XML Parsing:

KPI – 1: Parse xml data into structured format

Input:

```
<configuration>
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>
<property>
<name>hadoop.tmp.dir</name>
<value>/home/hadoop/hdata/hadoop-${user.name}</value>
</property>
</configuration>
```

Output:

```
fs.default.name,hdfs://localhost:9000
hadoop.tmp.dir,/home/hadoop/hdata/hadoop-${user.name}
```

Format of XML:

1. <configuration> tag is the root tag
2. <property> … </property> is a single record

Tasks:

1. Parse XML and convert it into structured data so that processed data can be stored in database
2. From XML data extract name and value tag and store them in comma delimited format
3. Develop a new input format that can read xml data according to specific tags