

實作課程 Lesson 04

蔡秉文

Cookie Tsai

You Will Learn

1. 認識 HBase
2. 認識 Hive
 1. Hive MetaStore
 2. Hive Server 2
3. 如何使用 HBase Shell
4. 如何使用 Beeline 和 Hive Shell

Big Data 的起源 Google 三篇論文

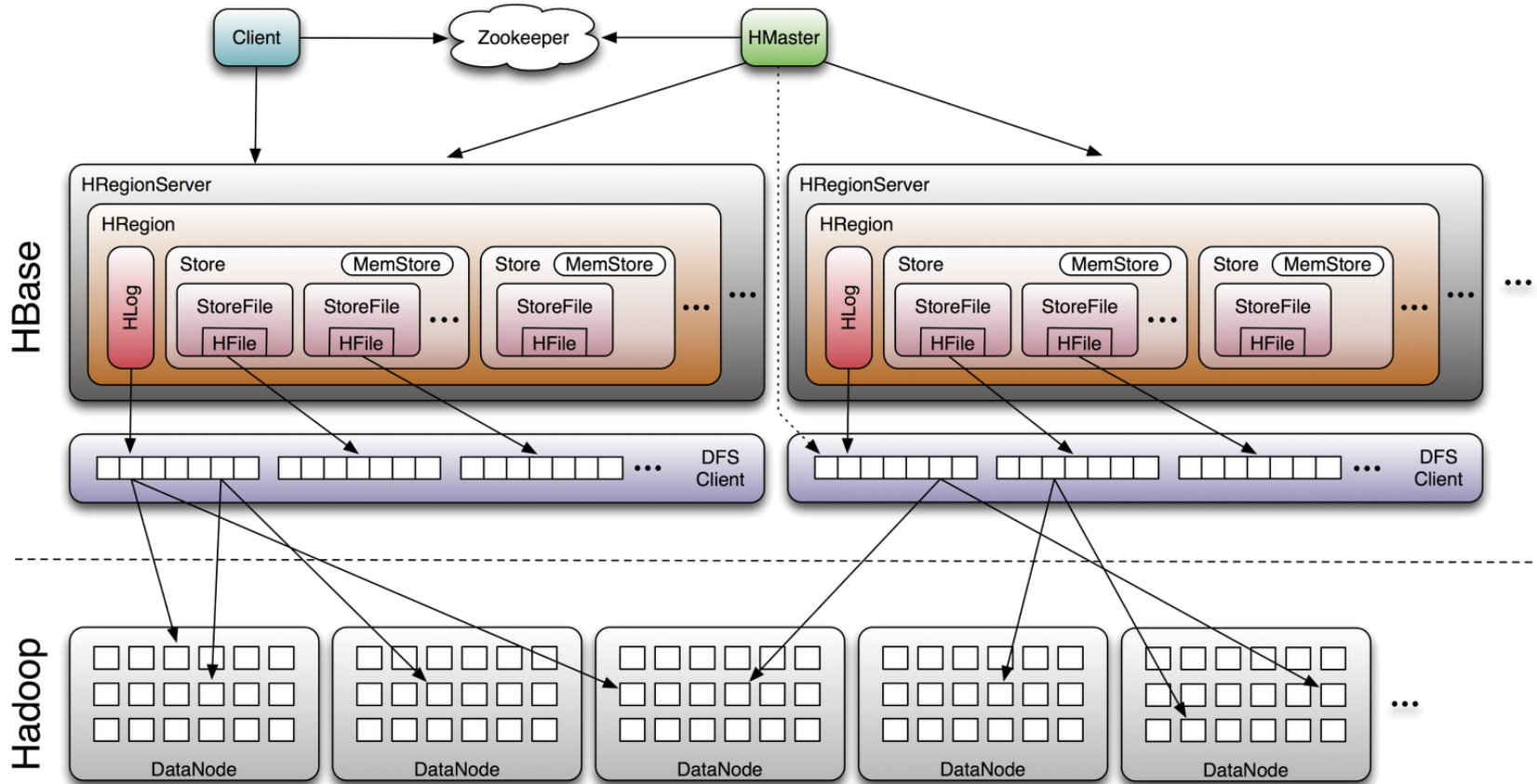
- The Google File System
<http://research.google.com/archive/gfs.html>
- MapReduce: Simplified Data Processing on Large Clusters
<http://research.google.com/archive/mapreduce.html>
- Bigtable: A Distributed Storage System for Structured Data
<http://research.google.com/archive/bigtable.html>

