Import a JSON file from the command line and apply actions with the data present in the JSON file

Open PowerShell as administrator and start hadoop services using the following commands:

- start-dfs.cmd
- start-yarn.cmd

```
PS C:\WINDOWS\system32> start-dfs.cmd
PS C:\WINDOWS\system32> start-yarn.cmd
starting yarn daemons
PS C:\WINDOWS\system32> jps
19344 ResourceManager
36800 DataNode
35268 Jps
35240 NodeManager
35864 NameNode
```

Add the .json file to HDFS using the command:

hdfs dfs -put C:/Semester7/DataAnalytics/Lab/Ex6/employee.json /user/hive/warehouse/employee_json/

```
C:\Windows\System32>hdfs dfs -put C:/Semester7/DataAnalytics/Lab/Ex6/employee.json /user/hive/warehouse/employee_json/
```

File contents

```
{"id": 1, "name": "John Doe", "age": 30, "salary": 50000}
{"id": 2, "name": "Jane Smith", "age": 25, "salary": 60000}
{"id": 3, "name": "Alice Johnson", "age": 28, "salary": 55000}
{"id": 4, "name": "Bob Brown", "age": 35, "salary": 70000}
{"id": 5, "name": "Charlie Davis", "age": 40, "salary": 80000}
{"id": 6, "name": "Eve White", "age": 22, "salary": 48000}
{"id": 7, "name": "Frank Black", "age": 32, "salary": 65000}
{"id": 8, "name": "Grace Green", "age": 27, "salary": 52000}
```

Run the following command to open Derby:

StartNetworkServer -h 0.0.0.0

```
PS C:\WINDOWS\system32> StartNetworkServer -h 0.0.0.0
Wed Sep 11 19:34:46 IST 2024 : Security manager installed using the Basic server security policy.
Wed Sep 11 19:34:46 IST 2024 : Apache Derby Network Server - 10.14.2.0 - (1828579) started and ready to accept connectio
ns on port 1527
```

```
C:\Windows\System32>jps
35348 NetworkServerControl
5988 ResourceManager
13176 Jps
39848 NodeManager
40728 DataNode
26940 NameNode
7532 RunJar
```

Go to Apache Hive's bin location with cd command and run the following command:

hive --service schematool -dbType derby –initSchema

```
PS C:\apache-hive-3.1.3-bin\bin> hive --service schematool -dbType derby -initSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop-3.3.6/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/apache-hive-3.1.3-bin/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
2024-09-11 19:47:10,283 INFO conf.HiveConf: Found configuration file null
2024-09-11 19:47:10,502 INFO tools.HiveSchemaHelper: Metastore connection URL:   jdbc:derby:;databaseName=metastore_db;create=true
Metastore connection URL:        jdbc:derby:;databaseName=metastore_db;create=true
2024-09-11 19:47:10,502 INFO tools.HiveSchemaHelper: Metastore Connection Driver :
        org.apache.derby.jdbc.EmbeddedDriver
```

```
Initialization script completed
schemaTool completed
```

Open Hive shell by typing:

hive

This will open the hive CLI.

Create a database

CREATE DATABASE IF NOT EXISTS employee_json;

SHOW DATABASES;

Switch to the new database:

use employee_json;

