# **VM installation**

安装vmware workstation；

出现错误：

http://predictionio.incubator.apache.org/system/

推荐引擎 http://actionml.com/docs/ur\_config

VMware Workstation unrecoverable error: (vcpu-0)

vcpu-0:VERIFY vmcore/vmm/main/cpuid.c:382 bugNr=1036521

A log file is available in "C:\Users\Elochukwu Ukwandu\Documents\Virtual Machines\Windows 8\vmware.log".

问题分析：

The failure indicates that you are using binary translation on a CPU that doesn't support CPUID spoofing, and the guest CPUID features don't match the host CPUID features.  I am a little surprised that this is happening, but it is an old processor, and it might not support CPUID masking.  I would like to see the vmware.log file if you can find it.

You should be able to work around this issue by changing the preferred execution mode to VT-x, assuming that VT-x is enabled on the host, and that the guest OS isn't one of the few that requires binary translation.

解决方案：

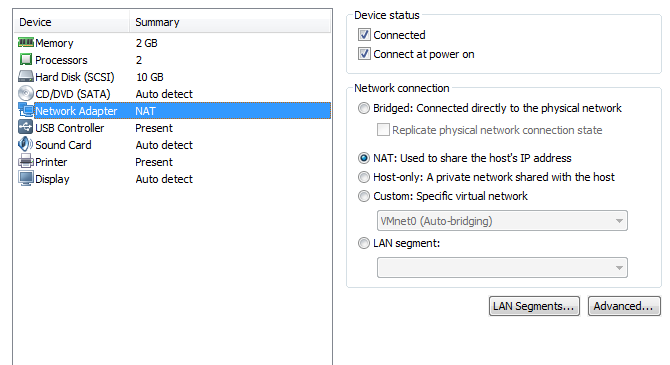
The tip from jmattson still counts.

Go into the BIOS/EFI from your HP laptop and enable VT-x/virtualisation.

## **Network setup**

### **虚机中设置成NAT模式**

在windows安装的vmware菜单里打开虚拟机网络设置：



### **虚机中设置IP地址**

[root@localhost network-scripts]# cd /etc/sysconfig/network-scripts

[root@localhost network-scripts]# cat ifcfg-eth0

DEVICE="eth0"

IPADDR=192.168.5.110

NETMASK=255.255.255.0

BOOTPROTO="static"

HWADDR="00:0C:29:8D:6A:61"

IPV6INIT="yes"

NM\_CONTROLLED="yes"

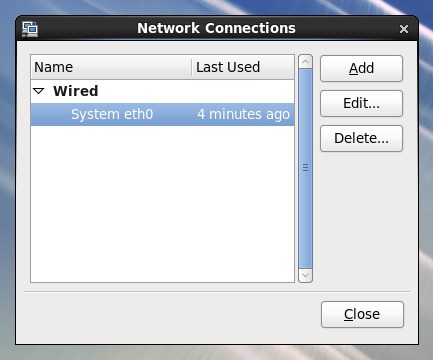
ONBOOT="yes"

TYPE="Ethernet"

UUID="d74d8f26-f638-4ac1-8635-504af72e5b92"

### **设置DNS**

在虚拟机的菜单preference->network connections中设置DNS和网关地址：



DNS: 202.96.199.133

GATEWAY : 192.168.5.2

### **重启网络服务**

[root@localhost init.d]# service network restart

Shutting down interface eth0: Device state: 3 (disconnected) [ OK ]

Shutting down loopback interface: [ OK ]

Bringing up loopback interface: [ OK ]

Bringing up interface eth0: Active connection state: activating

Active connection path: /org/freedesktop/NetworkManager/ActiveConnection/29

state: activated

Connection activated [ OK ]

1. 对于clone的需求，丢失eth0而变更为eth1的，修改70-persistent-net.rules文件：

[root@localhost udev]# cat cd /etc/udev/rules.d

# This file was automatically generated by the /lib/udev/write\_net\_rules

# program, run by the persistent-net-generator.rules rules file.

#

# You can modify it, as long as you keep each rule on a single

# line, and change only the value of the NAME= key.

# PCI device 0x8086:0x100f (e1000)

#SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?\*", ATTR{address}=="00:0c:29:8d:6a:61", ATTR{type}=="1", KERNEL=="eth\*", NAME="eth0"

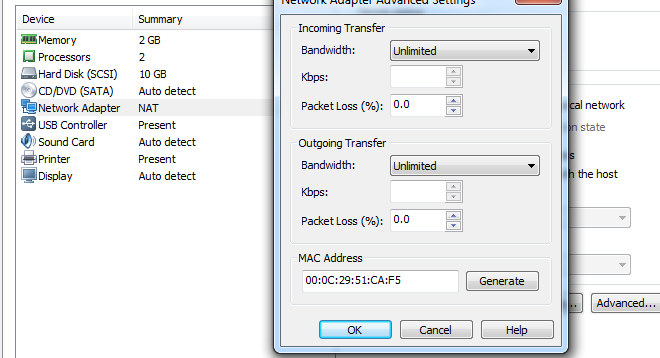
# PCI device 0x8086:0x100f (e1000)

SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?\*", ATTR{address}=="00:0c:29:62:7e:1f", ATTR{type}=="1", KERNEL=="eth\*", NAME="eth0"

将eth0的行注释掉，将NAME="eth1"修改为NAME="eth0"，

同样修改ifcfg-eth0，增加IPADDR, NETMASK, 并将HWADDR="00:0C:29:8D:6A:61"

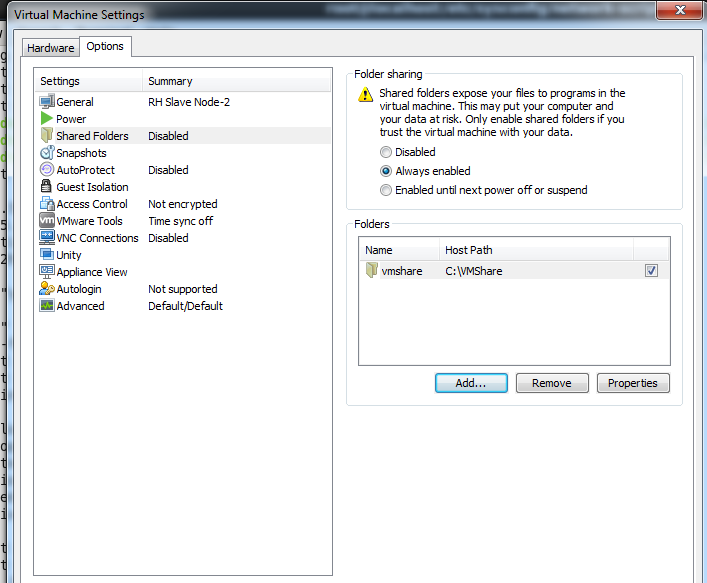
中的MAC地址修改为实际的MAC 地址。



修改完成后重新启动机器。

## **在虚机中访问宿主机目录**

选择虚机，选择设置：



将folder sharing改成always enabled. 并增加folders。

主机中可访问的目录在/mnt/hgfs/vmshare目录下。

## **其它**

1. 安装JDK到/opt/jdk1.7.0\_75，设置JAVA\_HOME.
2. 设置hostname:

[root@localhost sysconfig]# hostname master.hadoop

[root@localhost sysconfig]# hostname

master.hadoop

[root@localhost sysconfig]# cat network

NETWORKING=yes

HOSTNAME=master.hadoop

[root@localhost sysconfig]#

1. 修改etc/hosts文件，添加：

192.168.5.110 master.hadoop

192.168.5.118 datanode1.hadoop

192.168.5.119 datanode2.hadoop

重启网络服务：

[root@localhost sysconfig]# service network restart

# **打通SSH**

## **每台节点都生成秘钥**

[root@master .ssh]# ssh-keygen -t rsa

Generating public/private rsa key pair.

Enter file in which to save the key (/root/.ssh/id\_rsa):

Enter passphrase (empty for no passphrase):

Enter same passphrase again:

Your identification has been saved in /root/.ssh/id\_rsa.

Your public key has been saved in /root/.ssh/id\_rsa.pub.

