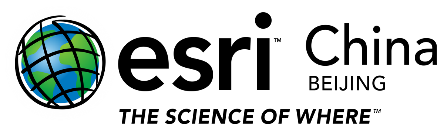
Esri大数据平台搭建指南之Hive



易智瑞中国信息技术有限公司

2017 年 3 月

**——制定及修订记录——**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **版本** | **完成日期** | **编写/修订纪要** | **编写者** | **备注** |
| **V1.0** | **2017-3-25** | **初稿** | **胡源** |  |
| **V1.1** | **2017-3-31** | **修订** | **胡源** | **内容完善** |
|  |  |  |  |  |
|  |  |  |  |  |

目录

[Esri大数据平台搭建指南之Hive 1](#_Toc479169654)

[1. Hive简介 4](#_Toc479169655)

[2. 搭建环境概览 4](#_Toc479169656)

[2.1 Hive元数据库 4](#_Toc479169657)

[2.2测试数据 4](#_Toc479169658)

[2.3软件介质 4](#_Toc479169659)

[3. Hive搭建 5](#_Toc479169660)

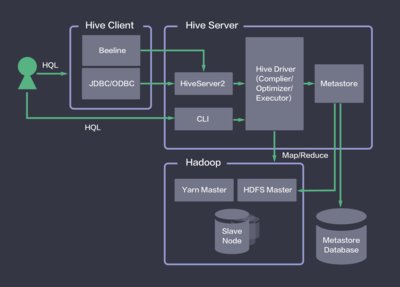
[4. 注册数据 7](#_Toc479169661)

[5. 执行分析 9](#_Toc479169662)

1. **Hive简介**

Hive是基于Hadoop的一个数据仓库工具，可以将结构化的数据文件映射为一张数据库表，并提供简单的SQL查询功能，可以将SQL语句转换为MapReduce任务进行运行。

Hive将元数据存储在RDBMS中，支持Derby、MySQL、Oracle和PostgreSQL,默认为Derby。Hive 将元数据存储在 RDBMS 中，我们可以通过Remote Server Mode模式（在服务器端启动一个 MetaStoreServer，客户端利用 Thrift 协议通过 MetaStoreServer 访问元数据库）在GeoAnalytics Server配置元数据的访问。Hive架构:



1. **搭建环境概览**

***注意****：Hive的搭建建立在HDFS的基础之上，HDFS的搭建请参考文档《Esri大数据平台搭建指南之HDFS》。*

2.1 Hive元数据库

Hive默认的Derby数据库（Hive包中已包含）

2.2测试数据

地震分布数据(earthquakes.csv)

