 修复了高安全集群重启HBase耗时过长的的问题。 - 修复了Spark 3.1.1版本与HBase集成使用时失败的问题。 - 优化了Graceful Stop流程。
Zookeeper	升级至社区3.6.3版本。
Presto	- 升级Presto至社区358版本。 - 支持UDF动态加载功能，详情请参见动态加载UDF。 - 支持数据湖分析。
Impala	- 修复了直接删除OSS分区目录出现list目录循环的问题。 - 修复了查询DLF元数据表时，提示 <code>no such method error</code> 的问题。
Zeppelin	升级Zeppelin至社区0.10.0版本。
Oozie	修复了HA情况下Oozie的Jetty Server依赖JAR包冲突，导致Jetty Server无法启动的问题。

发行版本信息

Hadoop集群

8.3.2. EMR-4.9.x版本说明

本文介绍EMR-4.9.x发行版本的发布日期和更新内容信息。

发布日期

EMR-4.9.0 2021年4月21日

更新内容

服务	变更点
SmartData	升级至3.5.0版本。 版本详情，请参见 SmartData 3.5.x版本简介 。
Spark	- 修复Adaptive Execution部分场景无法生效的问题。 - 修复统计聚合函数行为和Hive不一致的问题。 - 修复读取Hive ORC表char类型数据正确性的问题。
HDFS	支持国密SM4加密算法。
Hue	升级Hue至4.9.0版本。
Alluxio	升级Alluxio至2.5.0版本。
Livy	升级Livy至0.7.1版本。

发行版本信息

Hadoop集群

Zookeeper集群

Shuffle Service集群

Kafka集群

8.3.3. EMR-4.8.x版本说明

本文介绍EMR-4.8.x发行版本的发布日期和更新内容信息。

发布日期

EMR-4.8.0 2021年3月15日

更新内容

服务	变更点
SmartData	升级至3.4.0版本。 版本详情，请参见 SmartData 3.4.x版本简介 。
Spark	- 优化了部分默认配置。 - 性能优化：支持Window TopK下推。 - 增强Hive读写CSV或JSON表的兼容性。 - ANALYZE语句支持省略全表列名。 - 支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见管理LDAP认证。 - 改进Spark Beeline工具的易用性。
Hive	- 优化了部分默认配置。 - 性能优化：增强CBO。 - 支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见管理LDAP认证。
YARN	修复了Hadoop未授权访问Web UI时的高风险问题。即通过SSH Tunnel方式访问YARN Web UI时，需要在URL里显式指定user.name=name的问题。
Tez	优化了默认配置。
Ranger	- 修复Spark中Filter函数的下推问题。 - 修复Ranger Disable Presto后，Presto无法启动的问题。 - 支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见管理LDAP认证。
Hue	支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见 管理LDAP认证 。

服务	变更点
Impala	- 升级Impala至3.4.0版本。 - 升级Shiro至1.7.0版本。 - 支持DLF元数据。 - 支持查询Delta格式的数据。 - 支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见管理LDAP认证。 - 修复数据存储在OSS时，使用INSERT OVERWRITE命令的问题。
Hudi	- 支持SQL on Hudi功能。 - 修复了查询部分数据时准确性的问题。 - 在Spark查询Hudi的Copy On Write表时，支持分区裁剪。 - 支持分桶索引机制，提高写入性能。
Delta Lake	- 修复基于已存在Delta表无法同步元数据至Hive MetaStore的问题。 - 修复Merge命令无法解析 <code>*</code> 的问题。 - 修复基于Parquet格式的数据转换成Delta表，且创建表元数据时报错的问题。 - 修复当没有待compact的文件时，执行Optimize命令异常的问题。 - 支持Merge语法使用子查询作为Source命令。 - 使用Presto查询Delta表时，引入缓存机制，以提升查询效率。 - 支持Impala查询Delta表。
ESS	- 修复在Shuffle read阶段，出现closed channel异常、IndexOutOfBounds异常或堆外内存使用过多的问题。 - 修复开启Metrics之后，出现NPE异常的问题。
HAS	修复在HAS install流程产生了错误后admin.keytab无法重新init问题
Presto	支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见 管理LDAP认证 。
HBase	- 升级至2.2.6版本。 - 不再支持Ranger的权限控制。
Sqoop	支持导入Parquet格式的文件至OSS。
Superset	- 修复admin用户无法登录Web UI的问题。 - 数据集兼容Druid集群模式。 - 不再支持SparkSQL数据集。

服务	变更点
Knox	- 支持通过Knox访问Presto。 - 修复无法访问Druid Web UI的问题。 - 取消HTTP仅限制使用高安全Knox访问Ranger UI的限制。

发行版本信息

Hadoop集群

Shuffle Service集群

Kafka集群

8.3.4. EMR-4.6.x版本说明

本文介绍EMR-4.6.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-4.6.0 2021年1月15日

更新内容

服务	变更点
SmartData	升级至3.2.0版本。 详情请参见 SmartData 3.2.x版本简介 。
Spark	- 升级至2.4.7版本。 - 升级jQuery至3.5.1版本。 - 兼容Hive方式自动更新表和分区大小。 - 支持Spark元数据和作业运行信息输出至DataWorks。
Hive	- HCatalog支持Data Lake Formation。 - 支持Hive元数据和作业运行信息输出至DataWorks。
Metastore	- 新增Hive Statistics功能。 - HCatalog支持Data Lake Formation。 - 优化STSToken的获取方式。
HDFS	- 升级jQuery至3.5.1版本。 - 升级至3.2.1版本。
YARN	- 升级至3.2.1版本。 - 升级jQuery至3.5.1版本。 - 调整Fair Scheduler配置。 - 优化了Timeline Server。
Zeppelin	升级至0.9.0版本。

服务	变更点
OpenLDAP	- 增加审计功能。 - 默认开启SSL端口（10636）。 - 支持一键开启Presto。
Hue	支持Presto。
EMRHook	- 新增软件服务。 - hive-hook：支持Hive元数据和作业运行信息输出至DataWorks。 - spark-hook：支持Spark元数据和作业运行信息输出至DataWorks。

发行版本信息

Hadoop集群

Shuffle Service集群

Kafka集群

8.3.5. EMR-4.5.x版本说明

本文介绍EMR-4.5.x发行版本的发布日期和更新内容等信息。

EMR-4.5.1

- 发布日期
2020年12月13日
- 更新内容
 - 该版本修复了Hive和Presto查询分区表的问题。
 - 该版本目前只支持部分地域，包括华东1（杭州）、华东2（上海）和华北2（北京）。

EMR-4.5.0

- 发布日期
2020年12月7日
- 新增内容

服务	变更点
ESS（EMR Remote Shuffle Service）	支持1.0.0版本。 详情请参见 ESS 。
Hudi	支持0.6.0版本。
Delta Lake	支持0.6.1版本。

