**Distributed System**

|  |  |
| --- | --- |
| **审核人** |  |
| **重要性** | 中 |
| **紧迫性** | 中 |
| **拟制人** | 张包峰 |
| **提交日期** |  |

**作者：张包峰**

(版权所有,翻版必究)

**修改记录**

|  |  |  |
| --- | --- | --- |
| **更新时间** | **变更内容** | **变更理由** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

[**作者：张包峰** 1](#_Toc395136085)

[1.1 Paper 4](#_Toc395136086)

[1.1.1 SQL-on-Hadoop Full Circle Back to Shared Nothing DB Arch 4](#_Toc395136087)

[1.2 SQL on Hadoop 4](#_Toc395136088)

[1.2.1 Hive 4](#_Toc395136089)

[1.2.2 Shark 4](#_Toc395136090)

[1.2.3 Tajo (LinkedIn) 4](#_Toc395136091)

[1.2.4 Impala (Cloudera) 5](#_Toc395136092)

[1.2.5 Drill (MapR) 5](#_Toc395136093)

[1.2.6 Presto (Facebook) 5](#_Toc395136094)

[1.2.7 Hadapt 6](#_Toc395136095)

[1.2.8 PolyBase (Microsoft) 6](#_Toc395136096)

[1.2.9 HAWQ (Pivotal) 6](#_Toc395136097)

[1.3 Tez 6](#_Toc395136098)

[1.4 YARN 6](#_Toc395136099)

[1.5 Lambda Arch 6](#_Toc395136100)

[1.5.1 Summingbird 6](#_Toc395136101)

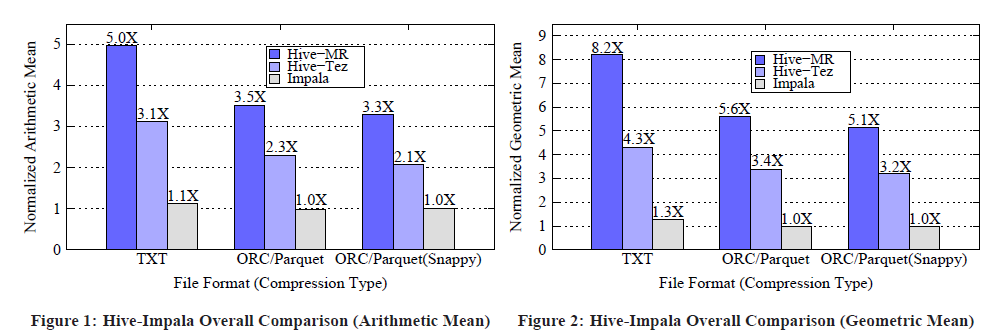
[1.5.2 Lambdoop 7](#_Toc395136102)

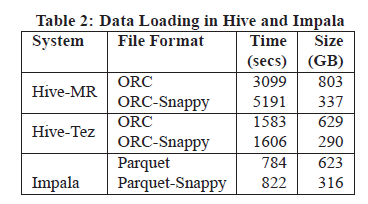
## Paper

### SQL-on-Hadoop Full Circle Back to Shared Nothing DB Arch

主要探讨Hive和Impala

TCP-H和TCP-DS测试





可以对impala做些了解和熟悉

## SQL on Hadoop

### Hive

### Shark

### Spark SQL

### Tajo (LinkedIn)

<http://tajo.apache.org/>

big data warehouse system on hadoop

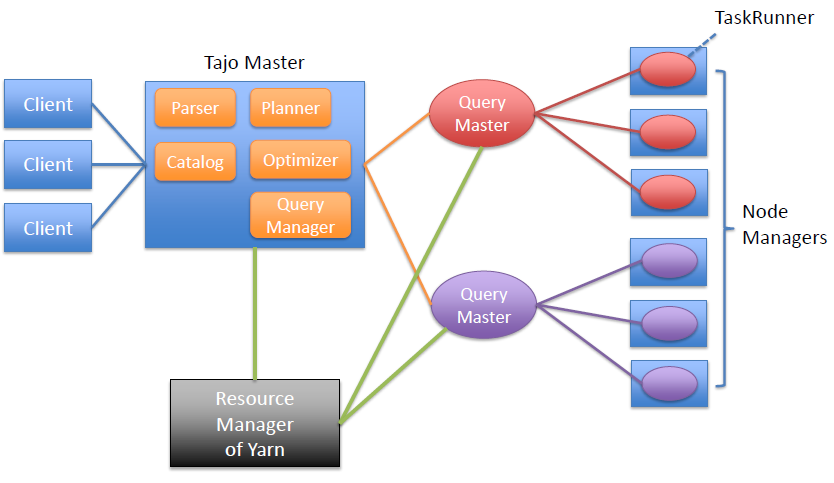
#### Get Started

<http://tajo.apache.org/docs/0.8.0/getting_started.html>

#### 设计

tajo\_intro.pdf

整体架构



### Impala (Cloudera)

### Drill (MapR)

Dremel开源实现

### Presto (Facebook)

### Hadapt

### PolyBase (Microsoft)

### HAWQ (Pivotal)

## Tez

## YARN

## Crunch

## Lambda Arch

<http://lambda-architecture.net/>

### Summingbird

### Lambdoop