大数据平台快速解决方案

SPEAKER

李锡铭

内容概要

搭建始末

技术概览

学习与使用路线

why?

关系数据库存储性能问题 IO

纯java计算能力遇瓶颈,要额外处理多线程 锁

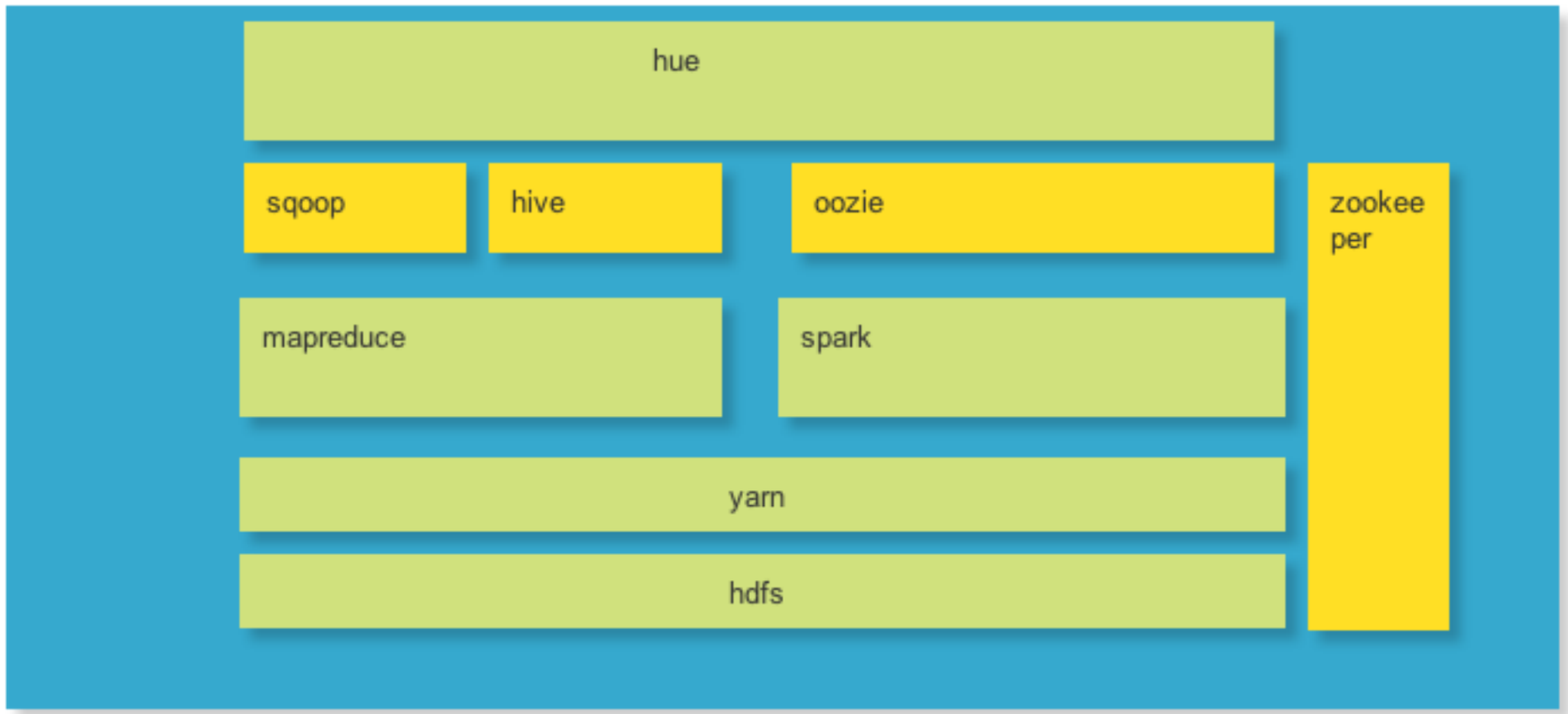
需求越做越慢

