

| 大数据技术-第八章：Hive数据仓库
Hive功能应用



CONTENTS

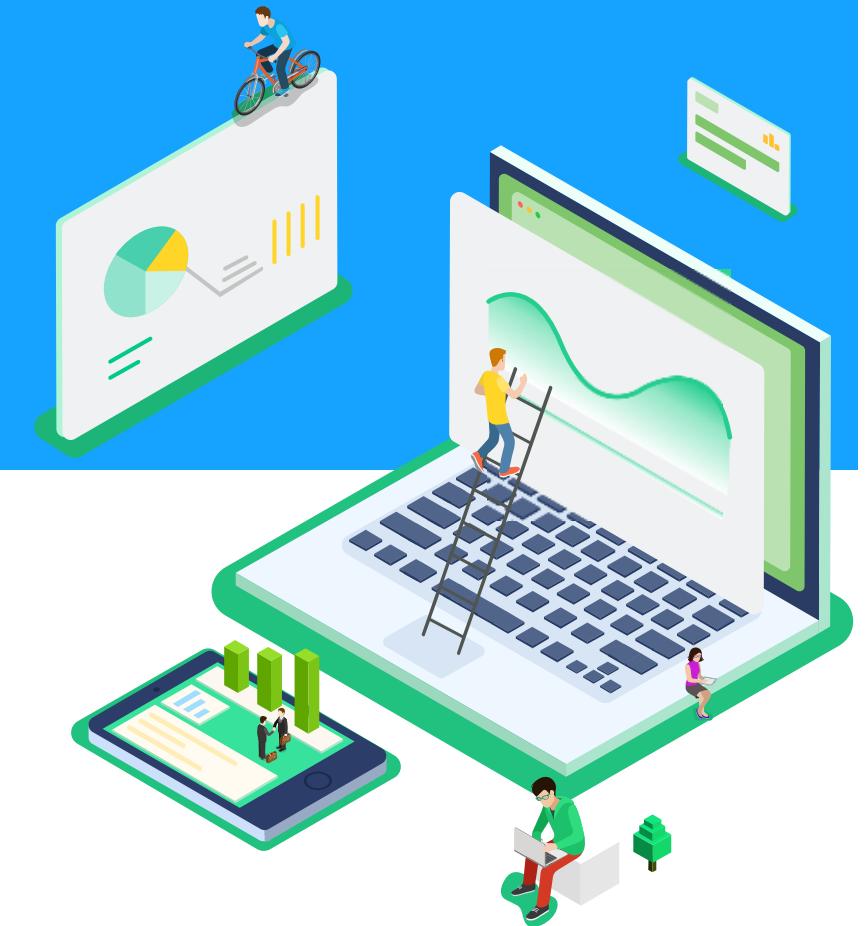
- 01. Hive与Hadoop的关系**
- 02. Hive与数据库的异同**

- 03. Hive的体系结构**
- 04. Hive的应用**



01

Hive与Hadoop的关系



» Hive与Hadoop的关系

Hive构建在Hadoop之上，HQL中对查询语句的解释、优化、生成查询计划是由Hive完成的。Hive读取的所有数据都是存储在Hadoop文件系统中。Hive查询计划被转化为MapReduce任务，在Hadoop中执行。

