



▶ 直播时间 11月10日（周四）晚8点-9点 | ⌚

大数据面试题干货盘点



大数据教育开拓者
尚硅谷最帅的男人（自称）

大海哥带你轻松收割名企Offer

- ▶ 大厂大数据面试题综合分类讲解
- ▶ 经典大数据框架的核心面试题盘点讲解
- ▶ 大数据重点项目的常见实战类面试题精讲
- ▶ 大数据新兴技术有哪些
- ▶ 大数据就业前景分析

获取资料群号：
376909204

课程咨询微信：
w13021299785

直播课20点准时开讲！

大数据面试题干货盘点—— 大海哥带你轻松收割名企Offer



江湖人称：大海哥

目录



一

经典大数据框架的核心面试题

二

大数据新兴技术有哪些

三

大数据重点项目的常见面试题

四

大厂大数据面试题综合分类

五

大数据就业前景分析

六

总结答疑



经典大数据框架的核心面试题



一、经典大数据框架的核心面试题



尚硅谷大数据技术之高频面D



尚硅谷大数据技术之高频p



1、Linux

1) Linux常用高级命令

2、Hadoop

1) HDFS读写流程

2) HDFS小文件危害及解决办法

3) Shuffle及其优化

4) Yarn工作机制

5) Yarn中各个调度器特点及生产环境中怎么选择

3、Zookeeper

1) Zookeeper非第一次选举机制

2) Zookeeper符合CAP法则中哪两个



4、Flume

- 1) 解释一下零点漂移产生的原因及解决办法（Flume）

5、Kafka

- 1) **Kafka**生产者发消息流程
- 2) **Kafka**的**Broker**工作流程
- 3) **Kafka**的消费者组消费流程
- 4) **Kafka**挂了如何处理
- 5) **Kafka**怎么保证数据不丢
- 6) **Kafka**数据重复如何处理
- 7) **Kafka**数据积压如何处理



5、Kafka

- 8) **Kafka**如何保证数据有序or怎么解决乱序
- 9) **Kafka**怎么做到的高效读写
- 10) **Kafka**如何提高吞吐量
- 11) **Kafka**消费策略（**Range**、**RoundRobin**、粘性）

6、Hive

- 1) **Hive**优化
- 2) **Hive**数据倾斜
- 3) 怎么将**Hive SQL**转换为可以执行的**MR**



7、Spark

- 1) Spark 转换算子10个
- 2) Spark 行动算子5个
- 3) Spark任务怎么切分
- 4) Spark提交流程
- 5) SortShuffle原理
- 6) 统一内存模型

8、MaxWell

- 1) MaxWell底层原理及为什么选择
- 2) MaxWell怎么产生的重复数据，如何解决



9、DataX

1) **DataX**在使用过程中遇到哪些问题，怎么解决的

10、Flink

1) **Flink**的架构有哪些角色

2) **Flink**与**Spark Streaming**的区别

3) 介绍一下时间语义，谈谈你对**Watermark**的理解

4) 窗口的分类、划分、生命周期

5) **Flink**如何解决乱序问题

6) 介绍一下**Flink**的状态和状态后端

7) 有没有状态比较大的作业，遇到过什么问题，怎么解决

8) **Flink**有没有遇到过反压，怎么发现、定位、分析、解决，效果如何



10、Flink

9) **Flink**如何保证数据精准一次

10) 你是如何理解**Flink**的**Checkpoint**

11) **Task**的重启策略

12) **Flink**有几种**Join**，详细说明特点

13) 你对**keyby**算子的理解

14) 你对**Interval Join**的理解

15) **Flink**的提交参数如何设置，怎么考虑的

16) **Flink**的提交流程

17) **Flink**的内存模型

18) **Flink**有没有遇到过数据倾斜，怎么发现、定位、分析、解决，效果如何



11、HBase

- 1) HBase存储结构
- 2) HBase的读、写流程
- 3) HBase热点问题如何解决

12、ClickHouse

- 1) ClickHouse的优势
- 2) Flink写入ClickHouse怎么保证一致性?
- 3) Clickhouse的常用引擎



大数据新兴技术有哪些



- 大规模并行分析（MPP）数据库：Doris
- 数据湖：Hudi、Iceberg、Data Lake
- 云原生：K8S
- 时序数据库：InfluxDB、TDengine