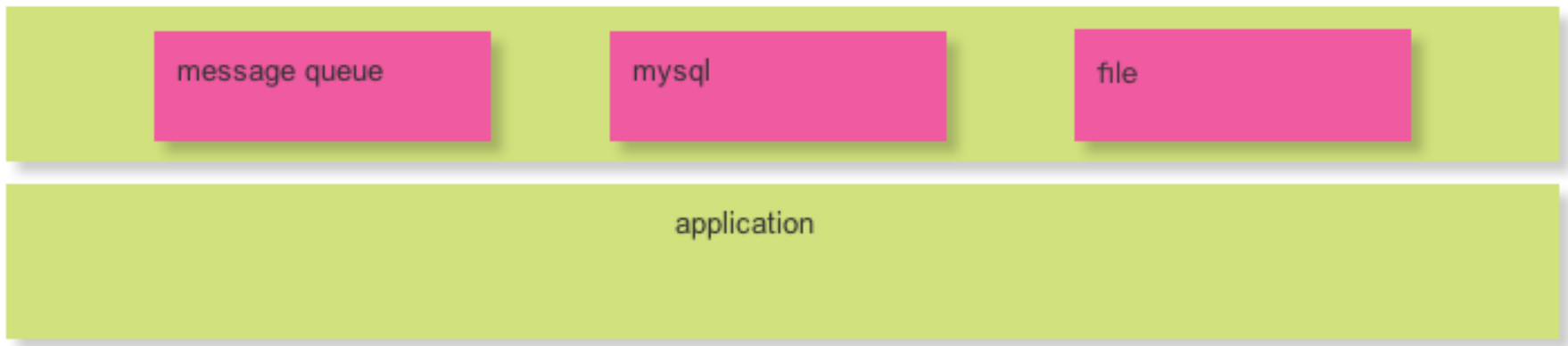
存储　计算

大数据

What Is Apache Hadoop?
The Apache™ Hadoop® project develops open-source software for reliable, scalable, distributed computing

