



FusionInsight HD
V100R002C60U10
SAS 对接指南

文档版本 02
发布日期 2016-09-30

华为技术有限公司



版权所有 © 华为技术有限公司 2016。 保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： <http://e.huawei.com>

目 录

1 配置对接 SAS.....1

2 测试 Hive 对接 SAS.....4

3 配置对接 SAS HPA.....8

4 配置对接 SAS EP.....14

1 配置对接 SAS

操作场景

系统管理员可以通过 SAS Access for hadoop 特性，配置 FusionInsight 对接 SAS 系统，并使用 SAS 系统分析 FusionInsight 中的数据。

前提条件

- 已部署 FusionInsight HD。
- 已安装 SAS 9.4_M2，并在本地 Windows 环境安装 SAS 客户端。

操作步骤

步骤 1 在 FusionInsight Manager 创建一个角色与“人机”用户，具体请参见《FusionInsight HD 管理员指南》的“创建用户”章节。角色需要根据业务需要授予权限，并将用户加入角色。

例如，创建用户“fi_sas”。

步骤 2 根据 SAS 服务端的 JDK 版本，在官方网站下载对应的 JCE 文件。JDK 1.7 对应 JCE 版本为 7，JDK 1.8 对应 JCE 版本为 8。

将获取的 zip 包在本地解压，可得到“local_policy.jar”和“US_export_policy.jar”。相关 jar 文件也可从 FusionInsight HD 的客户端中直接获取。

步骤 3 使用 WinSCP，以 SAS 用户将 jar 包上传到 SAS 服务端以下路径，文件权限与目录中其他文件保持一致：

“SAS 安装目录/SASPrivateJavaRuntimeEnvironment/9.4/jre/lib/security/”

步骤 4 在 FusionInsight Manager，选择“服务管理 > HDFS > 服务配置”，设置“参数类别”为“全部配置”，在“搜索”中输入“sas.plugin.enable”，在查找出的 sas.plugin.enable 配置项中，设置值为 true。点击“保存配置”，在弹出的对话框中选择“重新启动受影响的服务或实例。”，点击“确定”，输入密码后，等待该次操作完成。



注意

该次操作会重启 HDFS、Yarn 这两个服务，所以请在集群没有业务时执行。

步骤 5 在 SAS 服务端安装 FusionInsight HD 客户端，具体请参见《FusionInsight HD 管理员指南》的“基本操作”，“安装客户端”章节。

步骤 6 在 SAS 服务端配置“SAS_HADOOP_JAR_PATH”目录。

1. 使用 SAS 用户创建目录，例如“/opt/thirdparty/hadoop_jar”。
2. 在创建的目录中创建文件“core-site.xml”和“hdfs-site.xml”。

如果是安全模式集群，“core-site.xml”需要增加参数“hadoop.security.authentication”，参数值为“kerberos”，普通模式集群不需要设置。“hdfs-site.xml”无需设置。文件请保持 xml 结构。

安全版本“core-site.xml”示例：

```
<?xml version="1.0" encoding="UTF-8"?>
<configuration>
<property>
<name>hadoop.security.authentication</name>
<value>kerberos</value>
</property>
</configuration>
```

“hdfs-site.xml”示例：

```
<?xml version="1.0" encoding="UTF-8"?>
<configuration>
</configuration>
```

说明

- “hadoop.security.authentication”只能在“SAS_HADOOP_JAR_PATH”生效。
- “core-site.xml”和“hdfs-site.xml”必须保存在相同目录，否则 SAS 会报错。

3. 在 SAS 服务端将“FusionInsight 客户端目录/Hive/Beeline/lib/”和“FusionInsight 客户端目录/HDFS/hadoop/share/hadoop”中的全部 jar 包拷贝到“SAS_HADOOP_JAR_PATH”目录，文件权限为 755。

可以使用以下命令完成拷贝，其中\${CLIENT}为客户端的实际安装目录：

```
cp ${CLIENT}/Hive/Beeline/lib/*.jar SAS_HADOOP_JAR_PATH 目录
find ${CLIENT}/HDFS/hadoop/share/hadoop/ -name *.jar | xargs -i cp {}
SAS_HADOOP_JAR_PATH 目录
```

步骤 7 在 SAS 服务端配置“SAS_HADOOP_CONFIG_PATH”目录。

1. 使用 SAS 用户创建目录，例如“/opt/fi”。
2. 在 SAS 服务端将“FusionInsight 客户端目录/HDFS/hadoop/etc/hadoop/hdfs-site.xml”和“FusionInsight 客户端目录/HDFS/hadoop/etc/hadoop/core-site.xml”两个文件拷贝到“SAS_HADOOP_CONFIG_PATH”目录，文件权限为 755。

