

2. HIVE 体系结构.

元数据: MySQL, Oracle, derby.

HIVE		
学生 .. / student		
sid	sname	dname
1	Tom	it
成绩 .. / score		
cname	sid	score
大学物理	1	80

derby		
表		
表id	表名	表位置
1	学生	/Hdfs/student
2	成绩	/Hdfs/score
列信息:		
列id	列名	表id
1	sid	1
2	sname	1
...		
4	cname	2
...		

执行计划).

SQL >> explain ~~for~~ plan for select * from emp where deptno=10;

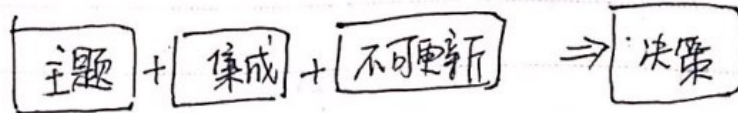
SQL >> select * from table (dbms_explain.display);

SQL >> creat index myindex on emp (deptno);

SQL >> explain plan for select * from emp where deptno=10;



1. 什么是HIVE



数据源	存储与管理	仓库引擎	展示
MySQL	抽取	SQL	web
Oracle	转换	HQL	
文本文档	存储	HIVE	

OLTP: 事务

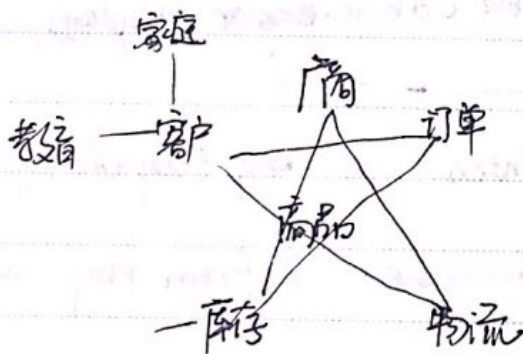
OLAP: 面向查询: 不更新删除

星型 \Rightarrow 雪花型

HIVE 数据



Hadoop 中的目录/文件.



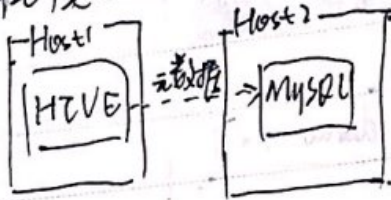
② 本地模式:



+ 连接。

MySQL
不同: 配置URL不同。

③ 远程模式:



安装:

1. hive.tar.gz.