2.3软件介质

Hive : apache-hive-1.2.1-bin.tar.gz

（测试数据和软件介质下载地址：<http://pan.baidu.com/s/1sldeqrr>）

1. **Hive搭建**
2. 创建Hive安装路径。

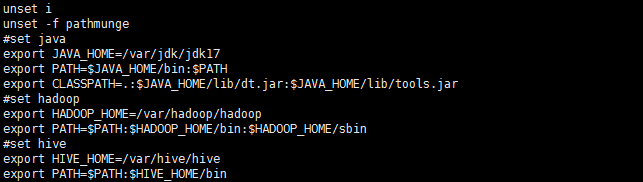
mkdir /var/hive

1. 通过Xftp软件向虚拟机传输Hive包到上一步指定目录，解压Hive并将文件夹重命名为hive。



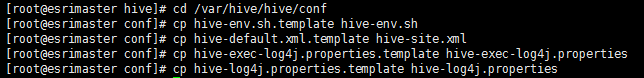
1. 配置Hive环境变量并使其生效。

vi /etc/profile



source /etc/profile

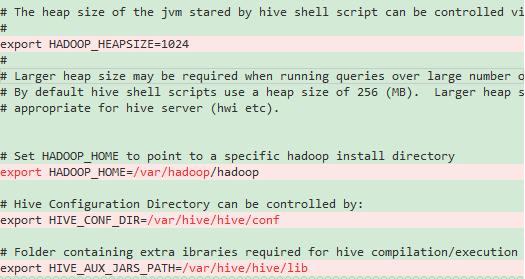
1. 复制需要的配置文件。



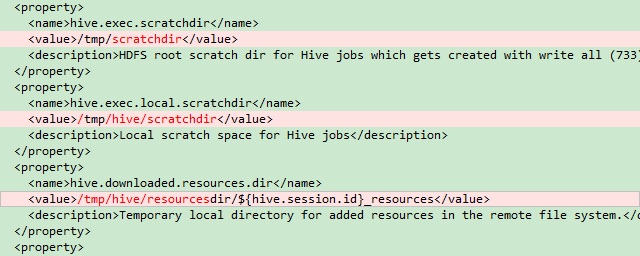
1. hive-env.sh文件和hive-site.xml文件配置。

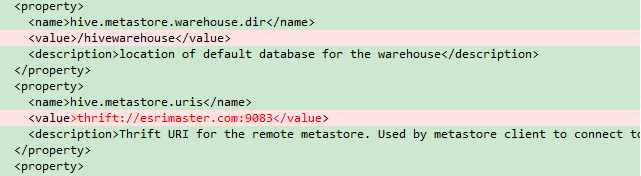
***提示****：由于这两个文件中配置参数比较难找，所以可以先下载到windows本机上通过notepad++搜索关键字进行修改，然后拷贝回centos。*

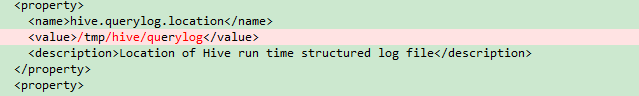
* hive-env.sh:

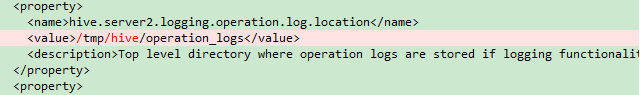


* hive-site.xml:

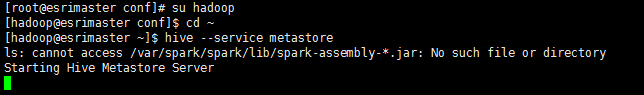




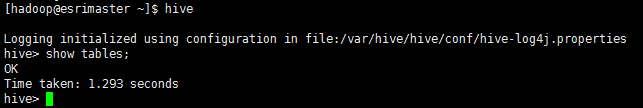




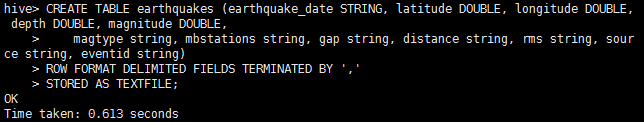
1. 切换到hadoop用户并启动Hive的metastore服务。



1. 启动另一个终端测试Hive。

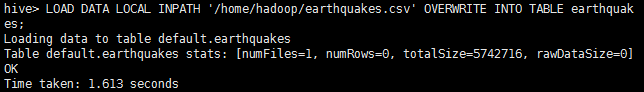


1. **注册数据**
2. 将earthquakes.csv文件通过Xftp软件上传到/home/hadoop目录。
3. 在Hive中创建earthquakes表。

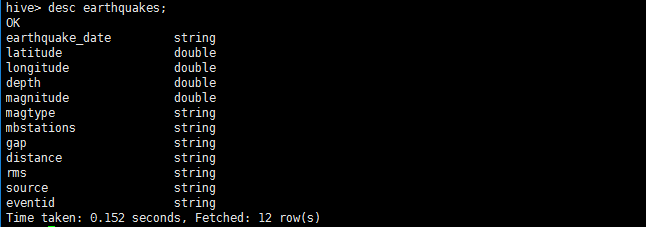


1. 将本地文件/home/hadoop/earthquakes.csv中数据导入earthquakes表。

***提示****：也可先将数据上传至HDFS，再通过load data inpath命令导入至Hive中数据表*



1. 查看表的信息。



1. 打开GeoAnalytics Server集群中主节点manager地址，进入 **/站点/Data Stores**，点击**注册**，选择**大数据文件共享**，弹出框中选择**Hive**，填写**名称**、**元数据URI**、**数据库名称**。

