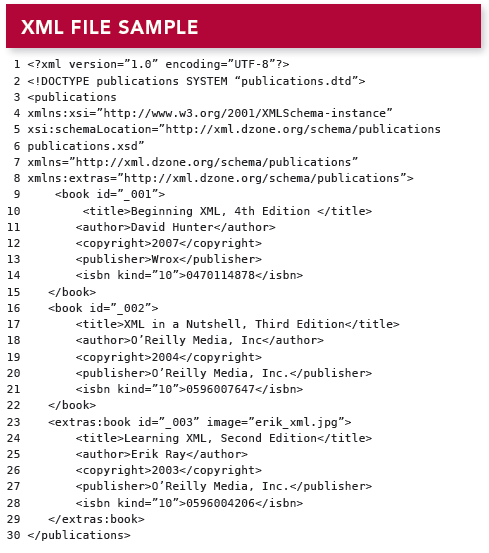
**Documentation for converting XML files to JSON and storing data in Mongo DB**

XML is a general-purpose specification for creating custom mark-up languages. It is classified as an extensible language because it allows its users to define their own elements. Its primary purpose is to help information systems share structured data, particularly via the Internet, and it is used both to encode documents and to serialize data. In the latter context, it is comparable with other text-based serialization languages such as JSON and YAML.



**NoSQL Big Data** systems such as document databases have emerged to address a broad set of applications, and other architectures, such as key-value stores, column family stores, and graph databases are optimized for **more** specific applications.

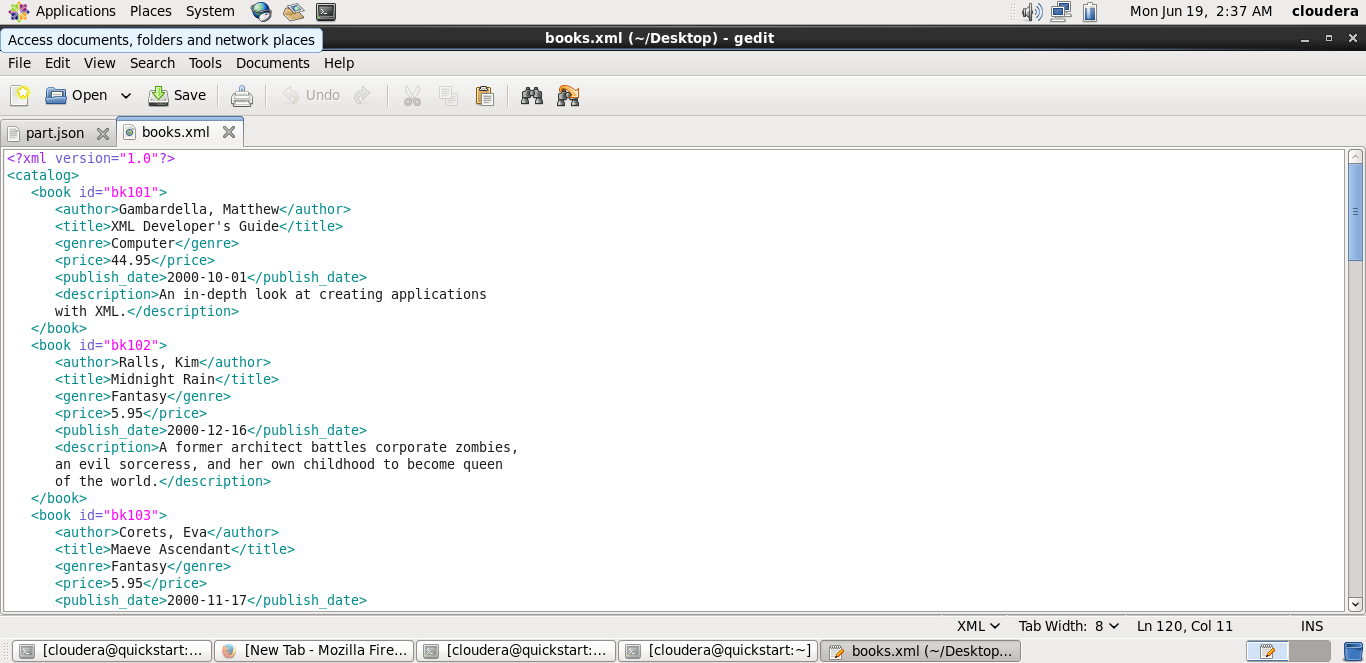
**Step 1:** Convert the XML file to json