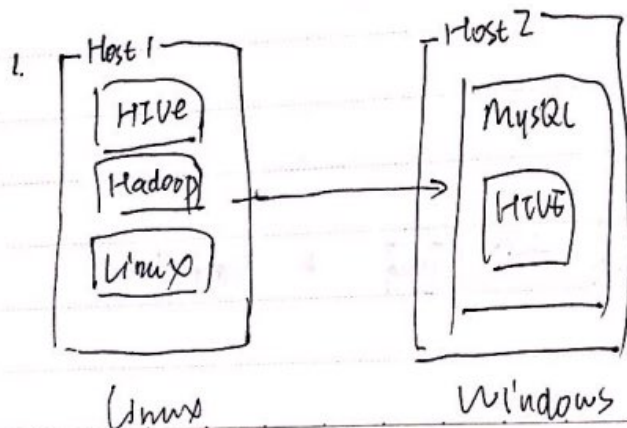
2. tar 开

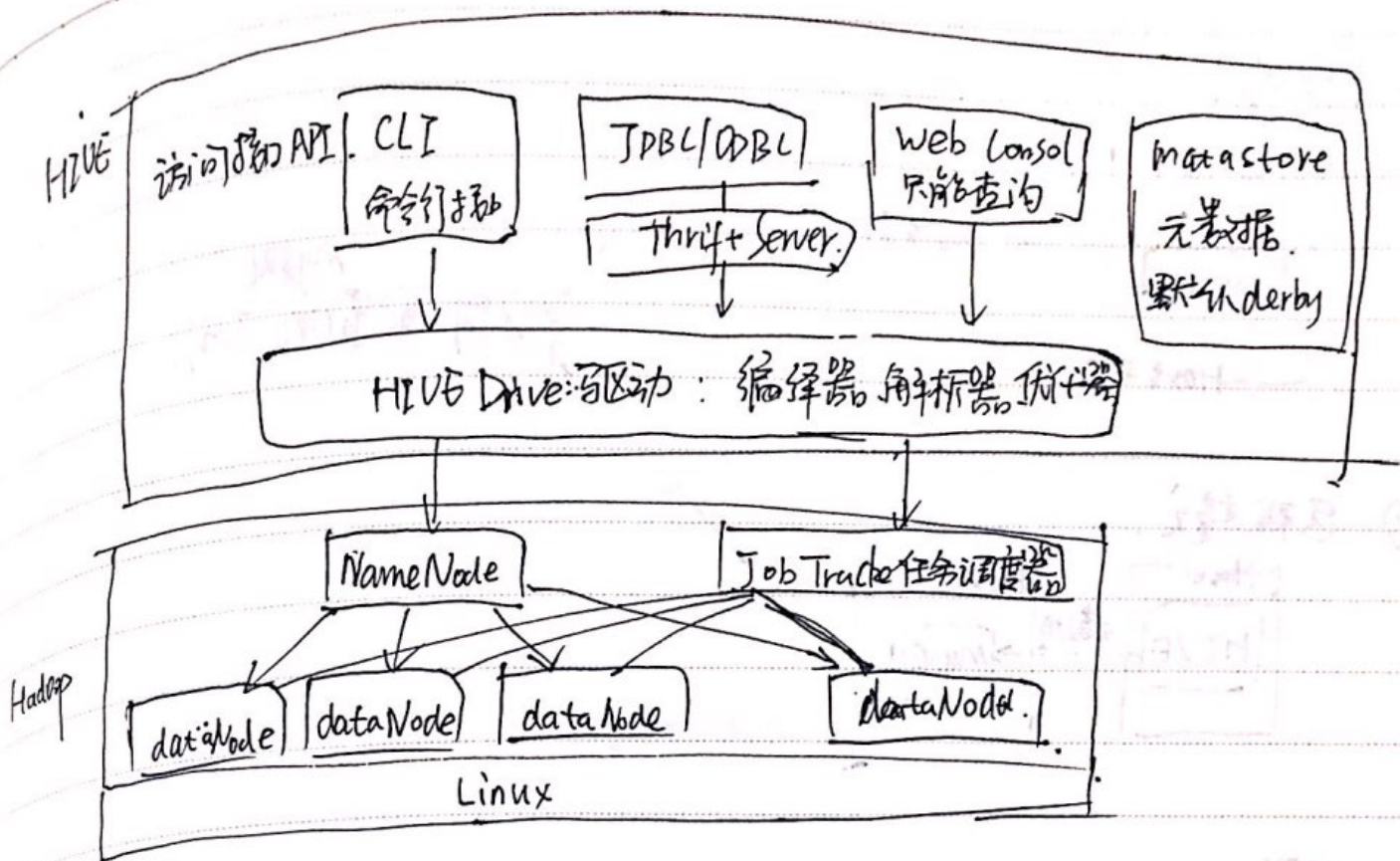
3. ./bin/hive 创建 derby db.

4. etc/profile
HIVE_HOME

export

安装模式。





3. HIVE安装.

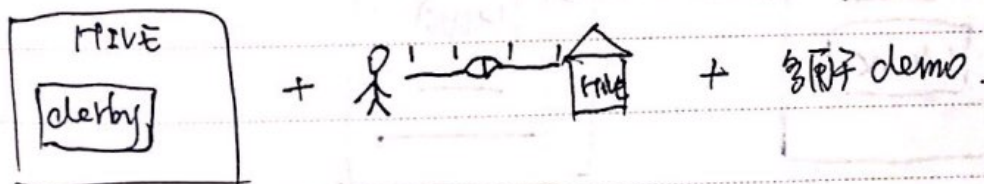
archive: apache.org/dist/Hive/hive-0.13.0

① hadoop 先安装

② Hive 后安装

三种安装模式

① 嵌入模式:



1. tar 开

2. 上传 MySQL-connector-java.jar → hive/lib 下.

3. 配置.

hive-site.xml.

<?xml ... >

<?xml-stylesheet ... >

<conf>

PURC {
 <prop>
 <name>{url}</name>
 <value>ip:port:db</value>
 </prop>

驱动 {

用户名 {

密码 {

</conf>

4. rec/profile
export.

5. \$ hive ts { }



4. HIVE 管理

① CLI: ① \$ hive quit
 ② \$ hive --server cli exit.

3. Ctrl + L / !clear

4. > show tables;

5. > show functions;

6. > desc test1;

7. > dfs -ls /user;

8. > !pwd.

> !ls

9. > select * from table1; -- ^{HQL} ~~和 set~~ 不转换 map/reduce
 > source 1.sql; -- 文件

> select name from test1; -- ^{转换为 m/R job} ~~表~~

10. \$ vi my.sql. < select * from test1;

& source ./my.sql;



④ 远程服务器启动方式.

port: 10000

启动: \$ hive --service hiveserver

注: 当 jdbc/odbc 登录 hive, 必须使用远程模式.

starting Hive Thrift server
远程服务.

5. hive 数据类型

① 基本数据类型

tinyint / smallint / int / bigint

float / double

boolean

string varchar / ~~char~~ char

② Hive Wiki

③ 复杂数据类型

Array : array <float>

Map : map <string, float>

Struct.

struct <name: string, age: int, sex: string>

[<'语文', 90>, <'数学', 100>]

array <map <string, float>>



11. 静默模式.

