Hive

重要的metastore属性P457

和sql的比较P460

Hive的一些特点

1、传统数据库大多是写时模式，即数据在写入的时候已经过滤了，不符合的不会被写入；

而hive却是读时模式，数据在查询的时候再对数据格式进行规定；

读时模式具有很高的数据加载效率

读时模式更为灵活，因为它允许数据有多种模式，以使得数据有多种分析方式

写时模式的数据查询更为快速，但是规定了数据的模式，比较死板

2、hive不支持更新或者删除，但是支持向文件末尾添加，这和hadoop文件系统是契合的，同时支持了 索引以便在某些情况下提升查询效率，包括紧凑型索引和位图型索引；

3、hive引入了表级和分区级的锁，有了锁可以防止一些产生意外的写操作；

4、含有有关hivql的语法https://cwiki.apache.org//confluence/display/Hive/LanguageManual

Hive的问题式学习：

1、数据类型

Hive开发什么时候需要注意数据类型？P461页列出了所有数据类型

所有对数据的查找处理的过程都需要用到数据类型，包括八大基本数据类型和三个复杂数据类型

2、Hive的操作和函数，Hive有哪些操作？和Mysql有何不同？Hive有哪些函数，如何使用?

Hive有关系操作（各种判断，匹配）、算术操作（数学意义上的操作）、逻辑操作P463讲解函数 的自学办法

3、Hive表的组织形式是怎样的，Hive的表有何特点？

1、可以分为外部表和托管表

2、用分区和桶的方式来组织，这是为了加速有限定条件的查询,如，把表分区为国家， 时间、气候等，然后在数据导入的时候，需要显示的标明所处的分区，这样就能被有效 的组织起来。

3、而桶则是根据现有数据中的某一列，如ID，通过哈希算法得到的值来决定它属于那 个桶，这样操作表一某一行数据的时候，如果涉及到表二的同一用户的数据，就直接获 取那个桶就可以了。

附录：

开放数据库连接（Open Database Connectivity，ODBC）

JAVA Date Object ：JDO