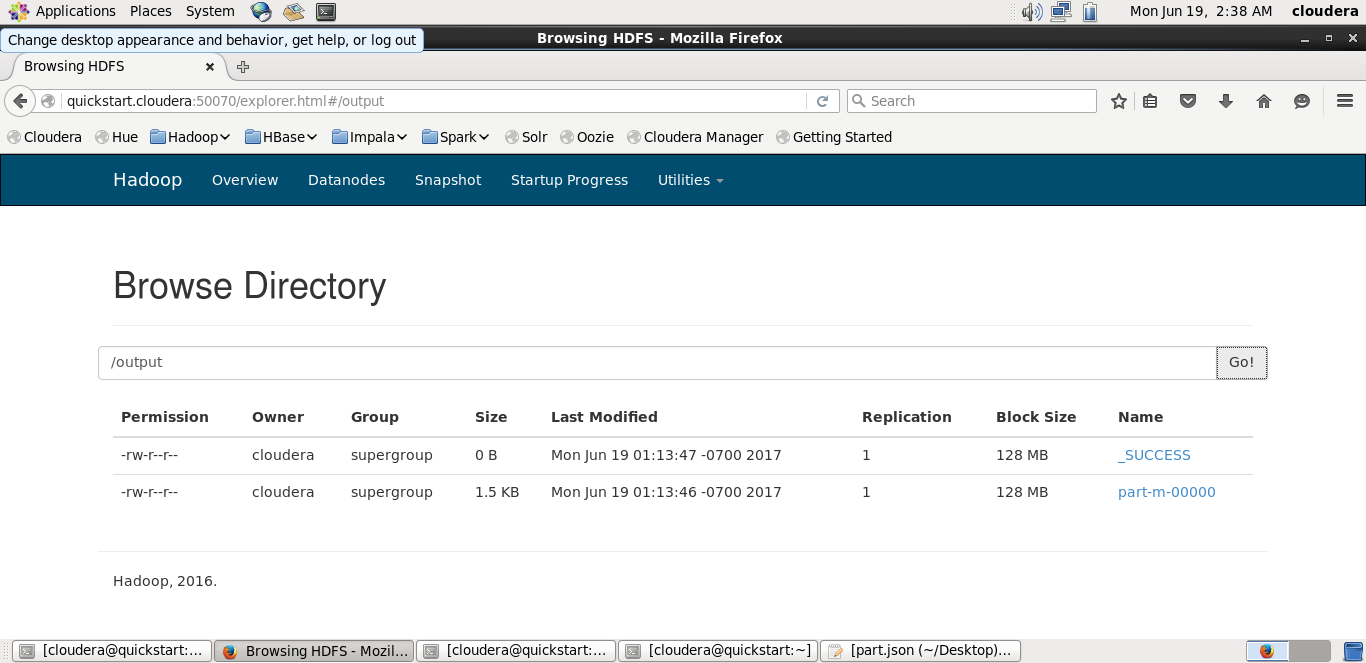
<https://drive.google.com/file/d/1R2qfavD_DvjMeU0hSutJSf5Z7aMhZaz-/view?usp=sharing>

Include necessary Hadoop JARs in the Eclipse IDE and name the JarFile xml\_parsing.jar

Build the JAR file and use it in terminal for processing XML file to JSON using the following command

