

Flume-0.9.4和Hbase-0.96整合(1)

这几天由于项目的需要，需要将Flume收集到的日志插入到Hbase中，有人说，这不很简单么？Flume里面自带了Hbase

sink，可以直接调用啊，还用说么？是

的，我在本博客的[《Flume-1.4.0和Hbase-0.96.0整合》](#)

文章中就提到如何用Flume和Hbase整合，从文章中就看出整个过程不太复杂，直接做相应的配置就行了。

那么为什么今天还要特意提一下Flume-0.9.4和Hbase-0.96整合？这是因为Flume-0.9.4和Hbase-0.96整合比Flume-1.4.0和Hbase-0.96整合麻烦多了！不是随便几个配置就能搞定的，里面涉及到修改Flume和Hadoop的源码！

先看下我公司的Hadoop、Hbase、Flume等的配置吧。2013年10月末，公司的Hadoop升级到2.2.0，Hbase升级到0.96,Zookeeper升级到3.4.5，但是Flume版本由于各种原因没有升级，还是用Flume-0.9.4，而Flume-0.9.4源码是基于Hadoop-0.20.2-CDH3B4、Hbase-0.90.1-cdh3u0开发的，Hadoop-0.20.2-CDH3B4和现在的Hadoop-2.2.0完全不一样的设计，而且直接用Hadoop-0.20.2-CDH3B4会使得Flume-0.9.4不能和Hbase-0.96.0通信，都不能通信了，何谈整合！但是经过几天的奋战，我们终于通过修改Flume和Hadoop的部分源码达到了Flume-0.9.4和Hbase-0.96整合，今天就分享一下我们是怎么修改的。

考虑到篇幅的问题，已经将本文拆分为二，为你带来的不便深感抱歉！下面是两篇文章的链接：

[《Flume-0.9.4和Hbase-0.96整合\(1\)》](#)、[《Flume-0.9.4和Hbase-0.96整合\(2\)》](#)

1、修改Flume-src根目录下的pom.xml文件中的部分依赖版本

(1)、Hadoop2x里面已经没有hadoop-core jar包，所以修改Hadoop的依赖包的版本：

```
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-core</artifactId>
  <version>${cdh.hadoop.version}</version>
</dependency>
```

修改为

```
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-mapreduce-client-core</artifactId>
  <version>2.2.0</version>
</dependency>
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-common</artifactId>
```

```
<version>2.2.0</version>
</dependency>
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-mapreduce-client-common</artifactId>
  <version>2.2.0</version>
</dependency>
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-mapreduce-client-jobclient</artifactId>
  <version>2.2.0</version>
</dependency>
```

(2)、修改Guava的版本

```
<dependency>
  <groupId>com.google.guava</groupId>
  <artifactId>guava</artifactId>
  <version>r07</version>
</dependency>
```

修改为

```
<dependency>
  <groupId>com.google.guava</groupId>
  <artifactId>guava</artifactId>
  <version>10.0.1</version>
</dependency>
```

(3)、修改flume-src\flume-core\pom.xml里面的以下配置

```
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-core</artifactId>
</dependency>
```

修改为

```
<dependency>
  <groupId>org.apache.hadoop</groupId>
```

```

    <artifactId>hadoop-mapreduce-client-core</artifactId>
    <version>2.2.0</version>
</dependency>
<dependency>
    <groupId>org.apache.hadoop</groupId>
    <artifactId>hadoop-common</artifactId>
    <version>2.2.0</version>
</dependency>
<dependency>
    <groupId>org.apache.hadoop</groupId>
    <artifactId>hadoop-mapreduce-client-common</artifactId>
    <version>2.2.0</version>
</dependency>
<dependency>
    <groupId>org.apache.hadoop</groupId>
    <artifactId>hadoop-mapreduce-client-jobclient</artifactId>
    <version>2.2.0</version>
</dependency>

```

(4)、修改flume-src\plugins\flume-plugin-hbasesink\pom.xml里面的以下配置

```

<dependency>
    <groupId>org.apache.hbase</groupId>
    <artifactId>hbase</artifactId>
    <version>${cdh.hbase.version}</version>
</dependency>

<dependency>
    <groupId>org.apache.hbase</groupId>
    <artifactId>hbase</artifactId>
    <version>${cdh.hbase.version}</version>
    <classifier>tests</classifier>
    <scope>test</scope>
</dependency>

<dependency>
    <groupId>org.apache.hadoop</groupId>
    <artifactId>hadoop-test</artifactId>
    <version>${cdh.hadoop.version}</version>
    <scope>test</scope>
</dependency>

```

修改为

```
<dependency>
  <groupId>org.apache.hbase</groupId>
  <artifactId>hbase-it</artifactId>
  <version>0.96.0-hadoop2</version>
</dependency>
```

考虑到篇幅的问题，已经将本文拆分为二，为你带来的不便深感抱歉！下面是两篇文章的链接：

[《Flume-0.9.4和Hbase-0.96整合\(1\)》](#)、[《Flume-0.9.4和Hbase-0.96整合\(2\)》](#)

本博客文章除特别声明，全部都是原创！

原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。

本文链接: [【】](#) ()