The key fingerprint is:

eb:fb:72:2b:1e:a2:3d:48:33:0a:4c:48:d2:f3:6a:bb root@master.hadoop

The key's randomart image is:

+--[ RSA 2048]----+

| . |

|..o |

|+ o |

|.. . |

|o . S |

|..o + . |

| o + +. o |

| o .o.oo.. |

| E.. .o+\*o. |

+-----------------+

注意: id\_rsa不要修改名字，缺省回车即可。

将所有节点生成的公钥.pub文件统一传到一个服务器上 。例如下面的118.pub是从服务器118传过来的。

[root@master .ssh]# ls

118.pub authorized\_keys id\_rsa id\_rsa.pub known\_hosts

[root@master .ssh]# cat id\_rsa.pub >> authorized\_keys

[root@master .ssh]# cat 118.pub >> authorized\_keys

将authorized\_keys 复制到118服务器上去。

[root@master .ssh]# scp authorized\_keys datanode1.hadoop:/root/.ssh/

确保每个节点上的文件权限：

[root@master .ssh]# chmod 600 authorized\_keys

[root@master root]# chmod 700 .ssh

设置sshd\_config:

cd /etc/ssh

[root@localhost ssh]# vi sshd\_config

修改如下几项：

RSAAuthentication yes

PubkeyAuthentication yes

AuthorizedKeysFile .ssh/authorized\_keys

重启动ssh服务：

service sshd restart

如果有错，尝试：

setenforce 0

grep /etc/

如果仍然有错，可以用ssh调试模式查看ssh过程：

ssh -vvvT master.hadoop

# **关闭防火墙**

[root@master .ssh]# service iptables stop

iptables: Flushing firewall rules: [ OK ]

iptables: Setting chains to policy ACCEPT: filter [ OK ]

iptables: Unloading modules: [ OK ]

[root@master .ssh]# chkconfig iptables off

# **安装NTP服务(failed)**

For the NTTP server:

[root@master ~]# rpm -qa | grep ntp

ntpdate-4.2.4p8-3.el6.x86\_64

fontpackages-filesystem-1.41-1.1.el6.noarch

ntp-4.2.4p8-3.el6.x86\_64

[root@master ~]# chkconfig --add ntpd

[root@master ~]# chkconfig ntpd on

[root@master ~]# cat /etc/ntp.conf

# For more information about this file, see the man pages

# ntp.conf(5), ntp\_acc(5), ntp\_auth(5), ntp\_clock(5), ntp\_misc(5), ntp\_mon(5).

driftfile /var/lib/ntp/drift

# Permit time synchronization with our time source, but do not

# permit the source to query or modify the service on this system.

restrict default kod nomodify notrap nopeer noquery

restrict -6 default kod nomodify notrap nopeer noquery

# Permit all access over the loopback interface. This could

# be tightened as well, but to do so would effect some of

# the administrative functions.

restrict 127.0.0.1

restrict -6 ::1

# Hosts on local network are less restricted.

#restrict 192.168.1.0 mask 255.255.255.0 nomodify notrap

# Use public servers from the pool.ntp.org project.

# Please consider joining the pool (http://www.pool.ntp.org/join.html).

server 0.rhel.pool.ntp.org

server 1.rhel.pool.ntp.org

server 2.rhel.pool.ntp.org

#broadcast 192.168.1.255 autokey # broadcast server

#broadcastclient # broadcast client

#broadcast 224.0.1.1 autokey # multicast server

#multicastclient 224.0.1.1 # multicast client

#manycastserver 239.255.254.254 # manycast server

#manycastclient 239.255.254.254 autokey # manycast client

# Undisciplined Local Clock. This is a fake driver intended for backup

# and when no outside source of synchronized time is available.

#server 127.127.1.0 # local clock

#fudge 127.127.1.0 stratum 10

# Enable public key cryptography.

#crypto

includefile /etc/ntp/crypto/pw

# Key file containing the keys and key identifiers used when operating

# with symmetric key cryptography.

keys /etc/ntp/keys

# Specify the key identifiers which are trusted.

#trustedkey 4 8 42

# Specify the key identifier to use with the ntpdc utility.

#requestkey 8

# Specify the key identifier to use with the ntpq utility.

#controlkey 8

# Enable writing of statistics records.

#statistics clockstats cryptostats loopstats peerstats

[root@master ~]# chkconfig --list ntpd

ntpd 0:off 1:off 2:on 3:on 4:on 5:on 6:off

[root@master ~]# service ntpd start

Starting ntpd: [ OK ]

[root@master ~]# chkconfig ntpd on

For the NTP client(data node):

[root@datanode1 ~]# ntpdate -u 192.168.5.110

8 Aug 03:38:42 ntpdate[2597]: no server suitable for synchronization found

# **安装JDK**

安装版本：1.8.0\_31

rpm -ivh jdk-8u73-linux-x64.rpm

# **build PredictionIO from source**

## **下载和解压**

tar –xvf apache-predictionio-0.11.0-incubating.tar.gz

## **Build**

参考安装文档：http://predictionio.incubator.apache.org/install/install-sourcecode/#building

[root@master predictionio-0.11.0]# ls

apache-predictionio-0.11.0-incubating.tar.gz build.sbt CONTRIBUTING.md DISCLAIMER e2 LICENSE.txt PMC.md RELEASE.md storage

assembly common core Dockerfile examples make-distribution.sh project sbt tests

bin conf data docs KEYS NOTICE.txt README.md scalastyle-config.xml tools

[root@master predictionio-0.11.0]# ./make-distribution.sh

Building binary distribution for PredictionIO 0.11.0-incubating...

+ sbt/sbt clean

Getting org.scala-sbt sbt 0.13.13 ...

downloading https://repo.typesafe.com/typesafe/ivy-releases/org.scala-sbt/sbt/0.13.13/jars/sbt.jar ...

[SUCCESSFUL ] org.scala-sbt#sbt;0.13.13!sbt.jar (11320ms)

downloading https://repo1.maven.org/maven2/org/scala-lang/scala-library/2.10.6/scala-library-2.10.6.jar ...

[SUCCESSFUL ] org.scala-lang#scala-library;2.10.6!scala-library.jar (26957ms)

downloading https://repo.typesafe.com/typesafe/ivy-releases/org.scala-sbt/main/0.13.13/jars/main.jar ...

[SUCCESSFUL ] org.scala-sbt#main;0.13.13!main.jar (25085ms)

downloading https://repo.typesafe.com/typesafe/ivy-releases/org.scala-sbt/compiler-interface/0.13.13/jars/compiler-interface.jar ...

…

PredictionIO-0.11.0-incubating/sbt/sbt

PredictionIO-0.11.0-incubating/RELEASE

PredictionIO binary distribution created at PredictionIO-0.11.0-incubating.tar.gz

解压该 gz即可获取该版本：

$ tar zxvf PredictionIO-0.11.0-incubating.tar.gz

# **安装依赖软件**

如果未在PredictionIO-0.11.0-incubating/vendors 下安装这些依赖软件，则需要修改PredictionIO-0.11.0-incubating/conf/pio-env.sh 中相关配置

## **安装Spark**

在predictionio安装目录下建立目录：

$mkdir PredictionIO-0.11.0-incubating/vendors

直接解压：

$tar zxvf spark-1.6.3-bin-hadoop2.6.tgz

注：单机版本不需要安装hadoop。

启动master:

$./sbin/start-master.sh

启动worker:

$./sbin/start-slave.sh <master-spark-URL>

## **安装Elastic Search**

在predictionio安装目录下建立目录，并解压

$cd ./PredictionIO-0.11.0-incubating/vendors

# tar xvf elasticsearch-1.7.6.tar.gz

启动elasticsearch:

[root@master bin]# ./elasticsearch

[2017-09-29 19:21:43,864][INFO ][node ] [Bloodhawk] version[1.7.6], pid[2760], build[c730b59/2016-11-18T15:21:16Z]

[2017-09-29 19:21:43,878][INFO ][node ] [Bloodhawk] initializing ...

[2017-09-29 19:21:44,204][INFO ][plugins ] [Bloodhawk] loaded [], sites []

[2017-09-29 19:21:44,324][INFO ][env ] [Bloodhawk] using [1] data paths, mounts [[/ (rootfs)]], net usable\_space [11.5gb], net total\_space [17.4gb], types [rootfs]

[2017-09-29 19:21:52,427][INFO ][node ] [Bloodhawk] initialized

[2017-09-29 19:21:52,428][INFO ][node ] [Bloodhawk] starting ...

[2017-09-29 19:21:52,574][INFO ][transport ] [Bloodhawk] bound\_address {inet[/0:0:0:0:0:0:0:0:9300]}, publish\_address {inet[/192.168.40.134:9300]}

[2017-09-29 19:21:52,626][INFO ][discovery ] [Bloodhawk] elasticsearch/Fxv8eLYyTJmeN6L8TxmShQ

[2017-09-29 19:21:56,463][INFO ][cluster.service ] [Bloodhawk] new\_master [Bloodhawk][Fxv8eLYyTJmeN6L8TxmShQ][master][inet[/192.168.40.134:9300]], reason: zen-disco-join (elected\_as\_master)

[2017-09-29 19:21:56,623][INFO ][http ] [Bloodhawk] bound\_address {inet[/0:0:0:0:0:0:0:0:9200]}, publish\_address {inet[/192.168.40.134:9200]}

[2017-09-29 19:21:56,623][INFO ][node ] [Bloodhawk] started

[2017-09-29 19:21:56,789][INFO ][gateway ] [Bloodhawk] recovered [1] indices into cluster\_state

测试该实例：

[root@master ~]# curl 'http://localhost:9200/?pretty'

{

"status" : 200,

"name" : "Bloodhawk",

"cluster\_name" : "elasticsearch",

"version" : {

"number" : "1.7.6",

"build\_hash" : "c730b59357f8ebc555286794dcd90b3411f517c9",

"build\_timestamp" : "2016-11-18T15:21:16Z",

"build\_snapshot" : false,

"lucene\_version" : "4.10.4"

},

"tagline" : "You Know, for Search"

