

大型架构及配置技术

NSD ARCHITECTURE **DAY05**

内容

上午	09:00 ~ 09:30	作业讲解和回顾
	09:30 ~ 10:20	大数据
	10:30 ~ 11:20	Hadoop
	11:30 ~ 12:00	
下午	14:00 ~ 14:50	Hadoop安装与配置
	15:00 ~ 15:50	
	16:10 ~ 17:10	HDFS
	17:20 ~ 18:00	总结和答疑



大数据



大数据介绍

大数据的由来

- 大数据
 - 随着计算机技术的发展，互联网的普及，信息的积累已经到了一个非常庞大的地步，信息的增长也在不断的加快，随着互联网、物联网建设的加快，信息更是爆炸是增长，收集、检索、统计这些信息越发困难，必须使用新的技术来解决这些问题

什么是大数据

知识讲解

- 大数据的定义
 - 大数据指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产
 - 是指从各种各样类型的数据中，快速获得有价值的信息



什么是大数据（续1）

知识讲解

- 大数据能做什么
 - 企业组织利用相关数据分析帮助他们降低成本、提高效率、开发新产品、做出更明智的业务决策等
 - 把数据集合并后进行分析得出的信息和数据关系性，用来察觉商业趋势、判定研究质量、避免疾病扩散、打击犯罪或测定即时交通路况等
 - 大规模并行处理数据库，数据挖掘电网，分布式文件系统或数据库，云计算平和可扩展的存储系统等



大数据特性

知识讲解



大数据特性（续1）

知识讲解

- 大数据的5V特性是什么？
 - (V)olume (大体量)
 - 可从数百TB到数十数百PB、甚至EB的规模
 - (V)ariety(多样性)
 - 大数据包括各种格式和形态的数据
 - (V)elocity(时效性)
 - 很多大数据需要在一定的时间限度下得到及时处理
 - (V)eracity(准确性)
 - 处理的结果要保证一定的准确性
 - (V)alue(大价值)
 - 大数据包含很多深度的价值，大数据分析挖掘和利用将带来巨大的商业价值