**Command:** hadoop jar /home/cloudera/ xml\_parsing.jar xml\_parsing input/XMLfile.xml output

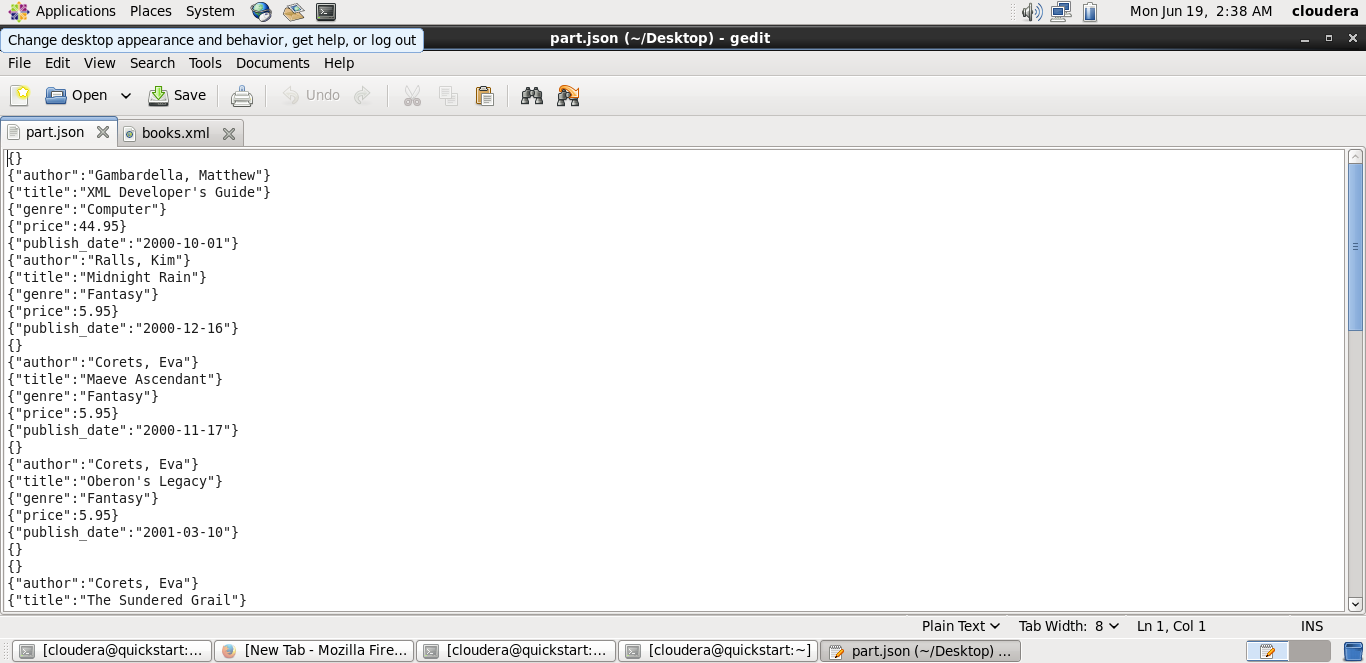




**Step 2:**

Rename the outputfile to part.json

**Command:** mv /home/user/Files/part-m-0000 /home/user/Files/part.json



**Step 3:**

Start MongoDB engine from /var/lib/mongo folder.

**Command:** sudo service mongodb start

**Step 4:**

Use MongoDB engine

**Command:** mongo

**Step 5:**

Create a new database if it doesn't exist, otherwise it will return the existing database.

**Command:** use <DATABASE\_NAME>

Eg. use test

This creates a new Database.

**Step 6:**

Create a collection, which is equivalent to a table in SQL.

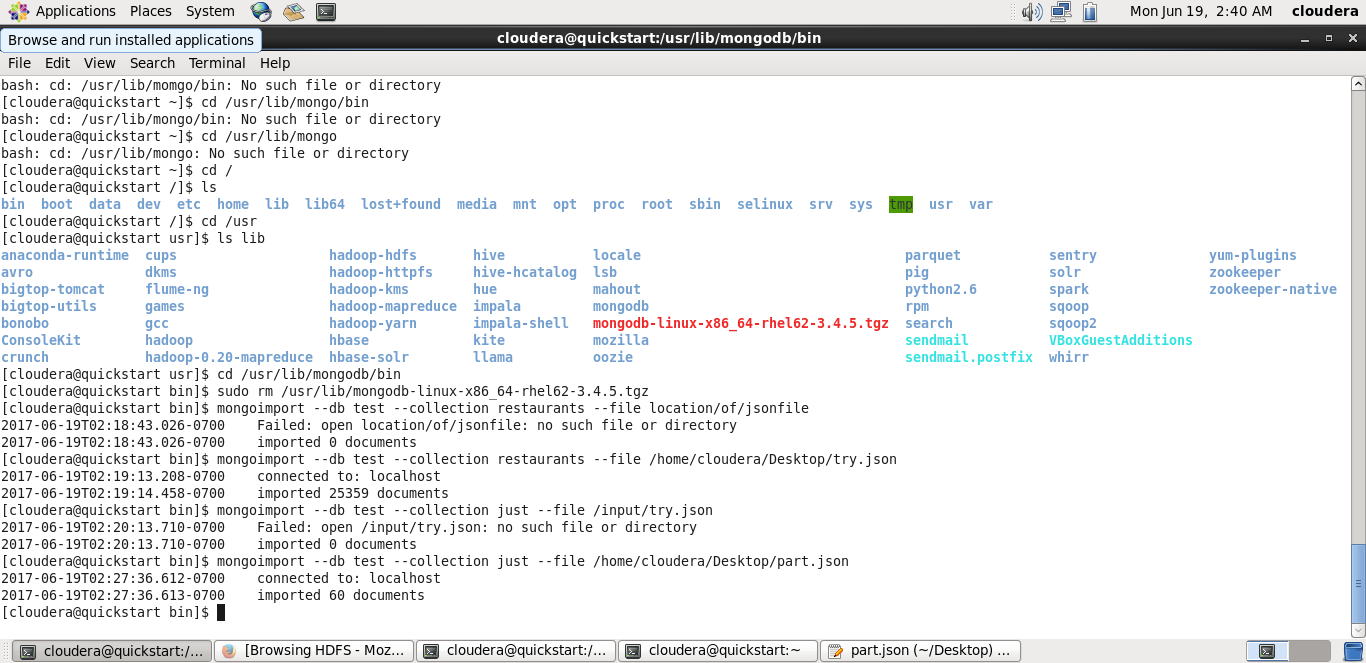
**Command:** db.createCollection("just")

