

# 特此声明

本视频由whoami博主（博客：[itweet.cn](http://itweet.cn)） Hadoop十周年  
庆开始公益录制，内容涉及

《Enterprise\_Hadoop\_Solutions》，声明此教程使用大部分  
内容均来自网络，版权属于广大网友；任何个人或企业未经  
授权不得作商业用途。持续关注，扫描微信公众号。





作者: whoami

# 现有几大Hadoop平台比较

## Hortonworks HDP 2

### - 特点

纯开源

原装Hadoop(做了rpm包)

版本跟进快

Hive+tez(主推)

企业级安全(不开源)

监控集成方案:

ganglia+nagios

Spark合作(原装)

贡献Hadoop源代码为主?

权限控制目前最好? 高度可视化?

### - 缺点

升级(坑,各种版本升级方式还不同?)

配置管理?

添加删除节点?

维护升级成本?

## Cloudera CDH 5

### - 特点

CM小白式安装

监控,使用可视化程度高

CM修改配置方便

小白式升级

资源分配设置可视化

文档非常详细

Impala(主推),维护成本低

Spark合作(功能阉割)

修复Hadoop缺陷为主(bug)

权限控制不够完善,可视化程度一般?

### - 缺点

CM不开源,企业级(Navigator付费)

版本跟进相对保守

社区版本不透明,

功能有些许小限制?

# 目录

## > HDP Cluster Install

- <http://www.itweet.cn/2015/08/31/hdp-install/>
- [http://docs.hortonworks.com/HDPDocuments/Ambari-2.2.0.0/bk\\_Installing\\_HDP\\_AMB/content/\\_ambari\\_repositories.html](http://docs.hortonworks.com/HDPDocuments/Ambari-2.2.0.0/bk_Installing_HDP_AMB/content/_ambari_repositories.html)

## › HDFS / MapReduce / Yarn / Hive / Impala / Oozie/sentry/zookeeper/Hue/sqoop2

- 进阶应用
- 集群资源分配和调优

## > Hive Tpcds

- TPC-Hive build package
- 生成测试数据

## > Hive/Impala SQL的强大功能

- Hive2.0/impala-kudu,逐渐向传统RDBMS挑战。

## > Hive vs Impala

- 几个Hints
  - Hive的Map Join
  - Impala的SHUFFLE Join ( partitioned join )

## > Presto vs sparksql vs Drill vs impala vs tez vs other

- 各种SQL-on-Hadoop的痛
- 各种限制,各种坑,稳定性成谜? 各种的营销噱头,没有解决实际问题?

# Thank you

提问时间?

**Blog:** <http://www.itweet.cn>

**PPT:** <https://github.com/itweet/course>



链接: <http://pan.baidu.com/s/1dDnpT6p> 密码: rg03

