

大数据概述

2018年8月17日 星期五 上午 9:08

大数据高并发基础 - 高并发 分布式

- NIO

- Concurrent

- 序列化 反序列化

- RPC

- Zookeeper

大数据离线分析 - 积攒了大量的数据后 一次性处理 得到结果

- Hadoop - HDFS MapReduce Yarn

- Hive

- Hbase

- Flume

- Sqoop

大数据实时分析 - 海量数据 实时处理 数据不停的过来 实时处理后 立即展示变化

- Kafka

- Storm

大数据内存计算 - 主要指以Spark为核心的 基于内存的高效的海量数据处理技术

- Spark Core

- Spark Sql

- Spark Streaming

- Spark Mllib

- Spark GraphX

爬虫&数据可视化

- Python语言

- Python爬虫开发

- 可视化 - Echars Davav HeatMap...

机器学习&算法

- R语言