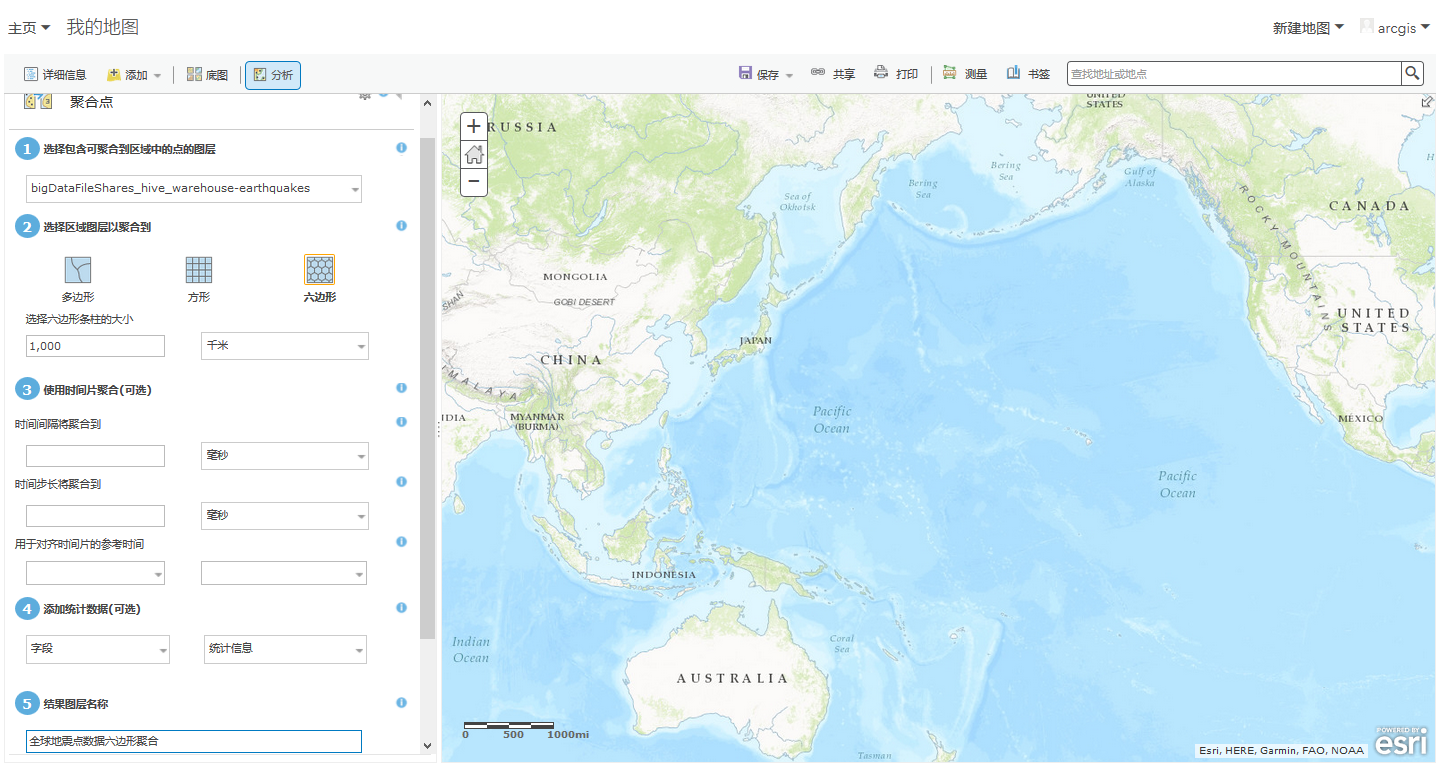
***注意：****此处数据库名称必填，就算是default也要填上default，亲测：如不填写，会报出“文件夹位置无效，无法访问此位置”的错误。*



