

2019

A low-angle, upward-looking shot of a modern glass skyscraper with a grid-like facade, set against a clear blue sky. The building's reflection is visible in the lower part of the frame.

企业级数据仓库实战

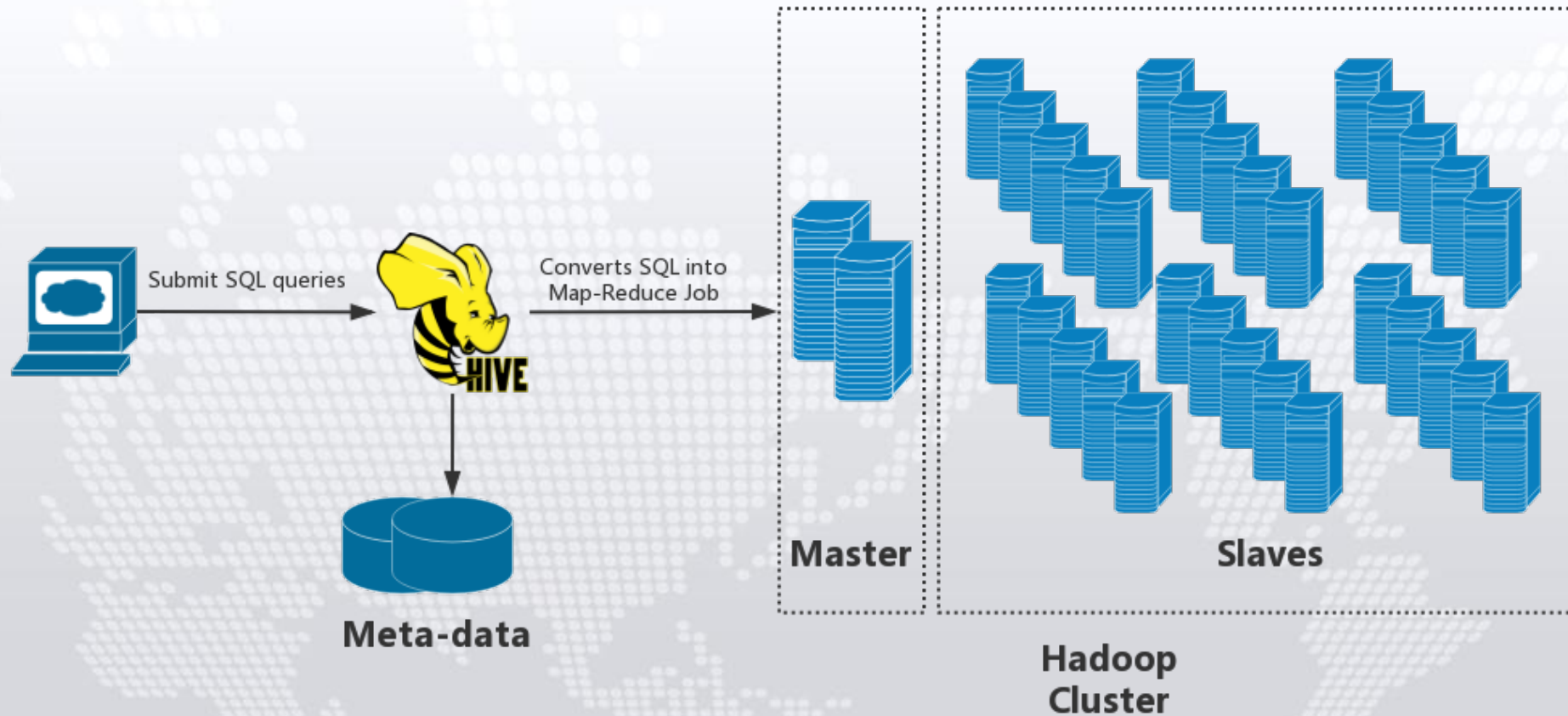
A light gray world map is centered in the background, composed of a grid of small dots. The map is partially obscured by the text and the building image above it.



工具篇-Hive介绍



The Apache Hive™ data warehouse software facilitates reading, writing, and managing large datasets residing in distributed storage using SQL. Structure can be projected onto data already in storage. A command line tool and JDBC driver are provided to connect users to Hive.