What is HBase

- A kind of NoSQL
- Manipulation in HDFS
- Using column family qualifier
- Each Row-Key is also a indexed column

Row-Key	Column		Timestamp	Value
	Family	Qualifier		
row1	cf	name	1442053885486	Tom
row2	cf	name	1442053885487	Mary
row2	cf	phone	1442053885487	0999XXXXXX
row3	cf	name	1442053885486	John

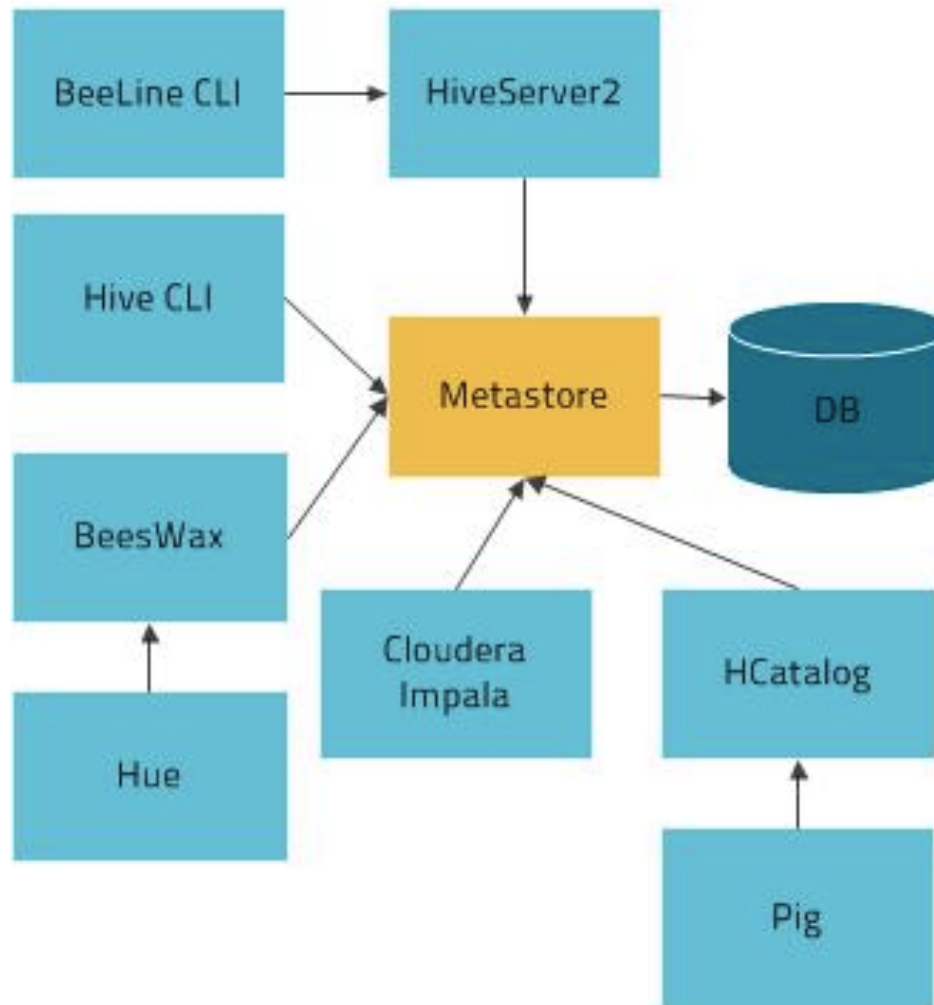
HBase Architecture



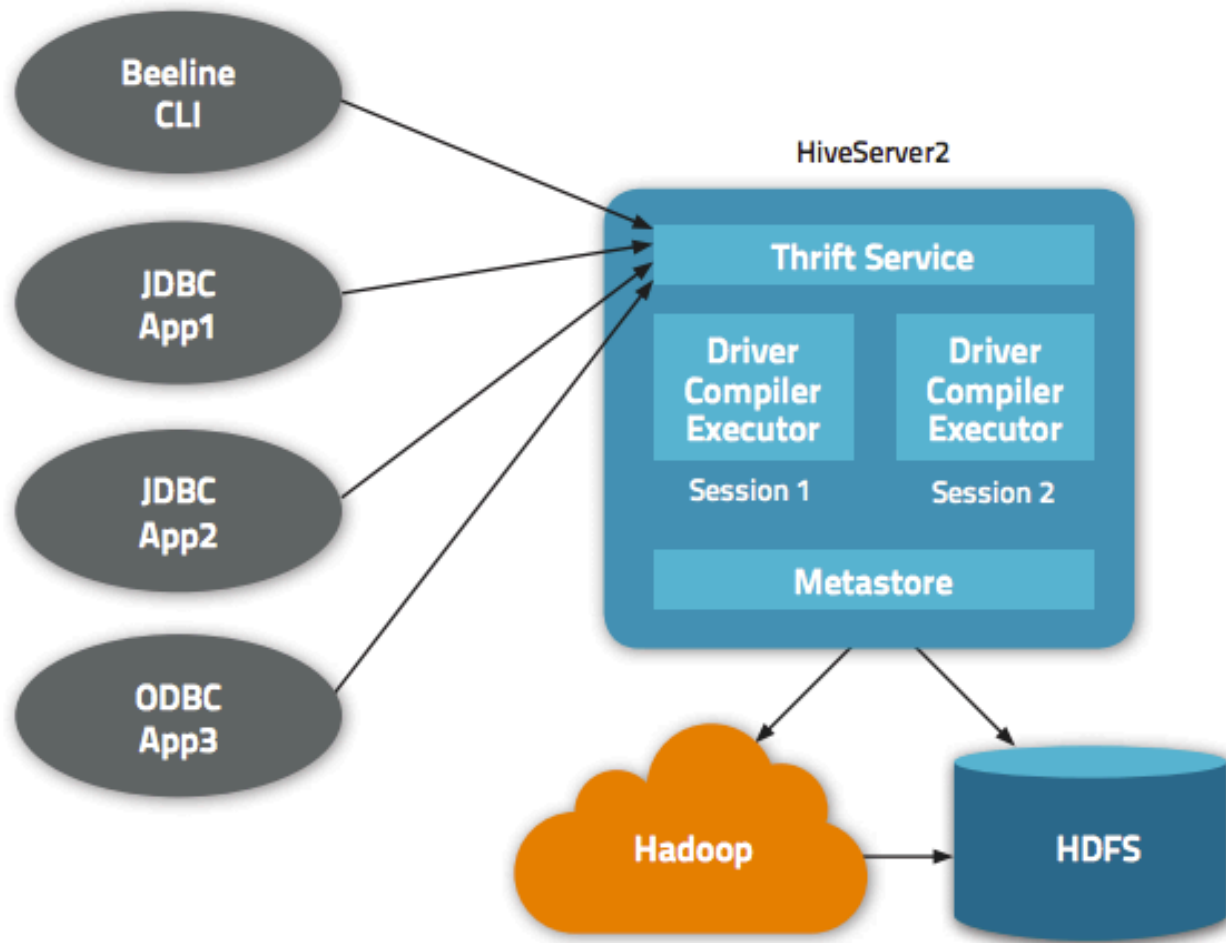
What is Hive

- Data warehouse software facilitates querying and managing large datasets residing in distributed storage.
- SQL-like language called HiveQL
- At the same time this language also allows traditional map/reduce

Hive Metastore Architecture



Hiveserver2



Reference

- HBase Getting Started
 - <https://hbase.apache.org/book.html#quickstart>
- Hive Getting Started
 - <https://cwiki.apache.org/confluence/display/Hive/GettingStarted>

Thank you for your listening