

第十八讲 总结



徐辰
cxu@dase.ecnu.edu.cn

华东师范大学

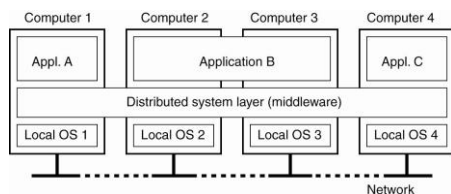


大纲

- 2
- 分布式计算
- 课程模块回顾
- 课程知识体系
- 发展趋势

分布式系统的定义

- 3
- A distributed system is: A collection of independent computers that appears to its users as a single coherent system



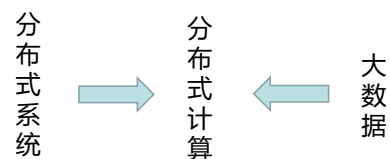
大数据4V



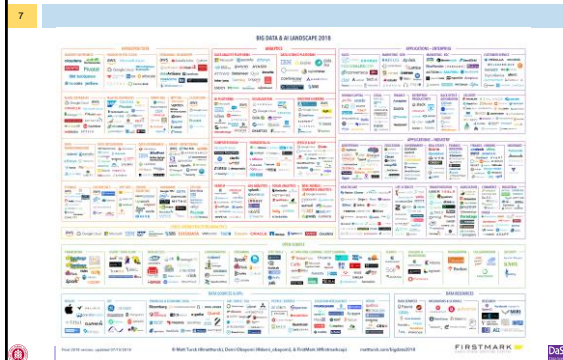
大数据5V



分布式系统遇到大数据



Big data landscape



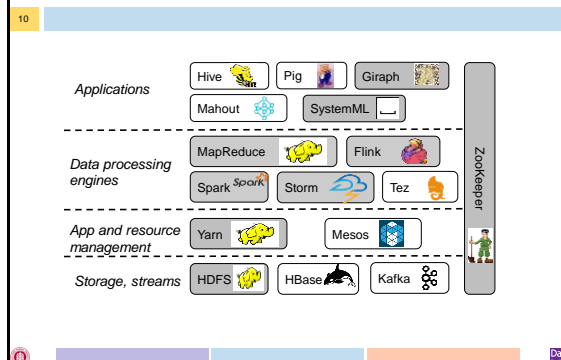
分布式计算

计算模式	解决问题	代表产品
批处理	针对大规模数据的批量处理	MapReduce、Spark等
流计算	针对流数据的实时计算	Storm、S4、Flume、Streams、Puma、DStream、Super Mario、银河流数据平台等
批流融合	批处理和流计算的融合	Flink、Google Dataflow等
图计算	针对大规模图结构数据的处理	Pregel、GraphX、Giraph、PowerGraph、Hama、GoldenOrb等
机器学习	支持大规模数据的机器学习	Mahout、SystemML、Parameter Server、Tensorflow等

