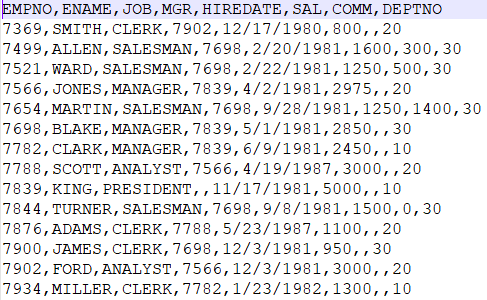
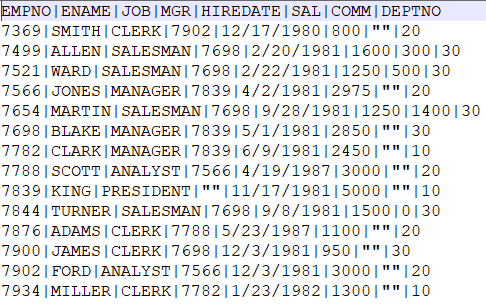
CSV to CSV:

Input: emp.csv



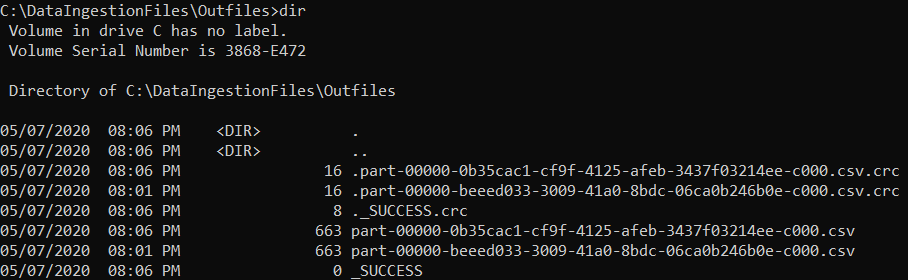
Output: part-00000-beeed033-3009-41a0-8bdc-06ca0b246b0e-c000.csv



Config file: config\_sample\_file2file.json

{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.csv",  
 "archivalPath": " ",  
 "selectCol" : ["\*"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF1",  
 "format": "csv",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]}

Append mode:

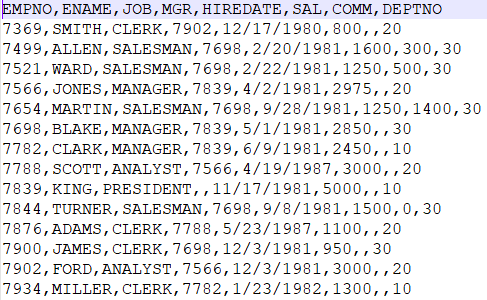


Command Executed:

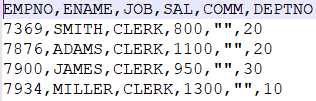
spark-submit C:\DataIngestion\ingestionframework-master\src\main.py C:\DataIngestion\ingestionframework-master\src\resources\config\_sample\_file2file.json

CSV to CSV with selected columns and filter:

Input:



Output:

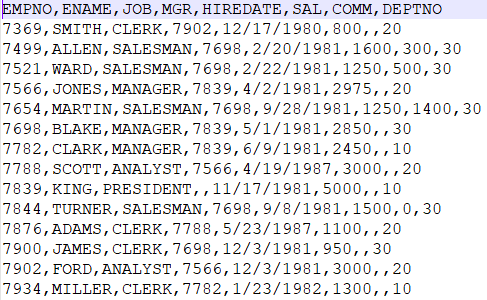


Config file:

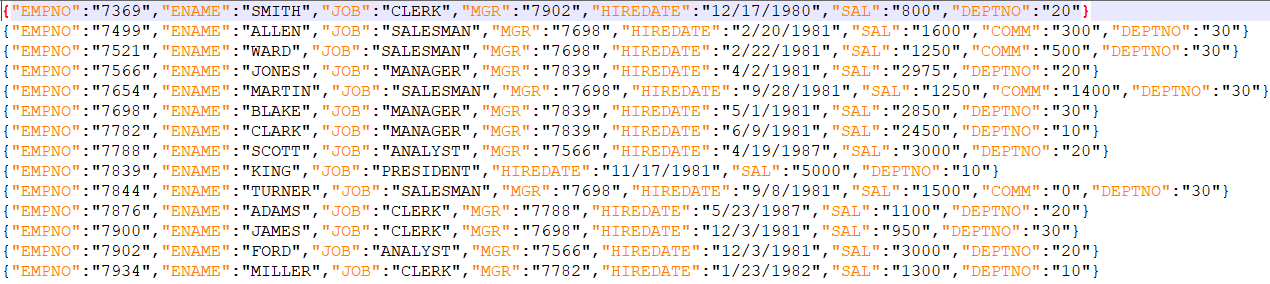
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.csv",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","SAL","COMM","DEPTNO"],  
 "filter" : "job='CLERK'",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF1",  
 "format": "csv",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]}

CSV to JSON:

Input: emp.csv



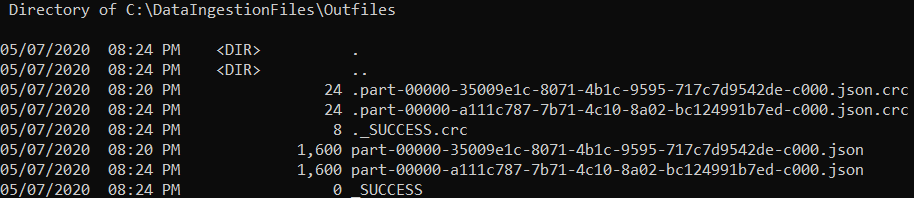
Output: part-00000-35009e1c-8071-4b1c-9595-717c7d9542de-c000.json



Config file: config\_sample\_file2file.json (changes only in fileOupts)

{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.csv",  
 "archivalPath": " ",  
 "selectCol" : ["\*"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF1",  
 "format": "json",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles",  
 "mode": "overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]}

Append mode:

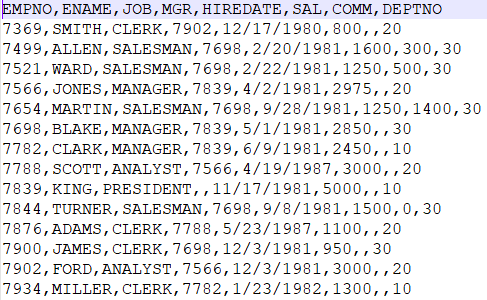


Command Executed:

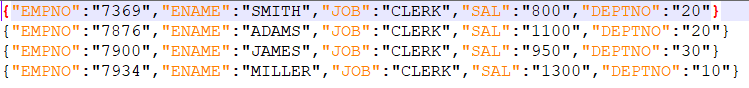
spark-submit C:\DataIngestion\ingestionframework-master\src\main.py C:\DataIngestion\ingestionframework-master\src\resources\config\_sample\_file2file.json

CSV to JSON with selected columns and filter:

Input:



Output:

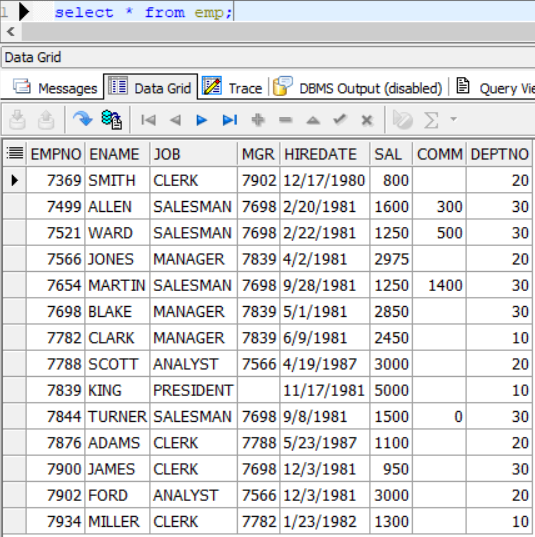


Config file:

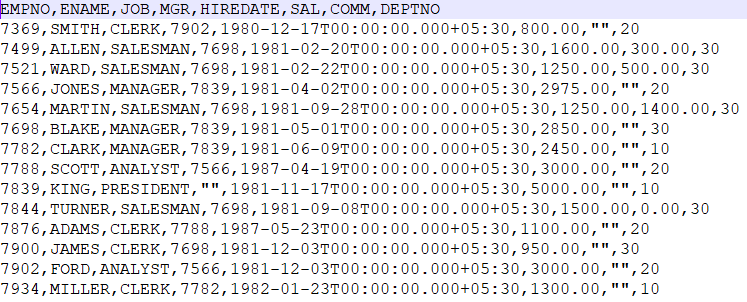
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.csv",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","SAL","COMM","DEPTNO"],  
 "filter" : "job='CLERK'",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF1",  
 "format": "json",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles",  
 "mode": "overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]}

Oracle to CSV:

Input: scott.emp



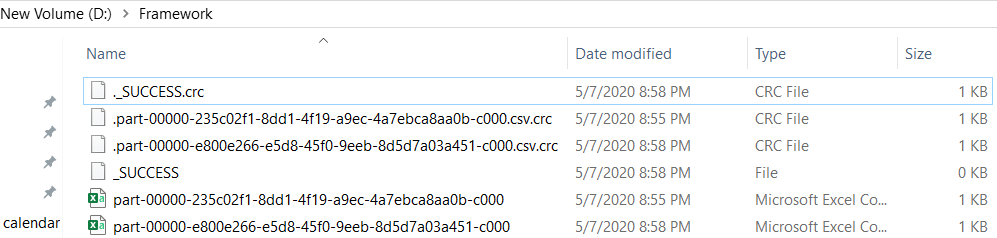
Output: part-00000-235c02f1-8dd1-4f19-a9ec-4a7ebca8aa0b-c000.csv



Config file: config\_sample\_oradb2file.json

{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Read from Database table",  
 "dbmsInputs":[  
 {  
 "dataFrameName": "DF3",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "scott",  
 "password": "tiger",  
 "dbtable": "emp",  
 "selectCol": ["\*"],  
 "filter": "None"  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF3",  
 "format": "csv",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles/",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]  
}

Append mode:

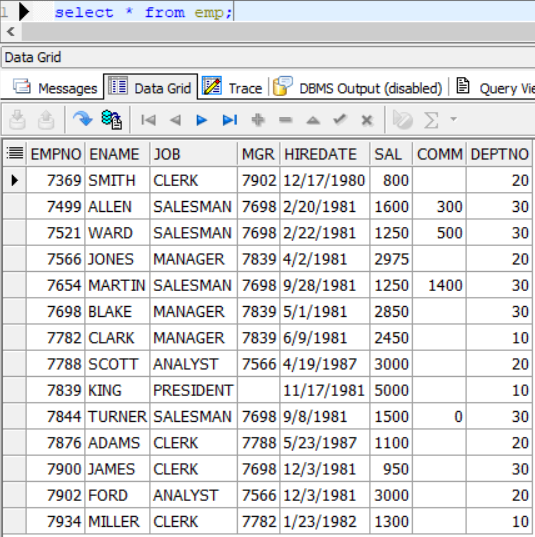


Command Executed:

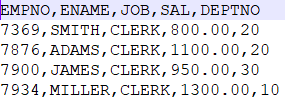
spark-submit --packages io.delta:delta-core\_2.11:0.5.0 --jars C:\DataIngestion\ingestionframework-master\JarDir\ojdbc6.jar C:\DataIngestion\ingestionframework-master\src\main.py C:\DataIngestion\ingestionframework-master\src\resources\config\_sample\_oradb2file.json

ORACLE to CSV with selected columns and filter:

Input:



Output:

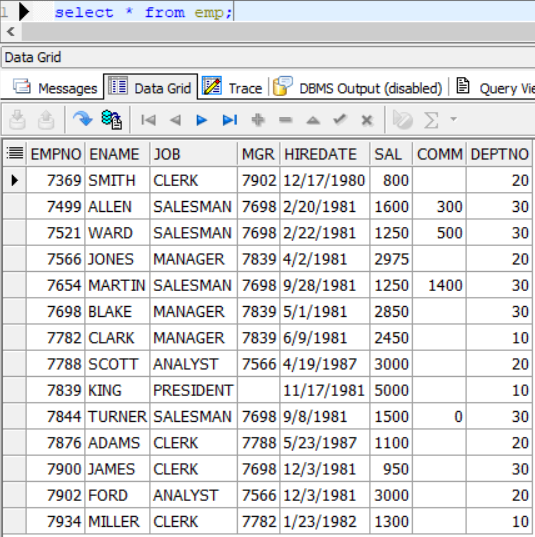


Config file:

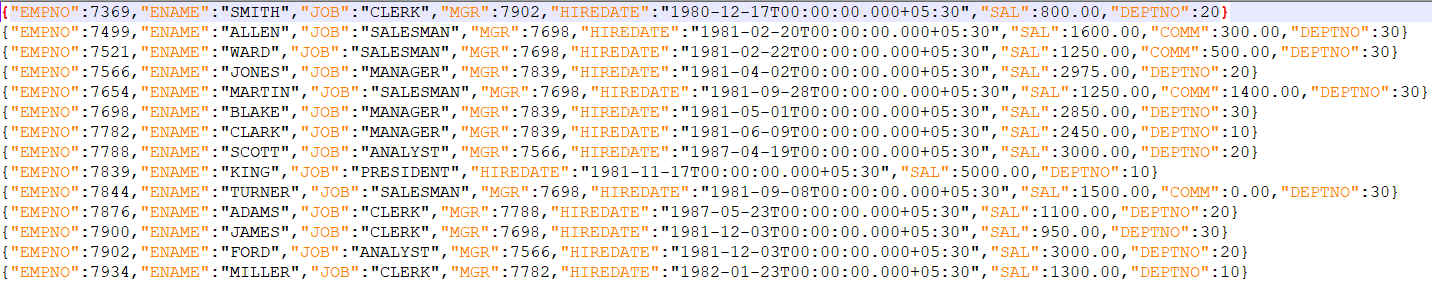
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Read from Database table",  
 "dbmsInputs":[  
 {  
 "dataFrameName": "DF3",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "scott",  
 "password": "tiger",  
 "dbtable": "emp",  
 "selectCol": ["EMPNO","ENAME","JOB","SAL","DEPTNO"],  
 "filter": "job='CLERK'"  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF3",  
 "format": "csv",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles/",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]  
}

Oracle to JSON:

Input: scott.emp



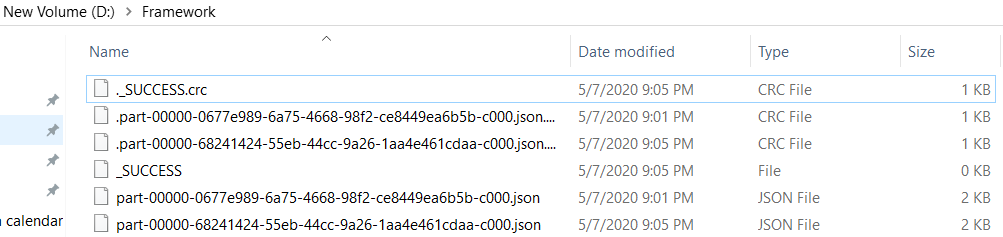
Output: part-00000-0677e989-6a75-4668-98f2-ce8449ea6b5b-c000.json