Create a Table using the following command:

```
CREATE TABLE employees (
        id INT,
        name STRING,
        age INT,
        salary DOUBLE
)
ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'
STORED AS TEXTFILE
LOCATION '/user/hive/warehouse/employee_json/';
```

```
hive> create table employees(
    > id INT,
    > name STRING,
    > age INT,
    > salary DOUBLE
    > )
    > ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'
    > STORED AS TEXTFILE
    > LOCATION '/user/hive/warehouse/employee_json/';
2024-09-11 23:26:57,813 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:26:57,814 INFO session.SessionState: Updating thread name to 15916fd5-0cc5-48bf-9cb8-a338d3b8713c main
2024-09-11 23:26:57,816 INFO ql.Driver: Compiling command(queryId=prath_20240911232657_a43d062a-1524-486a-b1de-f111c434d683): create table employees(
id INT,
name STRING,
age INT,
salary DOUBLE
)
ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'
STORED AS TEXTFILE
LOCATION '/user/hive/warehouse/employee_json/'
2024-09-11 23:26:57,839 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:26:57,839 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-09-11 23:26:57,840 INFO parse.CalcitePlanner: Creating table employee_json.employees position=13
2024-09-11 23:26:57,842 INFO metastore.HiveMetaStore: 0: get_database: @hive#employee_json
2024-09-11 23:26:57,842 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr
       cmd=get_database: @hive#employee_json
2024-09-11 23:26:58,226 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-11 23:26:58,227 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null)
2024-09-11 23:26:58,228 INFO ql.Driver: Completed compiling command(queryId=prath_20240911232657_a43d062a-1524-486a-b1de-f111c434d683); Time taken: 0.412 seconds
2024-09-11 23:26:58,228 INFO reexec.ReExecDriver: Execution #1 of query
2024-09-11 23:26:58,229 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:26:58,230 INFO ql.Driver: Executing command(queryId=prath_20240911232657_a43d062a-1524-486a-b1de-f111c434d683): create table employees(
id INT,
name STRING,
age INT,
salary DOUBLE
)
ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'
STORED AS TEXTFILE
LOCATION '/user/hive/warehouse/employee_json/'
2024-09-11 23:26:58,232 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-09-11 23:26:58,247 INFO metastore.HiveMetaStore: 0: create_table: Table(tableName:employees, dbName:employee_json, owner:prath, createTime:1726077418, lastAccessTime:0
, retention:0, sd:StorageDescriptor(cols:[FieldSchema(name:id, type:int, comment:null), FieldSchema(name:name, type:string, comment:null), FieldSchema(name:age, type:int, c
omment:null), FieldSchema(name:salary, type:double, comment:null)], location:hdfs://localhost:9000/user/hive/warehouse/employee_json, inputFormat:org.apache.hadoop.mapred.T
extInputFormat, outputFormat:org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat, compressed:false, numBuckets:-1, serdeInfo:SerDeInfo(name:null, serializationLib:or
```

SHOW TABLES;

```
hive> show tables;
2024-09-11 23:11:49,461 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:11:49,462 INFO session.SessionState: Updating thread name to 15916fd5-0cc5-48bf-9cb8-a338d3b8713c main
2024-09-11 23:11:49,465 INFO ql.Driver: Compiling command(queryId=prath_20240911231149_188eb98a-6506-4a41-8ed9-79dee35be0d8): show tables
2024-09-11 23:11:49,486 INFO metastore.HiveMetaStoreClient: Mestastore configuration metastore.filter.hook changed from org.apache.hadoop.hive.metastore.DefaultMetaStoreFil
terHookImpl to org.apache.hadoop.hive.ql.security.authorization.plugin.AuthorizationMetaStoreFilterHook
2024-09-11 23:11:49,487 INFO metastore.HiveMetaStore: 0: Cleaning up thread local RawStore...
2024-09-11 23:11:49,488 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=Cleaning up thread local RawStore...
2024-09-11 23:11:49,489 INFO metastore.HiveMetaStore: 0: Done cleaning up thread local RawStore
2024-09-11 23:11:49,489 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=Done cleaning up thread local RawStore
2024-09-11 23:11:49,489 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:11:49,495 INFO metastore.HiveMetaStore: 0: Opening raw store with implementation class:org.apache.hadoop.hive.metastore.ObjectStore
2024-09-11 23:11:49,495 WARN metastore.ObjectStore: datanucleus.autoStartMechanismMode is set to unsupported value null . Setting it to value: ignored
2024-09-11 23:11:49,495 INFO metastore.ObjectStore: ObjectStore, initialize called
2024-09-11 23:11:49,498 INFO metastore.MetaStoreDirectSql: Using direct SQL, underlying DB is DERBY
2024-09-11 23:11:49,498 INFO metastore.ObjectStore: Initialized ObjectStore
2024-09-11 23:11:49,499 INFO metastore.RetryingMetaStoreClient: RetryingMetaStoreClient proxy=class org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient ugi=prath
(auth:SIMPLE) retries=1 delay=1 lifetime=0
2024-09-11 23:11:49,501 INFO metastore.HiveMetaStore: 0: get_database: @hive#employee_json
2024-09-11 23:11:49,505 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_database: @hive#employee_json
2024-09-11 23:11:49,508 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-11 23:11:49,509 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
2024-09-11 23:11:49,511 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2024-09-11 23:11:49,512 INFO ql.Driver: Completed compiling command(queryId=prath_20240911231149_188eb98a-6506-4a41-8ed9-79dee35be0d8); Time taken: 0.047 seconds
2024-09-11 23:11:49,513 INFO reexec.ReExecDriver: Execution #1 of query
2024-09-11 23:11:49,514 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:11:49,516 INFO ql.Driver: Executing command(queryId=prath_20240911231149_188eb98a-6506-4a41-8ed9-79dee35be0d8): show tables
2024-09-11 23:11:49,520 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-09-11 23:11:49,525 INFO metastore.HiveMetaStore: 0: get_database: @hive#employee_json
2024-09-11 23:11:49,526 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_database: @hive#employee_json
2024-09-11 23:11:49,529 INFO metastore.HiveMetaStore: 0: get_tables: db=@hive#employee_json pat=.*
2024-09-11 23:11:49,529 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_tables: db=@hive#employee_json pat=.*
2024-09-11 23:11:49,559 INFO ql.Driver: Completed executing command(queryId=prath_20240911231149_188eb98a-6506-4a41-8ed9-79dee35be0d8); Time taken: 0.041 seconds
OK
2024-09-11 23:11:49,561 INFO ql.Driver: OK
2024-09-11 23:11:49,561 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:11:49,567 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-11 23:11:49,573 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_0:1,
employees
Time taken: 0.098 seconds, Fetched: 1 row(s)
2024-09-11 23:11:49,577 INFO CliDriver: Time taken: 0.098 seconds, Fetched: 1 row(s)
2024-09-11 23:11:49,577 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:11:49,578 INFO session.SessionState: Resetting thread name to  main
hive>
```