- 更新内容

服务	变更点
Ranger	- 升级至2.1.0版本。 - 支持Ownership权限。

服务	变更点
Presto	- 升级至338版本。 - 支持数据湖构建（DLF）元数据。
Zeppelin	升级至0.8.2版本。
SmartData	升级至3.1.0版本。 详情请参见 SmartData 3.1.x版本简介 。
Bigboot	升级至3.1.0版本。
Hive	- 支持数据湖构建（DLF）元数据。 - 支持Ranger Ownership权限。
Spark	支持数据湖构建（DLF）元数据。
DLF Metastore	- 修复高安全集群中Presto服务无法启动的问题。 - 增加支持Hive 3和Cache。 - 修复Hive和Presto查询数据的问题。
Impala	支持在EMR控制台自定义配置 <i>catalogd.flgs</i> 、 <i>impalad.flgs</i> 和 <i>statestored.flgs</i> 。
Tez	Tez UI的autoDeploy相关漏洞修复。
OpenLDAP	增加10389端口等待的判断。
Hue	MySQL Backend安全漏洞修复。
Kerberos	- 升级至Apache Kerby 2.0.1。 - 修复外部Kerberos集群的kadmin principal无法自定义的问题。
Sqoop	- 支持Parquet、AVRO和ORC等多种文件格式。 - 支持数据湖构建（DLF）元数据。

发行版本信息

Hadoop集群

Shuffle Service集群

Kafka集群

8.3.6. EMR-4.4.x版本说明

本文介绍EMR-4.4.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-4.4.1 2020年9月15日

更新内容

服务	变更点
YARN	- 删除软件栈 <code>yarn.application.classpath</code> 配置中的 <code>hadoop/tools/lib</code> 目录。 - 优化MR作业默认的参数配置。
Hive	优化默认的参数配置。
Tez	
Ranger	- 支持Impala权限控制。 - 升级jackson-databind版本。
Impala	- 支持集成Ranger。 - 升级Shiro至1.6.0版本。
SmartData	升级至2.7.301版本。
Bigboot	
Knox	- 支持Tez UI独立打开，支持YARN UI中的Tez。 - 升级Shiro至1.6.0版本。
EMRDOCTOR	修复时间配置文件为空时，导致不采集作业信息的问题。
Ganglia	增加HDFS Service RPC Port的端口探测。
Oozie	- 修复Web UI无法打开的问题。 - 升级jackson-databind版本。
Zookeeper	支持绑定内网IP启动服务端口。
Superset	修复启动脚本。
Livy	升级jackson-databind和fastjson版本。
Zepplin	升级jackson-databind和Shiro版本。
HAS	升级jackson-databind和fastjson版本。
Flume	升级fastjson版本。

发行版本信息

Hadoop集群

Zookeeper集群

Kafka集群

8.3.7. EMR-4.3.x版本说明

本文介绍EMR-4.3.x发行版本的发布日期和更新内容信息。

发布日期

EMR-4.3.0 2020年5月20日

更新内容

服务	变更点
Ranger	- 支持HDFS、Hive、Spark plugin自定义部署，在对应服务节点执行plugin enable操作。 - 支持在控制台配置ranger-admin和ranger-usersync。
Presto	升级Kudu Client。
Spark	- 升级至2.4.5版本。 - 升级关联的Delta Lake至0.6.0版本。 - 修复开启Ranger Hive后，Pyspark无法正常运行的缺陷。
HDFS	- 修复HDFS_NAMENODE_OPTS参数无法生效的缺陷。 - 支持自定义部署。
YARN	支持自定义部署。
Hive	支持自定义部署。
Knox	适配Hadoop 3.x中HDFS的NameNode UI。
Zeppelin	修复生成zeppelin.keytab时失败的缺陷。
Kafka	升级至2.4.1版本。
Kudu	升级至1.11.1版本。
Impala	修复haproxy问题。
Livy	修复xmllint问题。
HUE	- 支持Gateway安装HUE组件。 - 支持单个节点开启多个HUE实例。

发行版本信息

Hadoop集群

Zookeeper集群

Kafka集群

8.4. 版本说明（EMR-3.x系列）

8.4.1. EMR-3.41.x版本说明

本文介绍EMR-3.41.x版本的发布日期、更新内容和发行版本信息。

发布日期

EMR-3.41.0 2022年6月15日

更新内容

服务	变更点
JindoData	升级至4.3.1版本。
JindoSDK	升级至4.3.1版本。
Presto	- 修复了Presto集群无法查询OSS的问题。 - 支持HA。
Spark	- 支持选择Spark2或Spark3版本。 - 支持Spark ThriftServer和History监控指标。
Kafka	增加了Kafka Connect组件服务。
StarRocks	升级至2.1.6版本。
ZooKeeper	- 修复了服务启动时未能正确加载环境变量的问题。 - 修复了Heap size无法配置的问题。
Kyuubi	- 新增组件，版本为1.5.1。 - 支持Spark3 SQL。

发行版本信息

Data Lake集群

OLAP集群

DataFlow集群

8.4.2. EMR-3.40.x版本说明

本文介绍EMR-3.40.x版本的发布日期、更新内容和发行版本信息。

发布日期

EMR-3.40.0 2022年4月21日

更新内容

服务	变更点
JindoData	新增组件，版本为4.3.0，详情请参见 JindoData 4.3.0版本简介 。
JindoSDK	升级至4.3.0版本。
Spark	升级至3.2.1版本。
Hive	- 修复了TEZ开启Speculation后重复Commit的缺陷。 - 修复了必须Reload Function才能调用UDF的缺陷。
Presto	修复了Hadoop集群初始化完成后，添加Presto服务后，Presto服务无法启动的缺陷。
DeltaLake	修复了与Streaming SQL的兼容性问题。
Hudi	升级至0.10.1版本。

服务	变更点
Iceberg	升级至0.13.1版本。
YARN	- 增加了限制AM只运行在CORE组节点的功能配置。 - 修复了mareduce.map.java.opts配置缺少taihaodocor的问题。
Zookeeper	优化了JVM的参数配置。
Flink	适配JindoSDK 4.3.0。
Impala	
Flume	
Druid	
Sqoop	升级PostgreSQL版本。
Zeppelin	解决了JDBC Interpreter启动失败的问题。
Ranger	Ranger 1.2.0版本Spark Plugin支持Hudi。
Oozie	升级Log4j至2.17.2版本。
HBase	修复了HBase 1.4.9版本RegionServer无法启动的问题。
DLF-Auth	升级至2.0.0版本。

