

Java-Success.com

Prepare to fast-track, choose & go places with 800+ Java & Big Data Q&As with lots of code & diagrams.

[Home](#) [Why? ▾](#) [300+ Java FAQs ▾](#) [300+ Big Data FAQs ▾](#) [Courses ▾](#)[👤 Membership ▾](#) [Your Career ▾](#)

[Home](#) › [bigdata-success.com](#) › [Tutorials - Big Data](#) › [TUT - Pig](#) › 3. Apache Pig: XPath
for XML

3. Apache Pig: XPath for XML

 Posted on [February 6, 2016](#)

This extends the tutorial [1. Apache Pig Getting started](#) and [2. Apache Pig: Regex \(Regular expressions\)](#).

Input Data

scores.xml in folder: **/Users/arulk/projects**
representing marks of 4 students in 3 subjects:

```
1
2 <scores>
3   <subject>
4     <name>Science</name>
5     <marks>
6       <mark>80</mark>
```

300+ Java Interview FAQs

300+ Java FAQs



16+ Java Key Areas Q&As



150+ Java Architect FAQs



80+ Java Code Quality Q&As



150+ Java Coding Q&As



300+ Big Data Interview FAQs

300+ Big Data FAQs



Tutorials - Big Data



TUT -  Starting Big Data

TUT - Starting Spark & Scala

```

7      <mark>75</mark>
8      <mark>89</mark>
9      <mark>90</mark>
10     </marks>
11  </subject>
12  <subject>
13      <name>Maths</name>
14      <marks>
15          <mark>90</mark>
16          <mark>87</mark>
17          <mark>78</mark>
18          <mark>92</mark>
19      </marks>
20  </subject>
21  <subject>
22      <name>English</name>
23      <marks>
24          <mark>78</mark>
25          <mark>88</mark>
26          <mark>65</mark>
27          <mark>99</mark>
28      </marks>
29  </subject>
30 </scores>
31

```

TUT - Starting with Python

TUT - Kafka

TUT - Pig

TUT - Apache Storm

TUT - Spark Scala on Zeppelin

TUT - Cloudera

TUT - Cloudera on Docker

TUT - File Formats

TUT - Spark on Docker

TUT - Flume

TUT - Hadoop (HDFS)

TUT - HBase (NoSQL)

TUT - Hive (SQL)

TUT - Hadoop & Spark

TUT - MapReduce

TUT - Spark and Scala

TUT - Spark & Java

TUT - PySpark on Databricks

TUT - Zookeeper

Step 1: Start pig in local file system mode.

```

1
2 pig -x local
3

```

Step 2: Extract the “Subjects” from the input XML file.

```

1
2 grunt> SUBJECTS_EXTRACT = LOAD '/Users/arulk/proje
3

```

Dump the output:

```

1
2 grunt> dump SUBJECTS_EXTRACT;
3

```

```

1

```

800+ Java Interview Q&As

300+ Core Java Q&As



300+ Enterprise Java Q&As



150+ Java Frameworks Q&As



120+ Companion Tech Q&As



Tutorials - Enterprise Java



```
2 (<subject>          <name>Science</name>          <marks>  
3 (<subject>          <name>Maths</name>          <marks>  
4 (<subject>          <name>English</name>          <marks>  
5
```

Step 3: XPath expressions to extract the marks for each subject by 4 students. The **XPath** and **XPathAll** methods are defined in “piggybank.jar”.

```
1  
2 REGISTER /Users/arulk/pig-0.15.0/lib/piggybank.jar  
3 DEFINE XPath org.apache.pig.piggybank.evaluation.xr  
4 DEFINE XPathAll org.apache.pig.piggybank.evaluation  
5  
6 MARKS_FOR_SUBJECT = FOREACH SUBJECTS_EXTRACT GENER  
7
```

Dump the output:

```
1  
2 dump MARKS_FOR_SUBJECT;  
3
```

Output:

```
1  
2 ((Science),80,75,89,90)  
3 ((Maths),90,87,78,92)  
4 ((English),78,88,65,99)  
5
```

Step 4: Create a new file into a folder named “MARKS_FOR_SUBJECT.TXT” to store the output into a file named “part-m-00000”.

```
1  
2 store MARKS_FOR_SUBJECT into 'MARKS_FOR_SUBJECT.TXT'  
3
```

03: Java GC tuning for low latency applications ›

Disclaimer

The contents in this Java-Success are copyrighted and from EmpoweringTech Pty Ltd. The EmpoweringTech Pty Ltd has the right to correct or enhance the current content without any prior notice. These are general advice only, and one needs to take his/her own circumstances into consideration. The EmpoweringTech Pty Ltd will not be held liable for any damages caused or alleged to be caused either directly or indirectly by these materials and resources. Any trademarked names or labels used in this blog remain the property of their respective trademark owners. Links to external sites do not imply endorsement of the linked-to sites. [Privacy Policy](#)