	计算机基础	数据结构和算法计算机组成原理		
				Java 的面向对象  Java 语言的三大特征:封装、继承和多态  Java 语言数据类型
				Java 的自动类型转换,强制类型转换  String 的不可变性,虚拟机的常量池,String. intern() 的底层原理
			语言基础	Java 语言中的关键字: final、static、 transient、instanceof、volatile、 synchronized的底层原理
				Java 中常用的集合类的实现原理:ArrayList/ LinkedList/Vector、SynchronizedList/ Vector、HashMap/HashTable/ ConcurrentHashMap 互相的区别以及底层实现 原理
				动态代理  CAS、乐观锁与悲观锁、数据库相关锁机制、分布式锁、偏向锁、轻量级锁、重量级锁、
				锁优化、锁消除、锁粗化、自旋锁、可重入锁、 阻塞锁、死锁
			锁	死锁的原因 死锁的解决办法
				CountDownLatch、CyclicBarrier 和 Semaphore 三个类的使用和原理  并发和并行的区别
				线程与进程的区别 线程的实现、线程的状态、优先级、线程调度、 创建线程的多种方式、守护线程
			多线程	自己设计线程池、submit() 和 execute()、线程 池原理 为什么不允许使用 Executors 创建线程池
			シジが主	死锁、死锁如何排查、线程安全和内存模型的关系 ThreadLocal 变量
				Executor 创建线程池的方式: ThreadPoolExecutor 创建线程池、拒绝策略
				线程池关闭的方式  JUC 包中 List 接口的实现类: CopyOnWriteArrayList
<b>₩</b>		Java		JUC 包中 Set 接口的实现类: CopyOnWriteArraySet、 ConcurrentSkipListSet
基础篇			并发容器 (J.U.C)	JUC 包中 Map 接口的实现类: ConcurrentHashMap、 ConcurrentSkipListMap  JUC包中Queue接口的实现类:
				ConcurrentLinkedQueue、 ConcurrentLinkedDeque、 ArrayBlockingQueue、 LinkedBlockingQueue、 LinkedBlockingDeque
				JVM 内存结构 堆和栈
				Java 内存模型 垃圾回收 JVM 参数及调优
			JVM	Java 对象模型 虚拟机性能监控与故障处理工具
				即时编译器、编译优化 类加载机制
	编程语言			用户空间以及内核空间  Linux 网络 I/O 模型: 阻塞 I/O (Blocking I/O)、非阻塞 I/O (Non-Blocking I/O)、I/O 复用(I/O Multiplexing)、信号驱动的 I/O (Signal Driven
			NIO	I/O)、异步 I/O 灵拷贝(ZeroCopy) BIO、AIO、NIO 对比
				缓冲区 Buffer 通道 Channel
				反应堆 选择器
			RPC	RPC 的原理编程模型 常用的 RPC 框架: Thrift、Dubbo、 SpringCloud
				RPC 的应用场景和与消息队列的差别 RPC 核心技术点:服务暴露、远程代理对象、通信、序列化
		Python Scala	· 甘士·比·ᄼ	
	Linux基础	基础	基本指令 系统和网络命令 权限模型	
		基于MySQL了解常见的SQL语法,大数据领域 SQL化是未来的发展方向。	基础的Shell脚本	
	数据库入门	٥ ودا الدرجار بحرا المراجع الم		
	分布式理论基础和原理 10	Hadoop 体系是我们学习大数据框架的基石,尤 其是 MapReduce、HDFS、Yarn 三驾马车基本 垫定了整个数据方向的发展道路。		
			掌握 MapReduce 的工作原理 能用 MapReduce 手写代码实现简单的 WordCount 或者 TopN 算法	
		MapReduce	掌握 MapReduce Combiner 和 Partitioner的作用 熟悉 Hadoop 集群的搭建过程,并且能解决常见	
			的错误 熟悉 Hadoop 集群的扩容过程和常见的坑	
			如何解决 MapReduce 的数据倾斜  Shuffle 原理和减少 Shuffle 的方法  十分熟悉 HDFS 的架构图和读写流程	
		HDFS	十分熟悉 HDFS 的配置 熟悉 DataNode 和 NameNode 的作用	
			NameNode 的 HA 搭建和配置,Fsimage 和EditJournal 的作用的场景 HDFS 操作文件的常用命令	
			HDFS 的安全模式 Yarn 的产生背景和架构	
		Yarn	Yarn 中的角色划分和各自的作用 Yarn 的配置和常用的资源调度策略 Yarn 任务资源调度的过程	
			Hive 是一个数据仓库基础工具,在 Hadoop 中 用来处理结构化数据。它架构在 Hadoop 之上, 总归为大数据,并使得查询和分析方便。Hive 是 应用最广泛的 OLAP 框架。Hive SQL 也是我们	
	Hadoop 体系		进行 SQL 开发用的最多的框架。  HiveSQL 的原理:我们都知道 HiveSQL 会被翻译成 MapReduce 任务执行,那么一条 SQL 是如何翻译成 MapReduce 的?	
			Hive 和普通关系型数据库有什么区别?  Hive 支持哪些数据格式	
		Hive	Hive 在底层是如何存储 NULL 的 HiveSQL 支持的几种排序各代表什么意思(Sort By/Order By/Cluster By/Distrbute By)	