发行版本信息

说明

- OLAP集群和DataFlow集群需要在[新版控制台](#)查看。
- Druid集群和Presto集群需要在[旧版控制台](#)查看。

Hadoop集群

Druid集群

Presto集群

OLAP集群

DataFlow集群

8.4.3. EMR-3.39.x版本说明

本文介绍EMR-3.39.x版本的发布日期、更新内容和发行版本信息。

发布日期

版本	日期
EMR-3.39.2	2022年3月25日
EMR-3.39.1	2022年2月15日

更新内容

EMR-3.39.2

 说明 仅新版控制台的OLAP集群和DataFlow集群支持该版本。

服务	变更点
Flink	- 完善APM监控大盘，新增了部分监控指标。例如，sourceIdleTime。 - 支持云监控报警。
Kafka	- 支持SSL和SASL配置。 - 修改了部分参数的默认值。
Clickhouse	修改了部分参数的默认值。

EMR-3.39.1

服务	变更点
SmartData	组件下线。
BIGBOOT	
RSS	- 升级ESS服务为RSS，文档详情请参见RSS。 - 增强了服务的功能和稳定性。
JindoSDK	- 升级架构为JindoData。 - EMR首次集成JindoSDK 4.0，支持OSS和OSS-HDFS服务等，详情请参见JindoData 4.0.0版本简介。
Spark	- 优化了Hive on Spark。 - 适配JindoSDK。
Tez	适配JindoSDK。
Hive	适配JindoSDK。
Presto	- 支持UDF动态加载功能。 - Delta Lake表支持for ... as of语法的Time Travel查询。 - 增加独立的Delta Lake Catalog，提供默认的Delta连接器配置，并支持基于独立Catalog的ZOrder Data skip优化。 - 修复了Hudi连接器无法查询Hudi MOR表的问题，Hive连接器暂不支持查询Hudi MOR表。 - 适配JindoSDK。

服务	变更点
Delta Lake	- 元数据管理 - 使用Spark内置Catalog替代Hive CLI API同步元数据及分区信息。 - 自动上报表的统计信息（dataProfiling）到MetaStore。 - SQL - 支持Time Travel语法。 - 支持DropPartition SQL语法。 - 支持指定位置（FIRST和AFTER）的ADD COLUMN操作。 - 表管理能力增强 - 支持并默认开启可以根据表大小动态调整filesize。 - 支持并默认开启自动Vacuum，支持并发Vacuum。 - 优化了自动Compaction的逻辑，默认关闭。 - 新增Zorder语法，并加速了Zorder的处理过程。
Hudi	升级至0.10.0版本。
HDFS	适配JindoSDK。
YARN	适配JindoSDK。
Flume	适配JindoSDK。
Flink	- 默认将Flink的<code>/lib</code>目录上传到HDFS集群，方便用户通过<code>yarn.provided.lib.dirs</code>参数使用。 - 适配JindoSDK。
Impala	适配JindoSDK。
Ranger	- 修复Spark History Server启动失败的问题。 - 适配JindoSDK。
HBase	- 修复了默认参数存在的问题。 - 修复了GC日志日期格式问题。 - 修复了RS使用IP时存在的重启问题。
Druid	适配JindoSDK。
Clickhouse	- 新增事务功能，详情请参见事务使用。 - 优化了ClickHouse组件Stop时的处理逻辑。
Iceberg	- 升级至0.13.0版本。 - 隐藏默认配置项，提升用户使用体验。
DLF-Auth	修复了Spark History Server启动失败的问题。
StarRocks	新版控制台新增服务。 2.0.1版本上线，详情请参见 StarRocks概述 。

发行版本信息

 说明 OLAP集群需要在[新版控制台](#)查看。

Hadoop集群

Druid集群

Dataflow集群

ClickHouse集群

Presto集群

OLAP集群

8.4.4. EMR-3.38.x版本说明

本文介绍EMR-3.38.x版本的发布日期、更新内容和发行版本信息。

发布日期

版本	日期
EMR-3.38.3	2021年12月
EMR-3.38.2	2021年12月
EMR-3.38.1	2021年11月
EMR-3.38.0	2021年10月

更新内容

• EMR-3.38.3

修复了所有相关组件的Log4j安全漏洞，详细信息请参见[漏洞公告 | Apache Log4j2远程代码执行漏洞](#)。

服务	变更点
Presto	- 修复了高可用集群Presto查询Hudi表报错的问题。 - 修复了Elasticsearch连接器的Log4j漏洞。
DLF Metastore	- 将Metastore log默认开启改为关闭。 - 修复了Metastore gettablestats URI超长出错的问题。
Delta Lake	修复了变更Schema同步到Metastore的问题。
Flink	- 升级VVR到4.0.11版本，该版本支持了如下功能： - 发布Flink CDC商业功能： - 支持Schema Evolution。 - 支持整库同步的Flink SQL语义。 - 支持Gemini Statebackend将State存储到OSS上。 - 提供了企业版Hudi Connector，内置DLF作为元数据管理。
Sqoop	修复了Sqoop导入HCatalog表时，Decimal类型精度丢失的问题。

• EMR-3.38.2

服务	变更点
----	-----

服务	变更点
SmartData	- 升级SmartData至3.8.0版本, 详情请参见SmartData 3.8.x版本简介。 - 支持OSS基于Kerberos和Ranger的认证和授权管理。

● EMR-3.38.1

服务	变更点
SmartData	升级SmartData至3.7.3版本, 详情请参见 SmartData 3.7.x版本简介 。
Spark	- 移除了无效配置Log4j MetricsAppender。 - 修复了SparkContext启动过程中空指针异常的问题。
Presto	- 修复了Hadoop高可用集群, Presto需要配置host才能查询Hive表的问题。 - 修复了内存较小时, 默认配置下Presto无法启动的问题。 - 修复了修改<code>worker-jvm</code>配置无法生效的问题。 - 支持Ranger。
Impala	修复了查询DLF元数据表时, 提示 <code>no such method error</code> 的问题。
Ranger	- 支持Presto。 - 修复了Ranger Spark insert ORC和PARQUET表的权限问题。 - 修复了Ranger Hive role权限在开启Kerberos后无法生效的问题。
DLF-Auth	- 升级DLF-Auth至1.0.1版本。 - 支持DLF权限, 可以控制Presto权限。 - 修复了RAM用户缓存的问题。

● EMR-3.38.0

服务	变更点
SmartData	升级SmartData至3.7.2版本, 详情请参见 SmartData 3.7.x版本简介 。
Spark	- 升级Spark至2.4.8版本。 - 同时支持Spark 2.4.8和Spark 3.1.2。  说明 Spark3暂不支持Delta和Remote Shuffle Service。 - Spark 3.x系列, SparkSQL优化了Distinct计算性能, 即当聚合算子中包含多个count(distinct case ... when ...)时会触发优化功能。 - 修复了AQE在Stats缺失情况下数组越界的问题。 - 修复了AQE和Cache在特定场景下报错的问题。
Hive	升级Hive至2.3.9版本。

