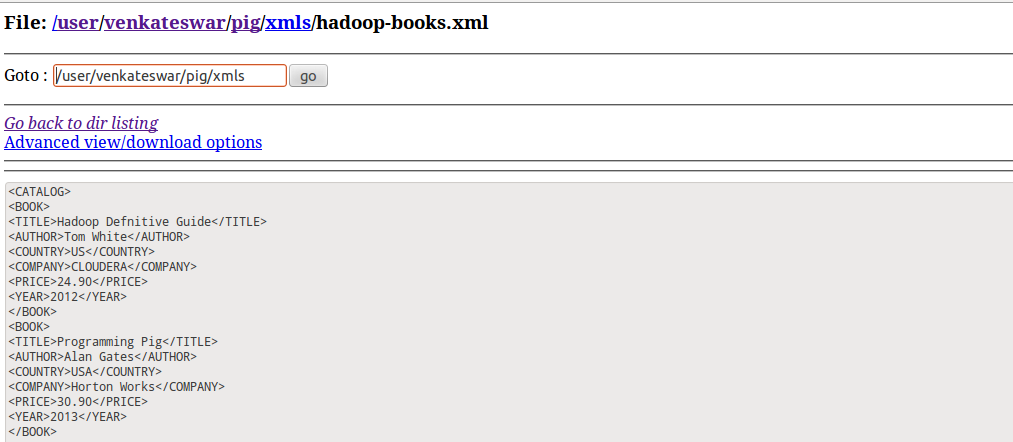
**Pig XML Processing :**

**Tools Used: (Pig, HDFS, Hive, Sqoop, Mysql)**

**HDFS:**

hadoop fs -put /home/venkateswar/work/pig/xmls/hadoop-books.xml user/venkateswar/pig/xmls/hadoop-books.xml



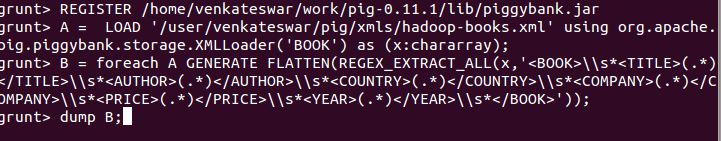
**PIG:**

REGISTER /home/venkateswar/work/pig-0.11.1/lib/piggybank.jar

A = LOAD '/user/venkateswar/pig/xmls/hadoop-books.xml' using org.apache.pig.piggybank.storage.XMLLoader('BOOK') as (x:chararray);

B = foreach A GENERATE FLATTEN(REGEX\_EXTRACT\_ALL(x,'<BOOK>\\s\*<TITLE>(.\*)</TITLE>\\s\*<AUTHOR>(.\*)</AUTHOR>\\s\*<COUNTRY>(.\*)</COUNTRY>\\s\*<COMPANY>(.\*)</COMPANY>\\s\*<PRICE>(.\*)</PRICE>\\s\*<YEAR>(.\*)</YEAR>\\s\*</BOOK>'));

dump B;



STORE B INTO '/user/venkateswar/pig/output2.txt' using PigStorage(';');



**HIVE**:

create table if not exists hadoop\_books

(TITLE string, AUTHOR string, COUNTRY string, COMPANY string, PRICE int, YEAR int)

row format delimited

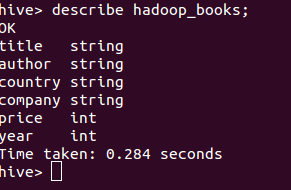
fields terminated by '\t'

lines terminated by '\n'

stored as textfile;

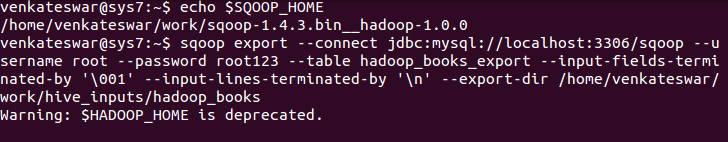
load data inpath 'user/venkateswar/pig/output2.txt' overwrite into table hadoop\_books;

INSERT OVERWRITE DIRECTORY '/home/venkateswar/work/hive inputs/hadoop\_books' SELECT \* FROM hadoop\_books;



**SQOOP:**

sqoop export --connect jdbc:mysql://localhost:3306/sqoop --username root --password root123 --table hadoop\_books\_export --input-fields-terminated-by '\001' --input-lines-terminated-by '\n' --export-dir /home/venkateswar/work/hive\_inputs/hadoop\_books



**MYSQL:**

CREATE TABLE hadoop\_books\_export (TITLE VARCHAR(64), AUTHOR VARCHAR(64), COUNTRY VARCHAR(64), COMPANY VARCHAR(64), PRICE int(32), YEAR int(32));