Add the contents of the JSON file to the table using the following command:

LOAD DATA INPATH '/user/hive/warehouse/employee_json/employee.json' INTO TABLE employees;

```
hive> load data inpath '/user/hive/warehouse/employee_json/employee.json' into table employees;
2024-09-11 23:31:16,134 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:31:16,135 INFO session.SessionState: Updating thread name to 15916fd5-0cc5-48bf-9cb8-a338d3b8713c main
2024-09-11 23:31:16,137 INFO ql.Driver: Compiling command(queryId=prath_20240911233116_1b318b29-630e-42a6-9077-3ffebe7bd4d2): load data inpath '/user/hive/warehouse/employe
e_json/employee.json' into table employees
2024-09-11 23:31:16,163 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:31:16,163 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.employee_json.employees
2024-09-11 23:31:16,164 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr
        cmd=get_table : tbl=hive.employee_json.employees
2024-09-11 23:31:16,208 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-11 23:31:16,208 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null)
2024-09-11 23:31:16,210 INFO ql.Driver: Completed compiling command(queryId=prath_20240911233116_1b318b29-630e-42a6-9077-3ffebe7bd4d2); Time taken: 0.073 seconds
2024-09-11 23:31:16,210 INFO reexec.ReExecDriver: Execution #1 of query
2024-09-11 23:31:16,212 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:31:16,213 INFO ql.Driver: Executing command(queryId=prath_20240911233116_1b318b29-630e-42a6-9077-3ffebe7bd4d2): load data inpath '/user/hive/warehouse/employe
e_json/employee.json' into table employees
2024-09-11 23:31:16,215 INFO ql.Driver: Starting task [Stage-0:MOVE] in serial mode
2024-09-11 23:31:16,216 INFO metastore.HiveMetaStore: 0: Cleaning up thread local RawStore...
2024-09-11 23:31:16,217 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr
        cmd=Cleaning up thread local RawStore...
2024-09-11 23:31:16,218 INFO metastore.HiveMetaStore: 0: Done cleaning up thread local RawStore
2024-09-11 23:31:16,219 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr
        cmd=Done cleaning up thread local RawStore
Loading data to table employee_json.employees
2024-09-11 23:31:16,221 INFO exec.Task: Loading data to table employee_json.employees from hdfs://localhost:9000/user/hive/warehouse/employee_json/employee.json
2024-09-11 23:31:16,227 INFO metastore.HiveMetaStore: 0: Opening raw store with implementation class:org.apache.hadoop.hive.metastore.ObjectStore
2024-09-11 23:31:16,227 WARN metastore.ObjectStore: datanucleus.autoStartMechanismMode is set to unsupported value null . Setting it to value: ignored
2024-09-11 23:31:16,229 INFO metastore.ObjectStore: ObjectStore, initialize called
2024-09-11 23:31:16,234 INFO metastore.MetaStoreDirectSql: Using direct SQL, underlying DB is DERBY
2024-09-11 23:31:16,236 INFO metastore.ObjectStore: Initialized ObjectStore
2024-09-11 23:31:16,238 INFO metastore.RetryingMetaStoreClient: RetryingMetaStoreClient proxy=class org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient ugi=prath
(auth:SIMPLE) retries=1 delay=1 lifetime=0
2024-09-11 23:31:16,239 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.employee_json.employees
2024-09-11 23:31:16,240 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr
        cmd=get_table : tbl=hive.employee_json.employees
2024-09-11 23:31:16,273 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.employee_json.employees
2024-09-11 23:31:16,273 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr
        cmd=get_table : tbl=hive.employee_json.employees
2024-09-11 23:31:16,328 INFO metastore.HiveMetaStore: 0: alter_table: hive.employee_json.employees newtbl=employees
2024-09-11 23:31:16,329 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr
        cmd=alter_table: hive.employee_json.employees newtbl=employees
2024-09-11 23:31:16,371 INFO ql.Driver: Starting task [Stage-1:STATS] in serial mode
2024-09-11 23:31:16,371 INFO metastore.HiveMetaStore: 0: Cleaning up thread local RawStore...
2024-09-11 23:31:16,372 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr
        cmd=Cleaning up thread local RawStore...
```