服务	变更点
Presto	- 发布成独立的Presto集群。 - 升级Presto至社区358版本。  注意 此版本暂不支持Ranger。 - 默认支持Hudi和MySQL等连接器，并更新了默认配置，详情请参见配置连接器。 - Presto集群支持弹性伸缩。  说明 如果需要弹性伸缩功能，可以提交工单处理。 - 支持数据湖分析。
DeltaLake	- 统一Hive 2和Hive 3的delta-connectors。 - 修复了delta-connectors查询多级分区表时的报错问题。
Hudi	- 升级Hudi至0.9.0版本。 - 修复了DeltaLake和Hudi的sql.extension的兼容性问题。
HDFS	NameNode预留空间默认参数自适应增加，保证磁盘空间不足时，NameNode及时进入SafeMode。
Flink	- 升级Flink至1.13-vvr-4.0.10版本，对应社区Flink 1.13.1。 - 添加了商业化的Flink Connector。例如，Hologres。 - 添加了相应的Metric Reporter，对接了APM大盘的监控。 - 针对Kafka Connector，添加了基于SchemaRegistry的Kafka Catalog，支持直接读写已存在的Kafka Topic，无需DDL。
Storm	组件下线。
Zeppelin	升级Zeppelin至社区0.10.0版本。
Ranger	当Presto为社区358版本时，该版本Ranger暂不支持Presto权限控制。
Hue	- 修复YARN Job Browser在部分情况下无法正常展示和终止作业的问题。 - 默认配置中放开YARN Job Browser。 - 默认配置中支持Presto协议。
Druid	修复了服务器掉电时Pid文件残留，导致节点重启失败的问题。
ClickHouse	- 更新了默认配置。 - 支持集群扩容，详情请参见扩容ClickHouse集群。 - 支持MetaChecker功能。 - 支持OSS表引擎和OSS表函数读数据。 - 支持表级别自定义ZooKeeper地址。

服务	变更点
Iceberg	新增组件。 版本为0.12.0-1.0.1，详情请参见 Iceberg概述 。
Knox	修复了Spark Task第一次访问时失败的问题。
DLF-Auth	新增组件。 支持DLF权限，可以控制Hive和Spark权限。版本为1.0.0，详情请参见 DLF-Auth 。
ESS	升级ESS至1.2.0版本。

发行版本信息

Hadoop集群

Druid集群

Dataflow集群

ClickHouse集群

Presto集群

8.4.5. EMR-3.37.x版本说明

本文介绍EMR-3.37.x版本的发布日期、更新内容和发行版本信息。

发布日期

版本	日期
EMR-3.37.1	2021年9月
EMR-3.37.0	2021年8月

更新内容

● EMR-3.37.1

服务	变更点
SmartData	升级SmartData至3.7.1版本。
Hue	修复了高安全集群无法使用Impala的问题。
Kudu	支持Kerberos。

● EMR-3.37.0

服务	变更点
SmartData	升级SmartData至3.7.0版本。
Spark	修复了与Delta Lake兼容性的问题。
DeltaLake	- 升级Delta-Connectors，支持使用StorageHandler语法建表并查询。 - 修复了INSERT OVERWRITE分区表时的问题。 - 修复了G-SCD场景下，Optimize将虚拟字段写入文件时的问题。

服务	变更点
YARN	- 节点Containers REST API增加了appId, CPU和Memory资源使用信息。 - 修复了弹性伸缩释放节点上AM日志无法查看的问题。 - 支持弹性伸缩Decommission后清理释放的节点。 - 完善了弹性伸缩Graceful Decommission的操作逻辑, 待NM进程结束后再标记下线完成。
Zookeeper	升级至社区3.6.3版本。
Flink	- 新增了SmartData组件。 - 修复了SSH到DataFlow-Flink集群中提交作业时, 无法免密访问OSS的问题。
Impala	修复了直接删除OSS分区目录出现list目录循环的问题。
Hue	修复了Hue与Oozie集成使用时界面显示的问题。
Kudu	升级至社区1.14.0版本。
Clickhouse	更新了默认配置。

发行版本信息

Hadoop集群

Druid集群

Flink集群

Kafka集群

ClickHouse集群

8.4.6. EMR-3.36.x版本说明

本文介绍EMR-3.36.x版本的发布日期、更新内容和发行版本信息。

发布日期

EMR-3.36.1 2021年7月16日

更新内容

服务	变更点
SmartData	升级SmartData至3.6.1版本。 版本详情, 请参见 SmartData 3.6.x版本简介 。
Hive	- 升级Hive至2.3.8版本。 - 修复使用DLF (DataLakeFormation) 元数据执行 <code>show create table</code> 命令时, 结果显示不正确的问题。 - 优化Hive默认参数, 以提升作业性能。 - 修改E-MapReduce控制台上, Hive服务配置页面的hive-env页签的配置项名称为大写, 便于用户使用。 - 优化文件系统与MetaStore不一致时写Hive表的报错信息。
HDFS	支持ZSTD (Zstandard) 压缩格式。

服务	变更点
Flink	升级Flink至1.12-vwr-3.0.2版本。 ❓ 说明 Hadoop集群中的Flink已移除。
Hudi	- 升级Hudi至0.8.0版本。 - 支持和Spark SQL集成。
Spark	- 优化E-MapReduce控制台上，Spark服务配置页面的spark-defaults页签的配置项名称。 - 优化输出日志性能。 - 支持ZSTD（Zstandard）压缩格式。
Impala	修复使用HDFS时报错提示Core Dump的问题。
Tez	优化Tez默认参数，以提升作业性能。
Knox	- 适配Kudu组件。 - 适配Impala组件。 - 适配Hbase组件。
Hue	修复HUE查询历史记录时，中文乱码的问题。
Phoenix	修复Hive和Spark SQL访问Phoenix表时，报未找到JDBC Driver的问题。
ClickHouse	上线APM监控报警。

发行版本信息

Hadoop集群	Druid集群	Flink集群	Kafka集群	ClickHouse集群
服务	版本			
HDFS	2.8.5			
YARN	2.8.5			
Hive	2.3.8			
Spark	2.4.7			
Knox	1.1.0			
Tez	0.9.2			
Ganglia	3.7.2			
Sqoop	1.4.7			
SmartData	3.6.1			

服务	版本
Bigboot	3.6.1
Hudi	0.8.0
OpenLDAP	2.4.44
Hue	4.9.0
HBase	1.4.9
Zookeeper	3.6.2
Presto	338
Impala	3.4.0
Zeppelin	0.9.0
Flume	1.9.0
Livy	0.7.1
Superset	0.36.0
Ranger	1.2.0
Storm	1.2.2
Phoenix	4.14.1
ESS	1.0.0
Alluxio	2.5.0
Kudu	1.10.0
Oozie	5.2.1