大数据重点项目的常见面试题

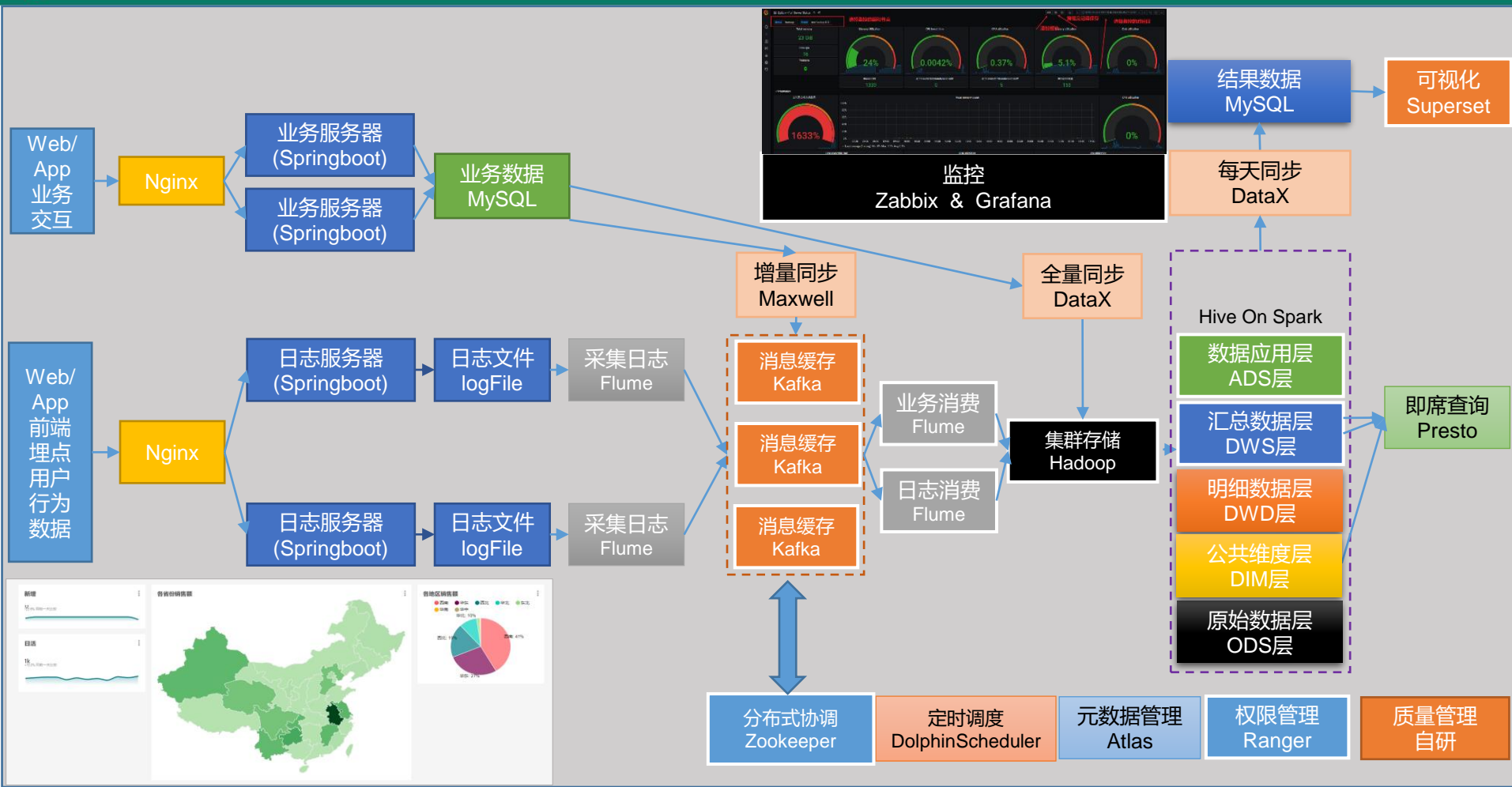


1、离线数仓项目

- 1) 为什么要做该项目?
- 2) 项目建模
- 3) 项目架构
- 4) 项目中遇到的问题及怎么解决
- 5) 项目数据量
- 6) 业务场景



三、大数据重点项目的常见面试题



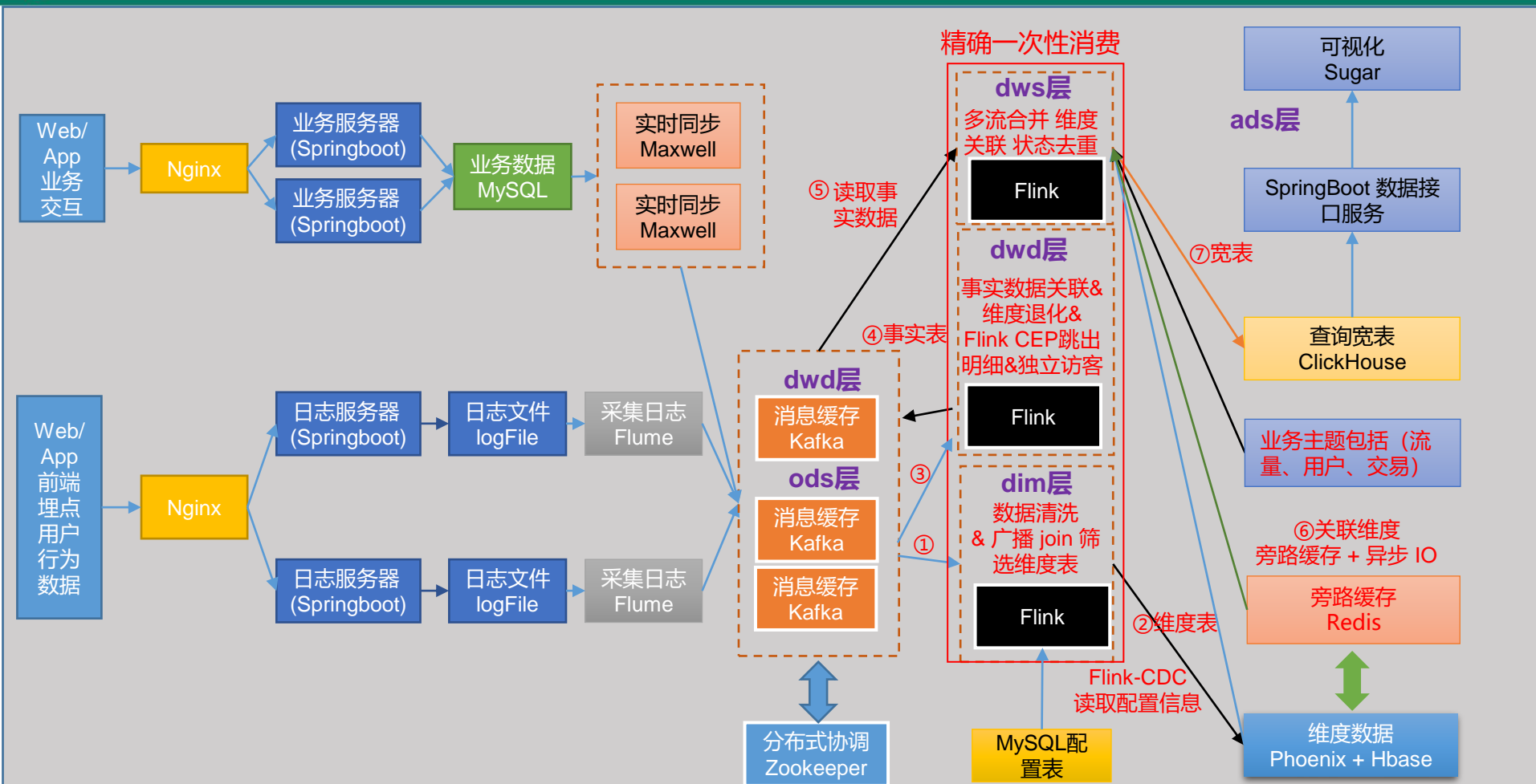


2、Flink实时数仓项目