Config file: config\_sample\_oradb2file.json (changes only in fileOupts)

{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Read from Database table",  
 "dbmsInputs":[  
 {  
 "dataFrameName": "DF3",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "scott",  
 "password": "tiger",  
 "dbtable": "emp",  
 "selectCol": ["\*"],  
 "filter": "None"  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF3",  
 "format": "json",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles/",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]  
}

Append mode:

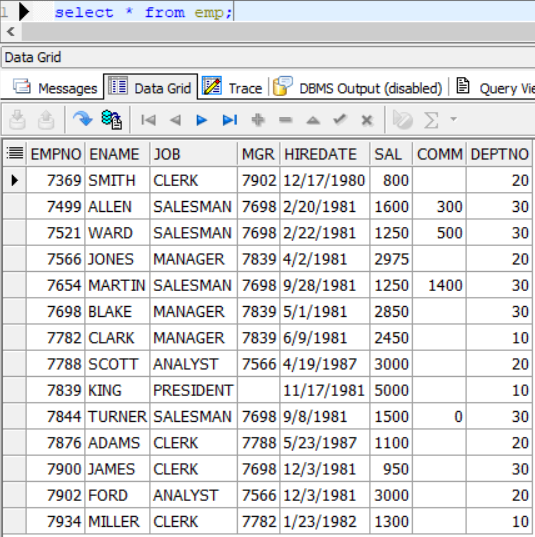


Command Executed:

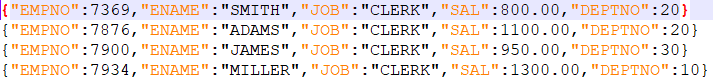
spark-submit --packages io.delta:delta-core\_2.11:0.5.0 --jars C:\DataIngestion\ingestionframework-master\JarDir\ojdbc6.jar C:\DataIngestion\ingestionframework-master\src\main.py C:\DataIngestion\ingestionframework-master\src\resources\config\_sample\_oradb2file.json

ORACLE to JSON with select columns and filter:

Input:



Output:

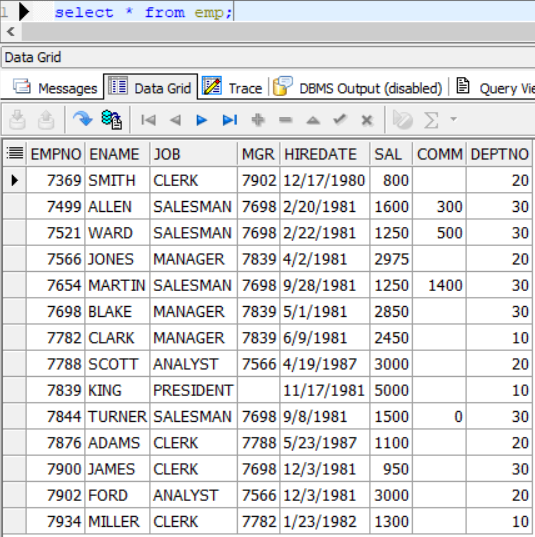


Config file:

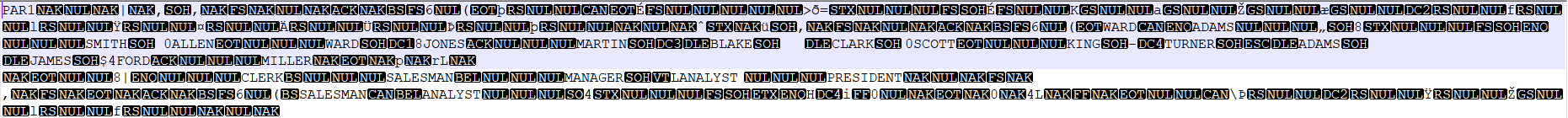
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Read from Database table",  
 "dbmsInputs":[  
 {  
 "dataFrameName": "DF3",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "scott",  
 "password": "tiger",  
 "dbtable": "emp",  
 "selectCol": ["EMPNO","ENAME","JOB","SAL","DEPTNO"],  
 "filter": "job='CLERK'"  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF3",  
 "format": "json",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles/",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]  
}

Oracle to PARQUET:

Input: scott.emp



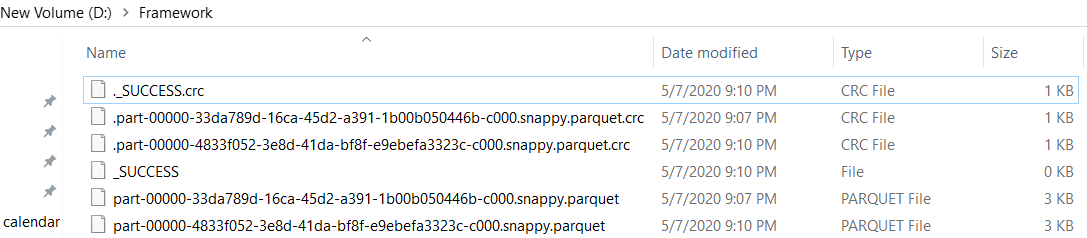
Output: part-00000-33da789d-16ca-45d2-a391-1b00b050446b-c000.snappy.parquet



Config file: config\_sample\_oradb2file.json (changes only in fileOupts)

{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Read from Database table",  
 "dbmsInputs":[  
 {  
 "dataFrameName": "DF3",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "scott",  
 "password": "tiger",  
 "dbtable": "emp",  
 "selectCol": ["EMPNO","ENAME","JOB","SAL","DEPTNO"],  
 "filter": "job='CLERK'"  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF3",  
 "format": "parquet",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles/",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]  
}

Append mode:

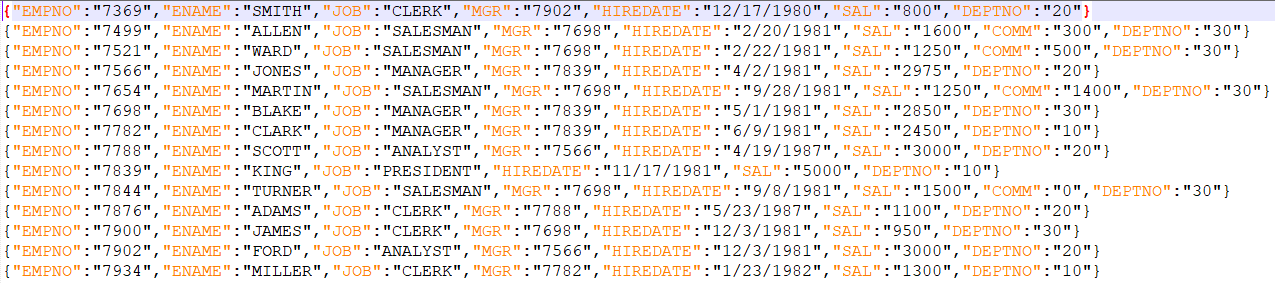


Command Executed:

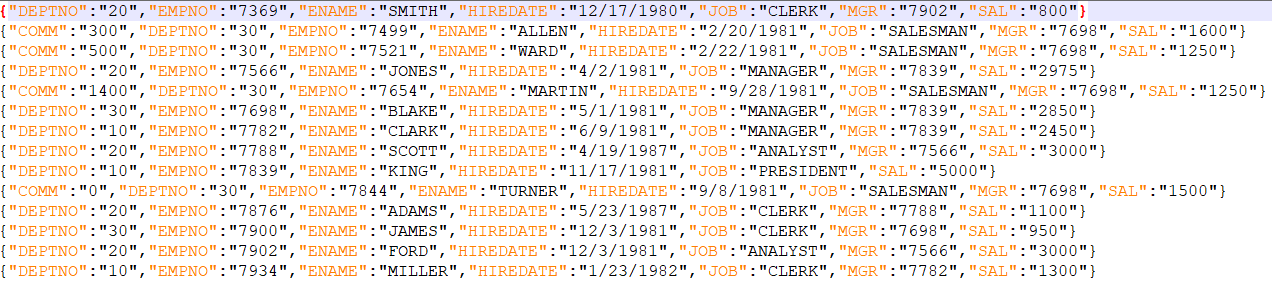
spark-submit --packages io.delta:delta-core\_2.11:0.5.0 --jars C:\DataIngestion\ingestionframework-master\JarDir\ojdbc6.jar C:\DataIngestion\ingestionframework-master\src\main.py C:\DataIngestion\ingestionframework-master\src\resources\config\_sample\_oradb2file.json

JSON to JSON:

Input: emp.json



Output: part-00000-71759c98-21fd-46df-879f-98bc17fcaba5-c000.json

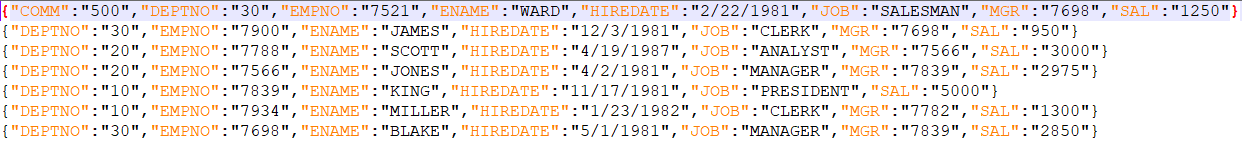


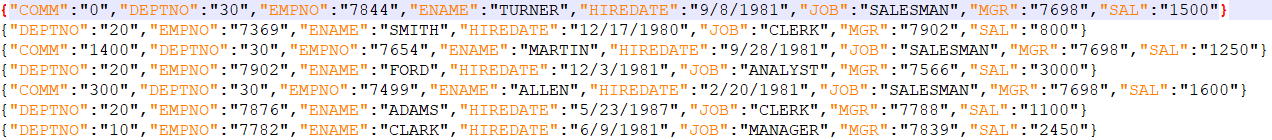
Config file: config\_sample\_oradb2file.json

{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "selectCol" : ["\*"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "json",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF1",  
 "format": "json",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles",  
 "mode": "overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]}

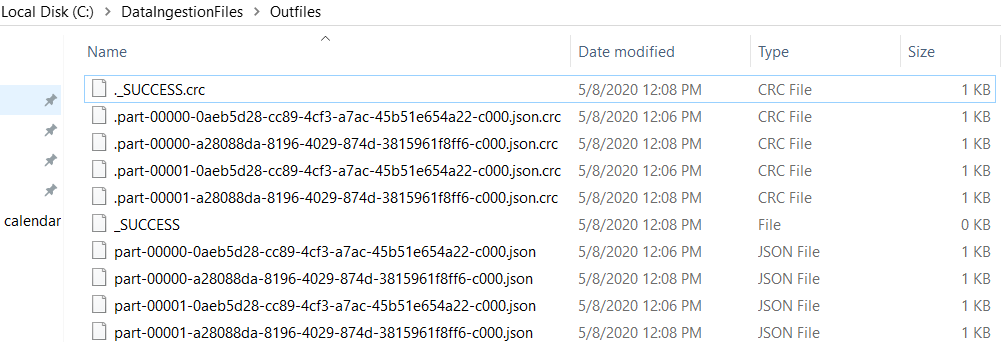
No of partitions:2

"fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "numPartitions": 2,  
 "dataFormat":{  
 "dataFormat": "JSON",  
 "multiLine": false





Append mode:

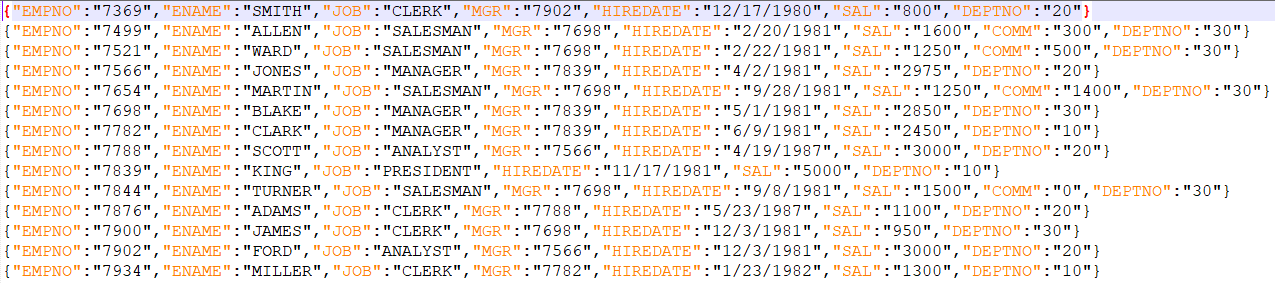


Command Executed:

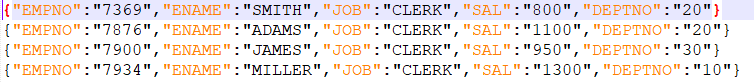
spark-submit C:\DataIngestion\ingestionframework-master\src\main.py C:\DataIngestion\ingestionframework-master\src\resources\config\_sample\_file2file.json

JSON to JSON with selected columns and filter:

Input:



Output:



Config file:

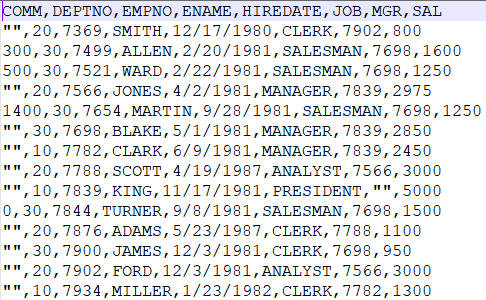
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","SAL","COMM","DEPTNO"],  
 "filter" : "job='CLERK'",  
 "dataFormat":{  
 "dataFormat": "json",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF1",  
 "format": "json",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles",  
 "mode": "overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]}

JSON to CSV:

Input: emp.json



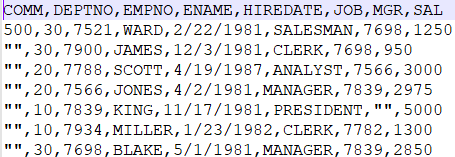
Output: part-00000-5be89f64-e558-4145-b124-a2a31b122413-c000.csv

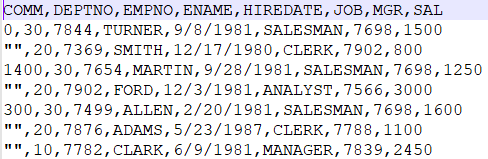


Config file: config\_sample\_oradb2file.json (changes only in fileOupts)