		Tilve	Hive 的动态分区 HQL 和 SQL 有哪些常见的区别	
			Hive 中的内部表和外部表的区别  Hive 表进行关联查询如何解决长尾和数据倾斜问题	
			HiveSQL 的优化(系统参数调整、SQL 语句优化)  HBase 是一个面向列的数据库,在表中它由行排	
			序。表模式定义只能列族,也就是键值对。一个表有多个列族以及每一个列族可以有任意数量的列。后续列的值连续地存储在磁盘上。 表中的每个单元格值都具有时间戳。总之,在一个 HBase:表是行的集合、行是列族的集合、列	
			族是列的集合、列是键值对的集合。  Hbase 的架构和原理  Hbase 的读写流程	
			Hbase 有没有并发问题?Hbase 如何实现自己的 MVVC 的?	•
		列式数据库 Hbase	Hbase 中几个重要的概念:HMaster、RegionServer、WAL 机制、MemStore  Hbase 在进行表设计过程中如何进行列族和RowKey 的设计	•
		7JLQXJIA/# I IDase	Hbase 的数据热点问题发现和解决办法 提高 Hbase 的读写性能的通用做法	•
			HBase 中 RowFilter 和 BloomFilter 的原理 Hbase API 中常见的比较器 Hbase 的预分区	•
			Hbase 的颁为区 Hbase 的 Compaction Hbase 集群中 HRegionServer 宕机如何解决	
		Spark 的一个主要特点是能够在内存中进行计算,即使依赖磁盘进行复杂的运算,Spark 依然比 MapReduce 更加高效。		
			Spark 简介 Spark 开发环境搭建 弹性式数据集 RDD	
		Spark Core	RDD 常用算子详解 Spark 运行模式与作业提交	
大数据篇	Spark体系		Spark 累加器与广播变量 基于 Zookeeper 搭建 Spark 高可用集群	
		Spark SQL :	DateFrame 和 DataSet  Structured API 的基本使用  Spark SQL 外部数据源	
			Spark SQL 常用聚合函数 Spark SQL JOIN 操作	
		Spark Streaming :	Spark Streaming 简介 Spark Streaming 基本操作 Spark Streaming 整合 Flume	
			Spark Streaming 整合 Flume  Spark Streaming 整合 Kafka  Flink 核心概念综述	
			Flink 开发环境搭建 Flink Data Source  Flink Data Transformation	
		Flink 概览	Flink Data Transformation  Flink Data Sink  Flink 窗口模型	
			Flink 状态管理与检查点机制 Flink Standalone 集群部署	
	Flink体系		Flink从入门到放弃(入门篇1)-Flink是什么 Flink从入门到放弃(入门篇2)-本地环境搭建&构建 第一个Flink应用	
			Flink从入门到放弃(入门篇3)-DataSetAPI Flink从入门到放弃(入门篇4)-DataStreamAPI	
			Flink集群部署 Flink重启策略	
		入门篇	Flink的分布式缓存 Flink中的窗口 Flink中的Time	
			Flink集群搭建的HA Flink中的时间戳和水印	
			Flink广播变量 Flink-Kafka-Connector	
			Flink-Table-&-SQL实战 15-Flink实战项目之实时热销排行 16-Flink-Redis-Sink	
			16-Flink-Redis-Sink 17-Flink消费Kafka写入Mysql  Kafka 是最初由 Linkedin 公司开发,是一个分布式、支持分区的(partition)、多副本的(	
			布式、支持分区的(partition)、多副本的(replica)的分布式消息系统,它的最大的特性就是可以实时的处理大量数据以满足各种需求场景  Kafka 的特性和使用场景	
			Kafka 中的一些概念:Leader、Broker、 Producer、Consumer、Topic、Group、 Offset、Partition、ISR	_
			Kafka 的整体架构 Kafka 选举策略 Kafka 读取和写入消息过程中都发生了什么	
	消息队列篇	kafka	Kafka 读取和与八肩思过程中都发生了什么  Kakfa 如何进行数据同步(ISR)  Kafka 实现分区消息顺序性的原理	- - -
			消费者和消费组的关系 消费 Kafka 消息的 Best Practice (最佳实践) 是怎样的	_
			Kafka 如何保证消息投递的可靠性和幂等性 Kafka 消息的事务性是如何实现的	_
			如何管理 Kafka 消息的 Offset  Kafka 的文件存储机制  Kafka 是如何支持 Exactly-once 语义的	- -
			Kafka 是如何支持 Exactly-once 语义的 通常 Kafka 还会要求和 RocketMQ 等消息中间 件进行比较	_
	Zookeeper	基础	Zookeeper 简介及核心概念  Zookeeper 单机环境和集群环境搭建  Zookeeper 常用 Shell 命令	
			Zookeeper ACL 权限控制	
	链接介绍: https://github.com/ wangzhiwubigdata/God-Of-BigData			

大数据架构

计算机网络

操作系统