

Please visit website: <http://cxyroad.com>

/apache/flink/blob/release-1.18.1/flink-core/src/main/java/org/apache/flink/core/fs/FileSystem.java”)类来进行操作。而这个类同样是`FileSystem SQL Connector`的底层实现类。既然如此，`Flink#FileSystem`应该支持访问HDFS、S3等其他文件系统，那其内部必然会使用`hadoop#FileSystem`的api，而`hadoop#FileSystem`自身如果支持`SFTP`，则此路可以走通，为了确认这一点，开始查看源码并查看类注释中的相关信息，发现一段有用信息： \*\*

```
...
/**
 * Flink implements and supports some file system types directly (for
 * example the default machine-local file system). Other file * system types
 * are accessed by an implementation that bridges to the suite of file
 * systems supported by Hadoop (such as for
 * example HDFS).
 */
```

// 翻译：Flink直接实现并支持一些文件系统类型（例如默认的机器本地文件系统）。其他文件系统类型由桥接到Hadoop支持的文件系统套件（例如HDFS）的实现访问。

```
...
4. 看到此处信心倍增，继续翻阅后发现在`Flink#FileSystem`有一个
[getUnguardedFileSystem](http://cxyroad.com/
"https://github.com/apache/flink/blob/a8c8b1c0e2c5e2e468f6de62f31a69e90af5c96e/flink-core/src/main/java/org/apache/flink/core/fs/FileSystem.java#L413")函
数，如下图：
```

![在这里插入图片描述](<https://p3-juejin.byteimg.com/tos-cn-i-k3u1fbpfcp/c5855a56943248c0b3a7c1c70046c693~tplv-k3u1fbpfcp-jj-mark:3024:0:0:0:q75.awebp>)

6. 该函数会检测文件`url`路径，如果路径是`file://`则会走`Flink#FileSystem`的内部实现，如果是`hdfs://`，`sftp://`这类前缀，则会调用[loadHadoopFsFactory](<http://cxyroad.com/>
"<https://github.com/apache/flink/blob/a8c8b1c0e2c5e2e468f6de62f31a69e90af5c96e/flink-core/src/main/java/org/apache/flink/core/fs/FileSystem.java#L1114>")函数，如下图：

![在这里插入图片描述](https://p3-juejin.byteimg.com/tos-cn-i-k3u1fbpfcp/7fc02eed316c4057ba7b62d4fdc77b63~tplv-k3u1fbpfcp-jj-mark:3024:0:0:0:q75.awebp)

7. 至此找到`Flink#FileSystem`与`Hadoop#FileSystem`的桥接处,在该函数中会先加载`hadoop#FileSystem`并构建`Flink#FileSystem`的子类[HadoopFileSystem](http://cxyroad.com/"https://github.com/apache/flink/blob/a8c8b1c0e2c5e2e468f6de62f31a69e90af5c96e/flink-fileSystems/flink-hadoop-fs/src/main/java/org/apache/flink/runtime/fs/hdfs/HadoopFileSystem.java#L37"),而在`HadoopFileSystem`类中使用`hadoop#FileSystem`提供的能力。

8. 那么`hadoop#FileSystem`是否提供了读写SFTP的能力呢?经过调研发现只有`Hadoop-2.8.x`版本以上才支持SFTP, [JIRA工单](http://cxyroad.com/"https://issues.apache.org/jira/browse/HADOOP-5732")

![在这里插入图片描述](https://p3-juejin.byteimg.com/tos-cn-i-k3u1fbpfcp/03db882ad4e2416dbc7c92d10624ce35~tplv-k3u1fbpfcp-jj-mark:3024:0:0:0:q75.awebp)

![在这里插入图片描述](https://p3-juejin.byteimg.com/tos-cn-i-k3u1fbpfcp/b02b53489302422c9f019c51964f4e45~tplv-k3u1fbpfcp-jj-mark:3024:0:0:0:q75.awebp)