- 1) 为什么要做该项目?
- 2) 项目建模
- 3) 项目架构
- 4) 项目中遇到的问题及怎么解决
- 5) 项目数据量
- 6) 业务场景



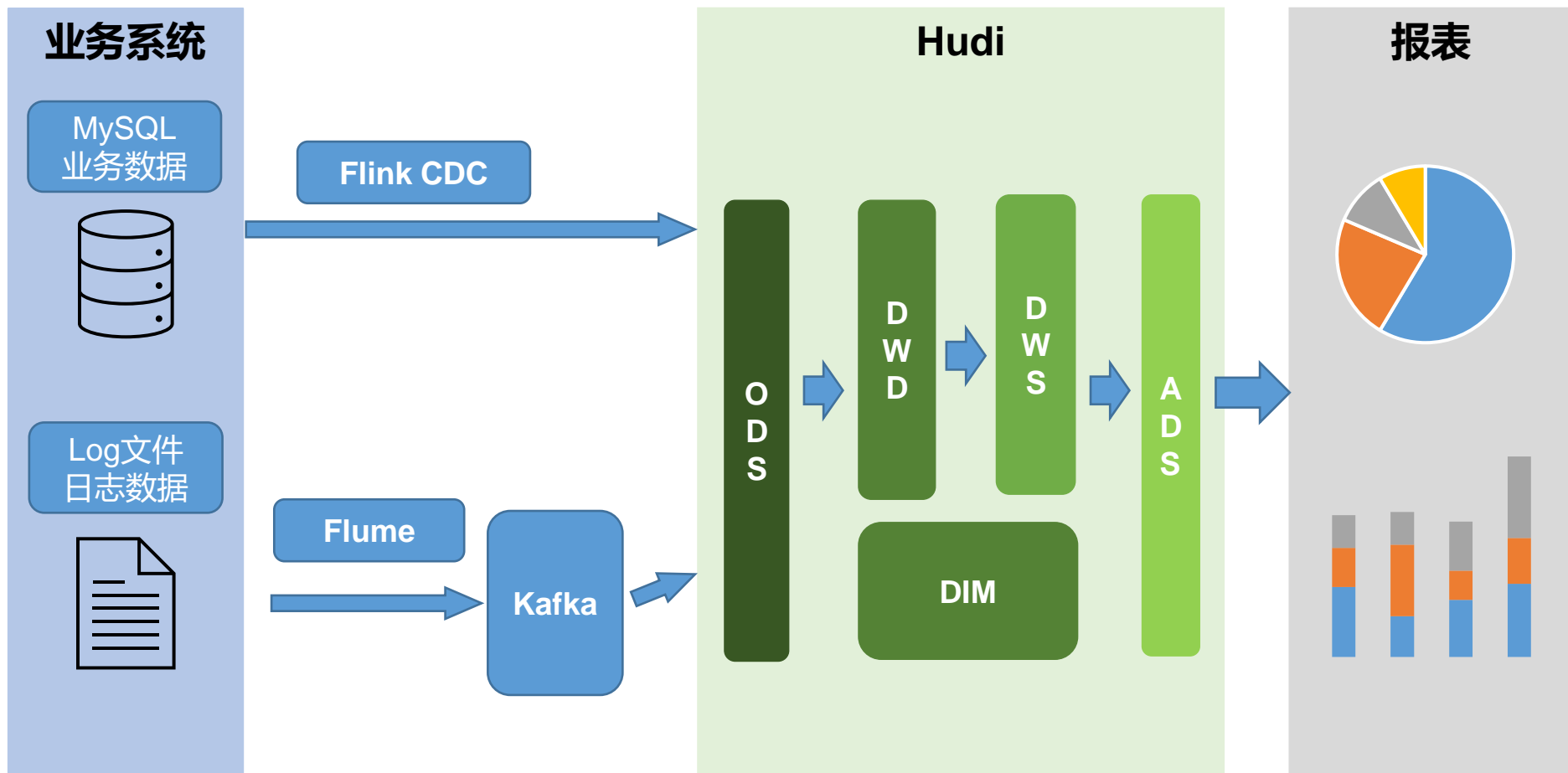
三、大数据重点项目的常见面试题





3、数据湖项目

- 1) 为什么要做该项目?
- 2) 项目建模
- 3) 项目架构
- 4) 项目中遇到的问题及怎么解决
- 5) 项目数据量
