

对象存储服务 OBSA-HDFS 使用指南

文档版本

01

发布日期

2020-03-16



华为技术有限公司



版权所有 © 华为技术有限公司 2020。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



HUAWEI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： <https://www.huawei.com>

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

目 录

1 OBSA-HDFS 简介.....1

2 下载 OBSA-HDFS.....2

3 安装.....3

4 配置.....4

4.1 Spark、MapReduce、Hive 环境配置.....4

4.2 HBase 环境配置.....4

5 修订记录.....6

1 OBSA-HDFS 简介

OBSA-HDFS插件基于OBS对象存储服务，提供HDFS的FileSystem语义。实现大数据计算Spark、MapReduce、Hive、HBase等组件与OBS对象存储服务的对接，为大数据计算提供“数据湖”存储。

应用场景一：华为云大数据服务MRS、DLI已在版本中包含OBSA-HDFS插件，无需下载安装，即可使用OBS作为大数据存储。

应用场景二：华为云租户基于IAAS资源，自建大数据计算平台，可下载并配置OBSA-HDFS插件，使用OBS作为大数据存储。

2 下载 OBSA-HDFS

下载OBSA-HDFS，地址如下（请获取最新版本）：

<https://github.com/huaweicloud/obsa-hdfs/tree/master/release>

说明

hadoop-huaweicloud-x.x.x-hw-y.jar包：前三位x.x.x为配套hadoop版本号；最后一位y为版本号，y值最大为最新版本。

如：hadoop-huaweicloud-2.8.3-hw-37.jar，2.8.3是配套hadoop版本号，37是OBSA-HDFS的版本号。

3 安装

在Hadoop、Spark、Hive、HBase等大数据组件中安装OBSA-HDFS，如下步骤：

步骤1 OBSA-HDFS软件包（版本号如：hadoop-huaweicloud-2.8.3-hw-37.jar），放入Hadoop、Spark、Hive、HBase等大数据组件的classpath目录下。

具体方法：在根目录下，find查找hadoop-aws-2.8.3.jar包所在目录，将hadoop-huaweicloud-2.8.3-hw-37.jar放入。

如：

Hadoop目录为

/home/modules/hadoop-2.8.3/share/hadoop/common/lib;

/home/modules/hadoop-2.8.3/share/hadoop/tools/lib;

/home/modules/hadoop-2.8.3/share/hadoop/httpfs/tomcat/webapps/webhdfs/
WEB-INF/lib;

/home/modules/hadoop-2.8.3/share/hadoop/hdfs/lib

Spark目录为

/home/modules/spark-2.3.0/jars/

Hive目录为

/home/modules/hive-2.3.3/lib

HBase目录为

/home/modules/hbase-2.1.0/lib

----结束

4 配置

4.1 Spark、MapReduce、Hive环境配置

4.2 HBase环境配置

4.1 Spark、MapReduce、Hive 环境配置

core-site.xml文件，OBS相关访问配置信息，如下：

```
<property>
<name>fs.obs.impl</name>
<value>org.apache.hadoop.fs.obs.OBSFileSystem</value>
</property>
<property>
<name>fs.obs.access.key</name>
<value>myaccesskey</value>
</property>
<property>
<name>fs.obs.secret.key</name>
<value>mysecretkey</value>
</property>
<property>
<name>fs.obs.endpoint</name>
<value>obsEndpoint</value>
</property>
<property>
<name>fs.obs.buffer.dir</name>
<value>/home/modules/data/buf</value>
</property>
```

4.2 HBase 环境配置

core-site.xml文件，OBS相关访问配置信息，如下：

```
<property>
<name>fs.obs.impl</name>
<value>org.apache.hadoop.fs.obs.OBSFileSystem</value>
</property>
<property>
<name>fs.obs.access.key</name>
<value>myaccesskey</value>
</property>
<property>
<name>fs.obs.secret.key</name>
<value>mysecretkey</value>
</property>
```

```
<property>
<name>fs.obs.endpoint</name>
<value>obsEndpoint</value>
</property>
<property>
<name>fs.obs.buffer.dir</name>
<value>/home/modules/data/buf</value>
</property>
<property>
<name>fs.obs.readahead.range</name>
<value>65536</value>
</property>
<property>
<name>fs.obs.read.transform.enable</name>
<value>false</value>
</property>
```

hbase-site.xml配置文件，实现hFile访问OBS，hLog访问HDFS。配置如下。

```
<property>
<name>hbase.rootdir</name>
<value>obs://bucketName/xxx</value>
</property>
<property>
<name>hbase.wal.provider</name>
<value>org.apache.hadoop.hbase.wal.FSHLogProvider</value>
</property>
<property>
<name>hbase.wal.dir</name>
<value>hdfs://node:8020/hbase</value>
</property>
```


5 修订记录

发布日期	修订记录
2019-10-15	第一次正式发布
2020-01-06	第二次修订