**Step 7:**

Import the file into MongoDB, thereby storing data in MongoDB with Hadoop framework.

**Command:** mongoimport --db dbName --collection collectionName --file fileName.json

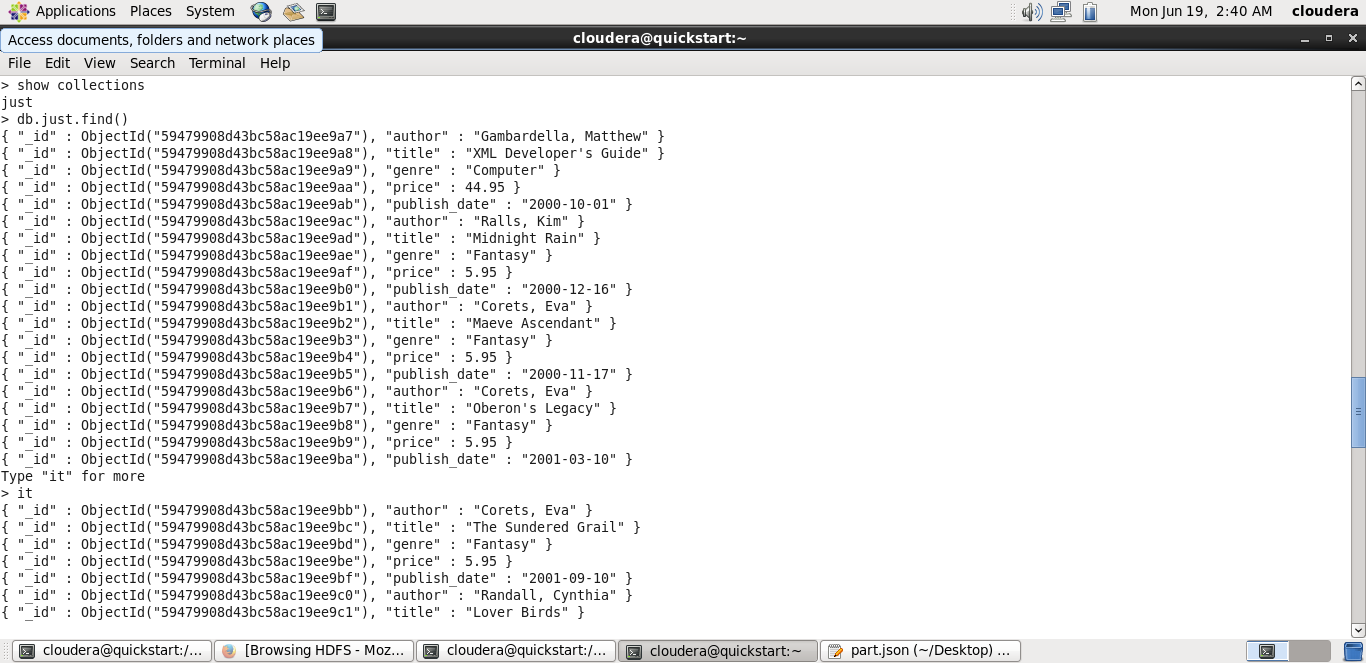
Eg. mongoimport --db test --collection just --file /home/cloudera/Desktop/part.json



**Step 8:**

To query data from MongoDB collection, you need to use MongoDB's **find()**method.

**Command:** db.just.find()



**Step 8:**

To display the results in a formatted way, you can use **pretty()** method.

**Command:** db.just.find().pretty()

**Step 9:**

Applying RDBMS clause equivalents in MongoDB

