1. Create RDD for above json file.

jsonFileRdd=spark.read.json("rdd\_json\_data\_file.json").rdd

2. display the contents in tabular format.

jsonFileRddToDf=jsonFileRdd.toDF()

jsonFileRddToDf.show()

3. Apply filter on id , gender to restrict output.

jsonFileRddToDf.filter("id<3").show()

jsonFileRddToDf.filter("gender='Male' ").show()

4. display contents in array format.

jsonFileRddToDf.collect()

5. classify whole data into 2 files based on gender and save them as json files.

jsonFileRddToDf.filter("gender='Male' ").coalesce(1).write.format('json').save('resultMale.json')

jsonFileRddToDf.filter("gender='Female' ").coalesce(1).write.format('json').save('resultFemale.json')

6. differentiate between show and collect when used with data frames.

show() will show only the content

Collect() will show content with structure/metadata

7. sort the given data based out of gender and store in a file.

jsonFileRddToDf.sort("gender").write.format('json').save('sortedGenderFile.json')

8. convert given json file into a text file.

jsonFileRddToDf.write.format('csv').save('jsonToText2')