步骤 8 配置 SAS 对接日志。

使用 SAS 用户修改 SAS 日志配置文件“SAS 安装目录/SASFoundation/9.4/misc/tkjava/sas.log4j.properties”，配置方式为 log4j，修改效果如下：

```
log4j.rootLogger=debug,A2
log4j.logger.com.sas=debug

log4j.appender.A2=org.apache.log4j.RollingFileAppender
log4j.appender.A2.File=/tmp/sas_log/sas_log.log
log4j.appender.A2.MaxFileSize=10MB
log4j.appender.A2.MaxBackupIndex=10
log4j.appender.A2.layout=org.apache.log4j.PatternLayout
log4j.appender.A2.layout.ConversionPattern=[%d] [%5p] [%t] (%F:%L) %l - %m%n
```

步骤 9 如果是安全模式集群，修改“SAS 安装目录/SASFoundation/9.4/sasv9.cfg”，在-JREOPTIONS 中新增：

```
-Djava.security.krb5.conf=FusionInsight 客户端目录
/KrbClient/kerberos/var/krb5kdc/krb5.conf
```

步骤 10 如果是安全模式集群，修改“SAS 安装目录/SASFoundation/9.4/sasv9.cfg”，在-JREOPTIONS 中新增：

```
-Dzookeeper.server.principal=zookeeper/hadoop.hadoop.com
-Djava.security.auth.login.config=FusionInsight 客户端目录
/ZooKeeper/zookeeper/conf/jaas.conf
```

步骤 11 如果是安全模式集群，执行以下命令，认证对接用户。

source FusionInsight 客户端目录/bigdata_env

kinit fi_sas

步骤 12 SAS 连接 FusionInsight。

在本地 Windows 环境运行 SAS 客户端，或者在 Linux 环境启动 SAS 命令行客户端：

```
SAS 安装目录/SASFoundation/9.4/bin/sas_en -nodms
```

编辑以下代码，将 SAS_HADOOP_JAR_PATH、SAS_HADOOP_CONFIG_PATH 和 uri 替换为实际的值，在 SAS 客户端输入并提交：

```
options set=SAS HADOOP JAR PATH="SAS HADOOP JAR PATH 目录";
options set=SAS HADOOP CONFIG PATH="SAS HADOOP CONFIG PATH 目录";
Libname hadoopdb hadoop server=hacluster subprotocol=hive2
uri='jdbc:hive2://10.39.235.143:24002,10.39.235.121:24002,10.39.235.103:24002/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM';
```

普通模式需要提供的连接的用户名，如下示例：

```
options set=SAS HADOOP JAR PATH="SAS HADOOP JAR PATH 目录";
options set=SAS HADOOP CONFIG PATH="SAS HADOOP CONFIG PATH 目录";
Libname hadoopdb hadoop server=hacluster user=fi_sas subprotocol=hive2
uri='jdbc:hive2://8.5.140.9:24002,8.5.140.12:24002,8.5.140.10:24002/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2';
```

界面提示以下信息表示 SAS 连接 Hive 成功：

```
NOTE: Libref HADOOPDB was successfully assigned as follows:
```

----结束

2 测试 Hive 对接 SAS

操作场景

系统管理员已经配置成功 FusionInsight 对接 SAS 系统，这里提供使用 SAS 系统分析 FusionInsight 中的数据示例。

前提条件

- 已部署 FusionInsight HD。
- 已安装 SAS 9.4_M2，并在本地 Windows 环境安装 SAS 客户端。
- 已配置 FusionInsight 成功对接 SAS 系统。

操作步骤

步骤 1 准备数据和表。执行以下命令：

```
### export FusionInsight 客户端环境变量
source $FusionInsightClient/bigdata_env
### 如果使用的是安全模式集群，以对接的用户进行登录
kinit fi_sas
```

执行 **beeline**，并执行以下 HQL：

```
DROP TABLE IF EXISTS employees info;
DROP TABLE IF EXISTS employees contact;
DROP TABLE IF EXISTS employees info extended;
DROP TABLE IF EXISTS employees like;
CREATE EXTERNAL TABLE IF NOT EXISTS employees info
(
  id INT,
  name STRING,
  usd flag STRING,
  salary DOUBLE,
  deductions MAP<STRING, DOUBLE>,
  address STRING,
  entrytime STRING
)
ROW FORMAT delimited fields terminated by ',' MAP KEYS TERMINATED BY '&'
STORED AS TEXTFILE;
CREATE EXTERNAL TABLE IF NOT EXISTS employees_contact
```

```
(
id INT,
tel_phone STRING,
email STRING
)
ROW FORMAT delimited fields terminated by ',' STORED AS TEXTFILE;

CREATE EXTERNAL TABLE IF NOT EXISTS employees_info_extended
(
id INT,
name STRING,
usd_flag STRING,
salary DOUBLE,
deductions MAP<STRING, DOUBLE>,
address STRING
)
PARTITIONED BY (entrytime STRING) STORED AS TEXTFILE;
ALTER TABLE employees_info_extended ADD COLUMNS (tel_phone STRING, email STRING);
CREATE TABLE employees_like LIKE employees_info;
```