8.4.7. EMR-3.35.x版本说明

本文介绍EMR-3.35.x版本的发布日期和更新内容信息。

发布日期

EMR-3.35.0 2021年4月21日

新增内容

服务	变更点
ClickHouse	- 支持从Spark导入数据至ClickHouse，详情请参见从Spark导入数据至ClickHouse。 - 支持从Flink导入数据至ClickHouse，详情请参见从Flink导入数据至ClickHouse。

更新内容

服务	变更点
SmartData	升级至3.5.0版本。 版本详情，请参见 SmartData 3.5.x版本简介 。
Spark	- 修复Adaptive Execution部分场景无法生效的问题。 - 修复统计聚合函数行为和Hive不一致的问题。 - 修复读取Hive ORC表char类型数据正确性的问题。
HDFS	支持国密SM4加密算法。
Hue	升级Hue至4.9.0版本。
Alluxio	升级Alluxio至2.5.0版本。
Druid	- 升级Druid至0.20.1版本。 - 增强了安全性。
Livy	升级Livy至0.7.1版本。

8.4.8. EMR-3.34.x版本说明

本文介绍EMR-3.34.x版本的发布日期和更新内容信息。

发布日期

EMR-3.34.0 2021年3月15日

更新内容

服务	变更点
SmartData	升级至3.4.0版本。 版本详情，请参见 SmartData 3.4.x版本简介 。
Spark	- 优化了部分默认配置。 - 性能优化：支持Window TopK下推。 - 增强Hive读写CSV或JSON表的兼容性。 - ANALYZE语句支持省略全表列名。 - 支持一键开启或关闭LDAP功能。 - 改进Spark Beeline工具的易用性。
Hive	- 优化了部分默认配置。 - 性能优化：增强CBO。 - 支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见管理LDAP认证。 - 升级Calcite版本至1.12.0。 - 增加参数hive.security.authorization.sqlstd.confwhitelist.append。

服务	变更点
Presto	支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见 管理LDAP认证 。
YARN	修复了Hadoop未授权访问Web UI时的高风险问题。即通过SSH Tunnel方式访问YARN WebUI时，需要在URL里显式指定user.name=name的问题。
Zookeeper	升级至3.6.2版本。
Flink	初始化时更新 <code>config.sh</code> 文件，修复HADOOP_CLASSPATH的问题。
Impala	- 升级Impala至3.4.0版本。 - 升级Shiro至1.7.0版本。 - 支持DLF元数据。 - 支持查询Delta格式的数据。 - 支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见管理LDAP认证。
Tez	优化了默认配置。
HAS	修复在HAS的install流程中产生错误后admin.keytab无法重新init问题。
Ranger	- 修复Spark中Filter函数的下推问题。 - 修复Ranger Disable Presto后，Presto无法启动的问题。 - 支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见管理LDAP认证。
Knox	修复Druid 0.20.0版本Knox链接的问题。
Hue	支持一键开启或关闭LDAP功能。 开启或关闭LDAP功能详情，请参见 管理LDAP认证 。
Hudi	- 支持SQL on Hudi功能。 - 修复了查询部分数据时准确性的问题。 - 在Spark查询Hudi的Copy On Write表时，支持分区裁剪。 - 支持分桶索引机制，提高写入性能。

服务	变更点
Delta Lake	- 修复基于已存在Delta表无法同步元数据至Hive Metastore的问题。 - 修复Merge命令无法解析 <code>*</code> 的问题。 - 修复基于Parquet格式的数据转换成Delta表，且创建表元数据时报错的问题。 - 修复当没有待compact的文件时，执行Optimize命令异常的问题。 - 支持Merge语法使用子查询作为source命令。 - 使用Presto查询Delta表时，引入缓存机制，以提升查询效率。 - 支持Impala查询Delta表。
Superset	- 修复admin用户无法登录Web UI的问题。 - 数据集兼容Druid集群模式。 - 不再支持SparkSQL数据集。
Sqoop	支持导入Parquet格式的文件至OSS。
Alluxio	升级至2.4.1版本。
Phoenix	Hive on Phoenix支持字段设置。
Pig	已移除。

8.4.9. EMR-3.33.x版本说明

本文介绍EMR-3.33.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-3.33.0 2021年1月15日

更新内容

服务	变更点
SmartData	升级至3.2.0版本。 详情请参见 SmartData 3.2.x版本简介 。
Spark	- 升级至2.4.7版本。 - 升级jQuery至3.5.1版本。 - 兼容Hive方式，自动更新表和分区大小。 - 支持Spark元数据和作业运行信息输出至DataWorks。
Hive	- 升级至2.3.7版本。 - HCatalog支持Data Lake Formation。 - 支持Hive元数据和作业运行信息输出至DataWorks。

服务	变更点
Metastore	- 新增Hive Statistics功能。 - HCatalog支持Data Lake Formation。 - 优化STToken的获取方式。
HDFS	升级jQuery至3.5.1版本。
YARN	- 升级jQuery至3.5.1版本。 - 调整Fair Scheduler配置。 - 优化了Timeline Server。
Zeppelin	升级至0.9.0版本。
Ranger	- 增加Hive的Audit日志配置。 - 增加Log4j Audit的配置。
OpenLDAP	- 增加审计功能。 - 默认开启SSL端口（10636）。 - 支持一键开启Presto。
Knox	- 修复Spring漏洞。 - 修复Spark UI中查看Executors页面的问题。 - 修复Oozie的Job状态页面的问题。
Hue	支持Presto。
Druid	升级至0.20.0版本。
EMRHook	- 新增软件服务。 - hive-hook：支持Hive元数据和作业运行信息输出至DataWorks。 - spark-hook：支持Spark元数据和作业运行信息输出至DataWorks。

8.4.10. EMR-3.32.x版本说明

本文介绍EMR-3.32.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-3.32.0 2020年11月23日

更新内容

服务	变更点
SmartData	升级至3.1.0版本。 详情请参见 SmartData 3.1.x版本简介 。

服务	变更点
Alluxio	- 支持Alluxio 2.4.0版本。 - 默认的参数配置，可以根据集群节点大小调整。 - 默认使用EMR集群内的HDFS作为底层的UnderFS，开箱即用。 - 增强Alluxio OSS UnderFS，适配OSS多版本等新功能。 - 适配Hadoop、Hive、Spark和Presto等引擎。
HUDI	支持HUDI 0.6.0版本。
Spark	JindoTable支持打开或关闭数据采集功能。
Hive	- 修复了HiveServer连接池泄漏的问题。 - JindoTable支持打开或关闭数据采集功能。 - 优化 <code>ADD COLUMN</code> 的性能。 - 修复了读取HUDI表时数据不正确的问题。 - 默认的参数配置，可以根据集群节点大小调整。
HDFS	支持了更高数量级的Snapshot。
YARN	默认的参数配置，可以根据集群节点大小调整。
Tez	默认的参数配置，可以根据集群节点大小调整。
Sqoop	修复了Avro格式的文件导入问题。

