1. 数据爬取模块：

* 爬取的网站：

图片包含 游戏机, 文字, 地图

描述已自动生成

图1 — 网站界面

* 爬取代码

图片包含 游戏机, 截图

描述已自动生成

图2 — 代码1

图片包含 游戏机, 截图, 鸟

描述已自动生成

图3 — 代码2

* 爬取结果csv文件：

图片包含 游戏机, 截图, 街道

描述已自动生成

图4 — 爬取结果CSV文件（部分）

1. Hadoop环境搭建：

* 虚拟机界面：

图片包含 电脑, 小, 监控, 蓝色

描述已自动生成

图5 — 虚拟机界面

* 虚拟机hostname设置：

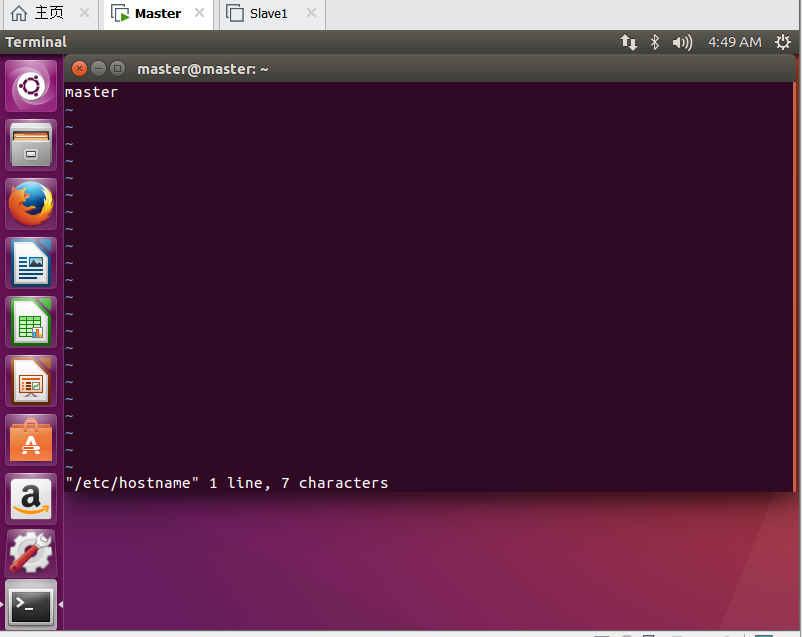


图6 — master

图片包含 游戏机, 截图, 电脑

描述已自动生成

图7 — slave（此处以slave1为例）

* Hosts设置

图片包含 游戏机, 截图

描述已自动生成

图8 — 以master的hosts为例

* Interfaces设置：

图片包含 游戏机, 截图

描述已自动生成

图9 — 以master的interfaces为例

* 全部公钥查看：