将以下数据写入 “employee_contact.dat” 文件：

```
1,135 2744 4655,lili@163.com
3,159 5667 2118,yoyo@gmail.com
4,186 1345 2269,lucy@qq.com
6,189 2245 9898,daisy@qq.com
8,134 7525 6998,candy@gmail.com
```

执行以下命令将 “employee_contact.dat” 文件上传到 HDFS 上：

```
hadoop fs -put employee_contact.dat /user/hive/warehouse/employees_contact
```

将以下数据写入 “employee_info.dat” 文件：

```
1,LiLi,R,8000.01,personal income tax&0.05,China:Shenzhen,2005
3,Yoyo,D,12000.02,personal income tax&0.09,America:NewYork,2005
4,Lucy,D,24000.03,personal income tax&0.09,America:Manhattan,2009
6,Daisy,D,36000.04,personal income tax&0.09,America:NewYork,2009
8,Candy,R,9000.05,personal income tax&0.05,China:Shanghai,2005
```

执行以下命令将 “employee_info.dat” 文件上传到 HDFS 上：

```
hadoop fs -put employee_info.dat /user/hive/warehouse/employees_info
```

执行完以上命令后，会在 Hive 中创建四张表：employees_contact, employees_info, employees_like, employees_info_extended，其中 employees_like 和 employees_info_extended 为空表，没有数据。

步骤 2 测试从 Hive 加载数据。在 SAS 连接上 Hive 后，执行以下脚本，执行前需将 SAS_HADOOP_JAR_PATH、SAS_HADOOP_CONFIG_PATH 和 uri 替换成实际的值：

```
options set=SAS HADOOP JAR PATH="/opt/sas/lib/";
options set=SAS HADOOP CONFIG PATH="/opt/sas/conf/";
Libname hadoopdb hadoop server=hacluster subprotocol=hive2
uri='jdbc:hive2://8.5.140.9:24002,8.5.140.12:24002,8.5.140.10:24002/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM';
libname test_sas '/tmp';
```



```
/*从Hive加载数据*/  
data test_sas.employees_info_sas;  
SET hadoopdb.employees_info;  
run;  
proc print data=test_sas.employees_info_sas;  
run;
```

执行完后能看到 SAS 正确打印了 employees_info 这张表的数据。

步骤 3 测试 SAS 加载数据到 Hive 表。在 SAS 连接上 Hive 后，执行以下脚本，执行前需将 SAS_HADOOP_JAR_PATH、SAS_HADOOP_CONFIG_PATH 和 uri 替换成实际的：

```
options set=SAS HADOOP JAR PATH="/opt/sas/lib/";  
options set=SAS HADOOP CONFIG PATH="/opt/sas/conf/";  
Libname hadoopdb hadoop server=hacluster subprotocol=hive2  
uri='jdbc:hive2://8.5.140.9:24002,8.5.140.12:24002,8.5.140.10:24002/;serviceDiscover  
yMode=zooKeeper;zooKeeperNamespace=hiveserver2;sasl.qop=auth-  
conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM';  
libname test_sas '/tmp';  
/*从hive加载数据*/  
data test_sas.employees_info_sas;  
SET hadoopdb.employees_info;  
run;  
  
/*加载数据到hive表*/  
DATA hadoopdb.employees_info_sas;  
SET test_sas.employees_info_sas;  
run;
```

执行完以上脚本后，登录 beeline 客户端查看 Hive 中的数据，可以看到新建了一张 employees_info_sas 表，表中的数据和 employees_info 一样。

步骤 4 测试 SAS append 数据到 Hive 表。在 SAS 连接上 Hive 后，执行以下脚本，执行前需将 SAS_HADOOP_JAR_PATH、SAS_HADOOP_CONFIG_PATH 和 uri 替换成实际的：

```
options set=SAS HADOOP JAR PATH="/opt/xy/sas/lib/";  
options set=SAS HADOOP CONFIG PATH="/opt/xy/sas/conf/";  
Libname hadoopdb hadoop server=hacluster subprotocol=hive2  
uri='jdbc:hive2://8.5.140.9:24002,8.5.140.12:24002,8.5.140.10:24002/;serviceDiscover  
yMode=zooKeeper;zooKeeperNamespace=hiveserver2;sasl.qop=auth-  
conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM';  
  
/*append 操作*/  
proc append base=hadoopdb.employees info sas  
data=hadoopdb.employees info;  
run;
```

