

```
1 https://www.bilibili.com/video/BV1ek4y117Yq?p=156
```

```
In [ ]:
```

```
1
```

英文名称为Data Warehouse，可简称为DW或DWH。数据仓库的目的是构建面向分析的集成化数据环境，为企业提供决策支持（Decision Support）。

数据仓库是存数据的，企业的各种数据往里面存，主要目的是为了分析有效数据，后续会基于它产出供分析挖掘的数据，或者数据应用需要的数据，如企业的分析性报告和各类报表等。

### 1.3. 数据库与数据仓库的区别

数据库与数据仓库的区别实际讲的是 OLTP 与 OLAP 的区别。

操作型处理，叫联机事务处理 OLTP（On-Line Transaction Processing，），也可以称面向交易的处理系统，它是针对具体业务在数据库联机的日常操作，通常对少数记录进行查询、修改。用户较为关心操作的响应时间、数据的安全性、完整性和并发支持的用户数等问题。传统的数据库系统作为数据管理的主要手段，主要用于操作型处理。

分析型处理，叫联机分析处理 OLAP（On-Line Analytical Processing）一般针对某些主题的历史数据进行分析，支持 管理决策。

首先要明白，数据仓库的出现，并不是要取代数据库。

- 数据库是面向事务的设计，数据仓库是面向主题设计的。
- 数据库一般存储业务数据，数据仓库存储的一般是历史数据。
- 数据库设计是尽量避免冗余，一般针对某一业务应用进行设计，比如一张简单的User表，记录用户名、密码等简单数据即可，符合业务应用，但是不符合分析。数据仓库在设计是有意引入冗余，依照分析需求，分析维度、分析指标进行设计。
- 数据库是为捕获数据而设计，数据仓库是为分析数据而设计。

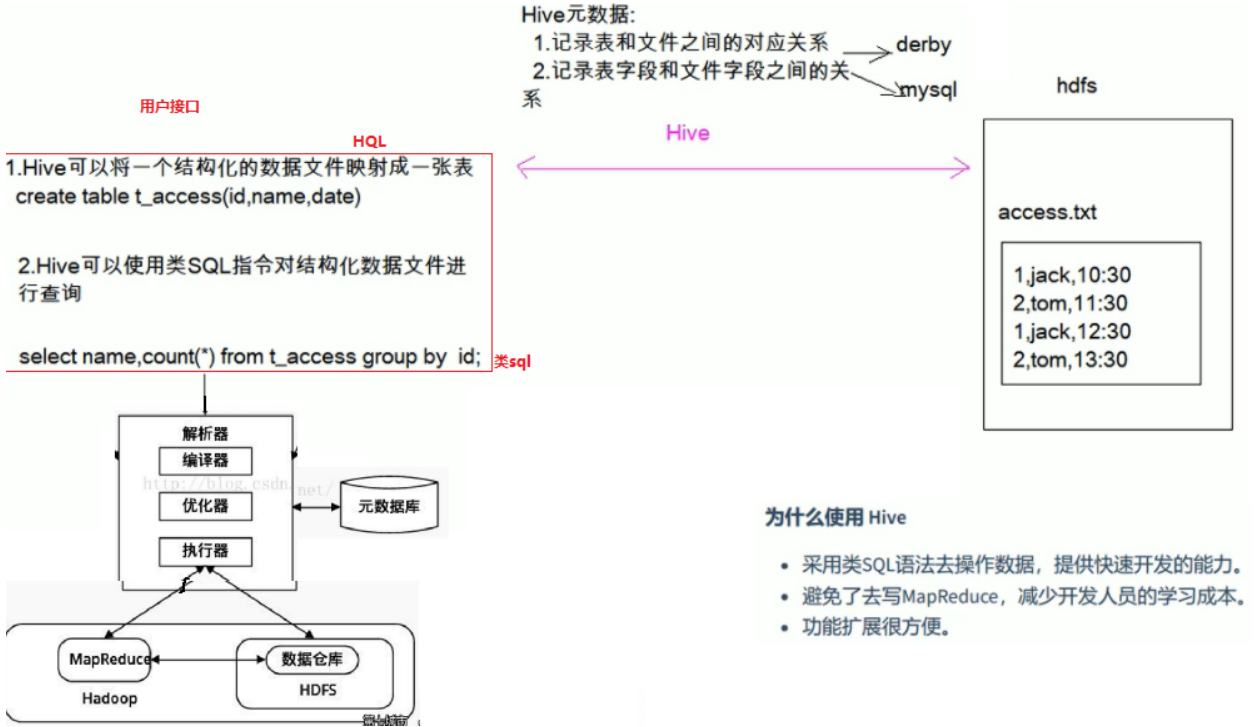
**数据仓库，是在数据库已经大量存在的情况下，为了进一步挖掘数据资源、为了决策需要而产生的，它决不是所谓的“大型数据库”。**

```
In [ ]:
```

```
1
```

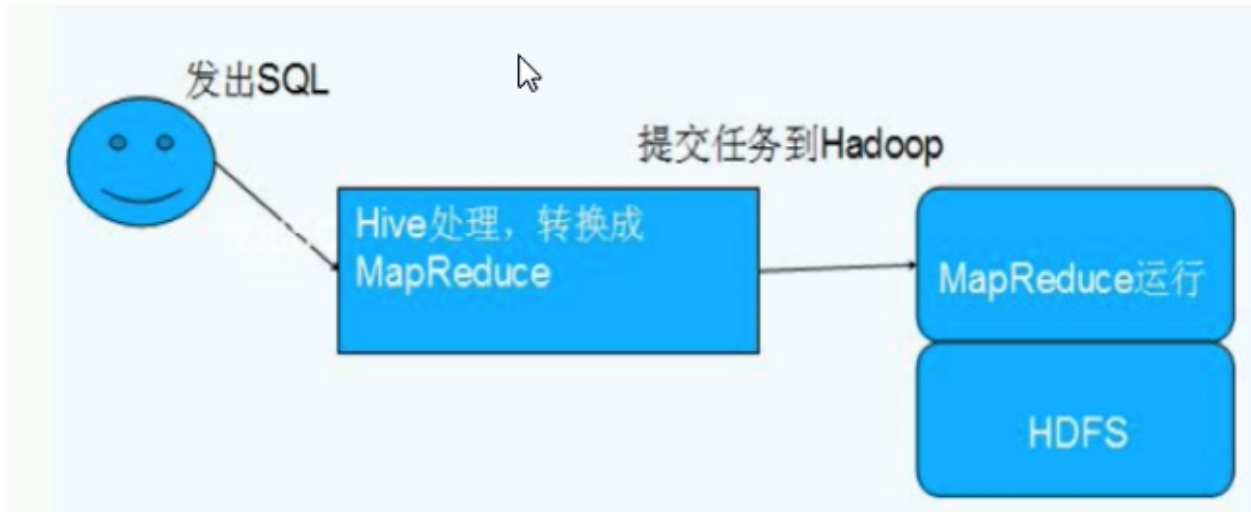
```
In [ ]:
```

```
1
```



hive与hadoop之间的关系

Hive利用HDFS存储数据，利用MapReduce查询分析数据



hive与传统数据库对比

	Hive	RDBMS 关系型数据库
查询语言	HQL	SQL
数据存储	HDFS	Raw Device or Local FS 本地磁盘
执行	MapReduce	Excutor
执行延迟	高	低
处理数据规模	大	小
索引	0.8版本后加入位图索引	有复杂的索引

总结：hive具有sql数据库的外表，但应用场景完全不同，hive只适合用来做批量数据统计分析

In [ ]: 1

In [ ]: 1

In [ ]: 1