}

## **安装Zookeeper(Optional:可用HBASE自带的)**

下载并解压。 这里使用版本zookeeper-3.4.10， 解压完成后进入其根目录。 此处配置为单机运行。修改文件zoo.cfg：

# The number of milliseconds of each tick

tickTime=2000

# The number of ticks that the initial

# synchronization phase can take

initLimit=10

# The number of ticks that can pass between

# sending a request and getting an acknowledgement

syncLimit=5

# the directory where the snapshot is stored.

# do not use /tmp for storage, /tmp here is just

# example sakes.

dataDir=/mnt/share/zookeeper-3.4.10/data

# the port at which the clients will connect

clientPort=2181

# the maximum number of client connections.

# increase this if you need to handle more clients

#maxClientCnxns=60

#

# Be sure to read the maintenance section of the

# administrator guide before turning on autopurge.

#

# http://zookeeper.apache.org/doc/current/zookeeperAdmin.html#sc\_maintenance

#

# The number of snapshots to retain in dataDir

#autopurge.snapRetainCount=3

# Purge task interval in hours

# Set to "0" to disable auto purge feature

#autopurge.purgeInterval=1

配置完成后即可启动zookeeper server:

[root@master bin]# ./zkServer.sh start

ZooKeeper JMX enabled by default

Using config: /mnt/share/zookeeper-3.4.10/bin/../conf/zoo.cfg

Starting zookeeper ... STARTED

[root@master bin]# jps

2516 QuorumPeerMain

2533 Jps

启动后连接到服务器：

[root@master bin]# ./zkCli.sh -server 127.0.0.1:2181

Connecting to 127.0.0.1:2181

2017-10-10 19:15:10,133 [myid:] - INFO [main:Environment@100] - Client environment:zookeeper.version=3.4.10-39d3a4f269333c922ed3db283be479f9deacaa0f, built on 03/23/2017 10:13 GMT

2017-10-10 19:15:10,139 [myid:] - INFO [main:Environment@100] - Client environment:host.name=master

2017-10-10 19:15:10,139 [myid:] - INFO [main:Environment@100] - Client environment:java.version=1.8.0\_73

2017-10-10 19:15:10,141 [myid:] - INFO [main:Environment@100] - Client environment:java.vendor=Oracle Corporation

2017-10-10 19:15:10,141 [myid:] - INFO [main:Environment@100] - Client environment:java.home=/usr/java/jdk1.8.0\_73/jre

2017-10-10 19:15:10,141 [myid:] - INFO [main:Environment@100] - Client environment:java.class.path=/mnt/share/zookeeper-3.4.10/bin/../build/classes:/mnt/share/zookeeper-3.4.10/bin/../build/lib/\*.jar:/mnt/share/zookeeper-3.4.10/bin/../lib/slf4j-log4j12-1.6.1.jar:/mnt/share/zookeeper-3.4.10/bin/../lib/slf4j-api-1.6.1.jar:/mnt/share/zookeeper-3.4.10/bin/../lib/netty-3.10.5.Final.jar:/mnt/share/zookeeper-3.4.10/bin/../lib/log4j-1.2.16.jar:/mnt/share/zookeeper-3.4.10/bin/../lib/jline-0.9.94.jar:/mnt/share/zookeeper-3.4.10/bin/../zookeeper-3.4.10.jar:/mnt/share/zookeeper-3.4.10/bin/../src/java/lib/\*.jar:/mnt/share/zookeeper-3.4.10/bin/../conf:

2017-10-10 19:15:10,142 [myid:] - INFO [main:Environment@100] - Client environment:java.library.path=/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib

2017-10-10 19:15:10,142 [myid:] - INFO [main:Environment@100] - Client environment:java.io.tmpdir=/tmp

2017-10-10 19:15:10,142 [myid:] - INFO [main:Environment@100] - Client environment:java.compiler=<NA>

2017-10-10 19:15:10,142 [myid:] - INFO [main:Environment@100] - Client environment:os.name=Linux

2017-10-10 19:15:10,142 [myid:] - INFO [main:Environment@100] - Client environment:os.arch=amd64

2017-10-10 19:15:10,143 [myid:] - INFO [main:Environment@100] - Client environment:os.version=3.10.0-327.22.2.el7.x86\_64

2017-10-10 19:15:10,143 [myid:] - INFO [main:Environment@100] - Client environment:user.name=root

2017-10-10 19:15:10,143 [myid:] - INFO [main:Environment@100] - Client environment:user.home=/root

2017-10-10 19:15:10,143 [myid:] - INFO [main:Environment@100] - Client environment:user.dir=/mnt/share/zookeeper-3.4.10/bin

2017-10-10 19:15:10,145 [myid:] - INFO [main:ZooKeeper@438] - Initiating client connection, connectString=127.0.0.1:2181 sessionTimeout=30000 watcher=org.apache.zookeeper.ZooKeeperMain$MyWatcher@799f7e29

Welcome to ZooKeeper!

2017-10-10 19:15:10,195 [myid:] - INFO [main-SendThread(127.0.0.1:2181):ClientCnxn$SendThread@1032] - Opening socket connection to server 127.0.0.1/127.0.0.1:2181. Will not attempt to authenticate using SASL (unknown error)

JLine support is enabled

2017-10-10 19:15:10,410 [myid:] - INFO [main-SendThread(127.0.0.1:2181):ClientCnxn$SendThread@876] - Socket connection established to 127.0.0.1/127.0.0.1:2181, initiating session

2017-10-10 19:15:10,581 [myid:] - INFO [main-SendThread(127.0.0.1:2181):ClientCnxn$SendThread@1299] - Session establishment complete on server 127.0.0.1/127.0.0.1:2181, sessionid = 0x15f05fcf3110000, negotiated timeout = 30000

WATCHER::

WatchedEvent state:SyncConnected type:None path:null

[zk: 127.0.0.1:2181(CONNECTED) 0]

可执行各种client端命令。用help列出可执行命令。

## **安装HBase**

### **解压**

# tar xvf hbase-1.2.6-bin.tar.gz

### **配置文件**

1. 设置JAVA\_HOME，修改文件JAVA\_HOME部分

[root@master conf]# vi hbase-env.sh

# The java implementation to use. Java 1.7+ required.

# export JAVA\_HOME=/usr/java/jdk1.6.0/

export JAVA\_HOME=/usr/bin

1. 修改*conf/hbase-site.xml*

Standalone模式下仅需要配置hbase和zookeeper写数据的本地文件路径。不需要自己创建hbase和zookeeper文件目录。

<configuration>

<property>

<name>hbase.rootdir</name>

<value>file:///mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/hbase-1.2.6/data</value>

</property>

<property>

<name>hbase.zookeeper.property.dataDir</name>

<value>/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/hbase-1.2.6/zookeeper</value>

</property>

</configuration>

1. 修改*conf/hbase-env.sh*

提供正确的JAVA\_HOME， 一般用/usr即可（rpm安装的JDK）

# The java implementation to use. Java 1.7+ required.

# export JAVA\_HOME=/usr/java/jdk1.6.0/

export JAVA\_HOME=/usr

1. 连接HBASE：

连接GUI

[http://localhost:16010](http://localhost:16010/)

启动HBASE，并连接服务端：

[root@master bin]# ./start-hbase.sh

starting master, logging to /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/hbase-1.2.6/bin/../logs/hbase-root-master-master.out

Java HotSpot(TM) 64-Bit Server VM warning: ignoring option PermSize=128m; support was removed in 8.0

Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=128m; support was removed in 8.0

[root@master bin]# ./hbase shell

2017-09-23 02:00:44,789 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

HBase Shell; enter 'help<RETURN>' for list of supported commands.

Type "exit<RETURN>" to leave the HBase Shell

Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

hbase(main):001:0>

1. 停止HBASE

[root@master bin]# ./stop-hbase.sh

stopping hbase....

# **配置和启动PredictionIO**

## **配置文件**

配置./conf/pio-env.sh, 将meta data的存储改成ES, 事件和模型的存储改成HBASE。

# Default is to use PostgreSQL

PIO\_STORAGE\_REPOSITORIES\_METADATA\_NAME=pio\_meta

PIO\_STORAGE\_REPOSITORIES\_METADATA\_SOURCE=ELASTICSEARCH

PIO\_STORAGE\_REPOSITORIES\_EVENTDATA\_NAME=pio\_event

PIO\_STORAGE\_REPOSITORIES\_EVENTDATA\_SOURCE=HBASE

PIO\_STORAGE\_REPOSITORIES\_MODELDATA\_NAME=pio\_model

PIO\_STORAGE\_REPOSITORIES\_MODELDATA\_SOURCE=LOCALFS

# Elasticsearch Example

# PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_TYPE=elasticsearch

# PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_HOSTS=localhost

# PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_PORTS=9200

# PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_SCHEMES=http

# PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_HOME=$PIO\_HOME/vendors/elasticsearch-5.2.1

# Elasticsearch 1.x Example

PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_TYPE=elasticsearch

#PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_CLUSTERNAME=<elasticsearch\_cluster\_name>

PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_HOSTS=localhost

PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_PORTS=9300

PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_HOME=$PIO\_HOME/vendors/elasticsearch-1.7.6

# Local File System Example

PIO\_STORAGE\_SOURCES\_LOCALFS\_TYPE=localfs

PIO\_STORAGE\_SOURCES\_LOCALFS\_PATH=$PIO\_FS\_BASEDIR/models

# HBase Example

PIO\_STORAGE\_SOURCES\_HBASE\_TYPE=hbase

PIO\_STORAGE\_SOURCES\_HBASE\_HOME=$PIO\_HOME/vendors/hbase-1.2.6

## **启动predicitonIO**

[root@master bin]# ./pio-start-all

Starting Elasticsearch...