DESC employees;

```
hive> desc employees;
2024-09-11 23:17:34,481 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:17:34,481 INFO session.SessionState: Updating thread name to 15916fd5-0cc5-48bf-9cb8-a338d3b8713c main
2024-09-11 23:17:34,486 INFO ql.Driver: Compiling command(queryId=prath_20240911231734_58fe66d8-593d-43d7-ae1c-35413f7f2756): desc employees
2024-09-11 23:17:34,511 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:17:34,517 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.employee_json.employees
2024-09-11 23:17:34,518 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr       cmd=get_table : tbl=hive.employee_json.employees
2024-09-11 23:17:34,539 INFO parse.DDLSemanticAnalyzer: analyzeDescribeTable done
2024-09-11 23:17:34,539 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-11 23:17:34,540 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:dat
a_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
2024-09-11 23:17:34,542 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2024-09-11 23:17:34,544 INFO ql.Driver: Completed compiling command(queryId=prath_20240911231734_58fe66d8-593d-43d7-ae1c-35413f7f2756); Time taken: 0.058 seconds
2024-09-11 23:17:34,544 INFO reexec.ReExecDriver: Execution #1 of query
2024-09-11 23:17:34,544 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:17:34,545 INFO ql.Driver: Executing command(queryId=prath_20240911231734_58fe66d8-593d-43d7-ae1c-35413f7f2756): desc employees
2024-09-11 23:17:34,545 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-09-11 23:17:34,546 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.employee_json.employees
2024-09-11 23:17:34,546 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr       cmd=get_table : tbl=hive.employee_json.employees
2024-09-11 23:17:34,580 INFO ql.Driver: Completed executing command(queryId=prath_20240911231734_58fe66d8-593d-43d7-ae1c-35413f7f2756); Time taken: 0.035 seconds
OK
2024-09-11 23:17:34,580 INFO ql.Driver: OK
2024-09-11 23:17:34,580 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:17:34,584 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-11 23:17:34,589 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_0:4,
id                      int                     from deserializer
name                    string                  from deserializer
age                     int                     from deserializer
salary                  double                  from deserializer
Time taken: 0.094 seconds, Fetched: 4 row(s)
2024-09-11 23:17:34,598 INFO CliDriver: Time taken: 0.094 seconds, Fetched: 4 row(s)
2024-09-11 23:17:34,598 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:17:34,599 INFO session.SessionState: Resetting thread name to  main
hive>
```