注意:

- 1、Hive不是数据仓库，Hive只是将SQL转化为MR的工具
- 2、Hive不存储数据，真实的数据存放在HDFS，相关统计信息存放在数据库中(一般为Mysql)
- 3、理论上Hive在执行效率上低于MR，但是开发难度比较低，所以实际生产中反而使用范围广泛
- 4、Hive SQL 语法类似标准的SQL语句，但是在实际用法上还是有所不同



学习资料：Hive官网 <https://hive.apache.org/index.html>

重点：

- 1、常见的Hive命令
- 2、Hive支持的文件格式
- 3、Hive支持的数据类型
- 4、常用的DDL操作
- 5、常用的DML操作
- 6、数据查询

包括Select、子查询、虚拟列、窗口分析函数、特殊聚合函数、执行计划



常见的Hive命名

```
-- 1.进入Hive
hive
-- 2.退出Hive
quit
exit
-- 3.设置变量
set <key>=<value>
-- 4.查看变量
set -- 打印由用户设置的变量
set -v -- 打印所有变量
-- 5.添加文件
add FILE[S] <filepath> <filepath>*
add JAR[S] <filepath> <filepath>*
add ARCHIVE[S] <filepath> <filepath>*
-- 6.罗列文件
list FILE[S]
list JAR[S]
list ARCHIVE[S]
```

```
-- 7.删除文件
delete FILE[S] <filepath>*
delete JAR[S] <filepath>*
delete ARCHIVE[S] <filepath>*
-- 8.执行shell命令
!<command>
-- 9.执行HDFS命令
dfs <dfs command>
-- 10.执行Hive SQL 查询
<query string>
-- 11.执行文件
source FILE <filepath>
-- 12.执行查询 string
hive -e '<query-string>'
-- 13.从文件执行一个或多个 SQL 查询。
hive -f <filepath>
```



数据格式-基本数据类型

分类	数据类型	类型名称	备注
数字型	整型	TINYINT	-128 to 127
		SMALLINT	-32,768 to 32,767
		INT	-2,147,483,648 to 2,147,483,647
		BIGINT	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807
	浮点型	FLOAT	4字节单精度浮点型
		DOUBLE	8字节双精度浮点型
		DECIMAL	类似DOUBLE，有版本限制
		NUMERIC	类似DECIMAL，有版本限制
日期时间类型	时间戳	TIMESTAMP	有版本限制、支持多种时间戳转换
	日期	DATE	有版本限制、支持多种日期转换
		INTERVAL	有版本限制
字符串型	字符串	STRING	使用单引号或者双引号表示
		VARCHAR	有版本限制、可变长度
		CHAR	有版本限制、不可变长度
其他类型	布尔型	BOOLEAN	使用IS关键字进行判断
	二进制	BINARY	有版本限制
复合数据类型	数组	ARRAY<data_type>	有版本限制
	映射	MAP<primitive_type,data_type>	
	结构体	STRUCT<col_name : data_type [COMMENT col_comment], ...>	有版本限制
	组合类型	UNIONTYPE<data_type, data_type, ...>	有版本限制



一般是生产环境中，常用的其实只有如下几种：

- 1、**bigint**：代表所有整型数据
- 2、**double/decimal**：代表所有浮点型数据
- 3、**string**：除以上两种类型外的所有类型
- 4、复合数据类型：array/map/struct

优点：

简单明了
易识别

缺点：

对于一些特殊的类型如布尔型、日期型等无法有效识别，在实际使用中需要特别注意



数据格式-隐式转换

	boolean	tinyint	smallint	int	<u>bigint</u>	float	<u>double</u>	decimal	<u>string</u>	varchar	timestamp	date	binary
bigint to	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
binary to	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
boolean to	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
date to	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
decimal to	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
double to	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
float to	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
int to	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
smallint to	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
string to	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
timestamp to	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
tinyint to	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE
varchar to	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE



类型	定义方法	描述
Map	MAP < primitive_type, data_type >	Map数据类型，主要是以K:V形式进行存储 可以通过 字段名['key'] 进行访问，将返回这个key对应的Value
Array	ARRAY < data_type >	Array数组，主要是以数组形式存储 可以通过 字段名[0] 进行访问，其中0表示的是数组中的第一位，以此类推
Struct	STRUCT < col_name : data_type [COMMENT col_comment], ...>	Struct结构体，主要以结构体形式进行存储 可以通过 字段名.列名 进行访问，其中列名就是定义中的col_name

THANK YOU FOR YOUR GUIDANCE.

谢谢