执行完以上脚本后，登录 beeline 客户端查看 Hive 中的数据，可以看到 employees_info_sas 表中的数据多了 5 行。

步骤 5 测试创建表并执行 INSERT 语句。在 SAS 连接上 Hive 后，执行以下脚本，执行前需将 SAS_HADOOP_JAR_PATH、SAS_HADOOP_CONFIG_PATH 和 uri 替换成实际的：

```
options set=SAS HADOOP JAR PATH="/opt/xy/sas/lib/";  
options set=SAS HADOOP CONFIG PATH="/opt/xy/sas/conf/";  
Libname hadoopdb hadoop server=hacluster subprotocol=hive2  
uri='jdbc:hive2://8.5.140.9:24002,8.5.140.12:24002,8.5.140.10:24002/;serviceDiscover
```

```
ryMode=zooKeeper;zooKeeperNamespace=hiveserver2;sasl.qop=auth-
conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM';

/*创建表*/
proc sql;
create table hadoopdb.employees_info_extended_sas
as select * from hadoopdb.employees_info_extended;

/*执行 insert 语句*/
INSERT into hadoopdb.employees_info_extended_sas
SELECT
a.id,
a.name,
a.usd_flag,
a.salary,
a.deductions,
a.address,
b.tel_phone,
b.email,
a.entrytime
FROM hadoopdb.employees_info a JOIN hadoopdb.employees_contact b ON (a.id = b.id)
WHERE a.entrytime = '2005';
quit;
```

执行完以上语句后，登录 beeline 客户端查看 Hive 中数据，可以看到 Hive 中新建了一张表 `employees_info_extended_sas`，且表中数据是正确的。

步骤 6 测试删除表。在 SAS 连接上 Hive 后，执行以下脚本，执行前需将 `SAS_HADOOP_JAR_PATH`、`SAS_HADOOP_CONFIG_PATH` 和 `uri` 替换成实际的：

```
options set=SAS HADOOP JAR PATH="/opt/xy/sas/lib/";
options set=SAS HADOOP CONFIG PATH="/opt/xy/sas/conf/";
Libname hadoopdb hadoop server=hacluster subprotocol=hive2
uri='jdbc:hive2://8.5.140.9:24002,8.5.140.12:24002,8.5.140.10:24002//;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2;sasl.qop=auth-
conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM';

/*删除表*/
proc sql;
drop table hadoopdb.employees info extended sas;
quit;
```

执行完以上语句后，登录 beeline 客户端，可以看到 `employees_info_extended_sas` 已经被删除。

----结束

3 配置对接 SAS HPA

操作场景

系统管理员可以通过 SAS HPA 特性，配置 FusionInsight 对接 SAS HPA，并使用 SAS 系统分析性能数据。

前提条件

- 已部署 FusionInsight HD。
- 在 FusionInsight Manager 创建一个角色与“人机”用户并下载 keytab 文件，具体请参见《FusionInsight HD 管理员指南》的“帐户管理”章节。角色需要根据业务需要授予权限，并将用户加入角色。
- 已安装 SAS 9.4_M2，并在本地 Windows 环境安装 SAS 客户端。
- 获取 SAS_High-Performance_Computing_Management_Console，VERSION 2.6。
- 获取 SAS High-Performance Analytics environment，VERSION 2.9。将文件保存到主 NameNode 节点新的目录中，例如“/opt/sas/TKGrid_2.9”。
- 配置 **root** 用户和 SAS 用户在 FusionInsight 集群所有节点的 SSH 互信登录。
- 将 SAS 服务端的主机名与 IP 对应关系加入 FusionInsight 集群每个节点的 host 文件。

操作步骤

步骤 1 在 FusionInsight Manager 查看主 NameNode 所在节点，使用 PuTTY 工具以 **omm** 用户登录主 NameNode 节点。然后执行以下命令，切换用户，并输入密码。

```
su - root
```

步骤 2 检查是否存在“/etc/gridhosts”文件，若不存在需手动创建，内容包含 SAS 服务端和 FusionInsight 集群所有 NameNode 和 DataNode 所在节点的主机名。其中主 NameNode 必需在第一行。

步骤 3 执行以下命令，安装 HPCMC，用于批量分发 SAS 操作命令。

```
cp SAS_High-Performance_Computing_Management_Console 保存目录  
/2_6/Linux_for_x64/sashpcmc-2.6.x86_64.rpm /tmp  
cd /tmp  
rpm -ivh --prefix=/opt/sas/hpcmc sashpcmc* --nodeps
```

```
/opt/sas/hpcmc/webmin/utilbin/setup

Provide the following details:
hpcadmin
yes
lnxsas
lnxsas
no

service sashpcmc status
service sashpcmc start

echo 'PATH=$PATH:/opt/sas/hpcmc/webmin/utilbin' >> ~/.bash_profile
echo 'export PATH' >> ~/.bash_profile
```

