## 关于应用系统HDFS的代码升级建议

大数据平台将从IDH升级到CDH，其中Hadoop组件将从1.0.3版本升级到2.5.1版本。以CDH5.2.1为例，本文主要介绍应用系统基于Hadoop2.5.0进行HDFS操作的建议。

下面详细介绍在使用JAVA API操作Hadoop2.5.0的HDFS文件时应注意的一些问题。

1. **Shell**

Shell层面，Hadoop 2.5.0与Hadoop 1.0.3相比，Shell命令有增加，主要有：

appendToFile、getfacl、getfattr、setfacl、setfattr。

具体用法请参考：

<http://hadoop.apache.org/docs/r2.5.2/hadoop-project-dist/hadoop-common/FileSystemShell.html>

1. **配置参数**

在Hadoop1.0.3版本中，通过Hadoop Java API 访问HDFS文件，只需要填写fs.default.name，如：

Configuration cfg = **new** Configuration();

cfg.set("fs.default.name", " hdfs://172.17.238.182:8020");

在Hadoop2.5.0版本中，通过Hadoop Java API 访问HDFS文件，只需要填写fs.defaultFS，如：

Configuration cfg = **new** Configuration();

cfg.set("fs.defaultFS", "hdfs://bB0103002:8020");

需要特别注意的是，在Cloudera Hadoop 中启用了NameNode 的HA，需要做如下设置：

Configuration cfg = **new** Configuration();

cfg.set("fs.defaultFS", "hdfs://nameservice1");

cfg. set("dfs.nameservices", "nameservice1");

cfg.set("dfs.ha.namenodes.nameservice1", "namenode573,namenode700");

cfg.set("dfs.client.failover.proxy.provider.nameservice1", "org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider");

cfg.set("dfs.namenode.rpc-address.nameservice1.namenode573", "bB0107001:8020");

cfg.set("dfs.namenode.rpc-address.nameservice1.namenode700", "bB0206001:8020");

1. **Jar包**

与Hadoop 1.0.3版本相比，通过Hadoop Java API 访问HDFS文件所需要的jar包也是不同的，具体如下：

|  |  |
| --- | --- |
| Hadoop 1.0.3 | Hadoop2.5.0 |
| commons-logging-1.1.1.jar | commons-cli-1.2.jar |
| commons-lang-2.5.jar | commons-collections-3.2.1.jar |
| commons-configuration-1.6.jar | commons-configuration-1.6.jar |
| guava-11.0.2.jar | commons-lang-2.6.jar |
| hadoop-core-1.0.3-Intel.jar | commons-logging-1.1.3.jar |
|  | guava-11.0.2.jar |
|  | hadoop-auth-2.5.0-cdh5.2.1.jar |
|  | hadoop-common-2.5.0-cdh5.2.1.jar |
|  | hadoop-hdfs-2.5.0-cdh5.2.1.jar |
|  | log4j-1.2.17.jar |
|  | protobuf-java-2.5.0.jar |
|  | slf4j-api-1.7.5.jar |
|  | slf4j-log4j12.jar |

通过JAVA API 对HDFS文件进行操作，暂时未发现Hadoop2.5.0与Hadoop1.0.3有任何不同之处，代码层面暂不需要进行修改。