

Hbase之phoenix篇 20:30开始

3月12日周末班2月26日全日制班欢迎您的到来！ 讲师：君临天下

需要代码、PPT、讨论技术可以加入以下

QQ群： 172599077 , 156927834

2016.2.26日脱产班

2016.3.12日周末班

2016.3.19日线上班

- Hbase的数据模型
- 安装hbase
- 安装phoenix
- Phoenix使用
- Phoenix客户端使用

- Hadoop生态系统



□

- HBase简介

- HBase – Hadoop Database , 是一个高可靠性、高性能、面向列、可伸缩、实时读写的分布式数据库
- 利用Hadoop HDFS作为其文件存储系统,利用Hadoop MapReduce来处理HBase中的海量数据,利用Zookeeper作为其分布式协同服务
- 主要用来存储非结构化和半结构化的松散数据 (列存 NoSQL 数据库)

- HBase数据模型

Row Key	Time Stamp	CF1	CF2	CF3
"com.cnn.www "	t6		CF2:q1=val3	CF3:q4=val4
	t5			
	t3	CF1:q2=val2		

- HBase数据模型

- Column Family列族 & qualifier列
- HBase表中的每个列都归属于某个列族，列族必须作为表模式(schema)定义的一部分预先给出。如 create ‘test’, ‘course’；
- 列名以列族作为前缀，每个“列族”都可以有多个列成员(column)；如 course:math, course:english, 新的列族成员（列）可以随后按需、动态加入；
- 权限控制、存储以及调优都是在列族层面进行的；
- HBase把同一列族里面的数据存储在同一目录下，由几个文件保存。

- HBase数据模型

- Timestamp时间戳

- 在HBase每个cell存储单元对同一份数据有多个版本，根据唯一的时间戳来区分每个版本之间的差异，不同版本的数据按照时间倒序排序，最新的数据版本排在最前面。

- 时间戳的类型是 64位整型。

- 时间戳可以由HBase(在数据写入时自动)赋值，此时时间戳是精确到毫秒的当前系统时间。

- 时间戳也可以由客户显式赋值，如果应用程序要避免数据版本冲突，就必须自己生成具有唯一性的时间戳。

- HBase数据模型
 - Cell单元格
 - 由行和列的坐标交叉决定；
 - 单元格是有版本的；
 - 单元格的内容是未解析的字节数组；
 - 由{*row key* , *column*(=<family> +<qualifier>), *version*} 唯一确定的单元。cell中的数据是没有类型的，全部是字节码形式存贮。

- HBase数据模型
 - ROW KEY
 - 决定一行数据
 - 按照字典顺序排序的。
 - Row key只能存储64k的字节数据

- Hbase安装
 - 本地模式安装（略）
 - 完全分布式安装：
 - 可用的zookeeper集群
 - 可用的hdfs集群

- 配置regionservers , 主机名
- 配置backup-masters
- 配置hbase-env.sh
 - 配置JAVA_HOME
 - 配置HBASE_MANAGERS_ZK=false
- 配置hbase-site.xml
 - Hbase.rootdir
 - Zookeeper集群主机
 - 启用hbase集群 : true
- 启动hbase
- 配置环境变量
 - export HBASE_HOME=/opt/sxt/soft/hbase-1.1.2
 - export HADOOP_CONF_DIR=\$HADOOP_HOME/etc/hadoop
 - export PATH=\$PATH:\$HBASE_HOME/bin

- 配置hbase-site.xml

- <property>
- <name>hbase.rootdir</name>
- <value>hdfs://chenkl/hbase</value>
- </property>
- <property>
- <name>hbase.cluster.distributed</name>
- <value>true</value>
- </property>
- <property>
- <name>hbase.zookeeper.quorum</name>
- <value>master,slave1,slave2</value>
- </property>

- Hbase的sql查询引擎
- **Hive , Tez , Impala , Shark/Spark , Phoenix**

- Phoenix
 - Phoenix查询引擎会将SQL查询转换为一个或多个HBase scan，并编排执行以生成标准的JDBC结果集。直接使用HBase API、协同处理器与自定义过滤器，对于简单查询来说，其性能量级是毫秒，对于百万级别的行数来说，其性能量级是秒

- Phoenix安装
 - bin/sqlline.py master:2181

- Phonenix使用
 - DDL
 - DML

- Create
 - CREATE TABLE my_table (id INTEGER not null primary key desc,
date DATE not null,
m.db_utilization DECIMAL, i.db_utilization)

- Drop
 - DROP TABLE my_schema.my_table;

- Alert
 - ALTER TABLE my_schema.my_table ADD d.dept_id char(10) VERSIONS=10
 - ALTER TABLE my_table ADD dept_name char(50), parent_id char(15) null primary key
 - ALTER TABLE my_table DROP COLUMN d.dept_id, parent_id;
 - ALTER VIEW my_view DROP COLUMN new_col;
 - ALTER TABLE my_table SET IMMUTABLE_ROWS=true,DISABLE_WAL=true;