- Hive依赖于HDFS 存储数据
- Hive依赖于MapReduce 处理数据

02

Hive与数据库的异同

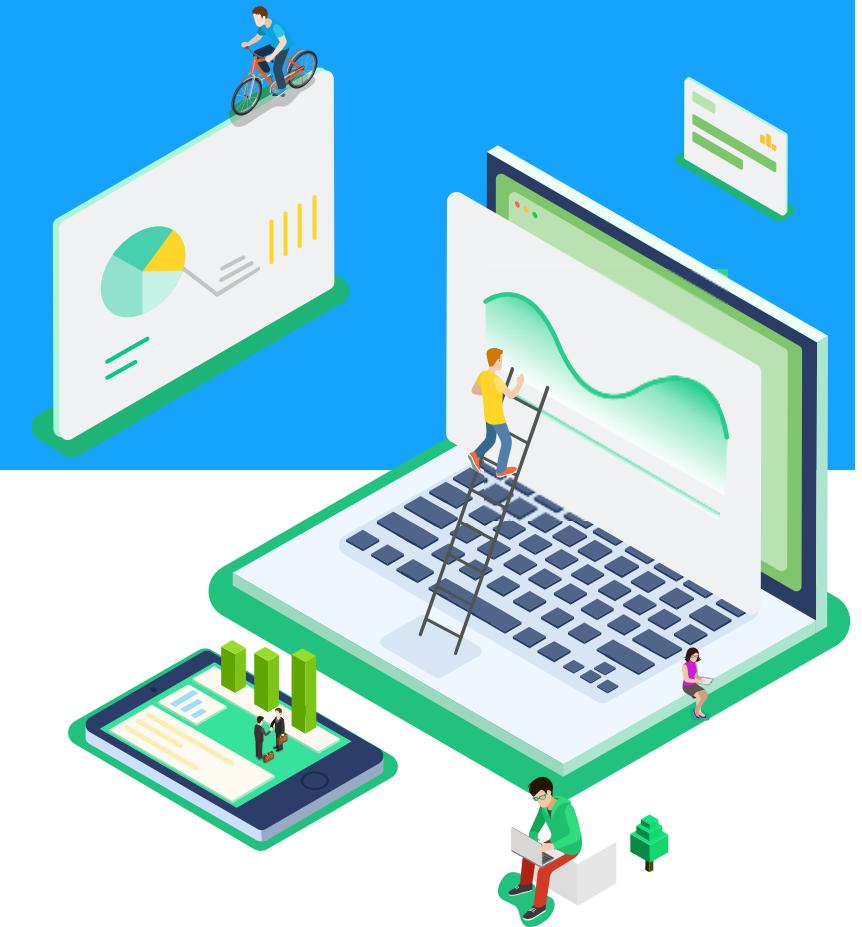


Hive在很多方面和传统的关系数据库类似，但是它的底层依赖的是HDFS和MapReduce，所以在很多方面又有别于传统数据库。

对比项目	Hive	传统数据库
数据插入	支持批量导入	支持单条和批量导入
数据更新	不支持	支持
索引	支持	支持
分区	支持	支持
执行延迟	高	低
扩展性	好	有限