底层概念　　衍生产品

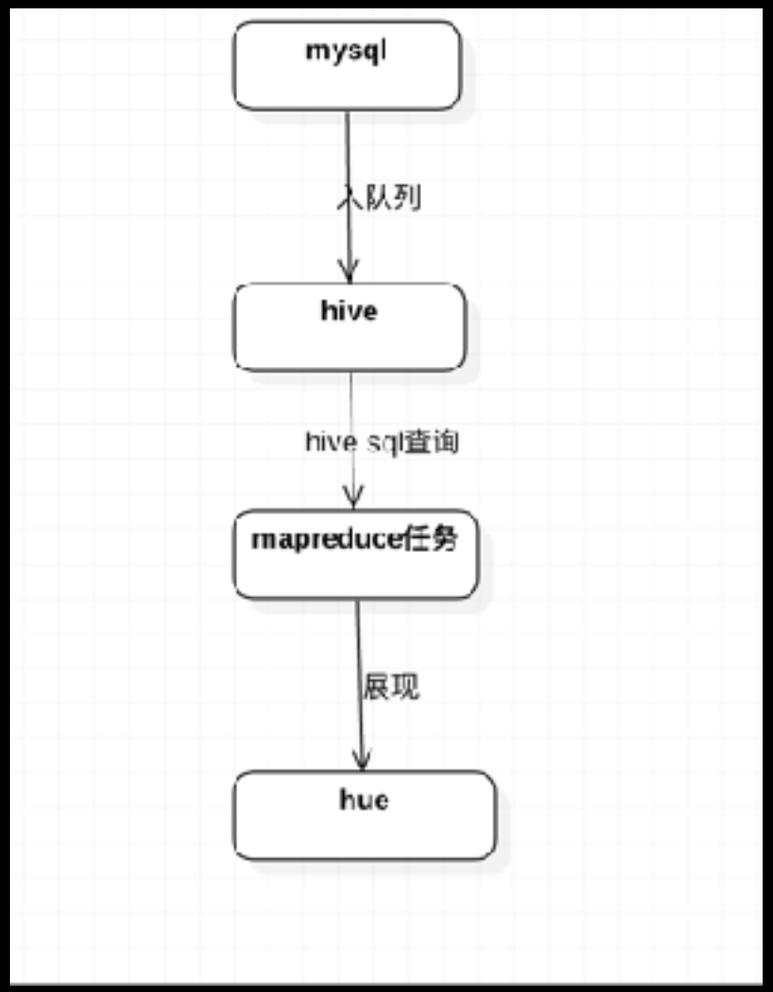


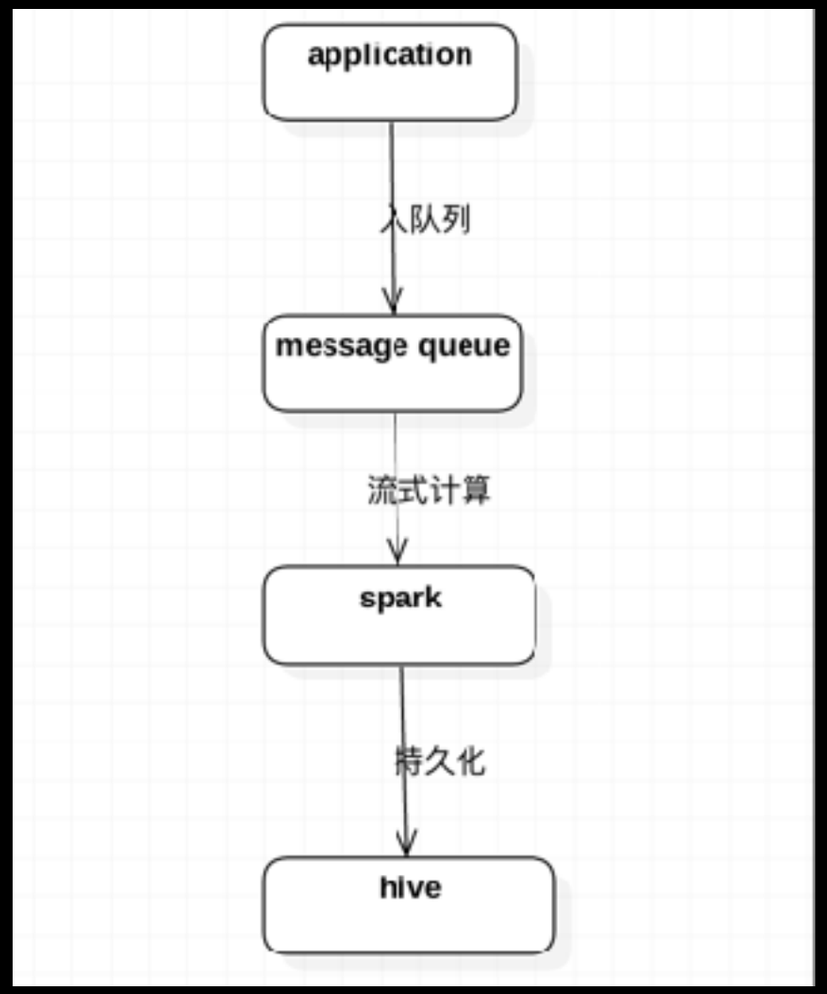


组件分类

