数据平台中心建设规划

数据仓库

现在

1. 主机分布

Node407-14-------8台主机

2. node40\*CDH集群角色

Node4014

NameNode

Impalad

Node4013

NameNode

Impalad

Node4012

NameNode

Impalad

Node4011

NameNode

Impalad

Node4010  
DataNode

Impalad

Node409

NameNode

Node407

（主master）

管理节点

Node408

NameNode

Yarn

3. 数据入仓：

1） mysql gemini库

2） mongoDB

3） ElasticSearch日志数据： 微信+app日志

4） mysql 线下数据库

4. 整理数据运行任务

Gabdp：数据靠gabdp-wf配置管理；

魔方：数据是shell脚本

5. 迁移集群方案：

进销存库

进销存

Gemini库

Mongo

Gabdp-log库

ElasticSearch  
日志

Gemini

Mysql+Canel

Gemini库

HIVE

具体实施步骤：

1. Node40\* CDH集群部署工作流调度系统；--- 邹有情（1天）
2. Demo工作流调度系统；（1h）
3. 迁移gabdp相关业务：修改新的库以及创建新的表；--- 刘光+邹有情（2-3天）
4. 配置魔方相关的数据业务 --- 邢成（2天）
5. 修改魔方涉及日志汇总统计任务 --- 李扬扬（1天）
6. 试运行：停原有shell脚本由工作流调度系统接管。（容灾就是shell脚本手动运行）
7. 持续开发sqlserver数据入仓的任务节点。--- 邹有情