大纲

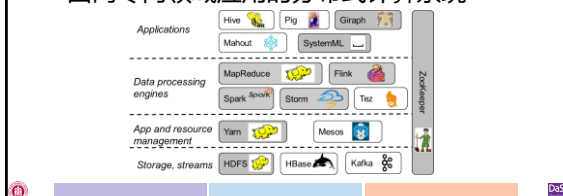
- 分布式计算
- 课程模块回顾
- 课程知识体系
- 发展趋势

分布式计算系统生态圈



模块设计

- 基础知识
- 支持分布式计算的底层系统
- 面向通用数据处理的分布式计算系统
- 面向专门领域应用的分布式计算系统



基础知识

- 简介
- 基础知识
 - ✚ 进程通信
 - ✚ 序列化与压缩

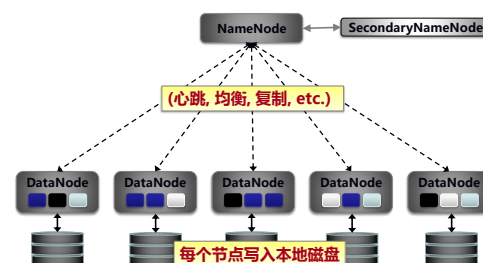
支持分布式计算的底层系统

13

- 协调服务系统
 - ✚ ZooKeeper
- 资源管理系统
 - ✚ Yarn
- 分布式文件系统
 - ✚ HDFS

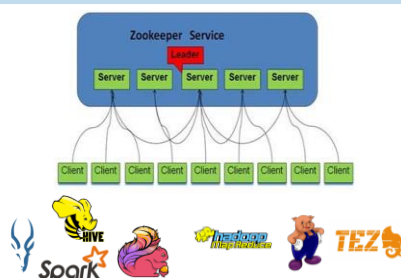
文件系统—HDFS

14



协调服务—ZooKeeper

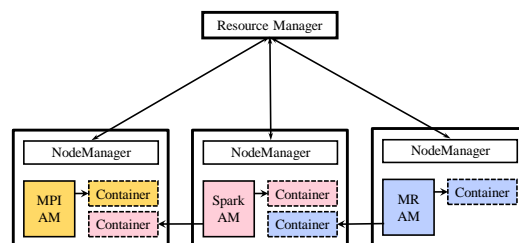
15



资源管理—YARN

16

- 一个资源管理平台运行多个计算框架



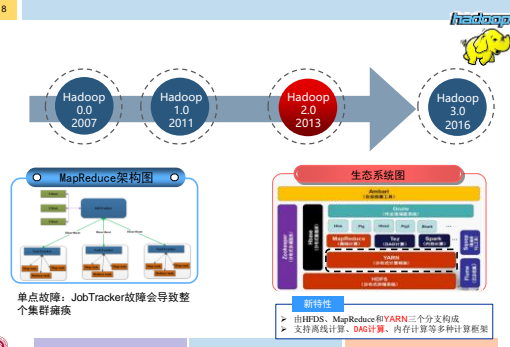
面向通用数据处理的分布式计算系统

17

- 批处理系统
 - ✚ MapReduce
 - ✚ Spark
- 流计算系统
 - ✚ Storm
 - ✚ Spark Streaming
- 批流融合系统
 - ✚ Google Dataflow
 - ✚ Flink

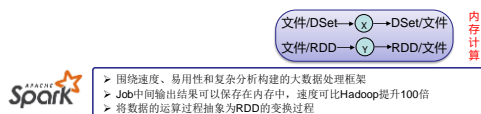
批处理—Apache Hadoop

18



批处理—Apache Spark

19



Spark特性

- 目前提供了针对Scala、Java和Python的API，2.0版本提供针对其它语言（例如R）的支持
- 可以和Hadoop生态系统和数据源（HDFS、Amazon S3、Hive、HBase、Cassandra等）进行集成
- 可以运行在由Hadoop YARN或者Apache Mesos管理的集群上，也可以运行在单独的集群上
- Spark Streaming可模拟流计算，响应时间可达百毫秒

流计算—Storm

20

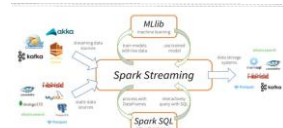


流计算—Spark Streaming

21

- 建立在Spark上的实时计算框架
- 将流式计算以时间片Δt（秒级）分解成一系列短小的批处理（Micro-batch），本质上是“更细粒度的批处理”
- 每块数据作为一个RDD，使用RDD操作处理每一小块数据，每个块都会生成一个Spark Job处理，最终结果返回多块

微批处理技术

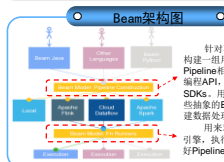


批流融合—Dataflow Model

22

仅是一个框架，没有计算引擎

- 基于谷歌的Dataflow模型创建的项目
- 按照如何对数据进行计算、如何对数据进行输出、何时将计算结果输出、迟到的数据如何处理四个维度调用具体的API
- 将流式数据当作一个无穷数据集，将离线数据当作一个有穷数据集