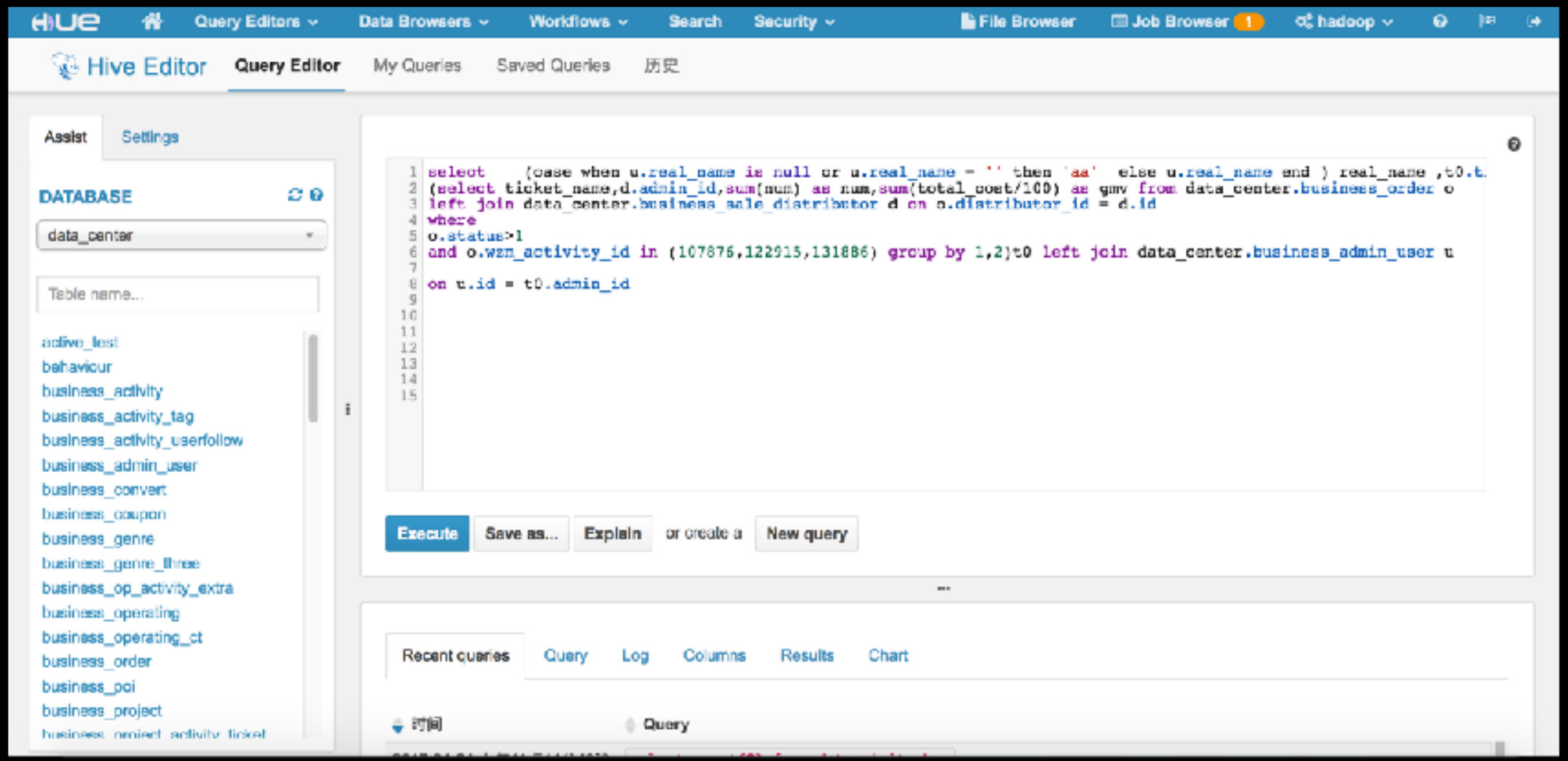
基础数据:mysql,file
大数据存储:hdfs,hive
大数据计算:mapreduce,spark,sqoop
大数据协调与调度:yarn,zookeeper,oozie
大数据展现:hue

