数据分析平台部署文档

### 日志采集模块

1. 使用cloudera manager（地址：<http://10.110.0.217:7180> ）将产生日志的服务器添加到HADOOP集群中；

步骤：

1. /etc/hosts下配置好自己的主机名，并添加：

10.110.0.202 nn01.hadoop nn01

10.110.0.203 nn02.hadoop nn02

10.110.0.204 dn01.hadoop dn01

10.110.0.215 dn02.hadoop dn02

10.110.0.216 dn03.hadoop dn03

10.110.0.217 dn04.hadoop dn04

1. 启动ntpd服务；
2. 点击页面“主机”->向群集添加新主机，使用步骤1配置的主机名查找；
3. 按向导添加主机，直到parcels部署到服务器；
4. 添加flume角色实例到新添加的服务器；

步骤：

1. 添加角色实例：点flume->实例->添加角色实例，选择新增服务器；
2. 修改新增实例配置：

将“代理名称”修改为“node”；

修改“配置文件”

参数“node.sources.iuniSrc.command”的值修改为实际日志文件的绝对路径，例：node.sources.iuniSrc.command = tail -fn0 /usr/local/nginx-1.4.4/logs/www.iuni.com\_access.log

参数“node.sources.iuniSrc.interceptors.inte.startDate”建议修改为第二天凌晨，表示只取第二天凌晨后产生的日志；如：node.sources.iuniSrc.interceptors.inte.startDate = 20150101000000

1. 将iunilog.lib.zip放到flume插件目录（如：/data/flume-ng/plugins.d/的iunilog子目录下，具体插件目录见flume配置），解压；
2. 将配置文件ipLib.properties放到iunilog子目录中；

示例配置文件：

ipLib.properties配置说明：

test.ipLibs = geo   
test.ipLibs.geo.type = GEO  
test.ipLibs.geo.lib = /home/data/flume-ng/plugins.d/iunilog/GeoLite2-City.mmdb  
每行开头的test为代理名称，可在flume的过滤器配置中指定，如：node.sources.iuniSrc.interceptors.inte.ipLibName = test，则说明flume代理node中数据源名称为iuniSrc的代理inte的IP库名称为test；

第1行，指这个代理中配置的IP的名称；

第2行，指这个IP库的类型，现提供4种类型：GEO,TAOBAO,IUNI,PURE；

第3行，指这个IP库的库文件位置； 使用不同类型的IP库，请指定其各自的IP库文件；

### 数据上报模块

1. 将da-ws.war包部署到tomcat容器下；
2. 数据上报接口说明详见《数据平台接口文档.docx》；
3. 将配置文件hbase.properties放到classpath下；

示例配置文件：

hbase.properties配置说明：

hbase.zookeeper.quorum = physical.hadoop, vir-vm1.hadoop, vir-vm2.hadoop  
page.dataTable.name = prd  
page.dataTable.cf.name = f  
page.dataTable.column.name = s

hbase.zookeeper.quorum指定hbase的zookeeper配置；

page.dataTable.name指定上报数据存入hbase的表名；

page.dataTable.cf.name指定列族名；

page.dataTable.column.name指定列名首字母（上例中表示：将上报的URL中参数名是以s开头的参数保存到hbase中，列名即为参数名）；

### 数据分析模块

* 1. 建表并初始化配置数据：执行如下三个初始化脚本：；
  2. 将da-app.jar文件放到任意一台可以执行JAVA程序的服务器；
  3. 数据库配置文件放置到其相同目录；
  4. 使用说明如下（也可直接执行java –jar da-app.jar查看使用说明）：

usage: analyze agent [--create-partition] [-e <arg>] -f <arg> [-h] [--manual] -n <arg> [--no-reload-conf] [-s <arg>] [-t <arg>]

--create-partition 需要创建分区

-e,--end <arg> 手动模式 - 结束时间，结束时间需大于开始时间

-f,--conf-file <arg> 指定配置文件

-h,--help 打印帮助

--manual 手动模式

-n,--name <arg> analyze代理名称

--no-reload-conf 不自动加载配置文件

-s,--start <arg> 手动模式 - 开始时间，开始时间需小于结束时间

-t,--ttype <arg> 手动模式 - 分析精度，按年(year)、月(month)、日(day)、时(hour)、分(minute)

### 数据展示模块

1. 将da-webapp.war部署到tomcat容器下；
2. 将配置文件”da-config.properties”和”da-dbconfig.properties”放到classpath下，示例配置文件：；
3. 配置文件da-config.properties为用户中心的配置，说明：

da.uuc.url=http://18.8.5.115:8480

da.uuc.app.id=107C569DCB304C7C9AFA07BCAC20F967

da.uuc.app.key=ECD4D165A4954F9BB062FC31F094EB5E

第1行：用户中心地址；

第2行：接入用户中心申请的ID；

第3行：接入用户中心时生成的密码；

1. 配置文件da-dbconfig.properties为数据库配置；