Starting HBase...

starting master, logging to /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/hbase-1.2.6/bin/../logs/hbase-root-master-master.out

Java HotSpot(TM) 64-Bit Server VM warning: ignoring option PermSize=128m; support was removed in 8.0

Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=128m; support was removed in 8.0

Waiting 10 seconds for Storage Repositories to fully initialize...

Starting PredictionIO Event Server...

[root@master bin]# jps -l

6769 org.apache.predictionio.tools.console.Console

6852 sun.tools.jps.Jps

6409 org.elasticsearch.bootstrap.Elasticsearch

6492 org.apache.hadoop.hbase.master.HMaster

[root@master bin]# ./pio status

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[INFO] [Management$] Inspecting PredictionIO...

[INFO] [Management$] PredictionIO 0.11.0-incubating is installed at /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating

[INFO] [Management$] Inspecting Apache Spark...

[INFO] [Management$] Apache Spark is installed at /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/spark-1.6.3-bin-hadoop2.6

[INFO] [Management$] Apache Spark 1.6.3 detected (meets minimum requirement of 1.3.0)

[INFO] [Management$] Inspecting storage backend connections...

[INFO] [Storage$] Verifying Meta Data Backend (Source: ELASTICSEARCH)...

[INFO] [Storage$] Verifying Model Data Backend (Source: LOCALFS)...

[INFO] [Storage$] Verifying Event Data Backend (Source: HBASE)...

[INFO] [Storage$] Test writing to Event Store (App Id 0)...

[INFO] [HBLEvents] The namespace pio\_event doesn't exist yet. Creating now...

[INFO] [HBLEvents] The table pio\_event:events\_0 doesn't exist yet. Creating now...

[INFO] [HBLEvents] Removing table pio\_event:events\_0...

[INFO] [Management$] Your system is all ready to go.

## **从Engine template创建新的engine**

参考：

<http://actionml.com/docs/ur_quickstart>

<http://predictionio.incubator.apache.org/templates/recommendation/quickstart/>

### **前置条件**

* python 2.7.9+
* pip, the package installer for python
* git, for version control and installing templates like the UR

### **安装python 2.7.11**

备份原有python（centOS7缺省安装2.7.5版本）

[root@master bin]cd /bin

[root@master bin]mv python python.bak

下载并编译python 2.7.11

[root@master engines]Wget https://www.python.org/ftp/python/2.7.11/Python-2.7.11.tgz

[root@master share]# tar -zxvf Python-2.7.11.tgz

[root@master share]# cd Python-2.7.11

[root@master Python-2.7.11]# ./configure --prefix=/usr/local/python/python2711

checking build system type... x86\_64-unknown-linux-gnu

checking host system type... x86\_64-unknown-linux-gnu

checking for --enable-universalsdk... no

checking for --with-universal-archs... 32-bit

checking MACHDEP... linux2

checking EXTRAPLATDIR...

。。。

creating Modules/Setup.local

creating Makefile

执行 ./configure --prefix=/usr/local/python/python3 命令。

./configure命令执行完毕之后创建一个文件creating Makefile，供下面的make命令使用 执行make install之后就会把程序安装到我们指定的目录中去。

Configure是一个可执行脚本，它有很多选项，在待安装的源码路径下使用命令./configure –help输出详细的选项列表。其中--prefix选项是配置安装的路径，如果不配置该选项，安装后可执行文件默认放在/usr /local/bin，库文件默认放在/usr/local/lib，配置文件默认放在/usr/local/etc，其它的资源文件放在/usr /local/share，比较凌乱。如果配置--prefix，如：./configure --prefix=/usr/local/test可以把所有资源文件放在/usr/local/test的路径中，不会杂乱。用了—prefix选项的另一个好处是卸载软件或移植软件。当某个安装的软件不再需要时，只须简单的删除该安装目录，就可以把软件卸载得干干净净；移植软件只需拷贝整个目录到另外一个机器即可（相同的操作系统）。当然要卸载程序，也可以在原来的make目录下用一次make uninstall，但前提是make文件指定过uninstall。

然后执行make、make install命令

[root@master Python-2.7.11]# make

….

running build\_scripts

creating build/scripts-2.7

copying and adjusting /mnt/share/Python-2.7.11/Tools/scripts/pydoc -> build/scripts-2.7

copying and adjusting /mnt/share/Python-2.7.11/Tools/scripts/idle -> build/scripts-2.7

copying and adjusting /mnt/share/Python-2.7.11/Tools/scripts/2to3 -> build/scripts-2.7

copying and adjusting /mnt/share/Python-2.7.11/Lib/smtpd.py -> build/scripts-2.7

changing mode of build/scripts-2.7/pydoc from 644 to 755

changing mode of build/scripts-2.7/idle from 644 to 755

changing mode of build/scripts-2.7/2to3 from 644 to 755

changing mode of build/scripts-2.7/smtpd.py from 644 to 755

/bin/install -c -m 644 ./Tools/gdb/libpython.py python-gdb.py

[root@master Python-2.7.11]# make install

[root@master bin]# cd /usr/local/python/python2711

安装成功之后进入到python2711文件夹可以看到Python的可执行文件目录、lib目录等相关目录。

[root@master bin]# ln -s /usr/local/python/python2711/bin/python /bin/python

[root@master bin]# python -V

Python 2.7.11

Python 2.7.11 版本会导致yum（基于2.7.5版本）不可用，因此修改yum配置文件(vi /usr/bin/yum)，把文件头部的#!/usr/bin/python改成#!/usr/bin/python2.7保存退出即可。

### **安装pip**

使用get-pip.py安装pip:

[root@master python]# curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

100 1558k 100 1558k 0 0 452k 0 0:00:03 0:00:03 --:--:-- 452k

[root@master python]# ls

get-pip.py python2711

[root@master python]# python get-pip.py

Collecting pip

Downloading pip-9.0.1-py2.py3-none-any.whl (1.3MB)

100% |████████████████████████████████| 1.3MB 189kB/s

Collecting setuptools

Downloading setuptools-36.5.0-py2.py3-none-any.whl (478kB)

100% |████████████████████████████████| 481kB 257kB/s

Collecting wheel

Downloading wheel-0.30.0-py2.py3-none-any.whl (49kB)

100% |████████████████████████████████| 51kB 271kB/s

Installing collected packages: pip, setuptools, wheel

Successfully installed pip-9.0.1 setuptools-36.5.0 wheel-0.30.0

[root@master python]#

将/usr/local/python/bin加入path

### **下载UR engine模板**

从github下载template

[root@master engines]# git clone https://github.com/actionml/universal-recommender.git ur

正克隆到 'ur'...

remote: Counting objects: 1667, done.

remote: Compressing objects: 100% (7/7), done.

remote: Total 1667 (delta 1), reused 2 (delta 0), pack-reused 1660

接收对象中: 100% (1667/1667), 2.54 MiB | 531.00 KiB/s, done.

处理 delta 中: 100% (1032/1032), done.

### **创建新engine(Universal Recommender)**

获取app ID和access key.

[root@master bin]# ./pio app new URec

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[INFO] [HBLEvents] The table pio\_event:events\_1 doesn't exist yet. Creating now...

[INFO] [App$] Initialized Event Store for this app ID: 1.

[INFO] [Pio$] Created a new app:

[INFO] [Pio$] Name: URec

[INFO] [Pio$] ID: 1

[INFO] [Pio$] Access Key: H0Dv88B2Z6A0097l8jr-5ER32E9f8HiYObOrhrawHArwkxMNZFkyjpFv7z17nnAs

[root@master bin]# ./pio app list

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[INFO] [Pio$] Name | ID | Access Key | Allowed Event(s)

[INFO] [Pio$] URec | 1 | H0Dv88B2Z6A0097l8jr-5ER32E9f8HiYObOrhrawHArwkxMNZFkyjpFv7z17nnAs | (all)

[INFO] [Pio$] Finished listing 1 app(s).

### **编辑engine.json.template并保存为engine.json**

完整的参数配置参考：<http://actionml.com/docs/ur_config>。最简单的缺省配置如下。可以先使用缺省配置，等熟悉后再逐步修改和优化现有参数。

确保engine.json的appName和前面定义的app name一致：

...

"datasource": {

"params" : {

"appName": "URec"

}

},

...

### **Build**

[root@master ur]# pio build –verbose

…

[INFO] [Engine$] [warn] Strategy 'first' was applied to a file

[INFO] [Engine$] [warn] Strategy 'rename' was applied to 5 files

[INFO] [Engine$] [info] SHA-1: bcff99e7b95bacdf803c7d90fad81876b1cd87aa

[INFO] [Engine$] [info] Packaging /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/engines/ur/target/scala-2.10/universal-recommender-assembly-0.6.0-deps.jar ...

[INFO] [Engine$] [info] Done packaging.

[INFO] [Engine$] [success] Total time: 1204 s, completed 2017-9-29 1:16:05

[INFO] [Engine$] Compilation finished successfully.

[INFO] [Engine$] Looking for an engine...

[INFO] [Engine$] Found universal-recommender\_2.10-0.6.0.jar

[INFO] [Engine$] Found universal-recommender-assembly-0.6.0-deps.jar

[INFO] [Engine$] Build finished successfully.