典型执行流程

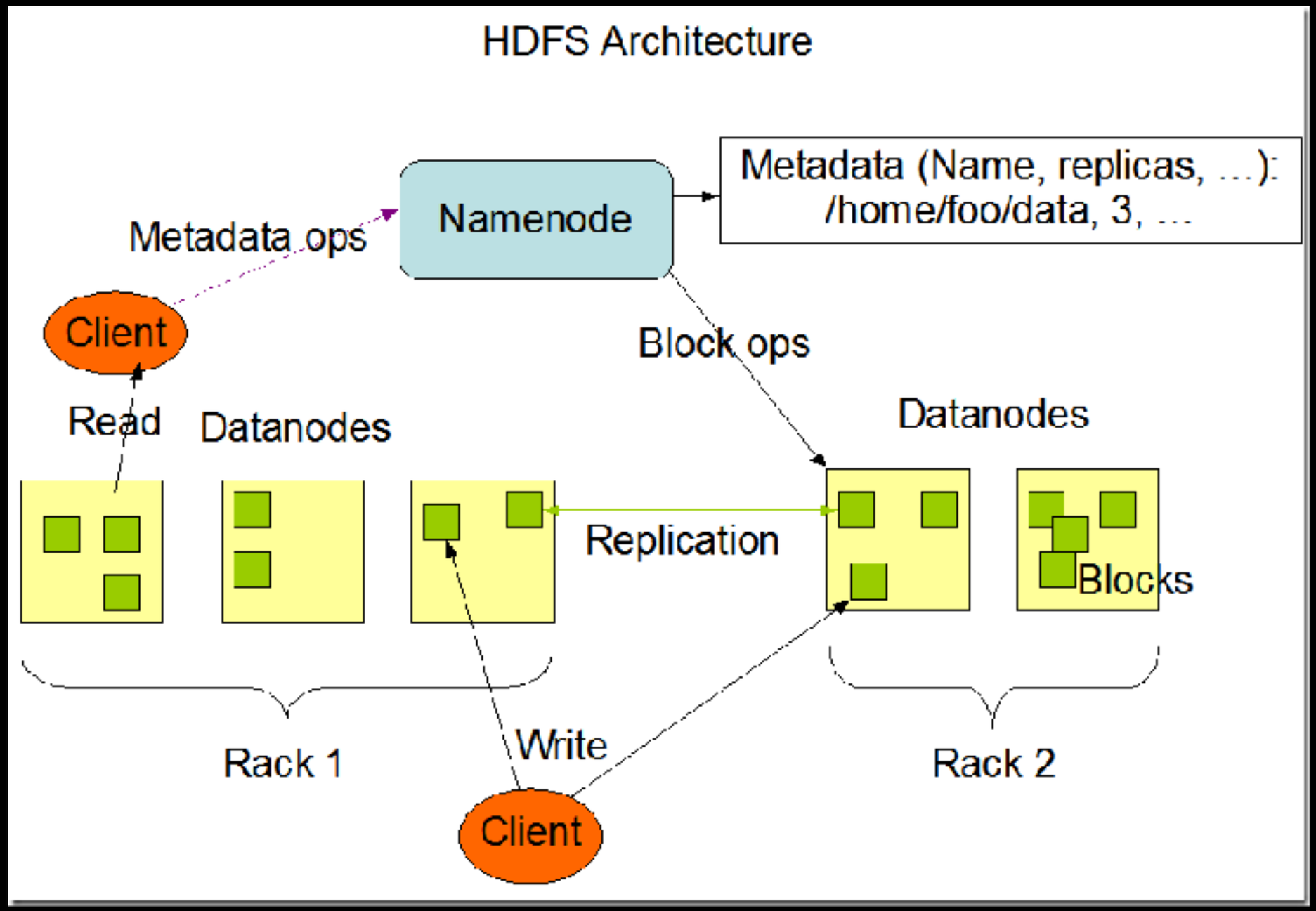

mysql
入队列
hive
hive sql查询
mapreduce任务
展现
hue


application
入队列
message queue
流式计算
spark
持久化
hive

Hue


HUE
Query Editors
Data Browsers
Workflows
Search
Security
File Browser
Job Browser
hadoop
Hive Editor
Query Editor
My Queries
Saved Queries
历史
Assist
Settings
DATABASE
data_center
Table name...
adlive_test
behavicur
business_activity
business_activity_tag
business_activity_userfollow
business_admin_user
business_convert
business_coupon
business_genre
business_genre_three
business_op_activity_extra
business_operating
business_operating_ct
business_order
business_poi
business_project
select (case when u.real_name is null or u.real_name = '' then 'aa' else u.real_name end) real_name ,t0.t
(select ticket_name,d.admin_id,sum(num) as num,sum(total_cost/100) as gmv from data_center.business_order o
left join data_center.business_sale_distributor d on o.distributor_id = d.id
where
o.status>1
and o.wsm_activity_id in (107876,122915,131886) group by 1,2)t0 left join data_center.business_admin_user u
on u.id = t0.admin_id
Execute
Save as...
Explain
or create a
New query
Recent queries
Query
Log
Columns
Results
Chart
时间
Query