9. 至此总结以下:虽然`Flink#FileSystem`原生并未支持`sftp`读写,但`Flink#FileSystem`中如果遇见不支持的文件前缀如:`hdfs://`或者`sftp FTP`相关信息配置好对应K,V,内容如下:

```
...
<configuration>
  <!-- 配置sftp实现类 -->
  <property>
    <name>fs.sftp.impl</name>
    <value>org.apache.hadoop.fs.sftp.SFTPFileSystem</value>
  </property>

  <!-- 配置sftp用户名 -->
  <property>
    <name>fs.sftp.user.101.230.65.134</name>
    <value>username</value>
  </property>
```

```
<!-- 配置sftp秘钥路径 -->
<property>
  <name>fs.sftp.keyfile</name>
  <value>D:\IdeaProjects\flink-
sftp\src\main\resources\uat_sftp_qadmin.rsa</value>
</property>

<!-- 配置sftp密码 -->
<property>
  <name>fs.sftp.password.101.230.65.134</name>
  <value>password</value>
</property>

</configuration>
...
```

9. 执行成功，如下图：

![在这里插入图片描述](https://p3-juejin.byteimg.com/tos-cn-i-k3u1fbpfcp/acc180ed4954409e9352f21a3d7eef21~tplv-k3u1fbpfcp-jj-mark:3024:0:0:0:q75.awebp)

10. 至此我们可以使用FlinkSQL#FileSystem已经写好的各种文件格式类型以及分区功能，还可以享受Hadoop#SFTPFileSystem读写sftp的能力，可以说完美的解决此需求。

#### 四、总结

-----

通过这次详细排查和研究，对Flink文件系统的实现有了更加深入的理解。起初，我对FlinkSQL是否支持sftp产生了疑问，然而，通过逐步追踪源代码，逐渐揭示了底层逻辑的实现机制：尽管原生的`Flink#FileSystem`并没有直接支持SFTP的读写操作，但它通过一个巧妙的桥接机制，将不支持的文件前缀（例如`hdfs://`或者`sftp://`）重新定向到`Hadoop#FileSystem`类来处理。而`Hadoop#FileSystem`底层提供了对多种文件类型的广泛支持，只要存在SFTP的实现类，就可以顺利进行操作。这个过程中，深刻体会到了理解底层原理和追踪代码的重要性。

这次经历让我明白，在后续的开发过程中，我们应该保持持续的好奇心，提出更多问题，积极思考，并深入探索底层实现原理，时刻保持探索精神，不断拓展我们的知识和技能，以便在开发过程中能够更加游刃有余地应对各种挑战。

## 五、相关资料

-----

- \* [Hadoop#JIRA工单](<http://cxyroad.com/https://issues.apache.org/jira/browse/HADOOP-5732>)
- \* [FileSystem SQL Connector](<http://cxyroad.com/https://nightlies.apache.org/flink/flink-docs-master/docs/connectors/table/filesystem/>)
- \* [org.apache.flink.core.fs.FileSystem](<http://cxyroad.com/https://github.com/apache/flink/blob/release-1.18.1/flink-core/src/main/java/org/apache/flink/core/fs/FileSystem.java>)
- \* [getUnguardedFileSystem](<http://cxyroad.com/https://github.com/apache/flink/blob/a8c8b1c0e2c5e2e468f6de62f31a69e90af5c96e/flink-core/src/main/java/org/apache/flink/core/fs/FileSystem.java#L413>)
- \* [loadHadoopFsFactory](<http://cxyroad.com/https://github.com/apache/flink/blob/a8c8b1c0e2c5e2e468f6de62f31a69e90af5c96e/flink-core/src/main/java/org/apache/flink/core/fs/FileSystem.java#L1114>)
- \* [HadoopFileSystem](<http://cxyroad.com/https://github.com/apache/flink/blob/a8c8b1c0e2c5e2e468f6de62f31a69e90af5c96e/flink-fileSystems/flink-hadoop-fs/src/main/java/org/apache/flink/runtime/fs/hdfs/HadoopFileSystem.java#L37>)

原文链接: <https://juejin.cn/post/7350917259718967323>