

尚硅谷大数据项目之 电商数仓



江湖人称：大海哥

目录



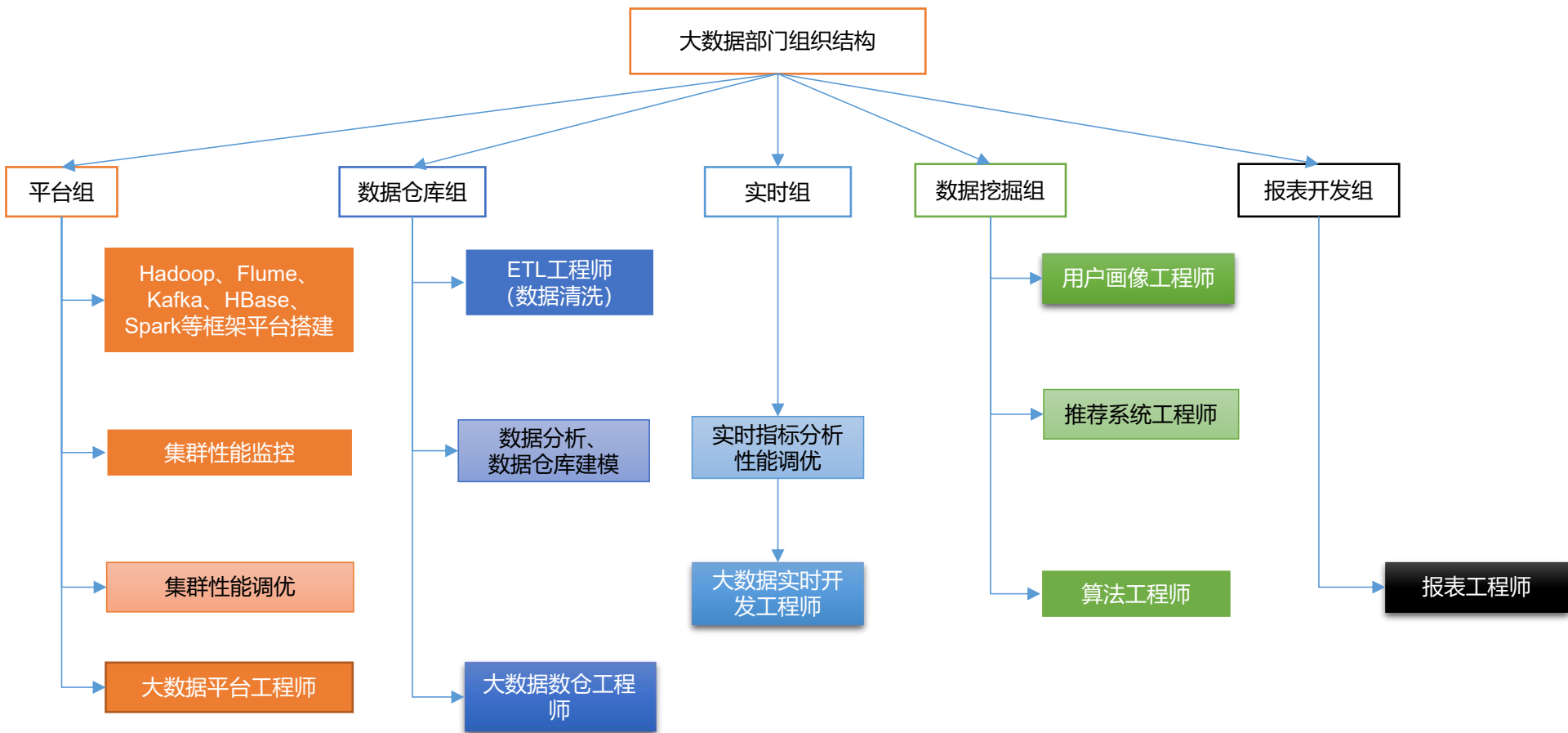
为什么要学习该项目



该项目有哪些亮点



一、为什么要学习该项目





➤ 一、覆盖当前主流框架

- 采集：Flume/Kafka/Sqoop
- 存储：MySQL/Hadoop/HBase
- 计算：Hive/Tez
- 查询：Presto/Druid/Kylin
- 可视化：Superset
- 任务调度：Azkaban
- 集群监控：Zabbix
- 元数据管理：Atlas
- 脚本：Shell

- 关注尚硅谷教育公众号，回复：大数据
(获取尚硅谷大数据全套免费资料)