- 统一：实现批处理、流处理统一
- 可移植：在底层计算平台升级的过程中无需重写上层已定义的Data Pipeline
- 可扩展：实现任意可以共享的Beam SDK、IO connector、Transform库

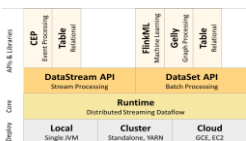
批流融合—Apache Flink

23

流计算引擎
状态管理
有界流—批处理

流计算毫秒级响应

- 一个面向分布式数据流处理和批量数据处理的开源计算平台
- 本质是一个流式计算引擎，运行时分别搭建了批处理(DataSet)和流式计算(DataStream)的编程接口和配套的生态系统
- 支持Redis、Kafka、Cassandra、Elasticsearch、Flume、Rabbitmq、Twitter、HDFS等数据存储形式



- 支持本地、集群、云等多种部署模式
- 运行时以JobGraph形式接收程序，DataStream API和DataSet API都会使用单独编译的处理方式生成JobGraph
- DataStream API和DataSet API支持Java、Scala和Python
- 提供了类似SQL的查询，图形化处理，以及机器学习库

面向专门领域应用的分布式计算系统

24

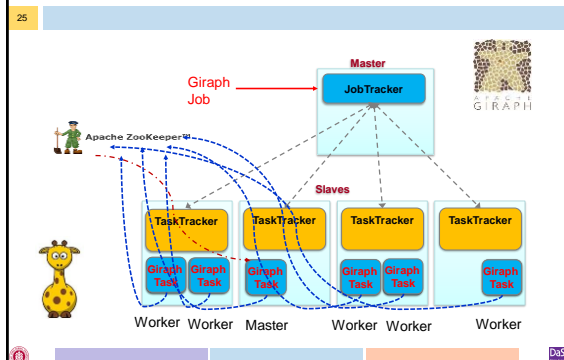
图处理系统

- Giraph/Hama
- GraphX/Gelly

机器学习系统

- SystemML/Mahout
- Parameter Server/Tensorflow

图处理—Giraph



大纲

- 26
- 分布式计算
 - 课程模块回顾
 - 课程知识体系
 - 发展趋势

知识体系

- 27
- 设计思想
 - 体系架构
 - 工作原理
 - 容错机制
 - 编程实例

设计思想

□ 数据模型

系统	数据模型
MapReduce	Key-value
Spark	RDD
Storm	Tuple
Spark Streaming	Dstream
Flink	DataStream
Giraph	Graph

设计思想

□ 计算模型

系统	计算模型
MapReduce	DAG
Spark	DAG
Storm	Topology (DAG)
Spark Streaming	DAG
Flink	DAG
Giraph	Vertex-centric

设计思想

□ 迭代模型

系统	迭代模型
MapReduce	
Spark	
Storm	
Spark Streaming	
Flink	√
Giraph	√

体系架构

31

- 架构图
- 应用程序执行流程

工作原理

32

- 数据输入、**数据转换**、数据输出
- 讨论主题
 - ✚ 执行引擎不同Task（运行某个或**某些**算子）之间如何进行数据交换？
 - ✚ Flink：如何基于执行引擎来实现Dataflow编程模型？

容错机制

33

- 故障类型
- 检查点（持久化）
- 故障恢复

编程实例

34

- MapReduce模型
- DAG模型
- 流计算模型
- Dataflow模型
- 图计算模型（Vertex-centric）
- 迭代模型
- 广播变量模型