安装后重新使用 **su** 命令切换 **root** 用户，并执行以下命令验证安装是否成功：

simsh hostname

步骤 4 在主 NameNode 节点，以 **root** 用户执行以下命令安装 TKGrid。

/opt/sas/TKGrid_2.9/TKGrid_Linux_x86_64.sh

界面提示以下信息，请分别输入确认结果：

表3-1 确认操作一览表

提示信息	确认操作
Shared install or replicate to each node? (Y=SHARED/n=replicated)	n
Enter additional paths to include in LD_LIBRARY_PATH, separated by colons (:)	无需额外操作
Enter remote process launcher command (Default is ssh).	无需额外操作
Enter additional options to mpirun.	-genvlist `env sed -e s/=.*/,/ sed /KRB5CCNAME/d tr -d '\n'^TKPATH,LD_LIBRARY_P ATH
Enter path to use for Utility files. (default is /tmp).	无需额外操作
Enter path to Hadoop. (default is Hadoop not installed).	/opt/huawei/Bigdata/hadoop/had oop
Force Root Rank to run on headnode? (y/N)	n
Enter full path to machine list. The head node '51-196- 21-3' should be listed first.	/etc/gridhosts
Enter maximum runtime for grid jobs (in seconds). Default 7200 (2 hours).	无需额外操作
Enter value for UMASK. (default is unset).	022

提示信息	确认操作
The install can now copy this directory to all the machines listed in '/opt/sas/TKGrid/grid.hosts' using scp, skipping the first entry. Perform copy?(YES/no)	yes

步骤 5 是否要使用网格计算功能？

- 是，执行步骤 6。
- 否，执行步骤 7。

步骤 6 在主 NameNode 节点，以 **root** 用户执行以下命令安装 TKTGDat。

/opt/sas/TKGrid_2.9/TKTGDat.sh

界面提示以下信息，请分别输入确认结果：

表3-2 确认操作一览表

提示信息	确认操作
Shared install or replicate to each node? (Y=SHARED/n=replicated)	n
Enter full path to machine list.	/etc/gridhosts
The install can now copy this directory to all the machines listed in '/etc/gridhosts' using scp, skipping the first entry. Perform copy?(YES/no)	yes

步骤 7 以 **root** 用户执行以下命令，在每个节点上建议软链接。

simsh ln -s /opt/sas/TKGrid_2.9/TKGrid /opt/sas/TKGrid

simsh ln -s /opt/sas/TKGrid_2.9/TKTGDat /opt/sas/TKTGDat

步骤 8 使用 SAS 用户执行以下命令查看安装是否成功。

/opt/sas/TKGrid/mpich2-install/bin/mpirun -f /etc/gridhosts hostname

界面将提示 SAS 服务器信息，例如：

```
you should see something like ...
sasserver01
sasserver02
sasserver03
sasserver04
```

步骤 9 使用 SAS 用户执行以下命令，查看获取“/etc/gridhosts”文件中服务器信息是否正确。

例如，主 NameNode 主机名为 sasserver01，需修改 GRIDHOST=sasserver01 并执行：

```
export TKPATH=/opt/sas/TKGrid/lib:/opt/sas/TKGrid/bin
export GRIDRSHCOMMAND=/opt/sas/TKGrid_2.9/TKGrid/bin/ssh.sh
export GRIDHOST=sasserver01
export GRIDINSTALLLOC=/opt/sas/TKGrid/
cd /opt/sas/TKGrid/bin
./tkgridmon
```

界面提示所有“/etc/gridhosts”文件中服务器的信息：

```
Here is sample output from the text version of gridmon.
It will continue to write to the session until the user interrupts the command
using Control-C.

Hostname = sasserver01
MemTotal = 16334464
Memfree= 15291460
NumCPUs= 21 % 0 %

/boot ( /dev/sdal ) -1 0 92846/470244
/dev/shm ( tmpfs ) -1 0 76/8167232
/ ( /dev/mapper/VolGroup-lv_root ) -1 0 8475228/97970080
hadoop 4 0 3492368/16334464
sasinst 29 2 1993620/16334464
rtkit 1 0 168456/16334464
gdm 14 0 3249528/16334464
nobody 1 0 12892/16334464
postfix 2 0 162876/16334464
ntp 1 0 30716/16334464
haldaemon 2 0 56832/16334464
rpcuser 1 0 23348/16334464
dbus 1 0 32196/16334464
rpc 1 0 18976/16334464
root 148 2 10747820/16334464

Hostname = sasserver02
MemTotal = 16334464
Memfree= 15582056
NumCPUs= 23 % 0 %

/boot ( /dev/sdal ) -1 0 92846/470244
/dev/shm ( tmpfs ) -1 0 80/8167232
/ ( /dev/mapper/VolGroup-lv_root ) -1 0 6970568/97970080
```

步骤 10 使用 PuTTY，以 **omm** 用户登录主管理节点，执行以下命令修改 FusionInsight 对接权限。

```
sh ${BIGDATA_HOME}/om-0.0.1/sbin/sas_adapter.sh -m
```

根据界面提示，输入“y”。

步骤 11 从“SAS_High_Performance_Deployment_for_Hadoop/2_6/Linux_for_x64/”获取“sashadoop.tar.gz”，并解压文件。