03

Hive体系结构

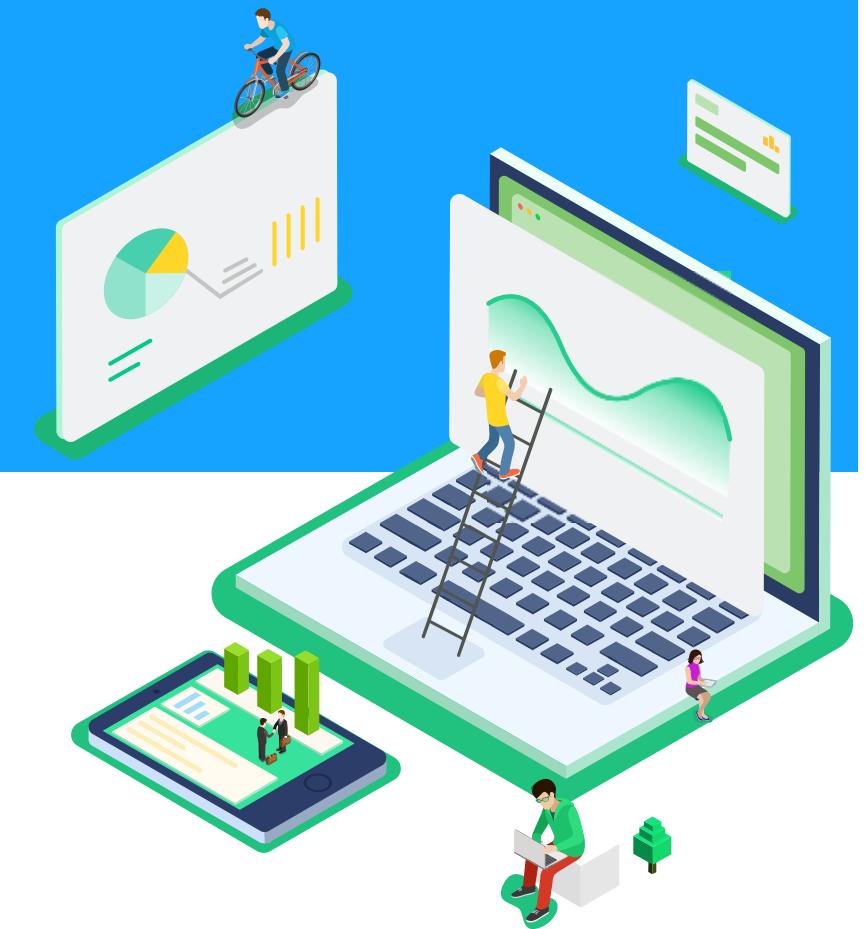


Hive的体系结构可以分为以下几部分：

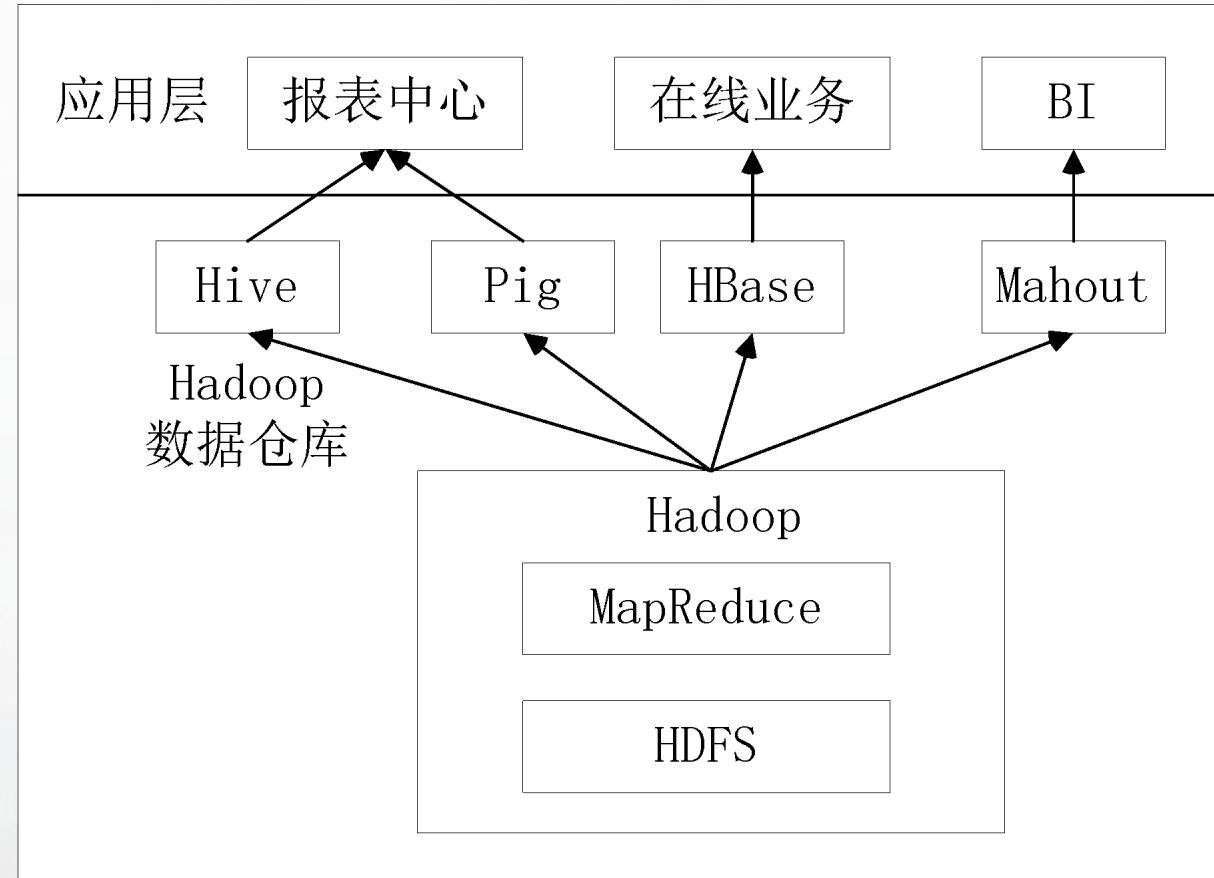
- (1) 用户接口主要有三个： CLI, Client和WUI。其中最常用的是CLI， CLI启动的时候，会同时启动一个Hive副本。Client是Hive的客户端，用户连接至Hive Server。在启动Client模式的时候，需要指出Hive Server所在节点，并且在该节点启动Hive Server。WUI是通过浏览器访问Hive。
- (2) Hive将元数据存储在数据库中，如MySQL、Derby。Hive中的元数据包括表的名字，表的列和分区及其属性，表的属性（是否为外部表等），表的数据所在目录等。
- (3) 解释器、编译器、优化器完成HQL查询语句从词法分析、语法分析、编译、优化以及查询计划的生成。生成的查询计划存储在HDFS中，并在随后有MapReduce调用执行。
- (4) Hive的数据存储在HDFS中，大部分的查询、计算由MapReduce完成。

04

Hive的应用



Hive在企业大数据分析平台中的应用



Turing AI 万维
图灵 | 大数据系列课程

大数据

BIG
DATA

智 / 能 / 科 / 技

放 / 眼 / 未 / 来

