

目录

目录	1
产品概述	2
基本概念	2
KMR的功能	2
大数据生态组件	2
产品优势	2
快速部署，操作易用	2
同步开源，稳定可靠	2
弹性伸缩，灵活计费	3
使用场景	3
日志分析	3
实时推荐应用	3
搜索场景	4
BI报告查询	4
产品术语表	4

产品概述

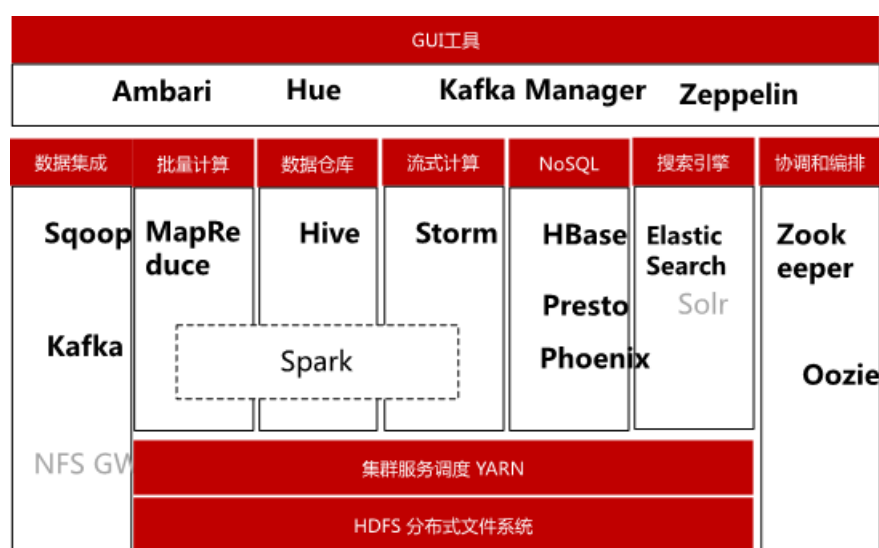
基本概念

KMR (Kingsoft MapReduce) 是构建在金山云上的大数据平台，提供海量数据的收集，储存，处理和展现的功能。KMR平台既可以帮助企业从大数据中提取知识，并且能够支撑企业的商业决策和数据应用。KMR平台完全兼容开源Hadoop/Spark社区，并提升了开源组件的可用性和稳定性，还实现了大数据平台的托管运维、弹性管理、安全管控。从而使企业专注业务，提升工作效率，降低企业建设的成本和周期。

KMR的功能

KMR (Kingsoft MapReduce) 是一个可伸缩的通用数据计算和分析平台，帮助您快速构建分布式数据分析系统。它以Apache Hadoop和Apache Spark两大数据计算框架为基础，同时，KMR实现计算与存储分离，KMR还可以方便的与金山云其他的云数据存储系统和数据库系统进行数据传输（例如，KS3等）。KMR提供了强大的扩展能力和弹性伸缩能力，解决了Hadoop安装部署成本和管理复杂性得问题，可以让您不必关注基础架构管理，而更加专注数据分析、处理和应用本身，任何的开发者或者公司只需要较低的成本就可以进行大规模的数据分析和处理工作。

大数据生态组件



KMR基于Hadoop生态系统，可无缝对接金山云KS3（对象存储）和RDS服务，着力于方便快捷的数据分析于存储，方便数据再多个系统间的共享和传输。KMR完全兼容开源系统，并在基础上强化，监控能力的整合加强。随着Hadoop社区发展生态组件不断扩充，KMR能力随之提升。