1. 点击创建，稍等片刻后即可看到数据注册成功，点击右侧铅笔按钮，可以查看并修改数据注册信息。

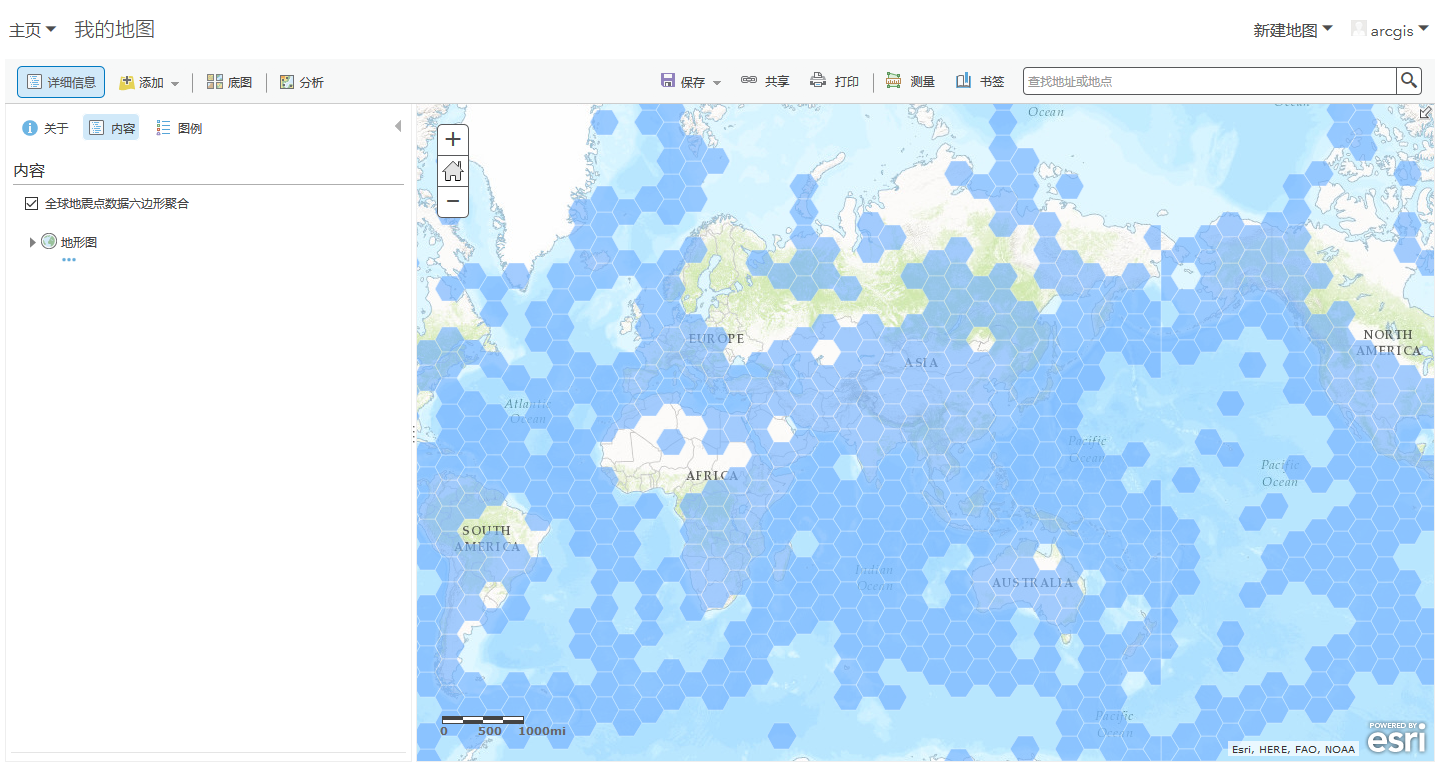


1. **执行分析**
2. 进入Portal for ArcGIS，在地图中点击**分析🡪GeoAnalytics工具🡪临近分析🡪聚合点**，在分析环境中将处理坐标系设置为102100，取消勾选“使用当前地图范围分析”。分析参数设置如下：



1. 点击运行分析，分析结果如下所示：

***注意****：分析前一定要确保Hive的元数据服务已经启动。*



1. 使用count字段对分析结果进行符号化。