[INFO] [Pio$] Your engine is ready for training.

### **安装SDK**

要使用python脚本（import\_eventserver.py）来导入一些测试数据，需要先安装Python的SDK。

[root@master data]# pip install predictionio

Collecting predictionio

Downloading PredictionIO-0.9.8.tar.gz

Collecting pytz>=2014.2 (from predictionio)

Downloading pytz-2017.2-py2.py3-none-any.whl (484kB)

100% |████████████████████████████████| 491kB 25kB/s

Building wheels for collected packages: predictionio

Running setup.py bdist\_wheel for predictionio ... done

Stored in directory: /root/.cache/pip/wheels/3e/69/70/71068da36c318aaeecf484514d056774300ca74e108a0f812c

Successfully built predictionio

Installing collected packages: pytz, predictionio

Successfully installed predictionio-0.9.8 pytz-2017.2

### **测试**

Build完成后，可以执行integration\_test来测试，先分别启动elastic search和HBase.

注意：如果例外中断，再启动时需要删除生成的user-engine.json。 integration\_test脚本在训练时指定4g的内存分配，如果没有那么多物理内存会抛出内存不足的错误。修改为1g即可。

integration\_test脚本：

…

echo ""

echo "Building and delpoying model"

pio build

pio train -- --driver-memory 1g --executor-memory 1g

echo "Model will remain deployed after this test"

nohup pio deploy > deploy.out &

echo "Waiting 30 seconds for the server to start"

sleep 30

…

执行结果如下。目前仍抛出zookeeper连接的错误。

[root@master ur]# ./examples/integration-test

Integration test for The Universal Recommender.

If some step fails check that your engine.json file has been restored

or look for it in 'user-engine.json'

Checking for needed files

Checking status, should exit if pio is not running.

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[INFO] [Management$] Inspecting PredictionIO...

[INFO] [Management$] PredictionIO 0.11.0-incubating is installed at /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating

[INFO] [Management$] Inspecting Apache Spark...

[INFO] [Management$] Apache Spark is installed at /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/spark-1.6.3-bin-hadoop2.6

[INFO] [Management$] Apache Spark 1.6.3 detected (meets minimum requirement of 1.3.0)

[INFO] [Management$] Inspecting storage backend connections...

[INFO] [Storage$] Verifying Meta Data Backend (Source: ELASTICSEARCH)...

[INFO] [Storage$] Verifying Model Data Backend (Source: LOCALFS)...

[INFO] [Storage$] Verifying Event Data Backend (Source: HBASE)...

[INFO] [Storage$] Test writing to Event Store (App Id 0)...

[INFO] [HBLEvents] The table pio\_event:events\_0 doesn't exist yet. Creating now...

[INFO] [HBLEvents] Removing table pio\_event:events\_0...

[INFO] [Management$] Your system is all ready to go.

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[ERROR] [App$] App handmade already exists. Aborting.

Checking to see if handmade app exists, should exit if not.

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[INFO] [Pio$] App Name: handmade

[INFO] [Pio$] App ID: 2

[INFO] [Pio$] Description:

[INFO] [Pio$] Access Key: zUJvV2TN6gTyph1NoXJxd2pX3zJv6QyEBd7xkgfW0ykoECVNpwS4leyi4ULGvYvy | (all)

Moving engine.json to user-engine.json if it exists

Moving examples/handmade-engine.json to engine.json for integration test.

Deleting handmade app data since the test is date dependent

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[INFO] [Pio$] Data of the following app (default channel only) will be deleted. Are you sure?

[INFO] [Pio$] App Name: handmade

[INFO] [Pio$] App ID: 2

[INFO] [Pio$] Description: None

[INFO] [HBLEvents] Removing table pio\_event:events\_2...

[INFO] [App$] Removed Event Store for the app ID: 2

[INFO] [HBLEvents] The table pio\_event:events\_2 doesn't exist yet. Creating now...

[INFO] [App$] Initialized Event Store for the app ID: 2

Importing data for integration test

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

Access key: zUJvV2TN6gTyph1NoXJxd2pX3zJv6QyEBd7xkgfW0ykoECVNpwS4leyi4ULGvYvy

Namespace(access\_key='zUJvV2TN6gTyph1NoXJxd2pX3zJv6QyEBd7xkgfW0ykoECVNpwS4leyi4ULGvYvy', file='data/sample-handmade-data.txt', url='http://localhost:7070')

Importing data...

Event: purchase entity\_id: u1 target\_entity\_id: Iphone 6 current\_date: 2017-10-09T14:34:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Iphone 5 current\_date: 2017-10-08T19:22:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Iphone 4 current\_date: 2017-10-08T00:10:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Ipad-retina current\_date: 2017-10-07T04:58:06.777117+00:00

Event: purchase entity\_id: U 2 target\_entity\_id: Nexus current\_date: 2017-10-06T09:46:06.777117+00:00

Event: purchase entity\_id: U 2 target\_entity\_id: Galaxy current\_date: 2017-10-05T14:34:06.777117+00:00

Event: purchase entity\_id: u-3 target\_entity\_id: Surface current\_date: 2017-10-04T19:22:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Iphone 5 current\_date: 2017-10-04T00:10:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Iphone 5 current\_date: 2017-10-03T04:58:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Iphone 5 current\_date: 2017-10-02T09:46:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Iphone 5 current\_date: 2017-10-01T14:34:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Iphone 4 current\_date: 2017-09-30T19:22:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Galaxy current\_date: 2017-09-30T00:10:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Phones current\_date: 2017-09-29T04:58:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Phones current\_date: 2017-09-28T09:46:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Phones current\_date: 2017-09-27T14:34:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Phones current\_date: 2017-09-26T19:22:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Phones current\_date: 2017-09-26T00:10:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Phones current\_date: 2017-09-25T04:58:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Mobile-acc current\_date: 2017-09-24T09:46:06.777117+00:00

Event: view entity\_id: U 2 target\_entity\_id: Phones current\_date: 2017-09-23T14:34:06.777117+00:00

Event: view entity\_id: U 2 target\_entity\_id: Tablets current\_date: 2017-09-22T19:22:06.777117+00:00

Event: view entity\_id: U 2 target\_entity\_id: Mobile-acc current\_date: 2017-09-22T00:10:06.777117+00:00

Event: view entity\_id: u-3 target\_entity\_id: Mobile-acc current\_date: 2017-09-21T04:58:06.777117+00:00

Event: view entity\_id: u-4 target\_entity\_id: Phones current\_date: 2017-09-20T09:46:06.777117+00:00

Event: view entity\_id: u-4 target\_entity\_id: Tablets current\_date: 2017-09-19T14:34:06.777117+00:00

Event: view entity\_id: u-4 target\_entity\_id: Soap current\_date: 2017-09-18T19:22:06.777117+00:00

Event: view entity\_id: u5 target\_entity\_id: Soap current\_date: 2017-09-18T00:10:06.777117+00:00

Event: category-pref entity\_id: u1 target\_entity\_id: phones current\_date: 2017-09-17T04:58:06.777117+00:00

Event: category-pref entity\_id: u1 target\_entity\_id: phones current\_date: 2017-09-16T09:46:06.777117+00:00

Event: category-pref entity\_id: u1 target\_entity\_id: phones current\_date: 2017-09-15T14:34:06.777117+00:00

Event: category-pref entity\_id: u1 target\_entity\_id: tablets current\_date: 2017-09-14T19:22:06.777117+00:00

Event: category-pref entity\_id: U 2 target\_entity\_id: tablets current\_date: 2017-09-14T00:10:06.777117+00:00

Event: category-pref entity\_id: U 2 target\_entity\_id: phones current\_date: 2017-09-13T04:58:06.777117+00:00

Event: category-pref entity\_id: u-3 target\_entity\_id: tablets current\_date: 2017-09-12T09:46:06.777117+00:00

Event: category-pref entity\_id: u-4 target\_entity\_id: phones current\_date: 2017-09-11T14:34:06.777117+00:00

Event: category-pref entity\_id: u-4 target\_entity\_id: phones current\_date: 2017-09-10T19:22:06.777117+00:00

Event: category-pref entity\_id: u5 target\_entity\_id: toiletries current\_date: 2017-09-10T00:10:06.777117+00:00

Event: $set entity\_id: Iphone 6 properties/categories: ['Phones', 'Electronics', 'Apple'] current\_date: 2017-09-09T04:58:06.777117+00:00

Event: $set entity\_id: Iphone 5 properties/categories: ['Phones', 'Electronics', 'Apple'] current\_date: 2017-09-08T09:46:06.777117+00:00

Event: $set entity\_id: Iphone 4 properties/categories: ['Phones', 'Electronics', 'Apple'] current\_date: 2017-09-07T14:34:06.777117+00:00

Event: $set entity\_id: Ipad-retina properties/categories: ['Tablets', 'Electronics', 'Apple'] current\_date: 2017-09-06T19:22:06.777117+00:00

Event: $set entity\_id: Nexus properties/categories: ['Tablets', 'Electronics', 'Google'] current\_date: 2017-09-06T00:10:06.777117+00:00

Event: $set entity\_id: Galaxy properties/categories: ['Phones', 'Electronics', 'Samsung'] current\_date: 2017-09-05T04:58:06.777117+00:00