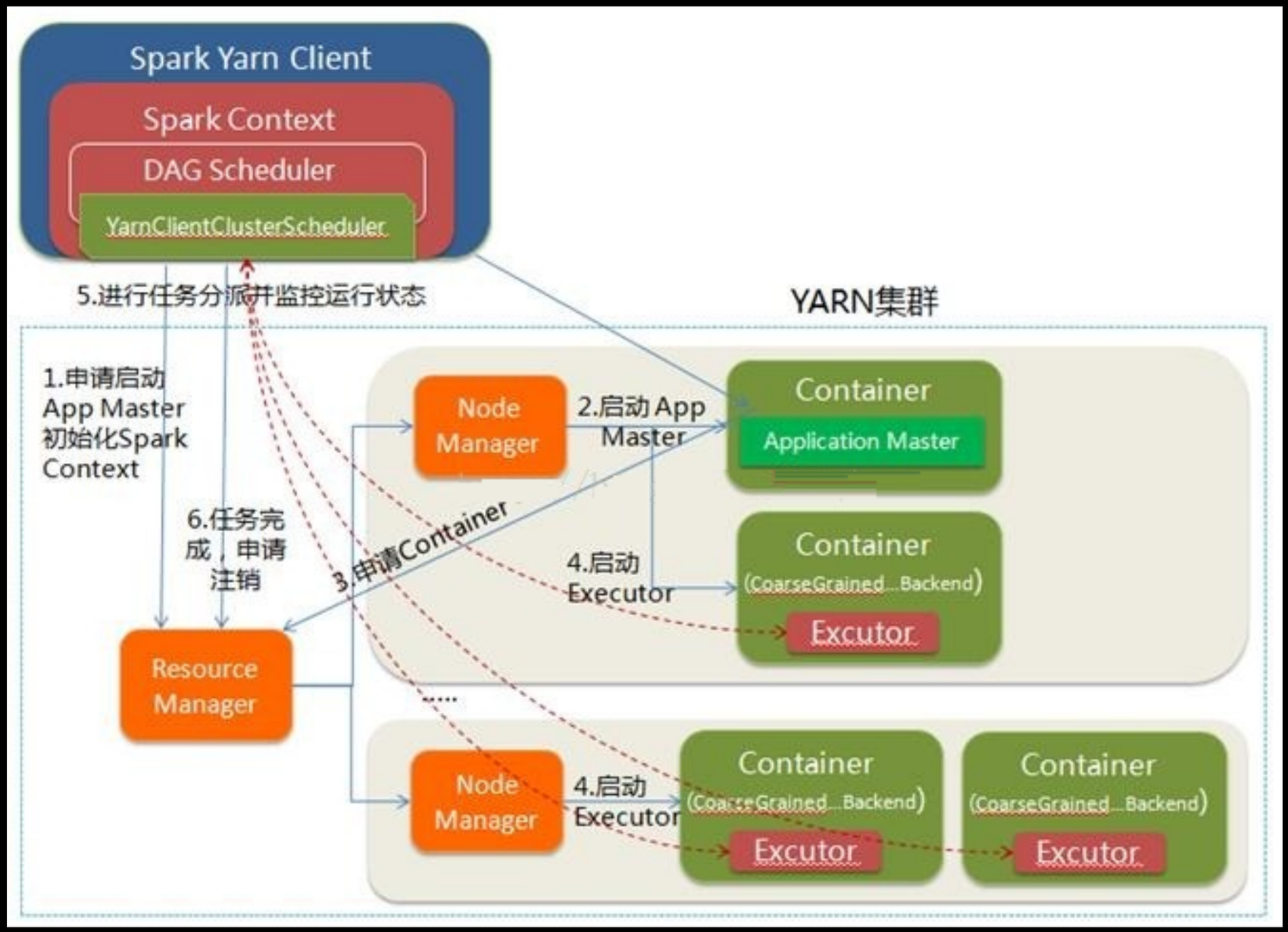
存储:hadoop hdfs:The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware



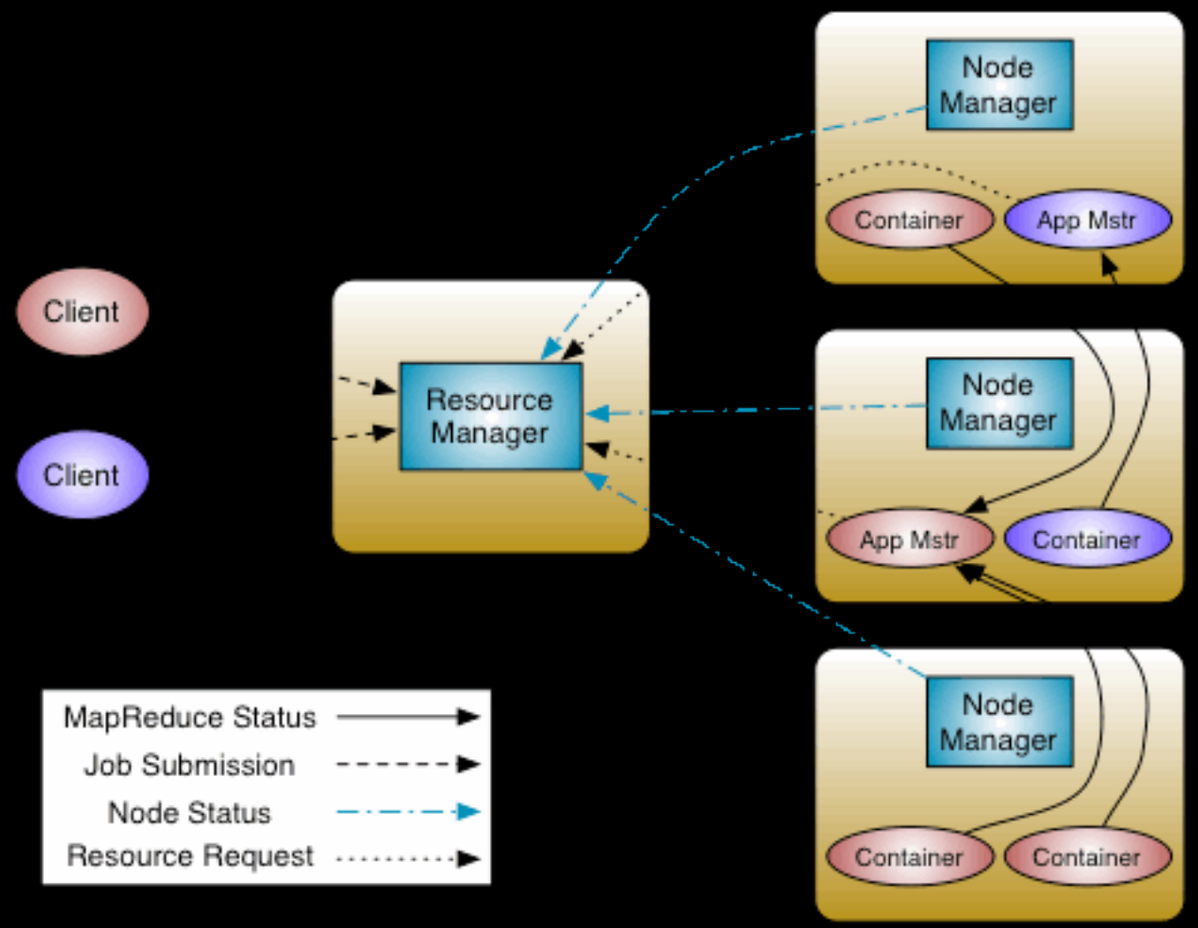
计算:mapReduce &spark

mapReduce:hadoop原生计算框架

spark:更全面,更快

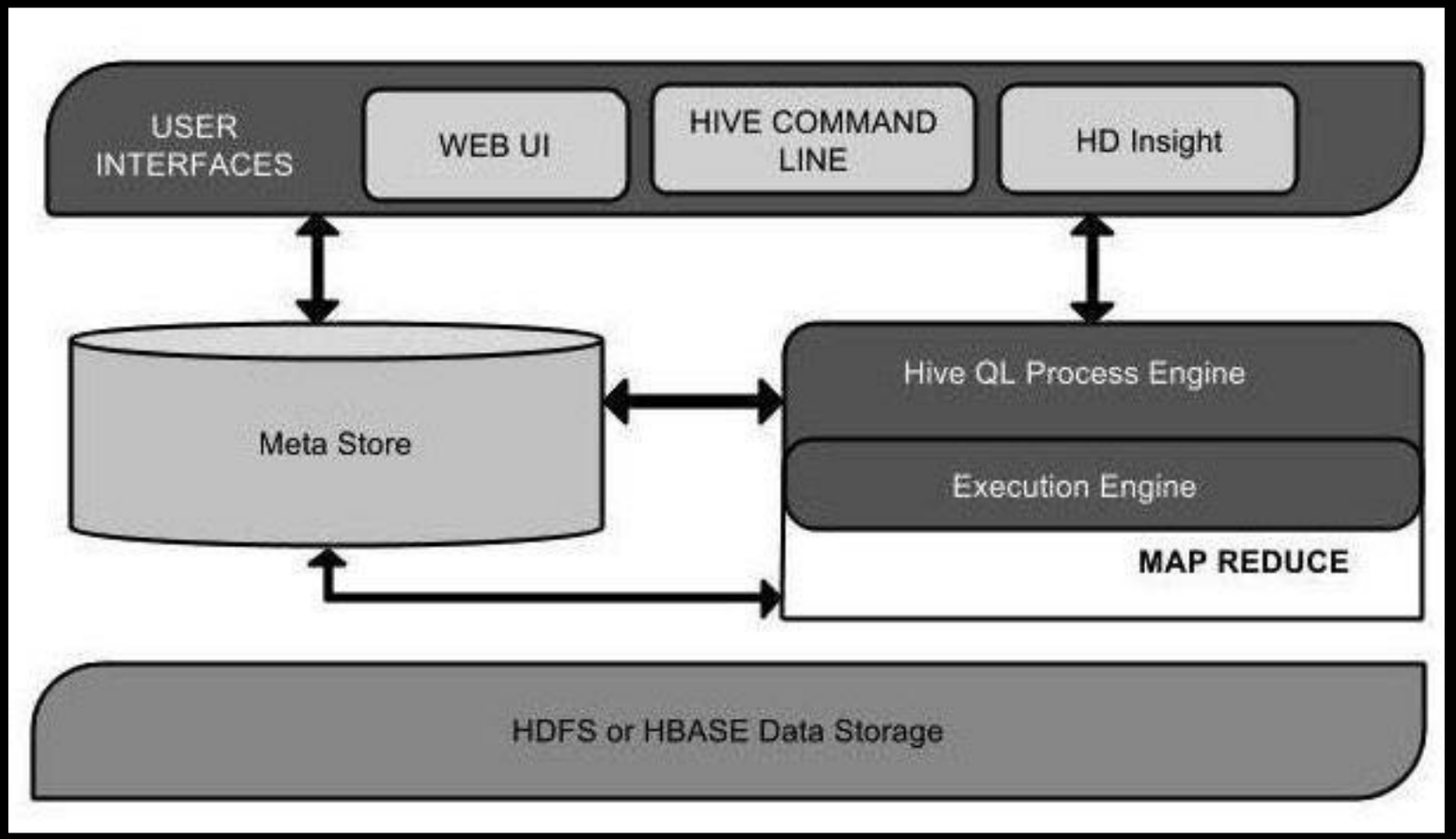


资源管理器:yarn,Apache Hadoop YARN （Yet Another Resource Negotiator）



hive

Hive是一个数据仓库基础工具在Hadoop中用来处理结构化数据,以关系型数据库的操作习惯操作hdfs



sqoop:主要用于在Hadoop(Hive)与传统的数据库(mysql、postgresql...)间进行数据的传递

oozie:大数据任务编排调度

学习与使用路线

掌握基础->快速实践->错误中前行->补充知识

Q&A