链滴

Spring for Apache Hadoop 1.0发布

作者：cat
原文链接：https://ld246.com/article/1363701695282
来源网站：链滴

 SpringSource发布了Spring for Apache Hadoop 1.0。开发者能够通过它编写基于Spring Framework的Hadoop应用，还很容易地与Spring Batch和Spring Integration集成。Spring for Apache Hadoop是Spring Data大项目的一个子项目，它基于开源的Apache 2.0许可发布。
 Hadoop应用通常是一个命令行工具、脚本和代码的集合。Spring for Apache Hadoop为Hadoop应用开发提供了一个一致性的编程模型和声明式配置模型。开发人员现在能够借助它使用Spring编模型（依赖注入、POJO和辅助模板）实现Hadoop应用，并且能够以标准的Java应用而不是命令行具的方式运行它。Spring for Apache Hadoop支持对HDFS的读写操作，支持运行MapReduce、流者级联工作，还能够与HBase、Hive和Pig交互。
 Spring for Apache Hadoop包含以下关键特性：

  支持声明式配置，能够创建、配置和参数化Hadoop连接，支持MapReduce、流、Hive、Pig级联工作。有不同的“runner”类执行不同的Hadoop交互类型，它们分别是JobRunner、ToolRunner、 JarRunner、 HiveRunner、 PigRunner、CascadeRunner和HdfsScriptRunner。
  全面的HDFS数据访问支持，可以使用所有基于JVM的脚本语言，例如Groovy、JRuby、Jython和Rhino。
  支持 Pig和Hive的模板类PigTemplate和HiveTemplate。这些辅助类提供了异常转化、资源管和轻量级对象映射功能。
  支持对HBase的声明式配置，同时为Dao层支持引入了HBaseTemplate。
  声明和编程支持Hadoop工具，包括文件系统Shell（FsShell）和分布式复制（DistCp）。

  安全支持。Spring for Apache Hadoop清楚运行Hadoop环境的安全约束，因此能够透明地从个本地开发环境迁移到一个完全Kerberos安全的Hadoop集群。
  支持Spring Batch。通过Spring Batch，多个步骤能够被调整为有状态的方式并使用REST API行管理。例如，Spring Batch处理大文件的能力就可以被用于向HDFS导入或者从HDFS导出文件。
  支持Spring Integration。Spring Integration允许对那些在被读取并写入HDFS及其他存储之能够被转换或者过滤的事件流进行处理。

 下面是配置示例和代码片段，大部分来自于Spring for Hadoop博客或者参考手册。
 MapReduce

<pre><!-- use the default configuration -->
<hdp:configuration />
<!-- create the job -->
<hdp:job id="word-count"
    input-path="/input/" output-path="/ouput/"
    mapper="org.apache.hadoop.examples.WordCount.TokenizerMapper"
    reducer="org.apache.hadoop.examples.WordCount.IntSumReducer" />
<!-- run the job -->
<hdp:job-runner id="word-count-runner" pre-action="cleanup-script" post-action="export-results" job="word-count" run-at-startup=quot;true" /></pre>
 HDFS

<pre><!-- copy a file using Rhino -->
<hdp:script id="inlined-js" language="javascript" run-at-startup=quot;true">
    importPackage(java.util)

name = UUID.randomUUID().toString()
scriptName = "src/main/resources/hadoop.properties"</pre>

```
// fs - FileSystem instance based on 'hadoopConfiguration' bean
fs.copyFromLocalFile(scriptName, name)

&lt;/hdp:script&gt;                    </pre> HBase

<p></p>
<p> </p>
<pre>&lt;!-- use default HBase configuration --&gt;
&lt;hdp:hbase-configuration /&gt;


&lt;!-- wire hbase configuration --&gt;

&lt;bean id=&quot;hbaseTemplate&quot;
class=&quot;org.springframework.data.hadoop.hbase.HbaseTemplate&q
ot; p:configuration-ref=&quot;hbaseConfiguration
quot; /&gt;

// read each row from HBaseTable (Java)

List rows = template.find(&quot;HBaseTable
quot;, &quot;HBaseColumn&quot;, new
owMapper() {

@Override

public String mapRow(Result result, int rowNum) throws Exception {

return result.toString();

}

}));</pre> Hive

<p></p>
<p> </p>
<pre>&lt;!-- configure data source --&gt;
&lt;bean id=&quot;hive-driver&quot; class=&quot;org.apache.hadoop.hive.jdbc.HiveDriver&
uot; /&gt;
&lt;bean id=&quot;hive-ds&quot; class=&quot;org.springframework.jdbc.datasource.Simple
riverDataSource&quot; c:driver-ref=&quot;hive-driver&quot; c:url=&quot;${hive.url}&quot; /
gt;


&lt;!-- configure standard JdbcTemplate declaration --&gt;

&lt;bean id=&quot;hiveTemplate&quot;
class=&quot;org.springframework.jdbc.core.JdbcTemplate&quot;
c:data-source-ref=&quot;hive-ds&quot;/
gt;</pre> Pig

<p></p>
<p> </p>
<pre>&lt;!-- run an external pig script --&gt;
&lt;hdp:pig-runner id=&quot;pigRunner&quot; run-at-startup=&quot;true&quot;&gt;
    &lt;hdp:script location=&quot;pig-scripts/script.pig&quot;/&gt;
&lt;/hdp:pig-runner&gt;</pre>
<p> 如果想要开始，可以<a href="http://www.springsource.com/download/community?projec
=Spring%20Data%20Hadoop">下载Spring for Apache Hadoop</a>或者使用<i>org.springfr
mework.data:spring-data-hadoop:1.0.0.RELEASE</i>  Maven构件。还可以获取Spring for
```

Hadoop的WordCount示例。在YouTube上还有介绍Spring Hadoop的网络会议。
 Spring for Apache Hadoop需要JDK 6.0及以上版本、Spring Framework 3.0及以上版本（推使用3.2）和Apache Hadoop 0.20.2 （推荐1.0.4）。现在并不支持Hadoop YARN、NextGen或 2.。支持所有的Apache Hadoop 1.0.x分布式组件，这些分布式组件包括vanilla Apache Hadoop、Cludera CDH3、CDH4和Greenplum HD等。
 想要获取更深入的信息，你可以阅读Spring for Apache Hadoop <u>参考手册</> 和 avadoc。Spring for Apache Hadoop 的源代码和<a href="https://github.com/SpringSource/spring-hadoop-samples/" 示例托管在GitHub上。
 查看英文原文： Spring for Apache Hadoop 1.0