8.4.11. EMR-3.30.x版本说明

本文介绍EMR-3.30.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-3.30.0 2020年10月26日

更新内容

服务	变更点
SmartData	升级至3.0.0。
Spark	- 支持阿里云DLF（Data Lake Formation）元数据。 - 升级HAS依赖至2.0.1。 - 修复Streaming SQL反引号问题。 - 移除Delta的JAR包，修改为Delta单独部署。 - 修改日志路径统一写至HDFS下。
Hive	- 支持阿里云DLF（Data Lake Formation）元数据。 - 解决了读Delta表空目录时写DUMMY文件问题。 - 升级HAS依赖至2.0.1。

服务	变更点
Presto	- 支持阿里云DLF（Data Lake Formation）元数据。 - 解决读Delta表的限制问题。 - 修复高安全模式下JVM配置缺失问题。 - 升级HAS依赖至2.0.1。
HDFS	- 支持热交换磁盘模式。 - 升级HAS依赖至2.0.1。
YARN	- 修复YARN RMZKStateStore的问题。 - 支持SLS输出的SNAPPY文件。 - 修改MapReduce Local模式目录配置，解决目录权限检查问题。 - 支持热交换磁盘模式。 - 日志路径统一写到HDFS下。 - 升级HAS依赖至2.0.1。
Zookeeper	- 支持绑定内网IP启动服务端口。 - 升级HAS依赖至2.0.1。
Flink-Vvp	- 升级至1.11-2.2.2版本。 - 支持SQL和Autopilot功能。 ? 说明 仅Dataflow集群支持Flink-Vvp，Hadoop集群暂不支持Flink-Vvp。
Flink	- 支持缓存模式写入OSS，结合Flink的Checkpoint与可重发的Source实现EXACTLY_ONCE语义。 - 同步了Flink社区1.11.1功能，SQL支持多路输出（MULTI INSERT）。 - 升级HAS依赖至2.0.1。
Impala	- 支持自定义配置 <i>catalogd.flgs</i>、<i>impalad.flgs</i> 和 <i>statestore.flgs</i>。 - 升级Shiro至1.6.0版本。 - 升级HAS依赖至2.0.1。
Tez	- 优化AM的默认内存参数。 - 升级HAS依赖至2.0.1。
HAS	
Storm	
Zeppelin	

服务	变更点
Ranger	升级HAS依赖至2.0.1。
OpenLDAP	
Oozie	
Knox	
Kafka	
HUE	
HBase	
Druid	

8.4.12. EMR-3.29.x版本说明

本文介绍EMR-3.29.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-3.29.0 2020年7月29日

更新内容

服务	变更点
Bigboot	- 升级至2.7.301版本。 - Jindo DistCp支持写入时按OSS归档或低频写入。 - 增强Fuse功能，支持多Namespaces。 - 完善Cache模式的元数据缓存功能。
Spark	- Spark升级至2.4.5.2.0。 - 支持第三方Metastore的功能。 - 增加datalake metastore-client。
Hive	- Hive升级至2.3.5.6.0。 - 支持第三方Metastore的功能。 - 增加datalake metastore-client。
Presto	升级至338版本。
Ranger	- 升级软件包至1.2.0-1.5.0。 - 支持Presto 338。 - 配置文件增加Description。
HDFS	自适应配置datanode reserved空间大小。
Knox	适配Impala、高版本Flink和PAI。

服务	变更点
Druid	升级至0.18.1版本。
SmartData	升级至2.7.301版本。

8.4.13. EMR-3.28.x版本说明

本文介绍EMR-3.28.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-3.28.0 2020年6月12日

新增内容

服务	变更点
Bigboot	- 发布首个JindoTable版本，基于表或分区的热度统计。 - 支持Block模式上完整的存储策略，支持分层存储策略，包括低频和归档等。 - 增加数据迁移工具Jindo DistCp。 - 完善和修复Jindo Fuse。 - 完善Cache模式中JFS Scheme在Hive引擎和Jindo JobCommitter上的集成。 - Block模式读路径上，设置比重可以直接读OSS，用来缓解和分摊读本地缓存的开销。 - JindoFS软件模块解耦，分为Bigboot（管控层）、Smartdata（分布式服务）和JindoFS SDK。每块独立升级维护。

更新内容

服务	变更点
Flink	已将开源Flink升级为企业版Veriverica Platform，基于开源Flink 1.10深度定制，提供自研存储引擎Gemini等增值功能。
Bigboot	升级至2.7.0版本。
Delta	- 升级至0.6.0版本。 - 解耦Delta与Spark代码。
Spark	- 升级至2.4.5版本。 - 兼容DataFactory的streaming-sql脚本。 - 支持Delta 0.6.0版本。
Hive	支持Delta 0.6.0版本。
Ranger	- 支持HDFS、Hive和Spark自定义部署。 - 支持在控制台配置ranger-admin-site和ranger-ugsync-site。

服务	变更点
HDFS	针对HDFS写入时无可用DataNode节点的异常，打印对应DataNode异常信息（HDFS-9023）。
Hue	- 支持Gateway集群安装Hue组件。 - 支持在单个节点部署多个Hue实例。
DataFactory	支持Delta 0.6.0版本。
Druid	升级至0.18.0版本。
Knox	- 升级至1.1.0-1.0.7版本。 - 适配HBase UI。

8.4.14. EMR-3.27.x版本说明

本文介绍EMR-3.27.x发行版本的发布日期和更新内容等信息。

发布日期

版本	日期
EMR-3.27.0	2020年4月29日
EMR-3.27.1	2020年5月8日
EMR-3.27.2	2020年5月20日

新功能

功能	变更点
组件自定义部署	支持对Master节点上的组件进行自定义部署，目前支持以下组件： - Hadoop - Spark - Hive - Zookeeper - Presto
弹性伸缩功能支持优雅下线	开启优雅下线后，节点不会被立即释放，而是在设置的时间段内等待任务执行完成后释放。

更新内容

服务	变更点
Spark	- CUBE中支持日期类型分区字段。 - 调大Spark-Submit的stack深度。

服务	变更点
Delta	- DDL相关语法增强，包括CREATE、SHOW、DESCRIBE等相关命令。 - 支持带ZOrder的Optimize语法。
Knox	- 适配Druid UI。 - 支持多Master部署。
Hive	- hcatalog表支持magic committer。 - 移除一些过时的默认配置。
Bigboot	- 升级至2.6.3版本。 - 支持多Master部署。
SmartData	- 升级至2.6.3版本。 - 支持多Master部署。
Ranger	- 支持Solr组件。 - 支持PrestoSQL 311版本。
Tez	支持scratchdir设置在OSS上。
Presto	升级至331版本。
Druid	升级至0.17.1版本。
Superset	升级至0.35.2版本。
Sqoop	- MySQL JDBC JAR包升级至5.1.48版本。 - MySQL direct导出模式支持通过 <code>--mysql-charset</code> 设置自定义编码。