- 6) 业务场景



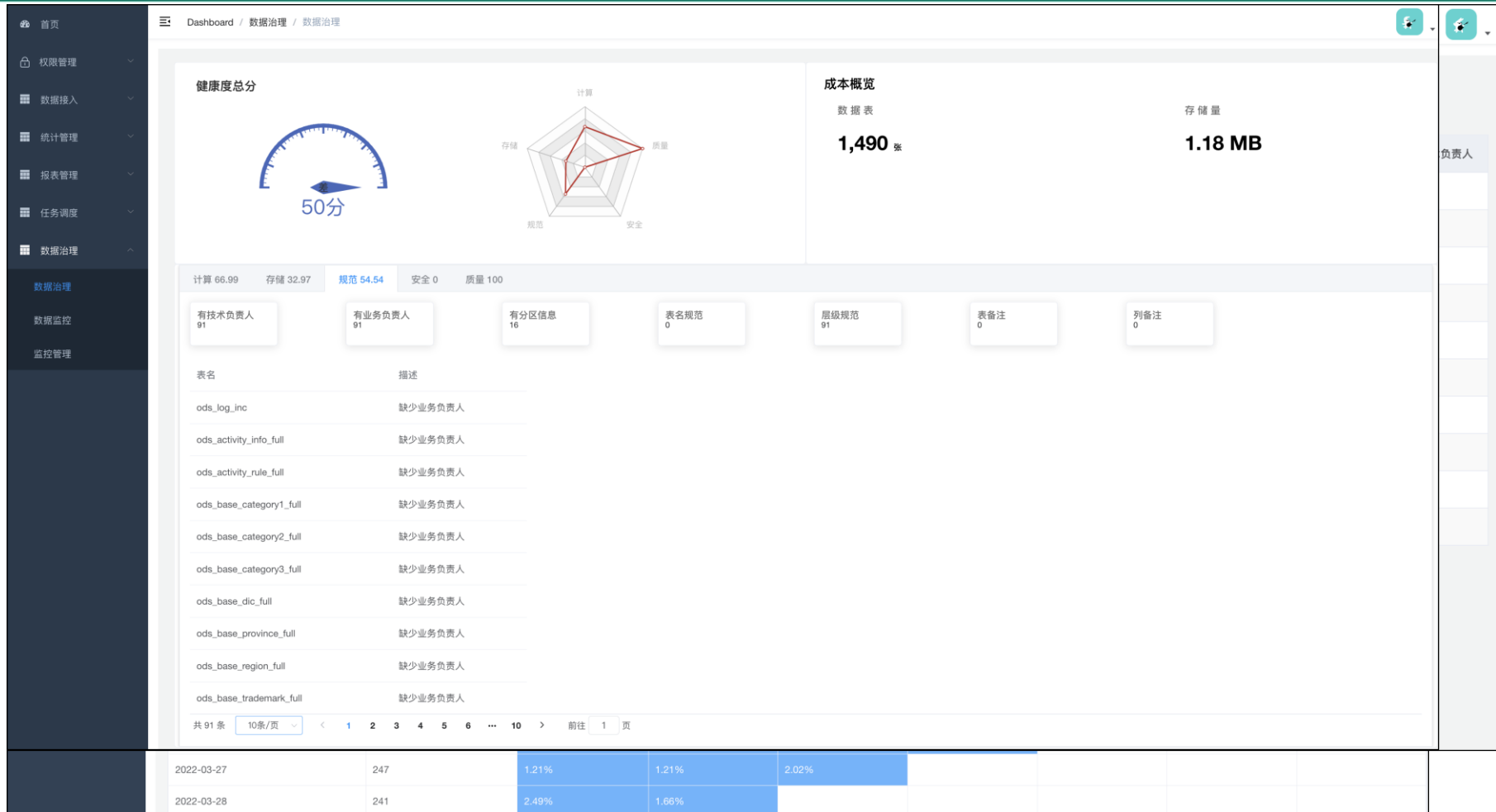


4、技术中台项目

- 1) 元数据管理
- 2) 数据质量监控
- 3) 权限管理
- 4) 数据治理



三、大数据重点项目的常见面试题





四

大厂大数据面试题综合分类



1、技术框架

2、项目介绍

3、项目遇到问题

4、业务场景题



2022年北京大数据面



2022年上海大数据面



2022年深圳大数据面

5、HQL题

6、算法题



五

大数据就业前景分析



送给正在努力的你

荀子-修身

道阻且长，行则将至。

行而不辍，未来可期。

岁月代伤，亦有光芒。

眼中有山河万里，何惧几分秋凉。



就业直播嘉年华

· 从简历到面试一路通关 ·

· 挑战高薪 冲击大厂 ·



获取录播视频及资料群号：

376909204

或加柳儿姐微信：



课程咨询微信：

w13021299785



谢谢观看