- 步骤 12 使用 WinSCP，以 **omm** 用户将获得的 jar 包复制到 FusionInsight 集群每个节点。如果主 NameNode 节点已安装 SAS HPCMC，可以上传 jar 包到主 NameNode，再执行以下命令分发到所有节点：

```
simcp sas.lasr.jar  
${BIGDATA_HOME}/FusionInsight/hadoop/hadoop/share/hadoop/common/lib
```

```
simcp sas.lasr.hadoop.jar  
${BIGDATA_HOME}/FusionInsight/hadoop/hadoop/share/hadoop/common/lib
```

- 步骤 13 使用 WinSCP，以 **omm** 用户将获得的 saslasrfd 文件复制到 FusionInsight 集群每个节点。如果主 NameNode 节点已安装 SAS HPCMC，可以上传 saslasrfd 文件到主 NameNode，再执行以下命令分发到所有节点：

```
simcp saslasrfd ${BIGDATA_HOME}/FusionInsight/hadoop/hadoop/bin
```

- 步骤 14 登录 FusionInsight Manager，选择“服务管理 > HDFS > 服务配置”，“参数类别”选择“全部配置”，修改如下参数：

表3-3 FusionInsight 服务端配置参数

参数名	key item	参数值
namenode.hdfs-site.customized.configs	com.sas.lasr.service.allow.put	true
	com.sas.lasr.hadoop.service.namenode.port	15452
	com.sas.lasr.hadoop.service.datanode.port	15453
dfs.namenode.plugins	无	com.sas.lasr.hadoop.NameNodeService
dfs.datanode.plugins	无	com.sas.lasr.hadoop.DataNodeService

单击“保存配置”，勾选“重启受影响的服务或实例”，单击“确定”重启服务。

- 步骤 15 在主 NameNode 节点安装 FusionInsight 客户端，并使用安装用户将客户端安装目录及子目录文件权限修改为 777。将下载的 keytab 文件上传到客户端安装节点，并授权 SAS 用户读取权限。

- 步骤 16 使用 SAS 用户登录安装客户端的节点，以及集群所有 NameNode 和 DataNode 所在节点，执行以下命令认证权限。

```
source /客户端安装目录/bigdata_env
```

```
kinit -kt <keytab 文件> <用户名>
```

- 步骤 17 使用 SAS 用户在 EG 客户端运行以下代码

在本地 Windows 环境运行 SAS 客户端，编辑以下代码，将各参数替换为实际正确的值，输入运行：

```
options set=GRIDHOST="cluster4-master1";  
options set=GRIDINSTALLLOC="/opt/sas/TKGrid";  
options set=GRIDRSHCOMMAND="/opt/sas/TKGrid_2.9/TKGrid/bin/ssh.sh";  
options set=TKPATH="/opt/sas/TKGrid/lib:/opt/sas/TKGrid/bin";  
libname hdfs sashdat path="/user/hpa/";
```

界面提示以下信息，对接 SAS HPA 完成：

```
NOTE: Libref HDFS was successfully assigned as follows:  
Engine:SASHDAT  
Physical Name: Directory '/user/hpa' of HDFS cluster on host 'cluster4-master1'
```

----结束

4 配置对接 SAS EP

操作场景

系统管理员可以通过 SAS EP 特性，配置 FusionInsight 对接 SAS EP，并使用 SAS 系统分析性能数据。

前提条件

- 已部署 FusionInsight HD。
- 在 FusionInsight Manager 创建一个角色与“人机”用户并下载 keytab 文件，具体请参见《FusionInsight HD 管理员指南》的“帐户管理”章节。角色需要根据业务需要授予权限，并将用户加入角色。
- 已安装 SAS 9.4_M2，并在本地 Windows 环境安装 SAS 客户端。
- 获取 SAS_High-Performance_Computing_Management_Console，VERSION 2.6。
- 获取 SAS High-Performance Analytics environment，VERSION 2.9。将文件保存到新的目录中，例如“/opt/sas/TKGrid_2.9”。
- 配置 **root** 用户和 SAS 用户在 FusionInsight 集群所有节点的 SSH 互信登录。
- 将 SAS 服务端的主机名与 IP 对应关系加入 FusionInsight 集群每个节点的 host 文件。
- FusionInsight 已配置通过 Access for hadoop 特性对接 SAS。
- FusionInsight 已配置通过 HPA 特性对接 SAS。

操作步骤

步骤 1 准备 TKGrid_REP。

在主 NameNode，以 **root** 用户执行以下命令，并将 jar 包权限修改为 777：

```
simsh mkdir /opt/hadoop jar
simsh find ${BIGDATA_HOME}/FusionInsight/FusionInsight-Hadoop-2.7.2/hadoop/ -name
*.jar | xargs -i cp {} /opt/hadoop_jar
```