图片包含 瓶子, 站, 男人, 田地

描述已自动生成

图10 — 公钥

* Ssh：

图片包含 截图, 照片, 电话, 屏幕

描述已自动生成

图11 — master连接slave1

图片包含 截图, 游戏机

描述已自动生成

图12 — slave1连接master

图片包含 游戏机, 电话

描述已自动生成

图13 — slave1连接slave2

* 结果文档结构及配置文件：

图片包含 截图, 游戏机

描述已自动生成

图14 — home目录

图片包含 游戏机

描述已自动生成

图15 — data文档

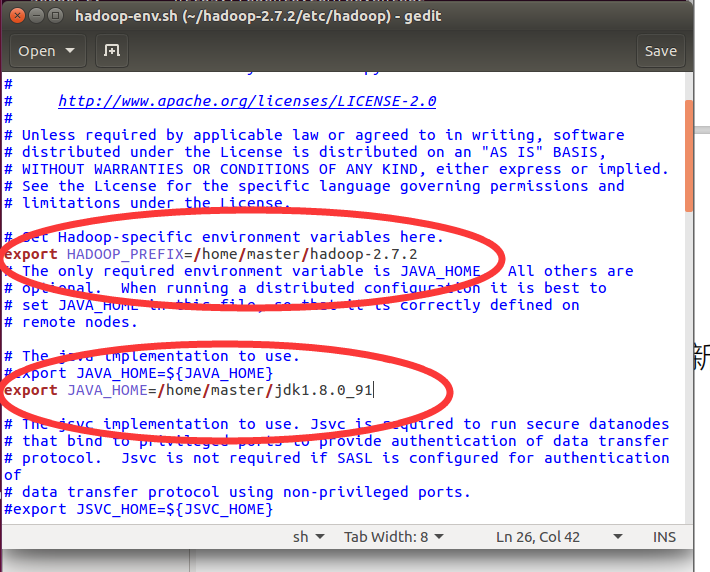


图16 — Hadoop-env.sh

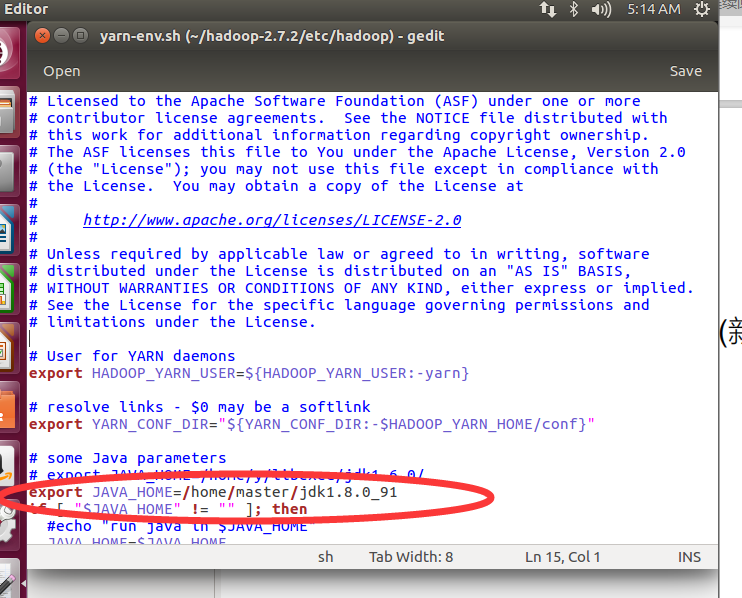


图17 — yarn-env.sh

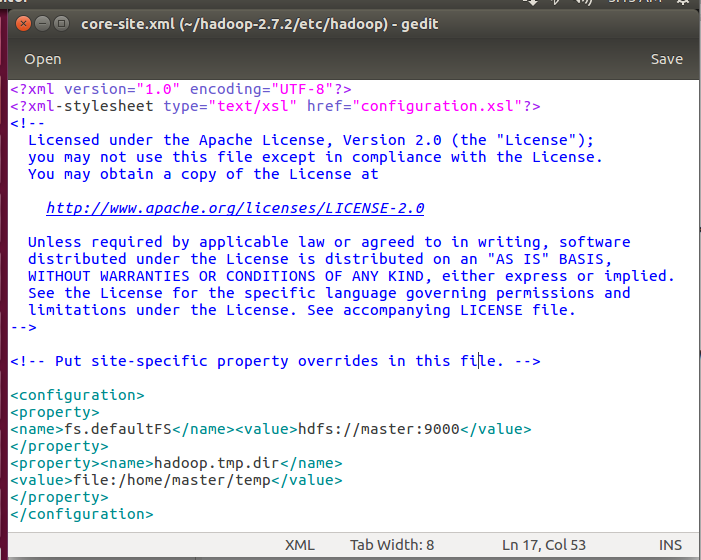


图18 — core-site.xml

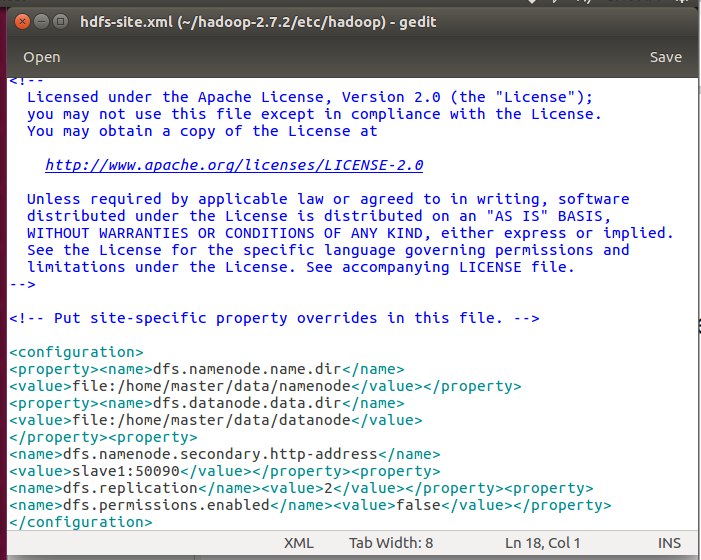


图19 — hdfs-site.xml