- select
 - SELECT * FROM TEST LIMIT 1000;
 - SELECT full_name FROM SALES_PERSON WHERE ranking >= 5.0
UNION ALL SELECT reviewer_name FROM
CUSTOMER_REVIEW WHERE score >= 8.0

- Upsert
 - UPSERT INTO TEST VALUES('foo','bar',3);
 - UPSERT INTO TEST(NAME,ID) VALUES('foo',123);

- Phoenix JDBC
 - 和一般的JDBC使用基本相同

```
public static void main( String[] args ) throws Throwable
{
    try {
        Class.forName("com.salesforce.phoenix.jdbc.PhoenixDriver");
    } catch(ClassNotFoundException ex) {
        ex.printStackTrace();
    }

    String url = "jdbc:phoenix:zkserver1,zkserver2:2181:/hbasePhoenix";
    Connection conn = DriverManager.getConnection(url);
    Statement statement = conn.createStatement();

    String sql = "select count(*) as cn from afs_afs_part";

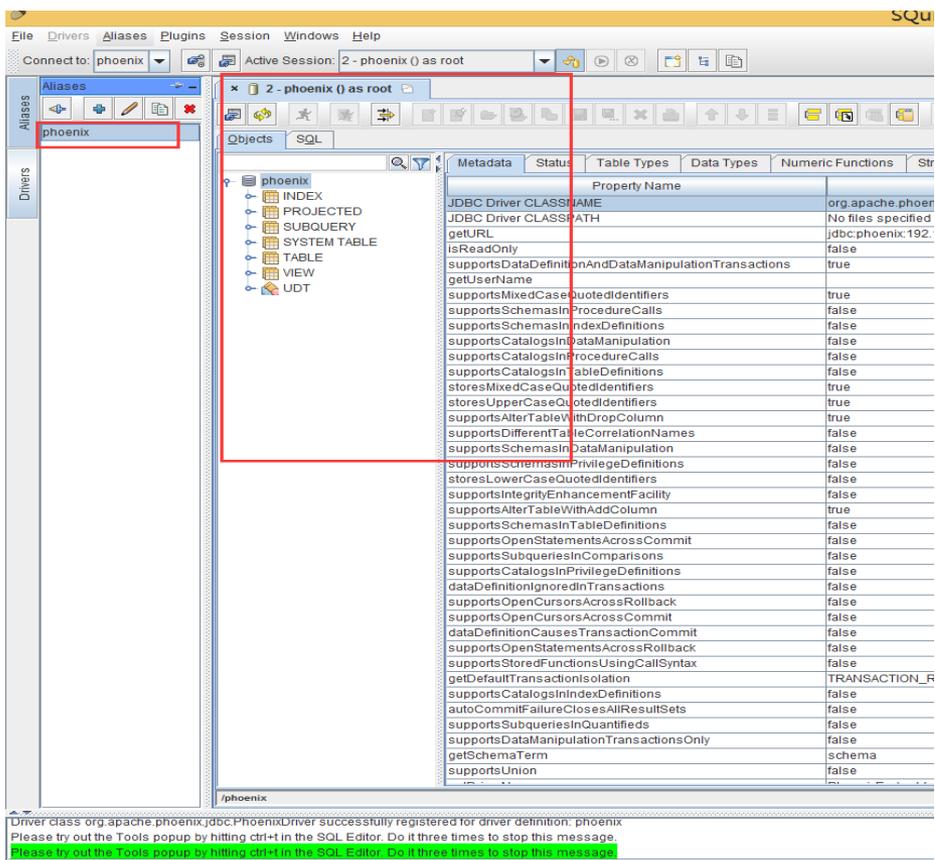
    long time = System.currentTimeMillis();
    ResultSet rs = statement.executeQuery(sql);

    while (rs.next()) {
        int count = rs.getInt("cn");
        System.out.println("row count is " + count);
    }

    long timeUsed = System.currentTimeMillis() - time;
    System.out.println("time " + timeUsed + "mm");
    rs.close();
    statement.close();
    conn.close();
}
```

- Phonenix客户端squirrel

- Squirrel安装使用
- <http://www.squirrelsql.org/>



- 1, 上课时间没有限制, 随时随地学习
- 2, 每节随堂笔记, 像今天一样该有图有图
- 3, 提供服务器配置, 搭建步骤说明
- 4, 有问题老师一对一辅导, QQ远程协助辅导
- 5, QQ群同学们良好的学习氛围
- 6, 终身免费重学
- 7, 课程升级免费学
- 8, 尚学堂科技有限公司是一个实体公司, 已做教育多年, 有着良好的口碑
- 9, 尚学堂大数据班老师均有丰富的授课经验, 线下线上课都有