➤ 二、包含的业务指标：近100个

日活、周活、月活、新增、沉默、回流、流失、留存数、留存率、最近连续三周活跃、最近七天连续三天活跃、漏斗分析、spu个数、sku个数、

商品销量排名、商品收藏排名、商品加入购物车排名、商品退款率排名、商品差评率、单日下单笔数、单日下单金额、单日下单用户数、

单日支付笔数、单日支付金额、单日支付人数、单日支付商品数、下单到支付的平均时长、单次复购率、多次复购率、购买人数、多次购买人数、


累积登录天数、最近30日登录天数、累积下单次数、累积下单金额、最近30日下单次数、最近30日下单金额、累积支付次数、累积支付金额、

最近30日支付次数、最近30日支付金额、累积退款次数、累积退款件数、累积退款金额、最近30日被加入购物车次数、最近30日被加入购物车件数、

累积被加入购物车次数。。。。。



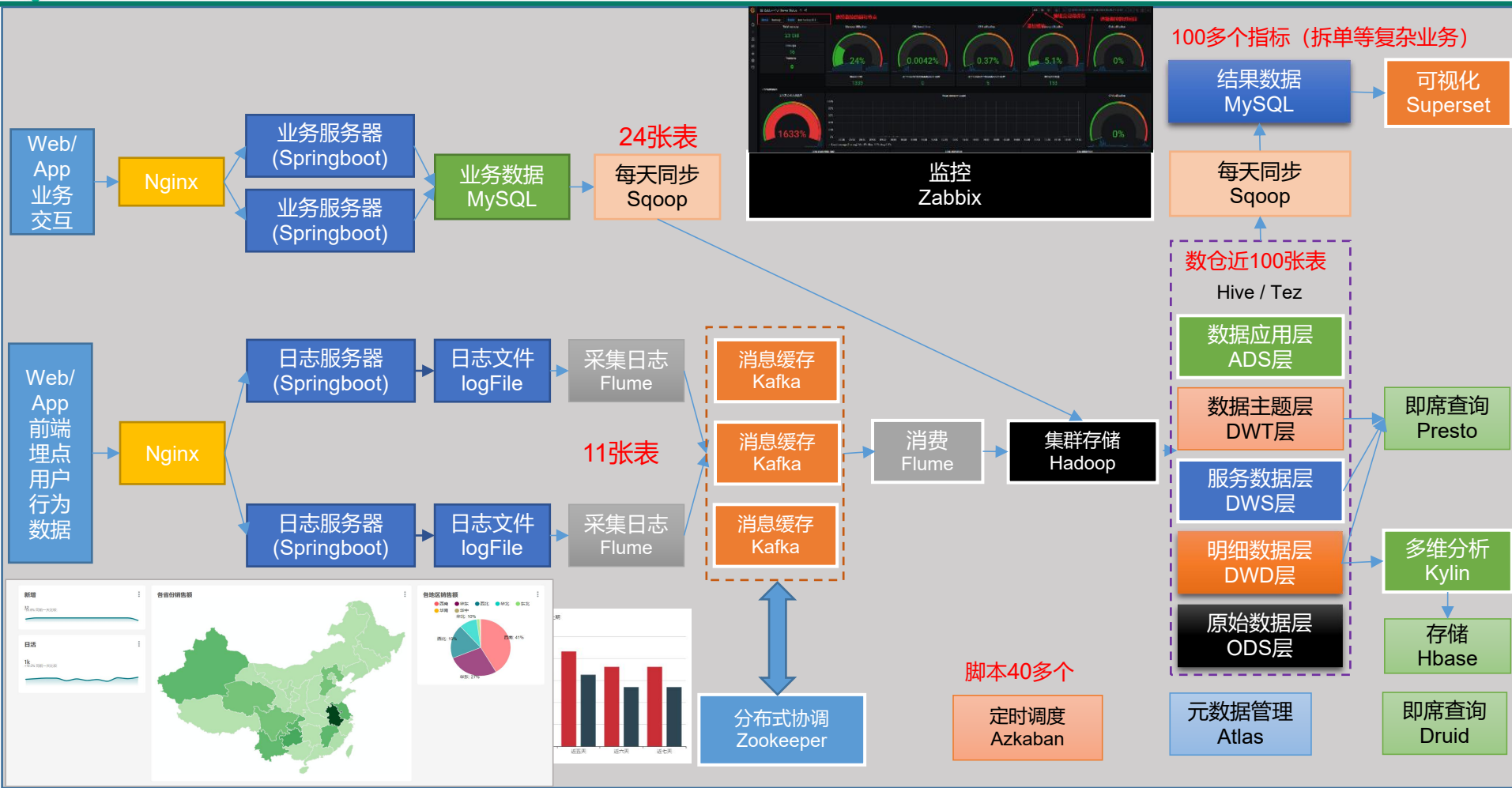
- 三、Shell脚本40多个
- 四、用户行为原始表：11张
- 五、业务表原始表：24张
- 六、数仓总表：近100张
- 七、标准的数仓建模理论（确定业务过程、声明粒度、确定维度、确定事实）
- 八、主流可视化技术：Superset
- 九、专业的集群监控组件：Zabbix
- 十、当前最流行的元数据管理：Atlas



名企收割机!
培训小清华!



二、该项目有什么特点





谢谢观看