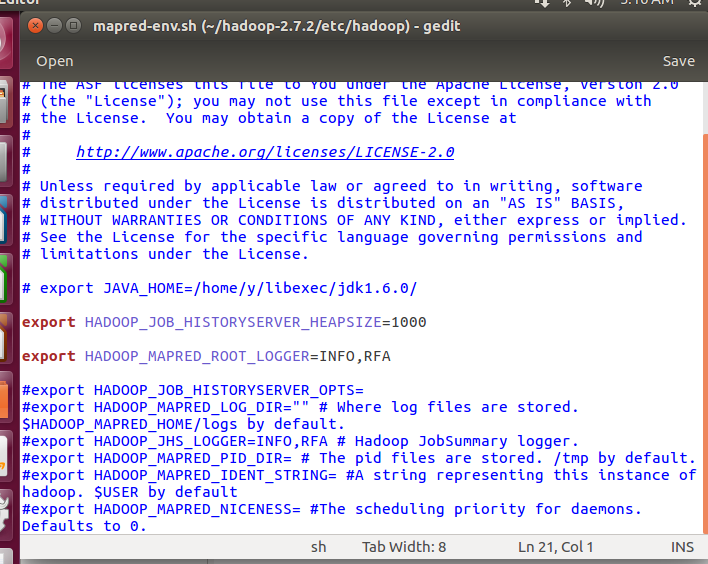


图20 — mapred-env.sh

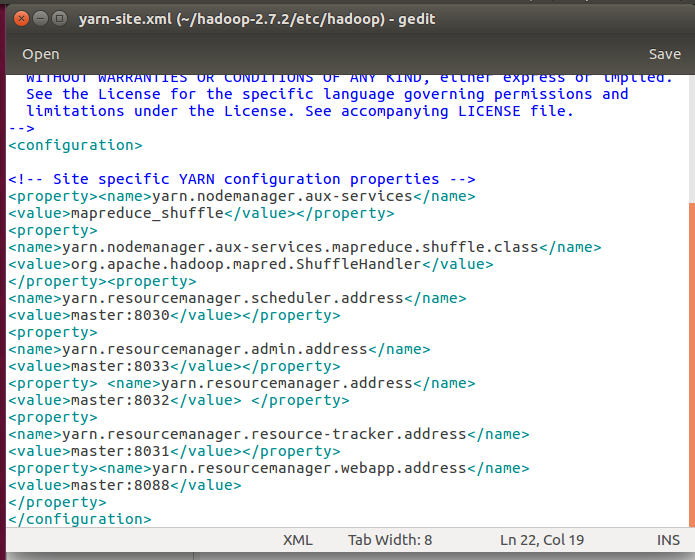


图21 — yarn-site.xml

图片包含 截图, 游戏机, 桌子

描述已自动生成

图22 — profile

* Jps测试

图片包含 游戏机, 标志

描述已自动生成

图23 — master jps

图片包含 游戏机, 画

描述已自动生成

图24 — slave1 jps

1. hive环境搭建：

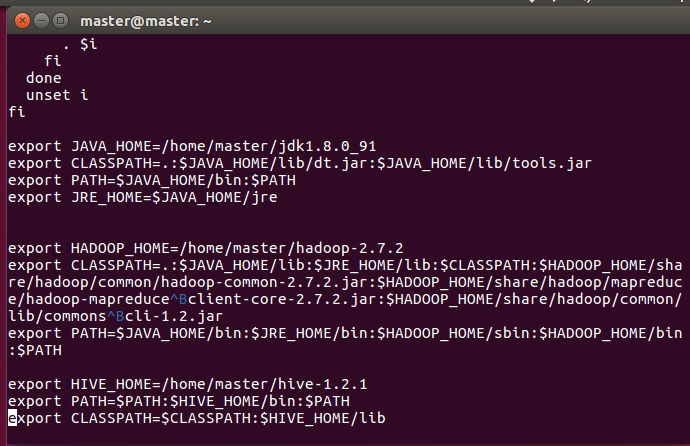


图25 — profile文件

* 启动hive：

图片包含 截图, 游戏机, 电话

描述已自动生成

图26 — hive的启动

* Hive数据库及表结构

图片包含 截图, 游戏机, 电话

描述已自动生成

图27 — hive 数据库及表结构

* Nycsv表内容：（上传的本地数据）

图片包含 游戏机, 电脑, 桌子

描述已自动生成

图28 — 上传数据内容（中文乱码问题转换编码方式即可解决）