产品优势

KMR具有云原生和大数据的双重优势，在易用性、可靠性、免运维、安全性具有显著的优势，另外KMR深度与其他金山云产品的深度整合。

快速部署，操作易用

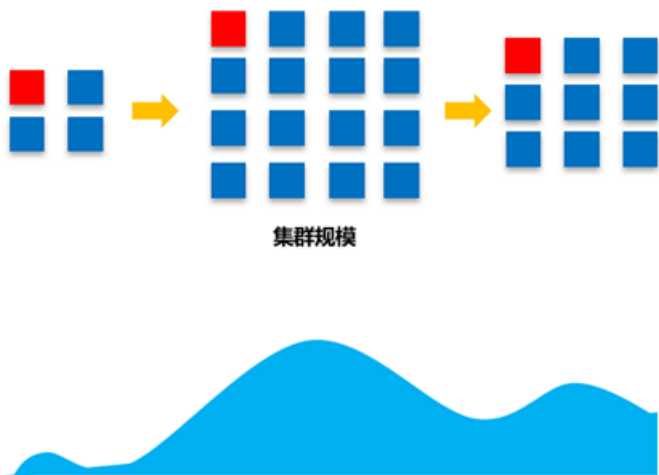
KMR提供分钟级启动集群的服务。KMR采用弹性计算服务（KEC）构建集群，通常情况下只需几分钟即可自动完成部署工作，您只需关心数据处理任务本身，无需关注硬件和底层系统的运维工作。

同步开源，稳定可靠

KMR完全兼容Hadoop、Spark等生态组件，以及Ambari、Hue等集群监控管理工具。KMR持续监控您的集群，会自动配置云主机防火墙设置以控制实例的网络访问权限，并且您可以在VPC中由您定义的逻辑隔离的网络中启动集群。VPC网络隔离等多个层面提供了服务可用性和数据安全保障。帮助您轻松构建复杂的大数据分析系统，满足批量计算、流式处理、消息队列、交互式查询、NoSQL等多种业务场景的需求。



弹性伸缩，灵活计费

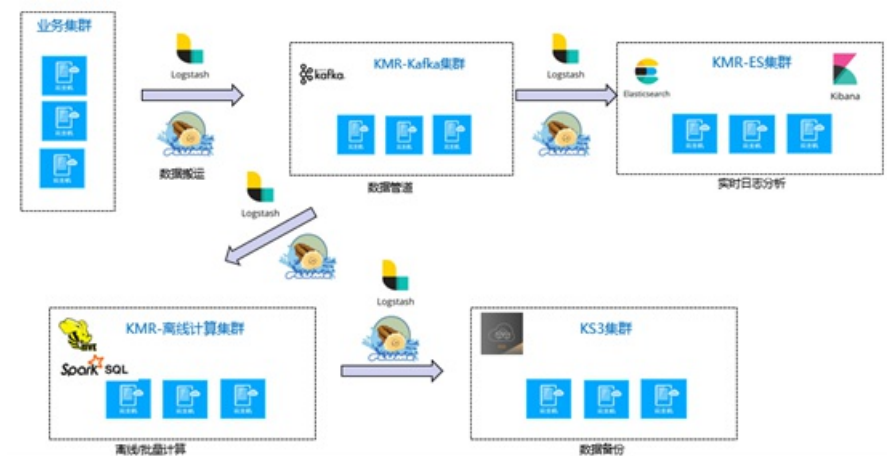


使用KMR，您可以按需创建、扩容和缩容集群，也可以按照离线任务的需求创建、释放集群。KMR实现了存储和计算分离，更加灵活的实现计算、存储的扩缩容，可以有效的减少您的使用成本，轻松应对多变的业务需求。