8.4.15. EMR-3.26.x版本说明

本文介绍EMR-3.26.x发行版本的发布日期和更新内容等信息。

发布日期

版本	日期
EMR-3.26.3	2020年4月16日

更新内容

服务	变更点
Bigboot	

服务	变更点
SmartData	- 升级到2.6.3版本。 - 支持OTS元数据和Namespace HA。
Hive	hcatalog表支持direct committer。
YARN	配置默认的committer为JindoOssCommitter。
HDFS	升级JindoFS相关配置。
Spark	配置默认的committer为JindoOssCommitter。

8.4.16. EMR-3.25.x版本说明

本文介绍EMR-3.25.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-3.25.0 2020年1月13日

新功能

Ranger服务：支持Ranger Presto操作。

更新内容

服务	变更点
Ranger	- 初始化HA集群RangerAdmin数据库。 - 修复RangerUserSync启动脚本时的安全性问题。
Spark	- 支持在控制台配置 <code>spark.sql.extensions</code> 等Delta相关参数。 - 支持Hive读取Delta table，避免set inputformat。 - 支持ALTER TABLE SET TBLPROPERTIES和UNSET TBLPROPERTIES语句。
Delta	
Hive	修复自动LOCAL模式下MR任务执行失败的问题。
Presto	- 升级至310版本。 - 升级joda-time版本至2.10.5。
Tez	- 升级至0.9.2版本。 - 修复tez-ui application进度无法正常显示的问题。 - 修复tez-ui application history无法查看的问题。
Impala	修复Impala无法访问lzo表的问题。
HDFS	移除mongo-hadoop的相关JAR包。
Zookeeper	升级至3.5.6版本。

服务	变更点
YARN	适配tez-ui, yarn-site页签支持添加配置项yarn.resourcemanager.system-metrics-publisher.enabled=true。
Bigboot	- 升级至2.2.3版本。 - OSS Cache模式下支持rename操作。
SmartData	
Knox	升级依赖包版本。
Oozie	升级依赖包版本。

8.4.17. EMR-3.24.x版本说明

本文介绍EMR-3.24.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-3.24.0 2019年11月18日

新功能

服务	变更点
Delta	- 支持SQL语法, 包括ALTER、CONVERT、CREATE、CTAS、DELETE、DESC、INSERT、MERGE、OPTIMIZE、UPDATE和VACUUM。 - 内置并优化Optimize。 - 支持Hive connector。 - 支持其他开源已有特性。
Grafana	新增组件 (Flink独立集群), 版本6.4.2。
Prometheus	新增组件 (Flink独立集群), 版本2.13.0。
AlertManager	新增组件 (Flink独立集群), 版本0.19.0。
TensorFlow on spark	- 支持TensorFlow框架置于Spark之上, 使得Spark与深度学习框架深度结合, 包括了任务调度和数据交换优化方案等, 为您提供从数据预处理到深度学习训练任务的一整套流程。 - 支持Streaming类型任务。

更新内容

服务	变更点
----	-----

服务	变更点
SmartData	• 优化JindoFS使用模式：Block模式使用方式不变；Cache模式不仅支持原有用法，还兼容了原有OSS文件系统的使用方式，支持数据缓存和元数据缓存，并可以通过配置分别控制开关（默认均关闭）。 • 优化Block模式和Cache模式读写性能。 • 优化磁盘清理，对本地磁盘上缓存的热数据实现更精确的统计和更及时的清理，并且能够严格保证磁盘使用率不会超过配额。 • 完善对Gateway集群的支持，能够在Gateway上使用Block模式和Cache模式。 • 支持一个存储集群与多个计算集群分离的部署方式。
Spark	• 增加Delta相关参数支持。 • 增加对Ranger spark plugin配置的支持。 • JindoCube升级到0.3.0版本。
Hive	• 增加SQL兼容性检查功能逻辑。 • Hive2.3.5+Hadoop2.8.5组合发布。 • 重启组件时不同步 <i>hiveserver2-site.xml</i> 中的内容至spark-conf下的 <i>hive-site.xml</i>。 • 支持使用MSCK命令添加增量目录。 • 修复Hive复用tez container时出现的bug。 • 支持使用MSCK命令优化列目录。
Bigboot	升级至2.2.1，修复Native代码支持在部分机型上的问题。
Ranger	• Spark plugin部署方式重构。 • 修复HA集群header2没有获取keytab的bug。
Kudu	修复启动逻辑。
Zookeeper	增加四字命令配置，默认开启。
HDFS	适配JindoFS。
YARN	• 修改默认配置 <code>yarn.scheduler.capacity.node-locality-delay</code> 为-1。 • 适配JindoFS。
Has	对接OpenLDAP做后端。
OpenLDAP	适配Has。
Presto	升级版本到0.228。
Kafka	移除D1坏盘。
Druid	升级至0.16.0。

服务	变更点
Flume	升级至1.9.0。
Flink	- 升级至1.9.1。 - 支持Flink独立集群（白名单发布）。

8.4.18. EMR-3.23.x版本说明

本文介绍EMR-3.23.x发行版本的发布日期和更新内容等信息。

发布日期

EMR-3.23.0 2019年09月18日

更新内容

服务	变更点
Druid	- 升级至0.15.1。 - 增加router组件。 - 升级fastjson到1.2.60。
Spark	- 更新spark thriftserver，解决class loader问题。 - 重构spark事务相关代码，提升稳定性。 - 解决升builtin hive至2.3版本后orc格式读写问题。 - 支持merge into语法。 - 支持scan和stream语法。 - Structured Streaming Kafka sink支持EOS。 - delta更新至0.4.0。
Hive	- 删除老版本的hive hook。 - 添加支持多个count distinct字段的数据倾斜处理优化。 - 解决join不同bucketversion的表时丢数据的问题。
Flink	升级至1.8.2。
Bigboot	- 更新小文件工具。 - 更新OSS jar，解决非daemon线程问题。
Kafka	- 新增感知Deploymen Set特性。 - 去掉fastjson依赖。
HDFS	- 优化SmartData OSS JAR包部署逻辑。 - 更新SmartData OSS JAR包。
Flume	升级fastjson至1.2.60。
Tensorflow on Spark	新增服务。

服务	变更点
Has	升级fastjson至1.2.60。
Livy	升级fastjson至1.2.60。

8.4.19. EMR-3.22.x版本说明

本文介绍EMR-3.22.x发行版本的发布日期、组件升级、新功能和更新内容等信息。

发布日期

EMR-3.22.0 2019年7月28日

新功能

服务	变更点
Kudu	- 新增组件，Kudu填补Hadoop生态圈的功能空白，可提供类似HBase快速数据插入以及随机存取的功能，允许用户进行数据修改，同时还提供类似HDFS或Parquet超大规模的数据分析以及查询的功能。 - 提供C++和Java API，以便用户进行二次开发。 - 提供Impala、Spark以及Hive Metastore的集成。 - Kudu版本基于开源社区Apache Kudu1.10.0版本。
OpenLDAP	- 新增组件，取代ApacheDS，ApacheDS下线。 - 高可用。