Event: $set entity\_id: Surface properties/categories: ['Tablets', 'Electronics', 'Microsoft'] current\_date: 2017-09-04T09:46:06.777117+00:00

Event: $set entity\_id: Iphone 4 properties/countries: ['United States', 'Canada', 'Estados Unidos Mexicanos'] current\_date: 2017-09-03T14:34:06.777117+00:00

Event: $set entity\_id: Ipad-retina properties/countries: ['United States', 'Estados Unidos Mexicanos'] current\_date: 2017-09-02T19:22:06.777117+00:00

Event: $set entity\_id: Nexus properties/countries: ['United States', 'Canada'] current\_date: 2017-09-02T00:10:06.777117+00:00

Event: $set entity\_id: Galaxy properties/countries: ['United States'] current\_date: 2017-09-01T04:58:06.777117+00:00

Event: $set entity\_id: Surface properties/countries: ['United States', 'Canada'] current\_date: 2017-08-31T09:46:06.777117+00:00

Event: $set entity\_id: Iphone 6 properties/categories: ['Cameras'] current\_date: 2017-08-30T14:34:06.777117+00:00

Event: $set entity\_id: Iphone 5 properties/categories: ['Cameras'] current\_date: 2017-08-29T19:22:06.777117+00:00

Event: $set entity\_id: Iphone 4 properties/categories: ['Cameras'] current\_date: 2017-08-29T00:10:06.777117+00:00

Event: $set entity\_id: Ipad-retina properties/categories: ['Computers'] current\_date: 2017-08-28T04:58:06.777117+00:00

Event: $set entity\_id: Nexus properties/categories: ['Computers'] current\_date: 2017-08-27T09:46:06.777117+00:00

Event: $set entity\_id: Galaxy properties/categories: ['Cameras'] current\_date: 2017-08-26T14:34:06.777117+00:00

Event: $set entity\_id: Surface properties/categories: ['Computers'] current\_date: 2017-08-25T19:22:06.777117+00:00

Event: $set entity\_id: Iphone 4 properties/countries: ['Cuba'] current\_date: 2017-08-25T00:10:06.777117+00:00

Event: $set entity\_id: Ipad-retina properties/countries: ['Cuba'] current\_date: 2017-08-24T04:58:06.777117+00:00

Event: $set entity\_id: Nexus properties/countries: ['Cuba'] current\_date: 2017-08-23T09:46:06.777117+00:00

Event: $set entity\_id: Galaxy properties/countries: ['Cuba'] current\_date: 2017-08-22T14:34:06.777117+00:00

Event: $set entity\_id: Surface properties/countries: ['Cuba'] current\_date: 2017-08-21T19:22:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Iphone 6 current\_date: 2017-08-21T00:10:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Iphone 5 current\_date: 2017-08-20T04:58:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Iphone 4 current\_date: 2017-08-19T09:46:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Ipad-retina current\_date: 2017-08-18T14:34:06.777117+00:00

Event: purchase entity\_id: U 2 target\_entity\_id: Nexus current\_date: 2017-08-17T19:22:06.777117+00:00

Event: purchase entity\_id: U 2 target\_entity\_id: Galaxy current\_date: 2017-08-17T00:10:06.777117+00:00

Event: purchase entity\_id: u-3 target\_entity\_id: Surface current\_date: 2017-08-16T04:58:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Iphone 5 current\_date: 2017-08-15T09:46:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Iphone 4 current\_date: 2017-08-14T14:34:06.777117+00:00

Event: purchase entity\_id: u-4 target\_entity\_id: Galaxy current\_date: 2017-08-13T19:22:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Phones current\_date: 2017-08-13T00:10:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Mobile-acc current\_date: 2017-08-12T04:58:06.777117+00:00

Event: view entity\_id: U 2 target\_entity\_id: Phones current\_date: 2017-08-11T09:46:06.777117+00:00

Event: view entity\_id: U 2 target\_entity\_id: Tablets current\_date: 2017-08-10T14:34:06.777117+00:00

Event: view entity\_id: U 2 target\_entity\_id: Mobile-acc current\_date: 2017-08-09T19:22:06.777117+00:00

Event: view entity\_id: u-3 target\_entity\_id: Mobile-acc current\_date: 2017-08-09T00:10:06.777117+00:00

Event: view entity\_id: u-4 target\_entity\_id: Phones current\_date: 2017-08-08T04:58:06.777117+00:00

Event: view entity\_id: u-4 target\_entity\_id: Tablets current\_date: 2017-08-07T09:46:06.777117+00:00

Event: view entity\_id: u-4 target\_entity\_id: Soap current\_date: 2017-08-06T14:34:06.777117+00:00

Event: view entity\_id: u5 target\_entity\_id: Soap current\_date: 2017-08-05T19:22:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Galaxy current\_date: 2017-08-05T00:10:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Galaxy current\_date: 2017-08-04T04:58:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Galaxy current\_date: 2017-08-03T09:46:06.777117+00:00

Event: purchase entity\_id: u1 target\_entity\_id: Galaxy current\_date: 2017-08-02T14:34:06.777117+00:00

Event: purchase entity\_id: U 2 target\_entity\_id: Iphone 5 current\_date: 2017-08-01T19:22:06.777117+00:00

Event: purchase entity\_id: U 2 target\_entity\_id: Iphone 5 current\_date: 2017-08-01T00:10:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Soap current\_date: 2017-07-31T04:58:06.777117+00:00

Event: view entity\_id: u1 target\_entity\_id: Soap current\_date: 2017-07-30T09:46:06.777117+00:00

Event: view entity\_id: u-3 target\_entity\_id: Mobile-acc current\_date: 2017-07-29T14:34:06.777117+00:00

Event: view entity\_id: u-4 target\_entity\_id: Mobile-acc current\_date: 2017-07-28T19:22:06.777117+00:00

Event: view entity\_id: u-4 target\_entity\_id: Mobile-acc current\_date: 2017-07-28T00:10:06.777117+00:00

Event: $set entity\_id: Iphone 6 properties/categories: ['Phones', 'Electronics', 'Apple'] current\_date: 2017-07-27T04:58:06.777117+00:00

Event: $set entity\_id: Iphone 5 properties/categories: ['Phones', 'Electronics', 'Apple'] current\_date: 2017-07-26T09:46:06.777117+00:00

Event: $set entity\_id: Iphone 4 properties/categories: ['Phones', 'Electronics', 'Apple'] current\_date: 2017-07-25T14:34:06.777117+00:00

Event: $set entity\_id: Ipad-retina properties/categories: ['Tablets', 'Electronics', 'Apple'] current\_date: 2017-07-24T19:22:06.777117+00:00

Event: $set entity\_id: Nexus properties/categories: ['Tablets', 'Electronics', 'Google'] current\_date: 2017-07-24T00:10:06.777117+00:00

Event: $set entity\_id: Galaxy properties/categories: ['Phones', 'Electronics', 'Samsung'] current\_date: 2017-07-23T04:58:06.777117+00:00

Event: $set entity\_id: Surface properties/categories: ['Tablets', 'Electronics', 'Microsoft'] current\_date: 2017-07-22T09:46:06.777117+00:00

Event: $set entity\_id: Iphone 4 properties/countries: ['United States', 'Canada', 'Estados Unidos Mexicanos'] current\_date: 2017-07-21T14:34:06.777117+00:00

Event: $set entity\_id: Ipad-retina properties/countries: ['United States', 'Estados Unidos Mexicanos'] current\_date: 2017-07-20T19:22:06.777117+00:00

Event: $set entity\_id: Nexus properties/countries: ['United States', 'Canada'] current\_date: 2017-07-20T00:10:06.777117+00:00

Event: $set entity\_id: Galaxy properties/countries: ['United States'] current\_date: 2017-07-19T04:58:06.777117+00:00

Event: $set entity\_id: Surface properties/countries: ['United States', 'Canada'] current\_date: 2017-07-18T09:46:06.777117+00:00

Event: $set entity\_id: Iphone 6 properties/defaultRank: ['7'] current\_date: 2017-07-17T14:34:06.777117+00:00

Event: $set entity\_id: Iphone 5 properties/defaultRank: ['6'] current\_date: 2017-07-16T19:22:06.777117+00:00

Event: $set entity\_id: Iphone 4 properties/defaultRank: ['5'] current\_date: 2017-07-16T00:10:06.777117+00:00

Event: $set entity\_id: Nexus properties/defaultRank: ['4'] current\_date: 2017-07-15T04:58:06.777117+00:00

Event: $set entity\_id: Ipad-retina properties/defaultRank: ['3'] current\_date: 2017-07-14T09:46:06.777117+00:00

Event: $set entity\_id: Galaxy properties/defaultRank: ['2'] current\_date: 2017-07-13T14:34:06.777117+00:00

Event: $set entity\_id: Surface properties/defaultRank: ['1'] current\_date: 2017-07-12T19:22:06.777117+00:00

All items: ['Iphone 6', 'Ipad-retina', 'Nexus', 'Surface', 'Iphone 4', 'Galaxy', 'Iphone 5']

Event: $set entity\_id: Iphone 6 properties/availableDate: 2017-10-05T04:58:06.777117+00:00 properties/date: 2017-10-07T04:58:06.777117+00:00 properties/expireDate: 2017-10-09T04:58:06.777117+00:00

