



# Hive 实验

## 显示数据库

---

- Hive 命令 : show
- 命令说明 : 查询显示已有数据库
- 命令实例 : show databases;

```
hive> show databases;  
OK  
default  
Time taken: 0.043 seconds, Fetched: 1 row(s)
```

## 创建数据库

---

- Hive 命令 : create database
- 命令说明 : 创建一个新的数据库
- 命令实例 : create database access\_log;

```
hive> create database access_log;  
OK  
Time taken: 0.126 seconds  
hive>
```

## 使用数据库

---

- Hive 命令：use
- 命令说明：激活使用数据库
- 命令实例：use access\_log;

```
hive> use access_log;  
OK  
Time taken: 0.045 seconds
```

## 创建表

---

- Hive 命令 : create table
- 命令说明 : 在数据库中创建表
- 命令实例 : create table access(time STRING, userid STRING, serverIP STRING, hostName STRING, spName STRING, upload INT, download INT) row format delimited fields terminated by '\t';

```
hive> create table access(time STRING,  
  >      userid STRING,  
  >      serverIP STRING,  
  >      hostName STRING,  
  >      spName STRING,  
  >      upload INT,  
  >      download INT) row format delimited fields terminated by '\t';  
OK  
Time taken: 0.214 seconds
```

## 将数据倒入表格

---

- Hive 命令 : load
- 命令说明 : 将已有数据倒入hive数据表中
- 命令实例 : load data inpath '/user/aisino/test/access.log' overwrite into table access;

```
hive> load data inpath '/user/fangcheng/access.log' overwrite into table access;
Loading data to table access_log.access
rmr: DEPRECATED: Please use 'rm -r' instead.
Moved: 'hdfs://Master:8020/user/hive/warehouse/access_log.db/access' to trash at: hdfs://Master:8020/user/fangcheng/.Trash/Current
chgrp: changing ownership of '/user/hive/warehouse/access_log.db/access/access.log': User does not belong to hive
Table access_log.access stats: [num_partitions: 0, num_files: 1, num_rows: 0, total_size: 9672471, raw_data_size: 0]
OK
Time taken: 1.742 seconds
```

## 查询数据

---

- Hive 命令 : select
- 命令说明 : 查询数据表中的数据
- 命令实例 : select userid, spname, sum(upload + download) from access group by userid, spname;

```
hive> select userid, spname, sum(upload + download) from access group by userid, spname;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
```