更新内容

组件	详细信息
----	------

组件	详细信息
JindoFileSystem	• 多种存储模式 ◦ Block模式：数据以Block形式存储在后端存储OSS上，本地Namespace服务维护元数据信息。在元数据性能和数据性能上，Block模式较优。Block模式支持不同的存储策略，包括WARM存储策略（本地-副本，OSS-副本）、COLD（仅OSS-副本）、HOT（本地多副本、OSS-副本）、TEMP（仅本地-副本）和ALL_HDD（本地多副本），默认为WARM，用户可以根据不同的应用场景对目录设置不同的存储策略。 ◦ Cache模式：该模式主要兼容现有OSS存储方式。在Cache模式下，文件以对象的形式存储在OSS上，每个文件根据实际访问情况会在本地进行数据和元数据的缓存，从而提高访问数据以及元数据的性能。Cache模式提供不同元数据同步策略以满足用户在不同场景下的需求。 • 外部客户端支持 ◦ 客户端SDK提供了E-MapReduce集群外访问E-MapReduce JindoFS 的文件系统的能力，通过客户端可以访问Block模式的Namespace，但外部客户端不能利用到E-MapReduce JindoFS在E-MapReduce集群内部构建的数据缓存，性能上相对于E-MapReduce集群内部使用也有一定的差距。 ◦ Cache模式则保留了原有OSS存储的语义，通过JindoFS在E-MapReduce集群内部实现了数据缓存加速，因此，E-MapReduce 集群外部可以直接通过OSS客户端访问数据，例如，OSS SDK或者E-MapReduce的OssFileSystem等。 • 生态组件支持 ◦ JindoFS目前已经支持E-MapReduce上的众多计算引擎，例如，Spark、Flink、Hive、MapReduce、Impala和Presto等。 ◦ 针对计算和存储分离的场景，也可以把作业日志存储在JindoFS上，例如，YARN Container log 和Spark Event log。 ◦ JindoFS可以作为HBase的HFile后端存储，扩展HBase的存储能力。
OssFileSystem	• OssFileSystem增加自动检测坏盘逻辑，修复OSS写入时，由于坏盘导致缓存写入失败的问题。 • 补全OssFileSytem相关配置。
Bigboot	• 升级到2.0.0版本。 • 包括多Namespace支持、本地数据块以大文件形式存储、多模存储支持和外部客户端支持等多项重大更新。 • 解决机器重启过程中Bigboot monitor状态不正确问题。 • 增加Kudu组件的服务spec。 • 增加各个服务spec的正确性检验。
Hadoop	• HDFS ◦ HDFS Federation适配，支持通过自定义配置和API创建HDFS Federation集群，避免创建Federation集群时的二次Format。 ◦ 优化坏盘检测逻辑，针对本地盘场景，可以通过dfsadmin触发DataNode blockreport时进行坏盘检测。 • YARN ◦ 修复MR作业Container日志存在JindoFS或OSS时，MapReduce JobHistory作业列表不更新的问题。

组件	详细信息
Spark	• Relational Cache 支持Relational Cache，Relational Cache通过预计算加速用户查询。用户可以创建Relational Cache对数据进行预计算，在执行用户查询时，Spark Optimizer自动发现合适的Cache，并改写SQL执行计划，基于Cache的数据继续计算，从而提升查询速度，适用于报表、Dashboard、数据同步和多维分析等场景。 ◦ 通过DDL，进行CACHE、UNCACHE、ALTER、SHOW等操作，Cache的数据支持Spark的所有数据源和数据格式。 ◦ 支持自动的Cache数据更新以及通过REFRESH命令更新Cache数据，支持基于分区的增量更新。 ◦ 支持基于Relational Cache的执行计划优化。 • Streaming SQL ◦ 规范Stream Query Writer的参数配置。 ◦ 优化Kafka数据表Schema兼容性检查。 ◦ Kafka数据表Schema不存在时自动创建到SchemaRegistry。 ◦ 优化Kafka Schema不兼容时的日志信息。 ◦ 修复查询结果写Kafka表时必须显式指定列名的问题。 ◦ 去掉流式SQL查询只支持Kafka和Loghub数据输入源的限制。 • Delta 新增Delta，用户可使用Spark创建Delta datasource，以支持流式数据写入、事务性读写、数据校验和数据回溯等应用场景。详情请参见Delta详细信息。 ◦ 支持使用DataFrame API从Delta读取数据或者写入数据到Delta。 ◦ 支持使用Structured Streaming API以Delta作为source或者sink进行数据的读或写。 ◦ 支持使用Delta API对数据进行update、delete、merge、vacuum、optimize等操作。 ◦ 支持使用SQL创建基于Delta的表、导入数据到Delta和读取Delta表等操作。 • Others ◦ constraint feature，支持主键和外键。 ◦ 解决servlet等jar冲突问题。
Flink	log4j日志回滚。
Kafka	• log4j日志回滚。 • 升级fastjson。
Zeppelin	升级依赖的commons-lang3包到3.7版本，修复pyspark无法写OSS的问题，详情请参见 Spark 2.4 incompatibility with commons-lang3 in Zeppelin 。
Ranger	增加Show grants支持。
Analytics-Zoo	修复Numpy安装错误问题。
Impala	兼容Apache Kudu 1.10.0版本。
Presto	升级到0.221版本。
ZooKeeper	升级到3.5.5版本。

8.4.20. EMR-3.22.x之前版本说明

本文介绍EMR-3.22.x之前的各个发行版本的发布日期、组件升级、新功能和更新内容等信息。

EMR-3.2.0

支持所有的服务在界面上进行起停操作，支持在界面对服务进行配置管理。

EMR-3.1.1

- OS升级到Cent OS 7.2。
- Spark升级到2.1.1版本。
- emr-core升级到1.2.6版本。
- 修复了OSS免AccessKey操作的缺陷。

EMR-3.0.2

- emr-core版本升级到1.2.5版本。
- OSS免AccessKey支持更多Region。
- 调整角色AccessKey的替换策略。
- 修复关于Hive和Hadoop的部分缺陷。

EMR-3.0.1

- 支持交互式，支持统一的表管理，使用外部统一数据库保存Hive meta，所有使用外部Hive meta的集群都共享同一份meta信息。
- 升级emr-core到1.2.4版本，优化了OSS读写的性能。
- Spark升级到2.0.2版本。

 说明 与EMR-3.0.0版本完全兼容。

EMR-3.0.0

第一次发布EMR版本。