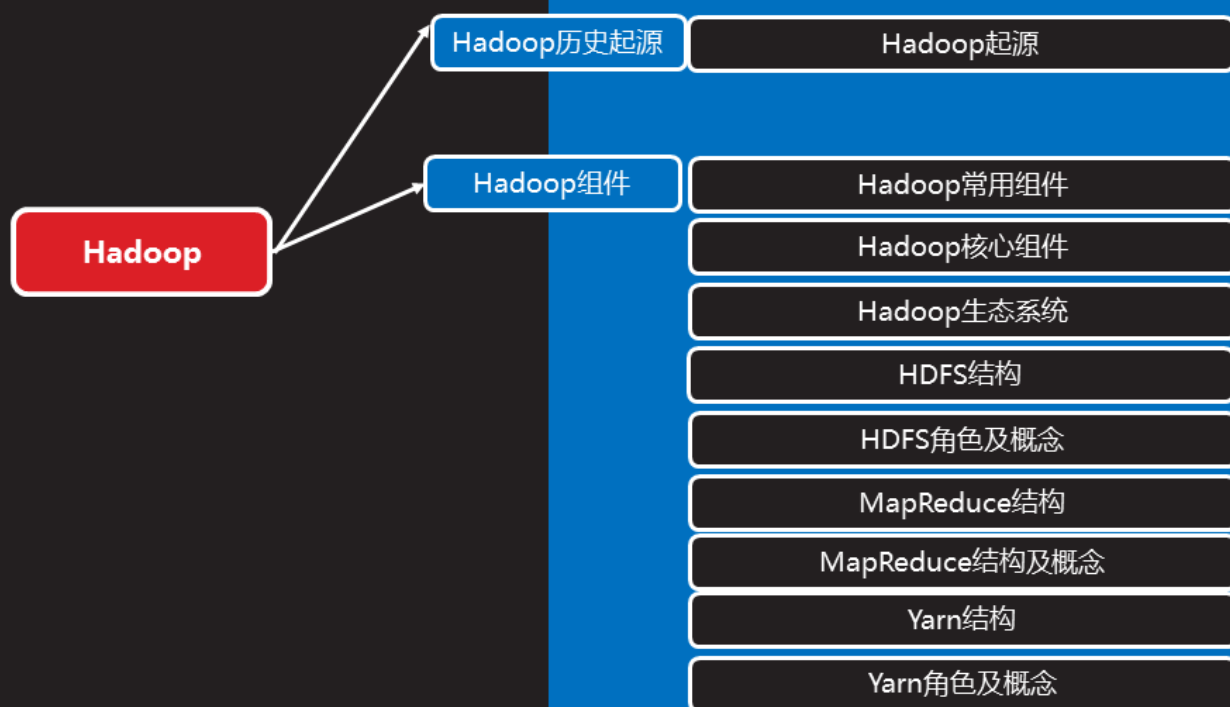
大数据与Hadoop

知识讲解

- Hadoop是什么
 - Hadoop是一种分析和处理海量数据的软件平台
 - Hadoop是一款开源软件，使用JAVA开发
 - Hadoop可以提供一个分布式基础架构
- Hadoop特点
 - 高可靠性、高扩展性、高效性、高容错性、低成本



Hadoop



Hadoop历史起源

Hadoop起源

- 2003年开始Google陆续发表了3篇论文
 - GFS , MapReduce , BigTable
- GFS
 - GFS是一个可扩展的分布式文件系统，用于大型的、分布式的、对大量数据进行访问的应用
 - 可以运行于廉价的普通硬件上，提供容错功能
- MapReduce
 - MapReduce是针对分布式并行计算的一套编程模型，由Map和Reduce组成，Map是映射，把指令分发到多个worker上，Reduce是规约，把worker计算出的结果合并



Hadoop起源 (续1)

知识讲解

- BigTable
 - BigTable是存储结构化数据
 - BigTable建立在GFS , Scheduler , Lock Service和MapReduce之上
 - 每个Table都是一个多维的稀疏图



Hadoop起源 (续2)

知识讲解

- GFS、MapReduce和BigTable三大技术被称为Google的三驾马车，虽然没有公布源码，但发布了这三个产品的详细设计论
- Yahoo资助的Hadoop，是按照这三篇论文的开源Java实现的，但在性能上Hadoop比Google要差很多
 - GFS - - -> HDFS
 - MapReduce - - -> MapReduce
 - BigTable - - -> Hbase



Hadoop组件



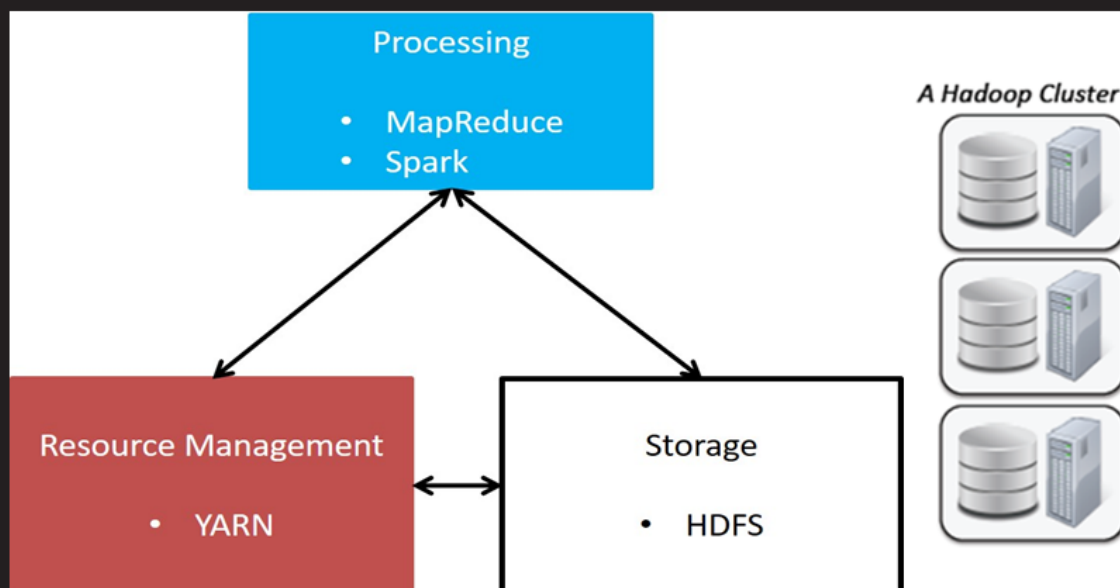
Hadoop常用组件

- HDFS : Hadoop分布式文件系统 (核心组件)
- MapReduce : 分布式计算框架 (核心组件)
- Yarn : 集群资源管理系统 (核心组件)

知

Hadoop核心组件

知识讲解



Hadoop生态系统

知识讲解



HDFS角色及概念



- Hadoop体系中数据存储管理的基础，是一个高度容错的系统，用于在低成本的通用硬件上运行
- 角色和概念
 - Client
 - Namenode
 - Secondarynode
 - Datanode

知识讲解