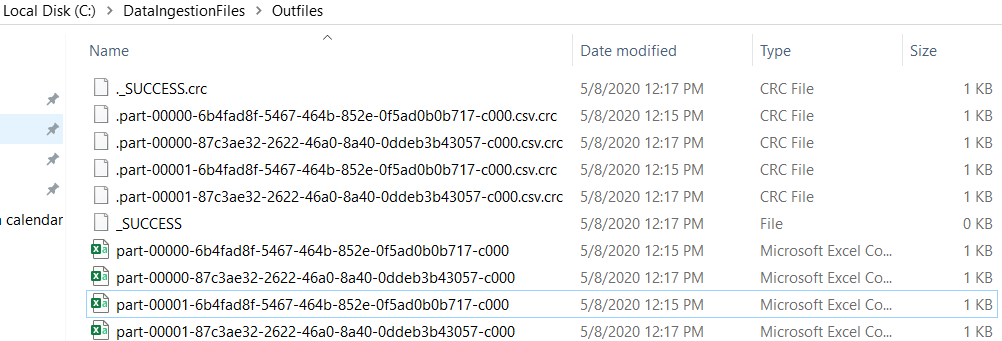
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "selectCol" : ["\*"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "json",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF1",  
 "format": "csv",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles",  
 "mode": "overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]}

No of partitions: 2





Append mode:



Command Executed:

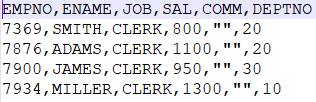
spark-submit C:\DataIngestion\ingestionframework-master\src\main.py C:\DataIngestion\ingestionframework-master\src\resources\config\_sample\_file2file.json

JSON to CSV with selected columns and filter:

Input:



Output:

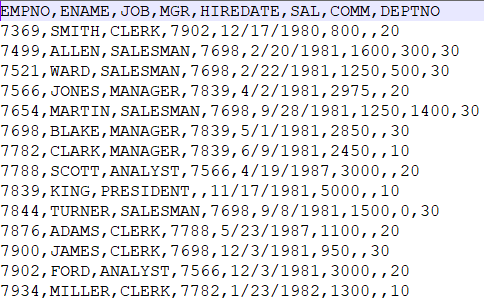


Config file:

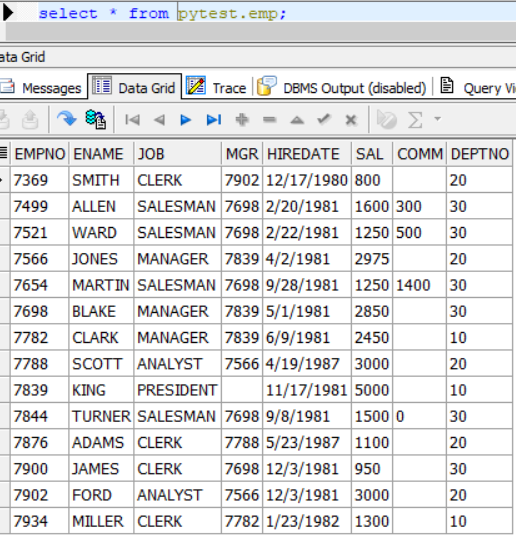
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","SAL","COMM","DEPTNO"],  
 "filter" : "job='CLERK'",  
 "dataFormat":{  
 "dataFormat": "json",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF1",  
 "format": "csv",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles",  
 "mode": "overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]}

CSV to ORACLE:

Input: emp.csv



Output: pytest.emp



Config file:

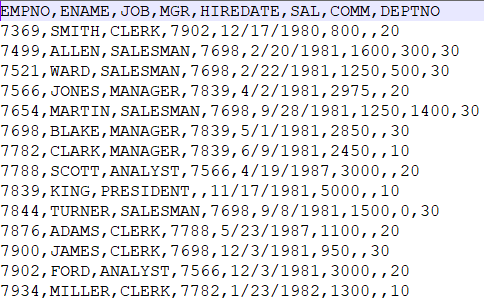
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.csv",  
 "archivalPath": " ",  
 "selectCol" : ["\*"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "dbmsOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "pytest",  
 "password": "pytest",  
 "dbtable": "emp",  
 "mode": "overwrite"  
 }  
 ]  
 }  
]  
}

Command Executed:

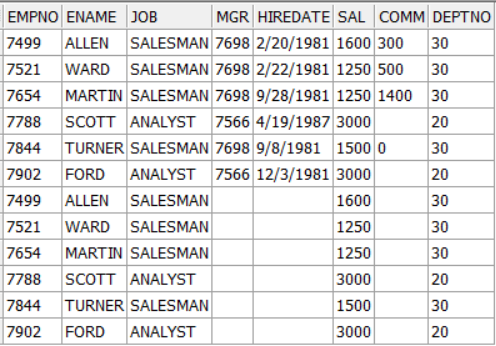
spark-submit --packages io.delta:delta-core\_2.11:0.5.0 --jars C:\DataIngestionFrameWorkNew\ingestionframework-master\JarDir\ojdbc6.jar C:\DataIngestionFrameWorkNew\ingestionframework-master\src\main.py C:\DataIngestionFrameWorkNew\ingestionframework-master\src\resources\config\_sample\_file2oradb.json

CSV to ORACLE with selected columns and filter:

Input:



Output:



Config file:

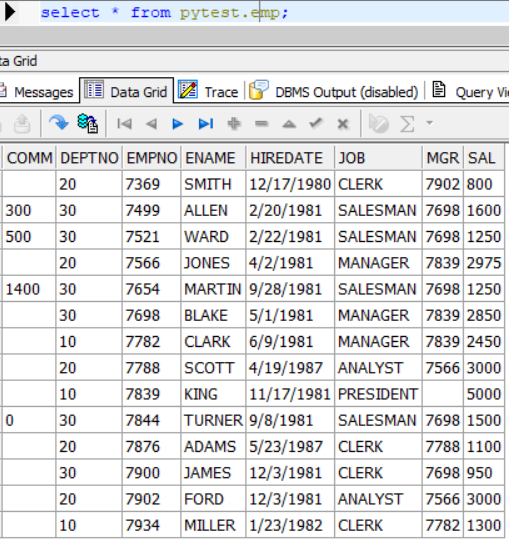
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.csv",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","SAL","DEPTNO"],  
 "filter" : "JOB IN ('ANALYST','SALESMAN')",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "dbmsOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "pytest",  
 "password": "pytest",  
 "dbtable": "emp",  
 "mode": "append"  
 }  
 ]  
 }  
]  
}

JSON to ORACLE:

Input: emp.json



Output: pytest.emp



Config file:

{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","MGR","HIREDATE","SAL","COMM","DEPTNO"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "json",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "dbmsOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "pytest",  
 "password": "pytest",  
 "dbtable": "emp",  
 "mode": "overwrite"  
 }  
 ]  
 }  
]  
}

Command executed:

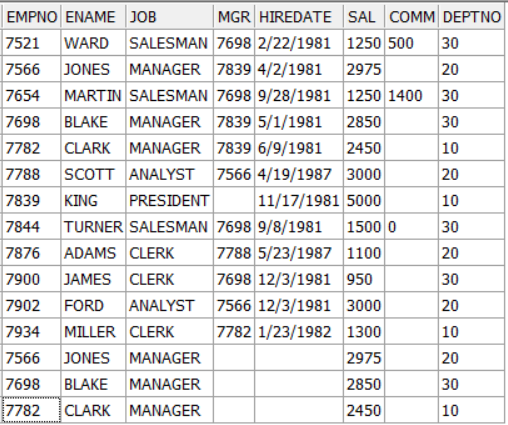
spark-submit --packages io.delta:delta-core\_2.11:0.5.0 --jars C:\DataIngestionFrameWorkNew\ingestionframework-master\JarDir\ojdbc6.jar C:\DataIngestionFrameWorkNew\ingestionframework-master\src\main.py C:\DataIngestionFrameWorkNew\ingestionframework-master\src\resources\config\_sample\_file2oradb.json