\$ hive -s # 不输出调试信息, 只看结果

⑦ 12. ^{linux下} \$ hive -e 'show tables';

\$ hive [^] -e 'select * from test1';

⑧ web.

错误 1. \$ hive --service hwi.

2. \$ tar xvf hive-src.tar.gz. & cd hwi.

3. \$ jar cvfM0 hive-hwi-0.13.0.war -c web/ . \$ ls

~~4~~ \$ hive-hwi-0.13.0.war. pom.xml src web.

4. cp hive-hwi-0.13.0.war hive/bin/下.

5 \$ vi hive-site.xml.

添加 @ hive-host

① port

② war-file.

7. \$ hive -service hwi

6. \$ cp jak/cib/tools.jar hive/hb.

8 浏览器 输入 ip:port/hwi



② 时间类型

时间类型
 - Date
 - Timestamp

v12

v8

'YYMM-DD'

时区无关, select unix-timestamp();
 时间戳

6. Hive 数据存储

② 基于 HDFS

Hive
表HDFS
目录

没有专门的数据存储格式 CSV 制表符

数据库、文件、表、视图

⇒ *.txt

建表时指定: 行分隔符与列分隔符

② hive 数据模型

表: - table 内部表

- partition 分区表

- external table 外部表

- bucket table 桶表

视图



内部表

hive
table

↔

hdfs
目录

user / hdfs / warehouse

仓库

创建表, 添加数据. 查看 hdfs.

\$ hdfs dfs -cat user/hive/warehouse/t4/0000-0

1 Tom M 60 80 96
2 Mary F 88 65 90
...

} 没有分隔符.

hive > creat table t5

row format delimited fields terminated by ','

分区表

hive 分区: ⇒ hdfs

gender='F'

gender='M'

分目录, 减少搜索范围.

hive > create table partition-table

> (sid int, name string)

> partitioned by (gender string)

> row format delimited fields terminated by ','

hive > desc partition-table;



桶表

hash 存放列不同桶!(文件中).

避免热块.

hive > create table bucket-table

> (sid int, sname string, age int)

> clustered by (sname) into 5 buckets;

hive desc table bucket-table.

视图

定义:

hive > create view empinfo:

> as

> select * ... -- 多表关联.

mysql → 物化视图 ← oracle

支持

支持

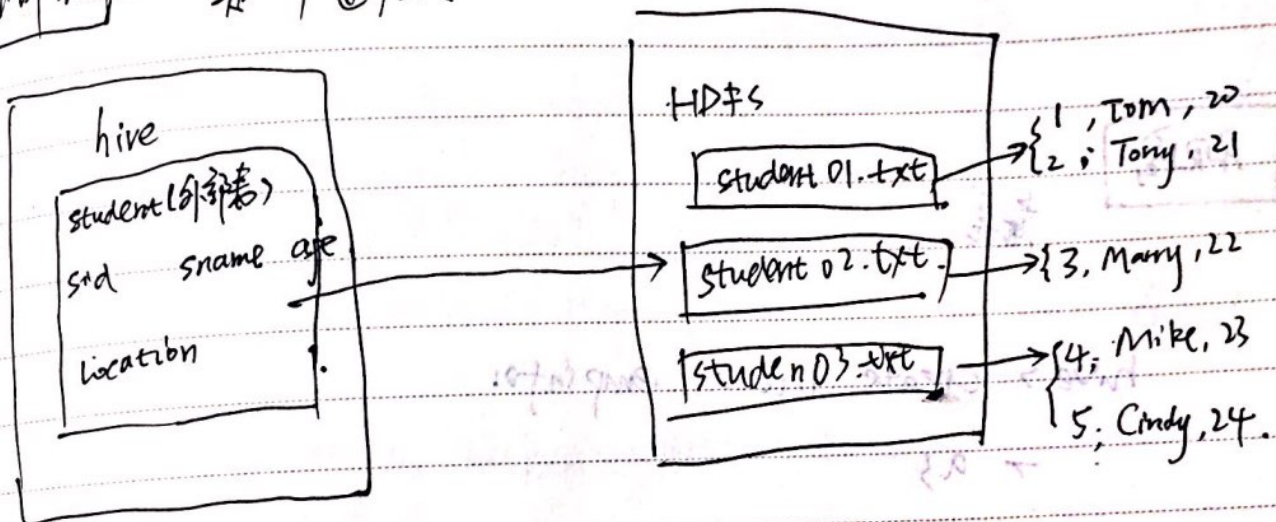
不支持.
hive

hive> insert into table partition-table partition (gender='M') select
sid, sname from sample-data where gender='M';

hive> explain select * from partition-table where gender='M';

注: Oracle 也有多种分区表.

外部表 : 是一个过程.



\$ hdfs dfs -put student 01.txt /input

hive> create ~~table~~ external table external-table
(sid int, sname string, age int)

row format delimited fields terminated by ','

location '/input';



7. 总结.

① 数据仓库

②. hive, 体系结构.

③. hive 安装配置

④. hive 数据模型.

