

一次请求的流程：

架构设计部署时的优化：

Impada damen 和 statestore， catalog 放在不同主机中。

因为impada demen执行计算，内存消耗大 对集群的监看检查，catlog完成 源数据信息同步。他俩放在一起，减少IO开销。

Impada damen 和datanode放在一起，快速。

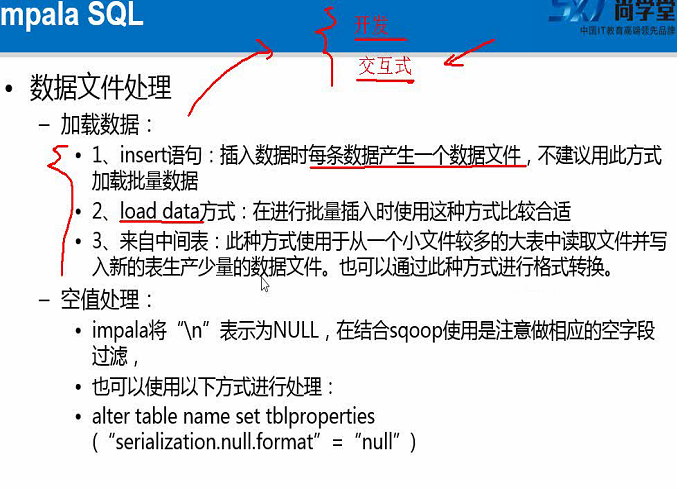
Query Planner sql解析节点。

Query Coordinato 中心协调节点。

Query Executor 计算节点

优化其他两点：分区，搜索内存优化！（具体的下课问同学）

优化，中间表，解决小文件问题。



空值的处理！



内部操作：

impala-shell

进入默认端口21000。然后再输入命令。

外部命令：

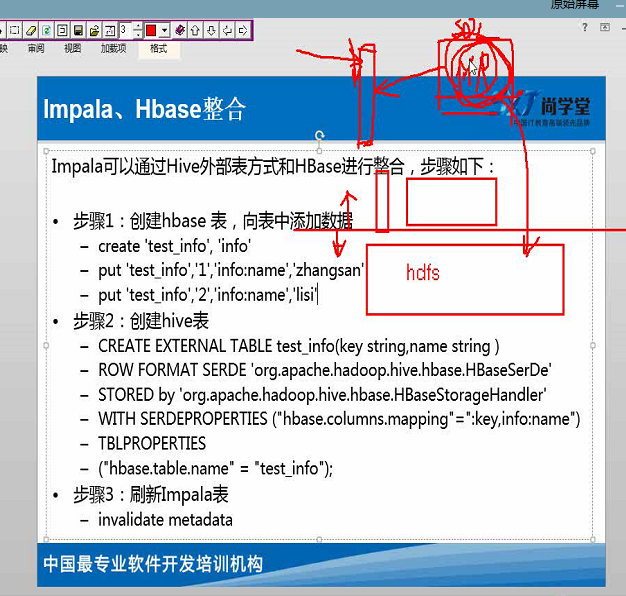
Impala shell - 在外部操作impala

1. 默认详细输出。 Impala-shell --quiet 关闭详细输出

1. -p 显示详细执行计划 （开启？
2. -r 刷新所有元数据。 新增一个表。Hive直接显示，但是impala需要外部命令刷新，才能显示。（新增数据是同步的，新增表，需要刷新）



Impala不支持以上三种。



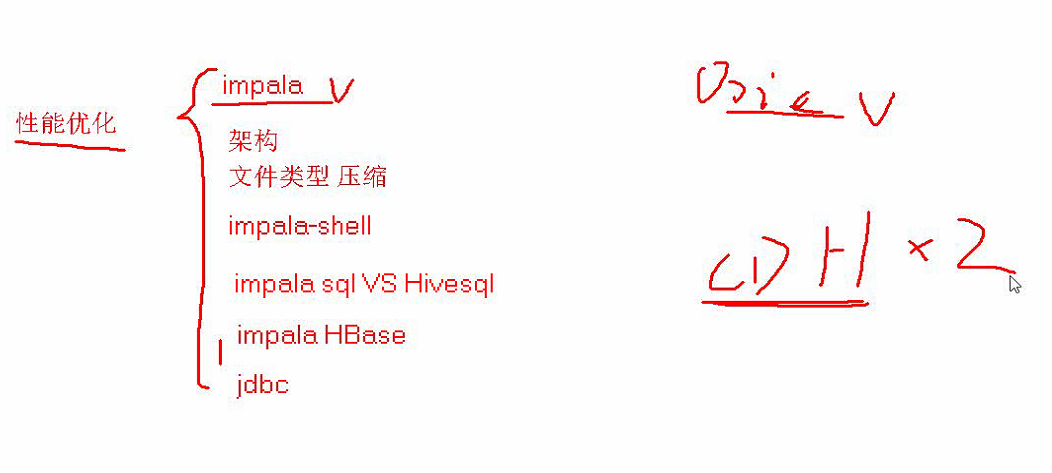
Hbase的操作一般是数据存储和查询。一些简单的操作。

对hbase上面数据进行分析，最好使用hive，它可以写一些函数，来对数据进行一些分析。

使用hive将impala和hbase整合。那么连分析速度，也提高了。

JDBC访问：

Oziee:



庞磊 18:05:00 (多人发送)

ansible

安装

1、配置源

http://mirrors.aliyun.com/help/centos

http://mirrors.aliyun.com/help/epel

2、安装 yum install ansible

1、模块方式

2、playbooks

文档

http://docs.ansible.com/ansible/

http://www.ansible.com.cn/