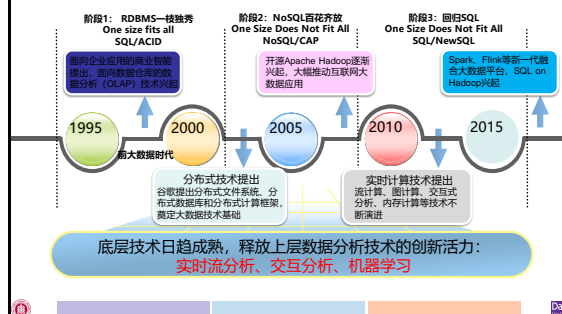
大纲

35

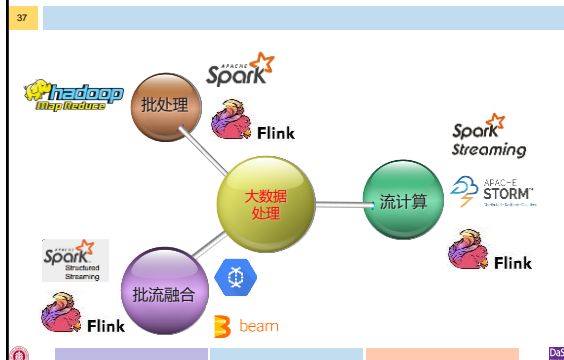
- 分布式计算
- 课程模块回顾
- 课程知识体系
- **发展趋势**

大数据管理技术的演进

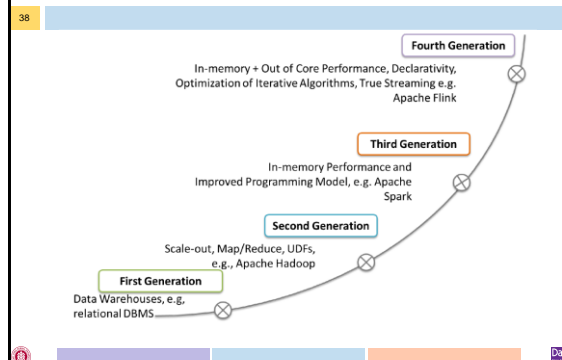
36



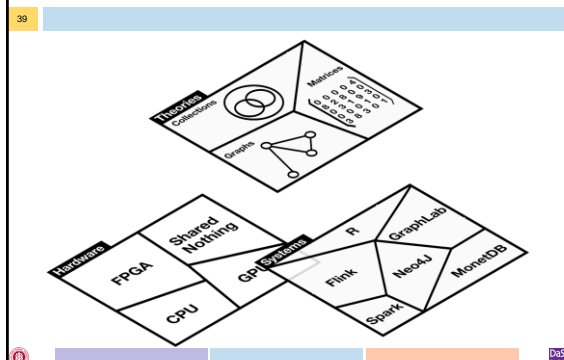
大数据处理系统



大数据处理系统的演进



Vision: Mosaics



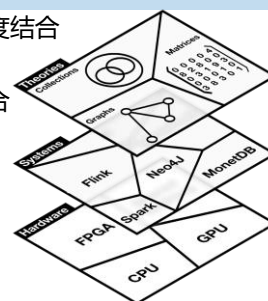
未来趋势

□ 算法：与系统的深度结合

□ 系统：批流深度融合

□ 硬件：新体系架构

- ✦ 运算：GPU
- ✦ 存储：NVM
- ✦ 网络：RDMA



期末考试

41

题型	每题分值	数量	单项总分
选择题	2	5	10
判断题	2	5	10
填空题	2	5	10
简答题	4-6	6-8	30
分析论述题	10-15	3	40

笔试安排

□ 本科生：分布式模型与编程

- ✦ 时间：2019年1月6日 10:30-12:30
- ✦ 地点：文附楼105

□ 研究生：大数据处理系统

- ✦ 时间：2019年1月6日 10:30-12:30
- ✦ 地点：文附楼118

Reminder : 本科生

43

□ 机考

🕒 时间: 2019年1月2日 18:00-20:30

📍 地点: 地理馆353

⚠️ 注意: **机考部分请务必准时参加**

Reminder : 研究生

44

□ 选择上课所讲述之外的系统, 从设计思想、体系结构、工作原理、容错机制等角度**与所学系统进行对比分析**, 谈谈你的理解

□ 格式、命名方式等要求详见大夏学堂

□ 提交方式:

📧 **2019.12.31**前发送到 dasebigdata@163.com

谢谢! Q&A



Credits

46

□ 王国仁

□ Volker Markl

□ 林子雨