Event: $set entity\_id: Ipad-retina properties/availableDate: 2017-10-06T00:10:06.777117+00:00 properties/date: 2017-10-08T00:10:06.777117+00:00 properties/expireDate: 2017-10-10T00:10:06.777117+00:00

Event: $set entity\_id: Nexus properties/availableDate: 2017-10-06T19:22:06.777117+00:00 properties/date: 2017-10-08T19:22:06.777117+00:00 properties/expireDate: 2017-10-10T19:22:06.777117+00:00

Event: $set entity\_id: Surface properties/availableDate: 2017-10-07T14:34:06.777117+00:00 properties/date: 2017-10-09T14:34:06.777117+00:00 properties/expireDate: 2017-10-11T14:34:06.777117+00:00

Event: $set entity\_id: Iphone 4 properties/availableDate: 2017-10-08T09:46:06.777117+00:00 properties/date: 2017-10-10T09:46:06.777117+00:00 properties/expireDate: 2017-10-12T09:46:06.777117+00:00

Event: $set entity\_id: Galaxy properties/availableDate: 2017-10-09T04:58:06.777117+00:00 properties/date: 2017-10-11T04:58:06.777117+00:00 properties/expireDate: 2017-10-13T04:58:06.777117+00:00

Event: $set entity\_id: Iphone 5 properties/availableDate: 2017-10-10T00:10:06.777117+00:00 properties/date: 2017-10-12T00:10:06.777117+00:00 properties/expireDate: 2017-10-14T00:10:06.777117+00:00

119 events are imported.

Building and delpoying model

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[INFO] [Engine$] Using command '/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/sbt/sbt' at /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/engines/ur to build.

[INFO] [Engine$] If the path above is incorrect, this process will fail.

[INFO] [Engine$] Uber JAR disabled. Making sure lib/pio-assembly-0.11.0-incubating.jar is absent.

[INFO] [Engine$] Going to run: /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/sbt/sbt package assemblyPackageDependency in /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/engines/ur

[INFO] [Engine$] Compilation finished successfully.

[INFO] [Engine$] Looking for an engine...

[INFO] [Engine$] Found universal-recommender\_2.10-0.6.0.jar

[INFO] [Engine$] Found universal-recommender-assembly-0.6.0-deps.jar

[INFO] [Engine$] Build finished successfully.

[INFO] [Pio$] Your engine is ready for training.

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[WARN] [WorkflowUtils$] Environment variable POSTGRES\_JDBC\_DRIVER is pointing to a nonexistent file /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/postgresql-42.0.0.jar. Ignoring.

[WARN] [WorkflowUtils$] Environment variable MYSQL\_JDBC\_DRIVER is pointing to a nonexistent file /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/mysql-connector-java-5.1.41.jar. Ignoring.

[INFO] [Runner$] Submission command: /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/spark-1.6.3-bin-hadoop2.6/bin/spark-submit --driver-memory 1g --executor-memory 1g --class org.apache.predictionio.workflow.CreateWorkflow --jars file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/engines/ur/target/scala-2.10/universal-recommender\_2.10-0.6.0.jar,file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/engines/ur/target/scala-2.10/universal-recommender-assembly-0.6.0-deps.jar,file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-elasticsearch1-assembly-0.11.0-incubating.jar,file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hbase-assembly-0.11.0-incubating.jar,file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar,file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-jdbc-assembly-0.11.0-incubating.jar,file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-localfs-assembly-0.11.0-incubating.jar --files file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/conf/log4j.properties,file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/hbase-1.2.6/conf/hbase-site.xml --driver-class-path /mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/conf:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/postgresql-42.0.0.jar:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/mysql-connector-java-5.1.41.jar:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/hbase-1.2.6/conf --driver-java-options -Dpio.log.dir=/root file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar --engine-id com.actionml.RecommendationEngine --engine-version 3825fe3681dab2d618ca7f46cda9f7ac65832e12 --engine-variant file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/engines/ur/engine.json --verbosity 0 --json-extractor Both --env PIO\_STORAGE\_SOURCES\_HBASE\_TYPE=hbase,PIO\_ENV\_LOADED=1,PIO\_STORAGE\_REPOSITORIES\_METADATA\_NAME=pio\_meta,PIO\_FS\_BASEDIR=/root/.pio\_store,PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_HOSTS=localhost,PIO\_STORAGE\_SOURCES\_HBASE\_HOME=/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/hbase-1.2.6,PIO\_HOME=/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating,PIO\_FS\_ENGINESDIR=/root/.pio\_store/engines,PIO\_STORAGE\_SOURCES\_LOCALFS\_PATH=/root/.pio\_store/models,PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_TYPE=elasticsearch,PIO\_STORAGE\_REPOSITORIES\_METADATA\_SOURCE=ELASTICSEARCH,PIO\_STORAGE\_REPOSITORIES\_MODELDATA\_SOURCE=LOCALFS,PIO\_STORAGE\_REPOSITORIES\_EVENTDATA\_NAME=pio\_event,PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_HOME=/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/vendors/elasticsearch-1.7.6,PIO\_FS\_TMPDIR=/root/.pio\_store/tmp,PIO\_STORAGE\_REPOSITORIES\_MODELDATA\_NAME=pio\_model,PIO\_STORAGE\_REPOSITORIES\_EVENTDATA\_SOURCE=HBASE,PIO\_CONF\_DIR=/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/conf,PIO\_STORAGE\_SOURCES\_ELASTICSEARCH\_PORTS=9300,PIO\_STORAGE\_SOURCES\_LOCALFS\_TYPE=localfs

[INFO] [RecommendationEngine$]

\_ \_ \_\_ \_\_ \_

/\ | | (\_) | \/ | |

/ \ \_\_\_| |\_ \_ \_\_\_ \_ \_\_ | \ / | |

/ /\ \ / \_\_| \_\_| |/ \_ \| '\_ \| |\/| | |

/ \_\_\_\_ \ (\_\_| |\_| | (\_) | | | | | | | |\_\_\_\_

/\_/ \\_\\_\_\_|\\_\_|\_|\\_\_\_/|\_| |\_|\_| |\_|\_\_\_\_\_\_|

[INFO] [Engine] Extracting datasource params...

[INFO] [WorkflowUtils$] No 'name' is found. Default empty String will be used.

[INFO] [Engine] Datasource params: (,DataSourceParams(handmade,List(purchase, view, category-pref),None,Some(3)))

[INFO] [Engine] Extracting preparator params...

[INFO] [Engine] Preparator params: (,Empty)

[INFO] [Engine] Extracting serving params...

[INFO] [Engine] Serving params: (,Empty)

[INFO] [Remoting] Starting remoting

[INFO] [Remoting] Remoting started; listening on addresses :[akka.tcp://sparkDriverActorSystem@192.168.40.134:36227]

[INFO] [DataSource]

╔════════════════════════════════════════════════════════════╗

║ Init DataSource ║

║ ══════════════════════════════════════════════════════════ ║

║ App name handmade ║

║ Event window None ║

║ Event names List(purchase, view, category-pref) ║

║ Min events per user Some(3) ║

╚════════════════════════════════════════════════════════════╝

[INFO] [URAlgorithm]

╔════════════════════════════════════════════════════════════╗

║ Init URAlgorithm ║

║ ══════════════════════════════════════════════════════════ ║

║ App name handmade ║

║ ES index name urindex ║

║ ES type name items ║

║ RecsModel all ║

║ Event names List(purchase, view, category-pref) ║

║ ══════════════════════════════════════════════════════════ ║

║ Random seed 26396769 ║

║ MaxCorrelatorsPerEventType 50 ║

║ MaxEventsPerEventType 500 ║

║ BlacklistEvents List(purchase) ║

║ ══════════════════════════════════════════════════════════ ║

║ User bias 1.0 ║

║ Item bias 1.0 ║

║ Max query events 3000 ║

║ Limit 4 ║

║ ══════════════════════════════════════════════════════════ ║

║ Rankings: ║

║ popular Some(popRank) ║

╚════════════════════════════════════════════════════════════╝

[INFO] [Engine$] EngineWorkflow.train

[INFO] [Engine$] DataSource: com.actionml.DataSource@65036e8d

[INFO] [Engine$] Preparator: com.actionml.Preparator@685d7ba5

[INFO] [Engine$] AlgorithmList: List(com.actionml.URAlgorithm@25533bba)

[INFO] [Engine$] Data sanity check is on.

[ERROR] [RecoverableZooKeeper] ZooKeeper exists failed after 1 attempts

[ERROR] [ZooKeeperWatcher] hconnection-0x2b9aeedb, quorum=localhost:2181, baseZNode=/hbase Received unexpected KeeperException, re-throwing exception

[WARN] [ZooKeeperRegistry] Can't retrieve clusterId from Zookeeper

[ERROR] [StorageClient] Cannot connect to ZooKeeper (ZooKeeper ensemble: localhost). Please make sure that the configuration is pointing at the correct ZooKeeper ensemble. By default, HBase manages its own ZooKeeper, so if you have not configured HBase to use an external ZooKeeper, that means your HBase is not started or configured properly.

[ERROR] [Storage$] Error initializing storage client for source HBASE

[ERROR] [Engine$] Error occured reading from data source. (Reason: Data source HBASE was not properly initialized.) Please see the log for debugging details.