步骤 2 执行以下命令，安装 TKGrid_REP。

```
sh /opt/sas/TKGrid_REP_x86_64.sh
```

界面提示以下信息，请分别输入确认结果：

表4-1 确认操作一览表

提示信息	确认操作
Do you want to configure remote access to Teradata? (yes/NO)	NO
Do you want to configure remote access to Greenplum? (yes/NO)	NO
Do you want to configure remote access to Hadoop? (yes/NO)	yes
Enter path of 64 bit JRE. i.e.: /usr/java/jdk1.7.0_09/jre	/opt/huawei/Bigdata/jdk/jre
Enter path of the directory containing the Hadoop client jars.	/opt/hadoop_jar
Do you want to configure remote access to Oracle?	no
Do you want to configure remote access to SAP HANA? (yes/NO)	no
Shared install or replicate to each node? (Y=SHARED/n=replicated)	n
Enter path to TKGrid install	/opt/sas/TKGrid
Enter additional paths to include in LD_LIBRARY_PATH, separated by colons (:)	-
The install can now copy this directory to all the machines listed in '/opt/sas/TKGrid/tkmpirsh.sh' using scp, skipping the first entry. Perform copy?(YES/no)	yes

步骤 3 切换到配置对接 SAS 准备的保存 Hadoop 配置文件目录，并修改 hdfs-site.xml，core-site.xml，yarn-site.xml 和 mapred-site.xml 文件。

- 删除 hdfs-site.xml，core-site.xml，yarn-site.xml 和 mapred-site.xml 中 value 为空的参数，例如<value/>或者<value />等。
- 修改 hdfs-site.xml，删除 hadoop.proxyuser.miner.groups 参数。
- 修改 yarn-site.xml，删除 clientPort 参数。
- 修改 core-site.xml，修改 fs.defaultFS 参数的值为主 NameNode 加端口号，端口号根据 dfs.namenode.rpc-address 的值决定，修改后例如：

```
<property>  
<name>fs.defaultFS</name>  
<value>hdfs://cluster4-master1:25000</value>  
</property>
```

步骤 4 创建“hadoop.xml”文件，并将“hdfs-site.xml”、“core-site.xml”、“yarn-site.xml”和“mapred-site.xml”文件修改后的内容复制到“hadoop.xml”文件中并保存。

步骤 5 执行以下命令，在主 NameNode 节点创建目录并修改权限。

```
mkdir /sasep
```

```
chmod -R 777 /sasep
```

步骤 6 执行以下命令，将 EP 的安装文件复制到目录中。

```
cp tkindbsrv-9.41_M2-2_lax.sh /sasep
```

```
cp hadoopmrjars-9.41_M2-2_lax.sh /sasep
```

步骤 7 执行以下命令，修改主 NameNode 节点环境变量，安装完 EP 后可以删除。

```
vi /etc/profile
```

```
#在末尾添加
export PATH=$PATH:${BIGDATA_HOME}/FusionInsight/FusionInsight-Hadoop-
2.7.2/hadoop/bin/
```

步骤 8 使用 SAS 用户登录安装客户端的节点，执行以下命令认证权限。

```
source /客户端安装目录/bigdata_env
```

```
kinit -kt <keytab 文件> <用户名>
```

步骤 9 以 root 用户执行以下命令，解压安装文件。

```
sh /sasep/tkindbsrv-9.41_M2-2_lax.sh
```

```
sh /sasep/hadoopmrjars-9.41_M2-2_lax.sh
```

步骤 10 执行以下命令，安装 EP。

```
sh sasep-servers.sh
```

```
-version <hadoop 版本>
```

```
-hadoopdir <hadoop core home directory>
```

```
-mrhome <the path to the MapReduce home >
```

```
-log <安装日志路径>
```

```
-host <需要安装 EP 的主机名列表>
```

例如：

```
sh sasep-servers.sh -add -version 2.1 -hadoopdir  
/opt/huawei/Bigdata/FusionInsight/hadoop/hadoop/hadoop -mrhome  
/opt/huawei/Bigdata/FusionInsight/hadoop/hadoop/hadoop -log /sasep/sasep_install.log -  
options "-trace 10" -host "10-196-21-3 10-196-24-4 10-196-24-5" -mrscript  
/sasep/hadoopmrjars-9.41_M2-2_lax.sh -epscript /sasep/tkindbsrv-9.41_M2-2_lax.sh
```

界面提示以下信息时，输入 y。

```
whether you want to start the SAS Embedded Process
```

 说明

SAS 9.4_M2 包含的“sasep-servers.sh”在官方介绍中最高支持 Hadoop 版本为 2.1，建议咨询服务供应商是否支持使用参数 2.1 支持高版本 Hadoop。

步骤 11 执行以下命令，重启 EP。

```
sh /sasep/SAS/SASTKInDatabaseServerForHadoop/9.41_M2-2/bin/sasep-servers.sh -stop
```

```
sh /sasep/SAS/SASTKInDatabaseServerForHadoop/9.41_M2-2/bin/sasep-servers.sh -start
```

步骤 12 修改依赖包路径。

```
simsh cp ${BIGDATA_HOME}/FusionInsight/FusionInsight-Hadoop-  
2.7.2/hadoop/lib/sas*.jar  
${BIGDATA_HOME}/FusionInsight/hadoop/hadoop/share/hadoop/common
```