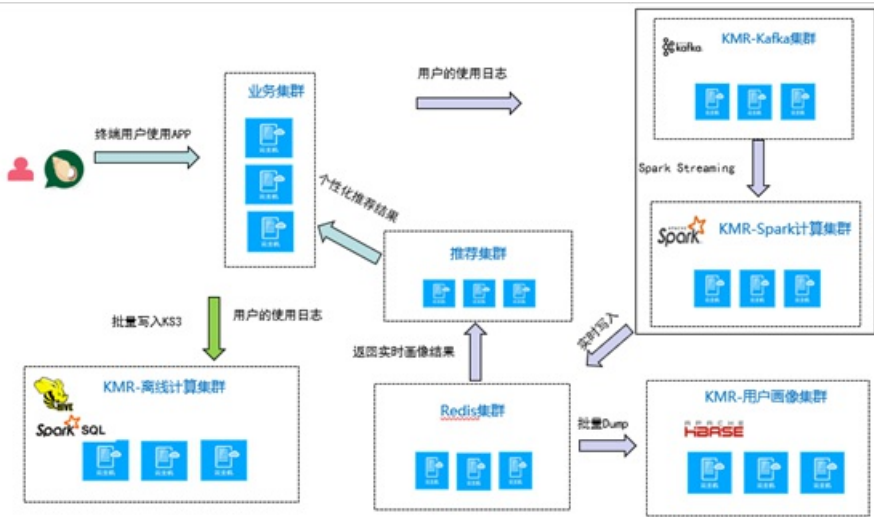
使用场景

KMR平滑对接Hadoop生态的应用场景，灵活集成金山云其他产品，ECS、KS3等，简化集群管理。KMR构建的分布式分析系统，帮助您对数据更好的收集、存储和处理，实现数据的价值最大化，主要应用场景如下：

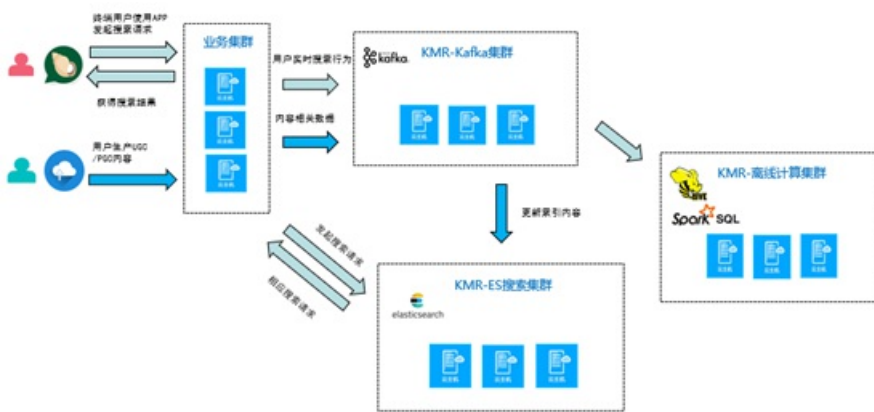
日志分析



实时推荐应用



搜索场景



BI报告查询



产品术语表

英文术语	简写/缩写	中文术语	术语描述
Kingsoft MapReduce	KMR		金山云平台上的Hadoop托管集群，可通过Web对外提供服务
KMR cluster	Cluster	KMR 集群	由若干金山云主机实例组成的Hadoop集群
KMR Master Node	Master	主节点	主要用于集群管理，并将计算程序和原始数据集分配到核心实例。此外，它还会跟踪每个计算作业的执行状态，监控实例的运行状况。KMR主节点与Hadoop系统的主节点相对应。一个KMR集群只有一个主节点。
KMR Core Node	Core	核心节点	主要用于执行各项集群计算作业，也做为hadoop分布式文件系统的数据节点存储数据。KMR核心节点与Hadoop系统的slave节点相对应。一个KMR集群可以有2至多个核心节点
KMR Job	Job	作业	一个作业是提交到集群中的一个工作单元。一个作业可能包含一个或多个Hadoop任务，或者包含安装或配置一个应用程序的指令。您可以对一个集群提交多达256个作业。
SSH KEY		SSH密钥	指用户在控制台上传的SSH 公钥

Kingsoft standard storage service	云存储	金山云标准存储服务
Hadoop file system	HDFS	Hadoop 分布式文件系统 (HDFS) 是一种分布式的、可扩展的文件系统, 供 Hadoop 使用。
MapReduce	MR	MapReduce 是一种用于分布式计算的编程模型, 用于大规模数据集的并行运算。
Hadoop		Apache Hadoop 是一种开源 Java 软件框架, 支持跨越一组服务器处理大量数据。它可以在一台服务器或成千上万台服务器上运行。Hadoop 使用名为 MapReduce 的编程模型在多个服务器之间分配处理工作。此外, 它还实施了一个名为 HDFS 的分布式文件系统, 在多个服务器之间存储数据。
Hue		Hue是一个开源的Apache Hadoop UI系统, 通过使用Hue我们可以在浏览器端的Web控制台上与Hadoop集群进行交互来分析处理数据, 例如操作HDFS上的数据, 运行MapReduce Job, 执行Hive的SQL语句, 浏览Hbase数据库等等。