

对象存储服务工具指南（OBSA-HDFS）

对象存储服务工具指南（OBSA-HDFS）

文档版本

01

发布日期

2019-10-14



版权所有 © 华为技术有限公司 2019。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：<http://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 OBSA-HDFS.....3

4 配置 OBSA-HDFS.....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

各版本的OBSA-HDFS，下载地址如下：

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

说明

hadoop-huaweicloud-x.x.x.x.jar的最大版本号为最新版本。

3 安装 OBSA-HDFS

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

步骤1 获取jar包，如：hadoop-huaweicloud-2.8.3.28.jar，OBS SDK相关的jar包（esdk-obs-java-3.19.7.1.jar、java-xmlbuilder-1.1.jar、okhttp-3.10.0.jar、okio-1.14.0.jar）。

步骤2 将jar包放入Spark、MapReduce、Hive、HBase等大数据组件的classpath目录下。

如：Spark目录为/home/modules/spark-2.3.0/jars/

注意

hadoop原目录中可能有低版本的java-xmlbuilder-0.4.jar包。如果有，需删除替换为java-xmlbuilder-1.1.jar，否则会有包冲突。

----结束

4 配置 OBSA-HDFS

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，配置如下：

```
<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>1 * 1024 * 1024</value>
</property>
```

4.2 HBase 环境配置

- core-site.xml文件，配置如下：

```
<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>64 * 1024</value>
</property>
```

- **hbase-site.xml**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