步骤 13 在 FusionInsight Manager，重启 YARN 服务。

步骤 14 在 SAS 客户端运行以下代码。

请根据实际配置环境修改路径与主 NameNode 主机名，例如：

```
data simulate1;  
array _a{8} _temporary_ (0,0,0,1,0,1,1,1);  
array _b{8} _temporary_ (0,0,1,0,1,0,1,1);  
array _c{8} _temporary_ (0,1,0,0,1,1,0,1);  
do obsno=1 to 100;  
x = rantrnbl(1,0.28,0.18,0.14,0.14,0.03,0.09,0.08,0.06);  
a = _a{x};  
b = _b{x};  
c = _c{x};  
x1 = int(ranuni(1)*400);  
x2 = 52 + ranuni(1)*38;  
x3 = ranuni(1)*12;  
lp = 6. -0.015*(1-a) + 0.7*(1-b) + 0.6*(1-c) + 0.02*x1 -0.05*x2 - 0.1*x3;  
y = ranbin(1,1,(1/(1+exp(lp))));  
output;  
end;  
run;  
options set=SPDE_HADOOP_WORK_PATH="/tmp";  
options spdeparallelread=yes;  
option set=SAS HADOOP RESTFUL 1;  
options msglevel=i;  
options set=SAS HADOOP JAR_PATH="/opt/thirdparty/hadoop jar";  
options set=SAS HADOOP CONFIG_PATH="/opt/fi";  
options set=GRIDHOST="sasserver01";  
options set=GRIDINSTALLLOC="/opt/sas/TKGrid REP";  
options set=GRIDRSHCOMMAND="/opt/sas/TKGrid 2.9/TKGrid/bin/ssh.sh";  
options set=TKPATH="/opt/sas/TKGrid REP/lib:/opt/sas/TKGrid REP/bin";  
libname spdelib spde "/sasdata" hdfshost=default;  
proc delete data=spdelib.simulate1;run;  
data spdelib.simulate1;  
SET simulate1;  
run;  
proc hplogistic data=spdelib.simulate1;  
class a b c;  
model y = a b c x1 x2 x3;  
run;
```

步骤 15 SAS 客户显示如下提示表示对接完成且连接正常。

The HPLOGISTIC Procedure			
Performance Information			
Host Node		Sasserver	
Execution Mode		Distributed	
Number of Compute Nodes		2	
Number of Threads per Node		32	
Data Access Information			
Data	Engine	Role	Path
SPDELIB.SIMULATE11SPDE InputParallel, Asymmetric			
Model Information			
Data Source		SPDELIB.SIMULATE11	
Response Variable		y	
Class Parameterization		GLM	
Distribution		Binary	
Link Function		Logit	
Optimization Technique		Newton-Raphson with Ridging	
Class Level Information			
Class	Levels	Values	
a	20	1	
b	20	1	
c	20	1	
Number of Observations Read100			
Number of Observations Used100			
Response Profile			

Ordered	Total
Value	Frequency
10	98
21	2

You are modeling the probability that y='0'.

Iteration History				
Iteration	Evaluations	Objective	Change	Max Gradient
		Function		
0		40.1111241636	.	12.07096
1		20.07627341820.03485075		4.848905
2		20.05846311930.01781030		2.165041
3		20.04784631050.01061681		0.974626
4		20.04184900980.00599730		0.39361
5		20.03973664160.00211237		0.120455
6		20.03920944620.00052720		0.025165
7		20.03905381530.00015563		0.006672
8	2	0.0389975660.00005625		0.002403
9		20.03897688970.00002068		0.000884
10		20.03896928530.00000760		0.000325
11		20.03896648810.00000280		0.00012
12		20.03896545910.00000103		0.000044
13		20.03896508050.00000038		0.000016
14		20.03896494130.00000014		5.962E-6
15		20.03896489010.00000005		2.193E-6
16		20.03896489010.00000000		1.976E-8

Convergence criterion (GCONV=1E-8) satisfied. Quasi-complete separation detected.

Dimensions	
Columns in X	10
Number of Effects	7
Max Effect Columns	2
Rank of Cross-product Matrix	7
Parameters in Optimization	7

Fit Statistics	
-2 Log Likelihood	7.7930
AIC (smaller is better)	21.7930
AICC (smaller is better)	23.0104
BIC (smaller is better)	40.0292

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11.8148	6	0.0662

Parameter Estimates					
Parameter	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	17.9437	946.37	Infty	0.020	0.9849
a 0	14.1547	732.29	Infty	0.020	0.9846
a 1	0
b 0	-15.2788	946.35	Infty	-0.020	0.9871
b 1	0
c 0	1.5291	2.3206	Infty	0.660	0.5099
c 1	0
x1	0.02905	0.02054	Infty	1.410	0.1573
x2	-0.02534	0.07360	Infty	-0.340	0.7307
x3	-0.2982	0.3996	Infty	-0.750	0.4555

----结束