Pig：Scripting脚本语言，可以将脚本转换成MapReduce作业在Hadoop集群上运行，使用类似SQL的面向数据流的语言Pig Latin，Pig Latin可以完成排序，过滤，求和，分组，关联等操作，可以支持自定义函数，Pig会将Pig Latin映射为MapReduce作业上传到集群运行，减少Java代码的编写，Pig的三种运行方式：Grunt shell，脚本方式，嵌入式

Pig由雅虎贡献给Apache基金会的

Linux使用set命令查看所有环境变量

Grunt：有自动补全机制（但只是命令补全机制），Autocomplete文件是大小写敏感的自动补全文件

Grunt shell命令：

help：查看命令列表

ls 目录：显示当前目录列表

cd 目录：进入指定目录

cat 文件名：查看文件

copyToLocal 文件名 本机系统全路径名：复制文件到本机系统

sh 本机系统全路径名：执行本机系统命令

Pig数据模型：

Bag：表

Tuple：行，记录

Field：属性

Pig不要求同一个Bag里面的各个Tuple有相同数量或者相同类型的Field，如果相同则成为关系即类似关系型数据库

Pig Latin常用语句

变量名 = LOAD 'hdfs路径文件名' AS (属性名:属性类型);：按照指定格式加载hdfs上的文件数据，默认文件中的数据分隔符为制表符Tab

USING PigStorage('分隔符');：指定分隔符

变量名 = FOREACH 变量名 GENERATE 属性名,处理函数（MAX(属性名.属性名)）;：逐行扫描进行某种处理

变量名 = FILTER 变量名BY 过滤条件（属性名 != 属性值 AND 属性名 = '属性值'）;：过滤指定变量名的数据结果

DUMP 变量名：将指定变量名的结果显示到屏幕上

STORE 变量名 INTO 'hdfs路径名'：将结果保存到文件

DESCRIBE 变量名：查看关系的结构

变量名= GROUP 变量名 BY 属性名;：按指定属性名将指定变量名的数据分组

文件读取默认分隔符是制表符Tab，可以设置指定制表符

Pig Latin支持使用Java、Python、JavaScript三种语言编写UDF