[root@master ur]#

### **训练Model TODO – 无法训练（无法连接zookeeper，单独安装zookeeper。。或者参考：**[**https://groups.google.com/forum/#!topic/predictionio-dev/s-rzaB8Sw7U**](https://groups.google.com/forum/#!topic/predictionio-dev/s-rzaB8Sw7U)**）**

In general, the PredictionIO bin/pio scripts wraps around Spark's spark-submit script. You can specify a lot of Spark configurations (i.e. executor memory, cores, master url, etc.) with it. You can supply these as pass-through arguments at the end of bin/pio command. http://predictionio.incubator.apache.org/resources/faq/#engine-training

[root@master ur]# pio train

…

[INFO] [Engine$] EngineWorkflow.train

[INFO] [Engine$] DataSource: com.actionml.DataSource@197ce367

[INFO] [Engine$] Preparator: com.actionml.Preparator@48ccbb32

[INFO] [Engine$] AlgorithmList: List(com.actionml.URAlgorithm@16d41725)

[INFO] [Engine$] Data sanity check is on.

[ERROR] [RecoverableZooKeeper] ZooKeeper exists failed after 1 attempts

[ERROR] [ZooKeeperWatcher] hconnection-0x6b09ce57, quorum=localhost:2181, baseZNode=/hbase Received unexpected KeeperException, re-throwing exception

[WARN] [ZooKeeperRegistry] Can't retrieve clusterId from Zookeeper

[ERROR] [StorageClient] Cannot connect to ZooKeeper (ZooKeeper ensemble: localhost). Please make sure that the configuration is pointing at the correct ZooKeeper ensemble. By default, HBase manages its own ZooKeeper, so if you have not configured HBase to use an external ZooKeeper, that means your HBase is not started or configured properly.

[ERROR] [Storage$] Error initializing storage client for source HBASE

[ERROR] [Engine$] Error occured reading from data source. (Reason: Data source HBASE was not properly initialized.) Please see the log for debugging details.

注意：需要启动HBASE。启动Hbase后重新执行。仍然失败。

[INFO] [Engine$] EngineWorkflow.train

[INFO] [Engine$] DataSource: com.actionml.DataSource@197ce367

[INFO] [Engine$] Preparator: com.actionml.Preparator@48ccbb32

[INFO] [Engine$] AlgorithmList: List(com.actionml.URAlgorithm@16d41725)

[INFO] [Engine$] Data sanity check is on.

[ERROR] [RecoverableZooKeeper] ZooKeeper exists failed after 4 attempts

[ERROR] [ZooKeeperWatcher] hconnection-0x4a2653a0-0x15f00bda356000a, quorum=localhost:2181, baseZNode=/hbase Received unexpected KeeperException, re-throwing exception

[ERROR] [HConnectionManager$HConnectionImplementation] Can't get connection to ZooKeeper: KeeperErrorCode = ConnectionLoss for /hbase

[ERROR] [RecoverableZooKeeper] ZooKeeper exists failed after 4 attempts

[ERROR] [ZooKeeperWatcher] hconnection-0x4a2653a0-0x15f00bda356000a, quorum=localhost:2181, baseZNode=/hbase Received unexpected KeeperException, re-throwing exception

[ERROR] [HConnectionManager$HConnectionImplementation] Can't get connection to ZooKeeper: KeeperErrorCode = ConnectionLoss for /hbase

[ERROR] [RecoverableZooKeeper] ZooKeeper exists failed after 4 attempts

[ERROR] [ZooKeeperWatcher] hconnection-0x4a2653a0-0x15f00bda356000a, quorum=localhost:2181, baseZNode=/hbase Received unexpected KeeperException, re-throwing exception

[ERROR] [HConnectionManager$HConnectionImplementation] Can't get connection to ZooKeeper: KeeperErrorCode = ConnectionLoss for /hbase

[ERROR] [RecoverableZooKeeper] ZooKeeper exists failed after 4 attempts

[ERROR] [ZooKeeperWatcher] hconnection-0x4a2653a0-0x15f00bda356000a, quorum=localhost:2181, baseZNode=/hbase Received unexpected KeeperException, re-throwing exception

[ERROR] [HConnectionManager$HConnectionImplementation] Can't get connection to ZooKeeper: KeeperErrorCode = ConnectionLoss for /hbase

[ERROR] [RecoverableZooKeeper] ZooKeeper exists failed after 4 attempts

[ERROR] [ZooKeeperWatcher] hconnection-0x4a2653a0-0x15f00bda356000a, quorum=localhost:2181, baseZNode=/hbase Received unexpected KeeperException, re-throwing exception

[ERROR] [HConnectionManager$HConnectionImplementation] Can't get connection to ZooKeeper: KeeperErrorCode = ConnectionLoss for /hbase

…

注意：需要启动HBASE。启动Hbase后重新执行。仍然失败。重新启动hbase, 并启动hbase shell，验证hbase命令行。再启动pio train。出现如下错误：

……

[INFO] [Engine$] AlgorithmList: List(com.actionml.URAlgorithm@25533bba)

[INFO] [Engine$] Data sanity check is on.

[WARN] [TableInputFormatBase] Cannot resolve the host name for master/192.168.40.134 because of javax.naming.NameNotFoundException: DNS name not found [response code 3]; remaining name '134.40.168.192.in-addr.arpa'

这是由于集群中的Reverse DNS无法正常工作， 执行

[root@master bin]# host 192.168.40.134

Host 134.40.168.192.in-addr.arpa. not found: 3(NXDOMAIN)

将/etc/resolv.conf中所有DNS项注释掉，确保所有项在/etc/hosts中正确。并执行dnsmasq

[root@master bin]# vi /etc/resolv.conf

[root@master bin]# vi /etc/hosts

[root@master bin]# dnsmasq

[root@master bin]# host 192.168.40.134

134.40.168.192.in-addr.arpa domain name pointer master.

重新执行pio train，出现如下错误：

[INFO] [Engine$] EngineWorkflow.train

[INFO] [Engine$] DataSource: com.actionml.DataSource@48ccbb32

[INFO] [Engine$] Preparator: com.actionml.Preparator@16d41725

[INFO] [Engine$] AlgorithmList: List(com.actionml.URAlgorithm@26d0ece6)

[INFO] [Engine$] Data sanity check is on.

[ERROR] [StorageClient] HBase master is not running (ZooKeeper ensemble: localhost). Please make sure that HBase is running properly, and that the configuration is pointing at the correct ZooKeeper ensemble.

[ERROR] [Storage$] Error initializing storage client for source HBASE

[ERROR] [Engine$] Error occured reading from data source. (Reason: Data source HBASE was not properly initialized.) Please see the log for debugging details.

JPS查看HMaster进程已经不在了。重新启动hbase, pio status查看，所有系统正常，重新执行pio train，仍然报错：

[INFO] [Engine$] EngineWorkflow.train

[INFO] [Engine$] DataSource: com.actionml.DataSource@65036e8d

[INFO] [Engine$] Preparator: com.actionml.Preparator@685d7ba5

[INFO] [Engine$] AlgorithmList: List(com.actionml.URAlgorithm@25533bba)

[INFO] [Engine$] Data sanity check is on.

[WARN] [HConnectionManager$HConnectionImplementation] This client just lost it's session with ZooKeeper, closing it. It will be recreated next time someone needs it

[WARN] [HConnectionManager$HConnectionImplementation] This client just lost it's session with ZooKeeper, closing it. It will be recreated next time someone needs it

[WARN] [HConnectionManager$HConnectionImplementation] This client just lost it's session with ZooKeeper, closing it. It will be recreated next time someone needs it

[WARN] [HeartbeatReceiver] Removing executor driver with no recent heartbeats: 137848 ms exceeds timeout 120000 ms

[WARN] [NettyRpcEndpointRef] Error sending message [message = Heartbeat(driver,[Lscala.Tuple2;@18f1211f,BlockManagerId(driver, localhost, 40221))] in 1 attempts

[ERROR] [TaskSchedulerImpl] Lost executor driver on localhost: Executor heartbeat timed out after 137848 ms

[WARN] [TaskSetManager] Lost task 0.0 in stage 3.0 (TID 2, localhost): ExecutorLostFailure (executor driver exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 137848 ms

[ERROR] [TaskSetManager] Task 0 in stage 3.0 failed 1 times; aborting job

[WARN] [NettyRpcEnv] Ignored message: true

[WARN] [NettyRpcEnv] Ignored message: true

[WARN] [NettyRpcEnv] Ignored message: true

[WARN] [NettyRpcEnv] Ignored message: HeartbeatResponse(true)

[WARN] [SparkContext] Killing executors is only supported in coarse-grained mode

[ERROR] [BlockManager] Failed to report broadcast\_3\_piece0 to master; giving up.

。。。

### **Deploying the Engine**

注意：需要在train完成后才能部署engine：

[root@master ur]# pio deploy

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/spark/pio-data-hdfs-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/mnt/share/predictionio-0.11.0/PredictionIO-0.11.0-incubating/lib/pio-assembly-0.11.0-incubating.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[ERROR] [Engine$] No valid engine instance found for engine com.actionml.RecommendationEngine 3825fe3681dab2d618ca7f46cda9f7ac65832e12.

Try running 'train' before 'deploy'. Aborting.

[root@master ur]#

### **Use the Engine**