JSON to ORACLE with selected columns and filter:

Input:



Output:

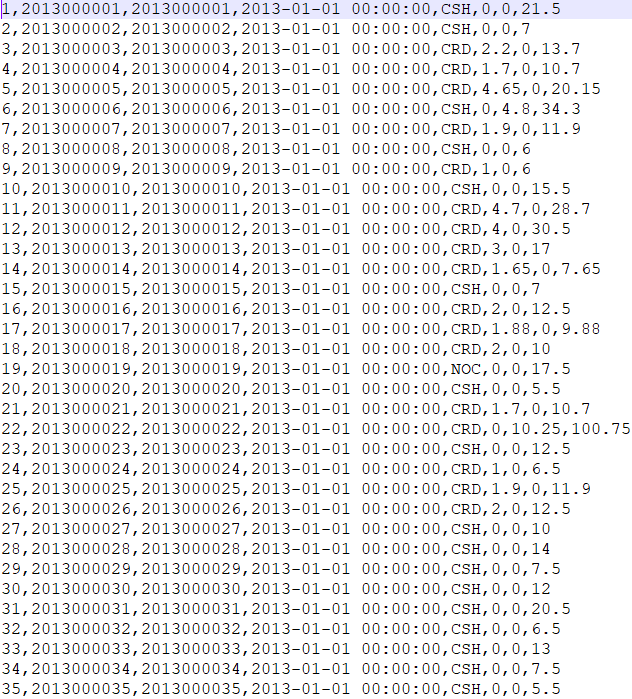


Config file:

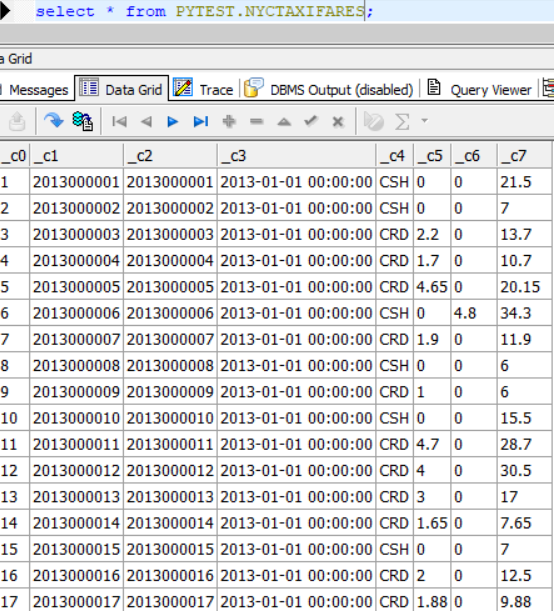
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","SAL","DEPTNO"],  
 "filter" : "job='MANAGER'",  
 "dataFormat":{  
 "dataFormat": "json",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "dbmsOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "pytest",  
 "password": "pytest",  
 "dbtable": "emp",  
 "mode": "append"  
 }  
 ]  
 }  
]  
}

CSV to ORACLE:

Input: nycTaxiFares.csv (Record count: 15 Lakhs)



Output: PYTEST.NYCTAXIFARES



Config file:

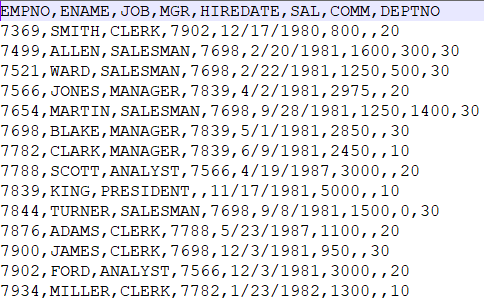
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/nycTaxiFares.csv",  
 "archivalPath": " ",  
 "selectCol" : ["\*"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "dbmsOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "url": "jdbc:oracle:thin:@localhost:1522:orcl1",  
 "driver": "oracle.jdbc.driver.OracleDriver",  
 "user": "pytest",  
 "password": "pytest",  
 "dbtable": "emp",  
 "mode": "overwrite"  
 }  
 ]  
 }  
]  
}

Command executed:

spark-submit --packages io.delta:delta-core\_2.11:0.5.0 --jars C:\DataIngestionFrameWorkNew\ingestionframework-master\JarDir\ojdbc6.jar C:\DataIngestionFrameWorkNew\ingestionframework-master\src\main.py C:\DataIngestionFrameWorkNew\ingestionframework-master\src\resources\config\_sample\_file2oradb.json

CSV to MongoDB:

Input: emp.csv



Output: myNewDB.employee



Config file:

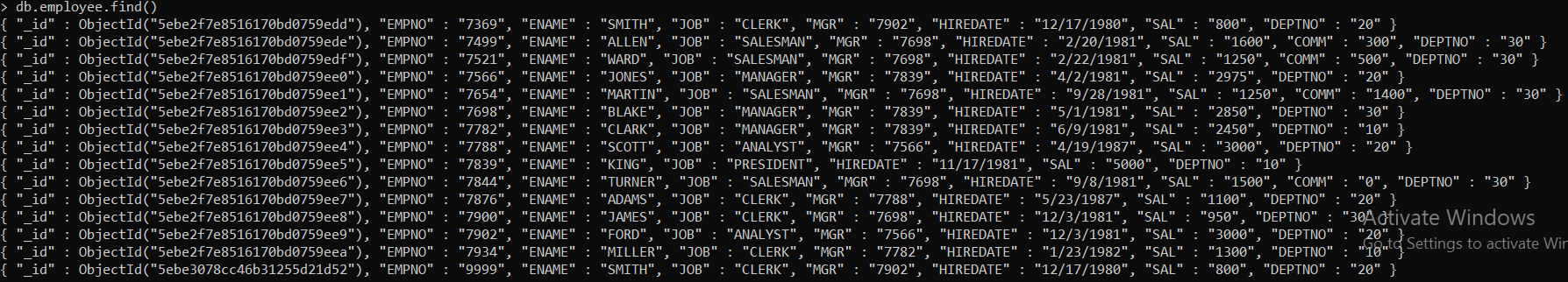
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.csv",  
 "archivalPath": " ",  
 "selectCol" : ["\*"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "mongoOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "uri": "mongodb://localhost:27017",  
 "database": "myNewDB",  
 "collection": "employee",  
 "mode": "overwrite"  
 }  
 ]  
 }  
]  
}

Append mode:

Input:



Output:

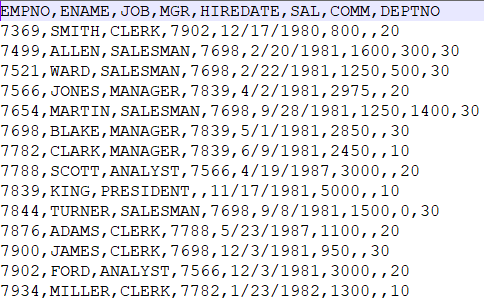


Command executed:

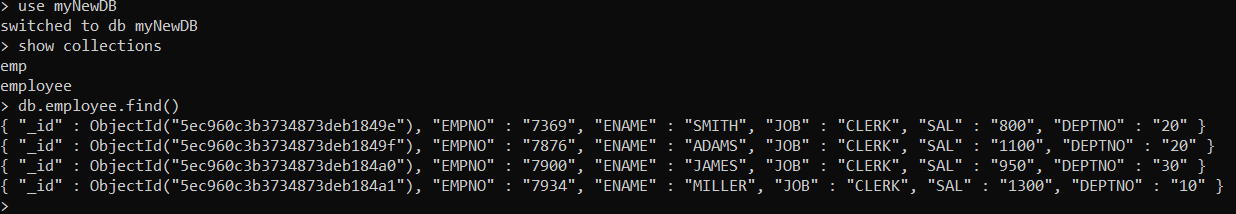
spark-submit --packages org.mongodb.spark:mongo-spark-connector\_2.11:2.4.1 C:\DataIngestionFrameWorkNew\ingestionframework-master\src\main.py C:\DataIngestionFrameWorkNew\ingestionframework-master\src\resources\config\_sample\_file2mongodb.json

CSV to MONGODB with selected columns and filter:

Input:



Output:

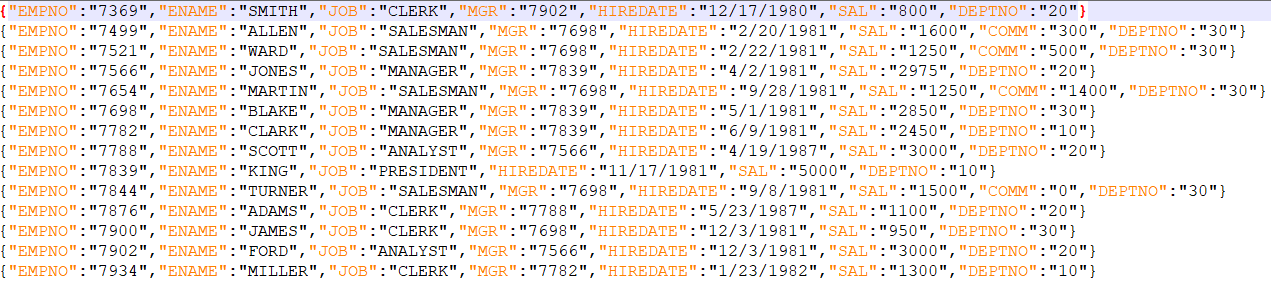


Config file:

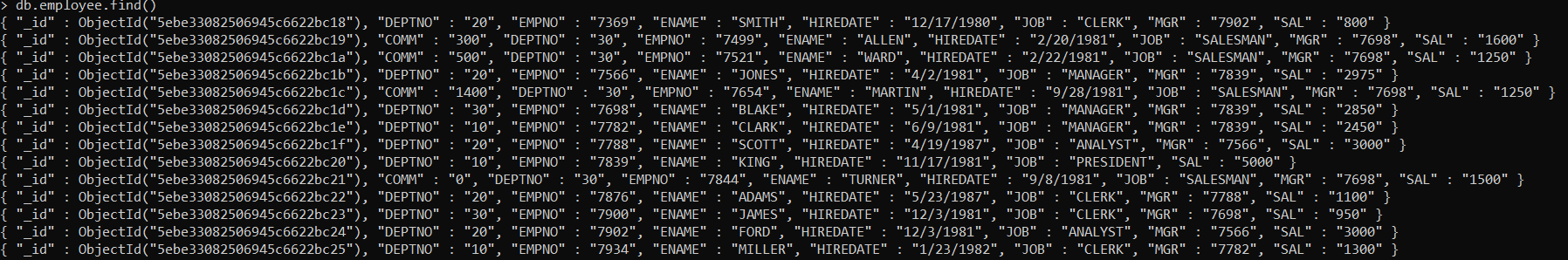
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.csv",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","SAL","DEPTNO"],  
 "filter" : "job='CLERK'",  
 "dataFormat":{  
 "dataFormat": "csv",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "mongoOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "uri": "mongodb://localhost:27017",  
 "database": "myNewDB",  
 "collection": "employee",  
 "mode": "overwrite"  
 }  
 ]  
 }  
]  
}

JSON to MongoDB:

Input: emp.json



Output: myNewDB.employee



Config file:

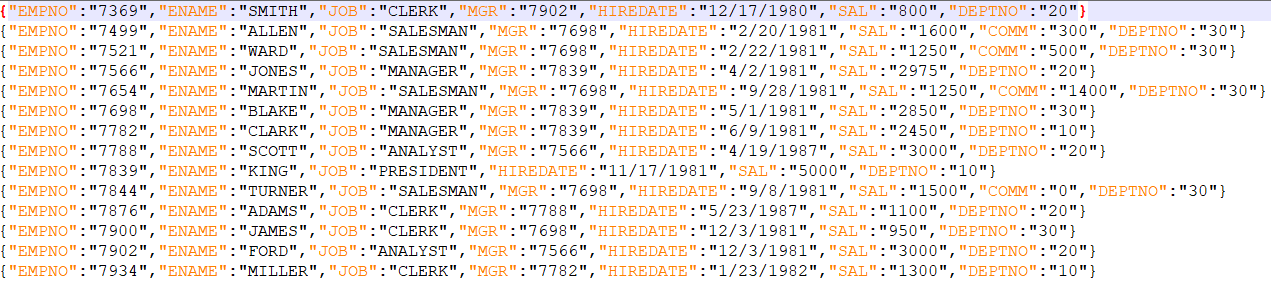
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "selectCol" : ["\*"],  
 "filter" : "None",  
 "dataFormat":{  
 "dataFormat": "json",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "mongoOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "uri": "mongodb://localhost:27017",  
 "database": "myNewDB",  
 "collection": "employee",  
 "mode": "overwrite"  
 }  
 ]  
 }  
]  
}

Command executed:

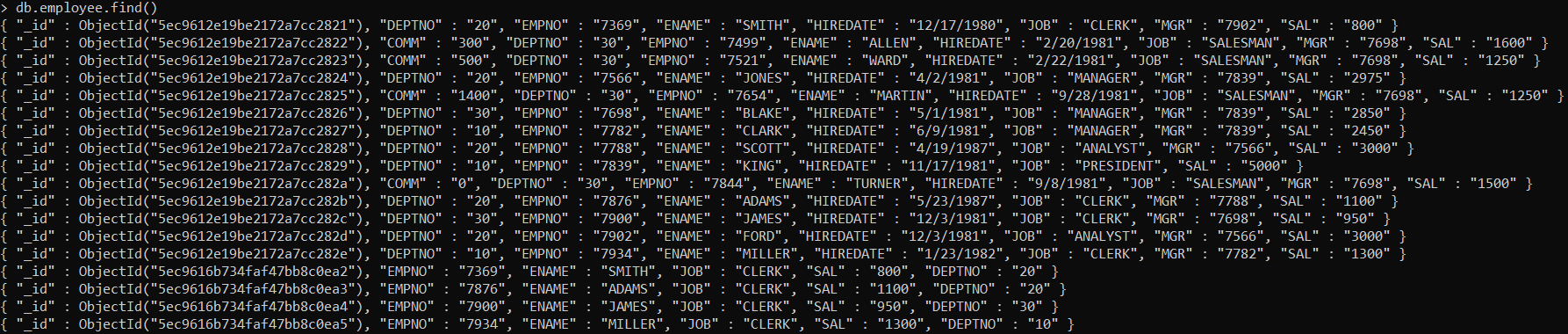
spark-submit --packages org.mongodb.spark:mongo-spark-connector\_2.11:2.4.1 C:\DataIngestionFrameWorkNew\ingestionframework-master\src\main.py C:\DataIngestionFrameWorkNew\ingestionframework-master\src\resources\config\_sample\_file2mongodb.json

JSON to MONGODB with selected columns and filter:

Input:



Output:

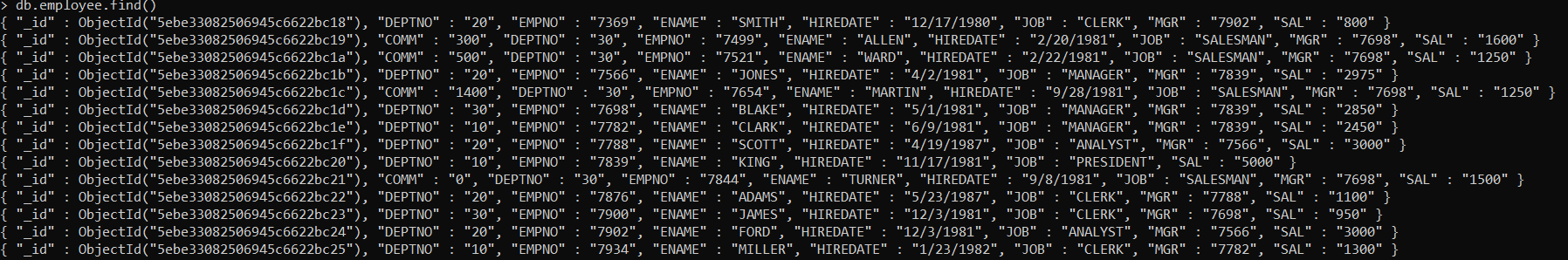


Config file:

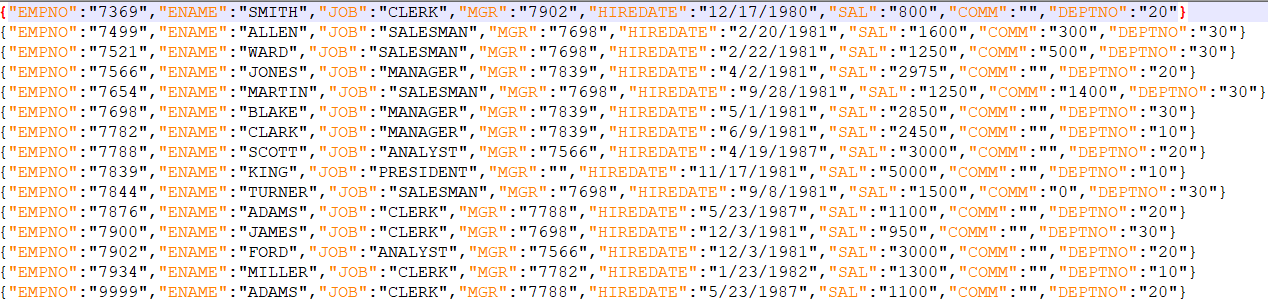
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Write into MONGO Database table",  
 "fileInputs":[  
 {  
 "dataFrameName": "DF1",  
 "fileInPath": "C:/DataIngestionFiles/Infiles/emp.json",  
 "archivalPath": " ",  
 "selectCol" : ["EMPNO","ENAME","JOB","SAL","DEPTNO"],  
 "filter" : "job='CLERK'",  
 "dataFormat":{  
 "dataFormat": "json",  
 "header": true,  
 "delimiter": ",",  
 "inferSchema": false,  
 "multiLine": false  
 }  
 }  
 ],  
 "mongoOutputs":[  
 {  
 "dataFrameName": "DF1",  
 "uri": "mongodb://localhost:27017",  
 "database": "myNewDB",  
 "collection": "employee",  
 "mode": "append"  
 }  
 ]  
 }  
]  
}

MongoDB to JSON:

Input: myNewDB.employee



Output: employee.json



Config file:

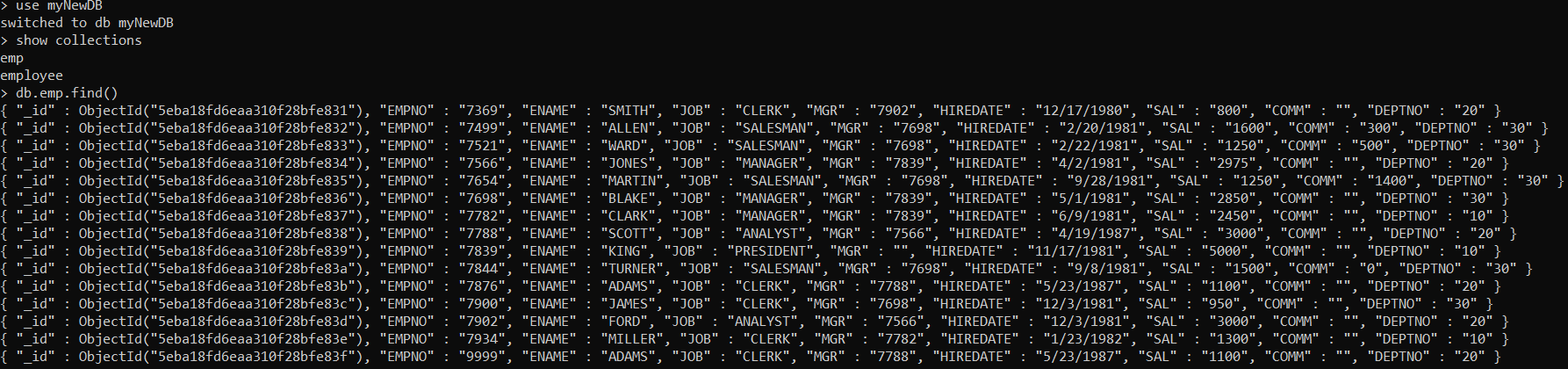
{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Read from MONGO Database table",  
 "mongoInputs":[  
 {  
 "dataFrameName": "DF3",  
 "uri": "mongodb://localhost:27017",  
 "database": "myNewDB",  
 "collection": "emp",  
 "selectCol" : ["\*"],  
 "filter" : "None"  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF3",  
 "format": "json",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles/",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]  
}

Command executed:

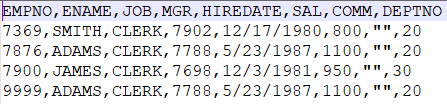
spark-submit --packages org.mongodb.spark:mongo-spark-connector\_2.11:2.4.1 C:\DataIngestionFrameWorkNew\ingestionframework-master\src\main.py C:\DataIngestionFrameWorkNew\ingestionframework-master\src\resources\config\_sample\_mongodb2file.json

MongoDB to CSV:

Input: myNewDB.emp



Output:



Config file:

{  
"mode": "batch",  
"jobs": [  
 {  
 "name": "job2",  
 "description": " Read from MONGO Database table",  
 "mongoInputs":[  
 {  
 "dataFrameName": "DF3",  
 "uri": "mongodb://localhost:27017",  
 "database": "myNewDB",  
 "collection": "emp",  
 "selectCol" : ["EMPNO","ENAME","JOB","MGR","HIREDATE","SAL","COMM","DEPTNO"],  
 "filter" : "job='CLERK' AND sal<=1100"  
 }  
 ],  
 "fileOutputs": [  
 {  
 "dataFrameName": "DF3",  
 "format": "csv",  
 "fileOutPath": "C:/DataIngestionFiles/Outfiles/",  
 "mode": "Overwrite",  
 "partitions": 2,  
 "header": true,  
 "delimiter": ","  
 }  
 ]  
 }  
]  
}

Command executed:

spark-submit --packages org.mongodb.spark:mongo-spark-connector\_2.11:2.4.1 C:\DataIngestionFrameWorkNew\ingestionframework-master\src\main.py C:\DataIngestionFrameWorkNew\ingestionframework-master\src\resources\config\_sample\_mongodb2file.json