SELECT * FROM employees;

```
hive> select * from employees;
2024-09-11 23:31:56,997 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:31:56,998 INFO session.SessionState: Updating thread name to 15916fd5-0cc5-48bf-9cb8-a338d3b8713c main
2024-09-11 23:31:57,001 INFO ql.Driver: Compiling command(queryId=prath_20240911233156_e676c4b6-6b10-4b8d-8054-22387b88f219): select * from employees
2024-09-11 23:31:57,024 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:31:57,025 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-09-11 23:31:57,025 INFO parse.CalcitePlanner: Completed phase 1 of Semantic Analysis
2024-09-11 23:31:57,025 INFO parse.CalcitePlanner: Get metadata for source tables
2024-09-11 23:31:57,026 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.employee_json.employees
2024-09-11 23:31:57,027 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_table : tbl=hive.employee_json.employees
2024-09-11 23:31:57,038 INFO parse.CalcitePlanner: Get metadata for subqueries
2024-09-11 23:31:57,038 INFO parse.CalcitePlanner: Get metadata for destination tables
2024-09-11 23:31:57,060 INFO ql.Context: New scratch dir is hdfs://localhost:9000/tmp/hive/prath/15916fd5-0cc5-48bf-9cb8-a338d3b8713c/hive_2024-09-11_23-31-57_023_482717281
8626818109-1
2024-09-11 23:31:57,061 INFO parse.CalcitePlanner: Completed getting MetaData in Semantic Analysis
2024-09-11 23:31:57,070 INFO metastore.HiveMetaStore: 0: get_not_null_constraints : tbl=hive.employee_json.employees
2024-09-11 23:31:57,073 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_not_null_constraints : tbl=hive.employee_json.employees
2024-09-11 23:31:57,080 INFO metastore.HiveMetaStore: 0: get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:31:57,081 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:31:57,086 INFO metastore.HiveMetaStore: 0: get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:31:57,091 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:31:57,094 INFO metastore.HiveMetaStore: 0: get_unique_constraints : tbl=hive.employee_json.employees
2024-09-11 23:31:57,097 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_unique_constraints : tbl=hive.employee_json.employees
```

```
2024-09-11 23:31:57,328 INFO exec.TableScanOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_TS_0:10,
2024-09-11 23:31:57,328 INFO exec.SelectOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_SEL_1:10,
2024-09-11 23:31:57,329 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_3:10,
1       John Doe        30      50000.0
2       Jane Smith      25      60000.0
3       Alice Johnson   28      55000.0
4       Bob Brown       35      70000.0
5       Charlie Davis   40      80000.0
6       Eve White       22      48000.0
7       Frank Black     32      65000.0
8       Grace Green     27      52000.0
9       Henry Gold      29      59000.0
10      Isabel Blue     33      73000.0
Time taken: 0.272 seconds, Fetched: 10 row(s)
2024-09-11 23:31:57,353 INFO CliDriver: Time taken: 0.272 seconds, Fetched: 10 row(s)
2024-09-11 23:31:57,357 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:31:57,358 INFO session.SessionState: Resetting thread name to  main
hive>
```

Perform various other operations on the data in the table

WHERE:

SELECT id,name,age,salary from employees where salary>60000;

```
hive> select id,name,age,salary from employees where salary>60000;
2024-09-11 23:42:26,983 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:42:26,983 INFO session.SessionState: Updating thread name to 15916fd5-0cc5-48bf-9cb8-a338d3b8713c main
2024-09-11 23:42:26,986 INFO ql.Driver: Compiling command(queryId=prath_20240911234226_f869c858-cd64-4c32-b6c7-d5f4941b38d9): select id,name,age,salary from employees where
 salary>60000
2024-09-11 23:42:27,009 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:42:27,009 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-09-11 23:42:27,010 INFO parse.CalcitePlanner: Completed phase 1 of Semantic Analysis
2024-09-11 23:42:27,010 INFO parse.CalcitePlanner: Get metadata for source tables
2024-09-11 23:42:27,010 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.employee_json.employees
2024-09-11 23:42:27,010 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_table : tbl=hive.employee_json.employees
2024-09-11 23:42:27,028 INFO parse.CalcitePlanner: Get metadata for subqueries
2024-09-11 23:42:27,029 INFO parse.CalcitePlanner: Get metadata for destination tables
2024-09-11 23:42:27,056 INFO ql.Context: New scratch dir is hdfs://localhost:9000/tmp/hive/prath/15916fd5-0cc5-48bf-9cb8-a338d3b8713c/hive_2024-09-11_23-42-27_003_586755435
9709350645-1
2024-09-11 23:42:27,056 INFO parse.CalcitePlanner: Completed getting MetaData in Semantic Analysis
2024-09-11 23:42:27,063 INFO metastore.HiveMetaStore: 0: get_not_null_constraints : tbl=hive.employee_json.employees
2024-09-11 23:42:27,063 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_not_null_constraints : tbl=hive.employee_json.employees
2024-09-11 23:42:27,066 INFO metastore.HiveMetaStore: 0: get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:42:27,066 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:42:27,069 INFO metastore.HiveMetaStore: 0: get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:42:27,075 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:42:27,077 INFO metastore.HiveMetaStore: 0: get_unique_constraints : tbl=hive.employee_json.employees
2024-09-11 23:42:27,078 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_unique_constraints : tbl=hive.employee_json.employees
2024-09-11 23:42:27,080 INFO metastore.HiveMetaStore: 0: get_foreign_keys : parentdb=null parenttbl=null foreigndb=employee_json foreigntbl=employees
2024-09-11 23:42:27,081 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_foreign_keys : parentdb=null parenttbl=null foreigndb=employee_json foreign
tbl=employees
2024-09-11 23:42:28,303 INFO metastore.HiveMetaStore: 0: get_databases: @hive#
2024-09-11 23:42:28,303 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_databases: @hive#
2024-09-11 23:42:28,308 INFO metastore.HiveMetaStore: 0: get_materialized_views_for_rewriting: db=@hive#default
2024-09-11 23:42:28,308 INFO HiveMetaStore.audit: ugi=prath      ip=unknown-ip-addr      cmd=get_materialized_views_for_rewriting: db=@hive#default
```

```
2024-09-11 23:42:28,493 INFO exec.SelectOperator: RECORDS_OUT_OPERATOR_SEL_2:4, RECORDS_OUT_INTERMEDIATE:0,
2024-09-11 23:42:28,493 INFO exec.ListSinkOperator: RECORDS_OUT_OPERATOR_LIST_SINK_5:4, RECORDS_OUT_INTERMEDIATE:0,
4       Bob Brown       35      70000.0
5       Charlie Davis   40      80000.0
7       Frank Black     32      65000.0
10      Isabel Blue     33      73000.0
Time taken: 1.477 seconds, Fetched: 4 row(s)
2024-09-11 23:42:28,515 INFO CliDriver: Time taken: 1.477 seconds, Fetched: 4 row(s)
2024-09-11 23:42:28,516 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:42:28,518 INFO session.SessionState: Resetting thread name to  main
hive>
```

PROJECTION:

select id,name from employees;

```
hive> select id,name from employees;
2024-09-11 23:45:04,194 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:45:04,195 INFO session.SessionState: Updating thread name to 15916fd5-0cc5-48bf-9cb8-a338d3b8713c main
2024-09-11 23:45:04,198 INFO ql.Driver: Compiling command(queryId=prath_20240911234504_c4ee5eed-d75a-4d42-aa72-47fd0036dd27): select id,name from employees
2024-09-11 23:45:04,222 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-11 23:45:04,223 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-09-11 23:45:04,223 INFO parse.CalcitePlanner: Completed phase 1 of Semantic Analysis
2024-09-11 23:45:04,223 INFO parse.CalcitePlanner: Get metadata for source tables
2024-09-11 23:45:04,225 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.employee_json.employees
2024-09-11 23:45:04,226 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_table : tbl=hive.employee_json.employees
2024-09-11 23:45:04,243 INFO parse.CalcitePlanner: Get metadata for subqueries
2024-09-11 23:45:04,244 INFO parse.CalcitePlanner: Get metadata for destination tables
2024-09-11 23:45:04,284 INFO ql.Context: New scratch dir is hdfs://localhost:9000/tmp/hive/prath/15916fd5-0cc5-48bf-9cb8-a338d3b8713c/hive_2024-09-11_23-45-04_220_258752934
1977641656-1
2024-09-11 23:45:04,285 INFO parse.CalcitePlanner: Completed getting MetaData in Semantic Analysis
2024-09-11 23:45:04,299 INFO metastore.HiveMetaStore: 0: get_not_null_constraints : tbl=hive.employee_json.employees
2024-09-11 23:45:04,302 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_not_null_constraints : tbl=hive.employee_json.employees
2024-09-11 23:45:04,311 INFO metastore.HiveMetaStore: 0: get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:45:04,312 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:45:04,321 INFO metastore.HiveMetaStore: 0: get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:45:04,325 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_primary_keys : tbl=hive.employee_json.employees
2024-09-11 23:45:04,329 INFO metastore.HiveMetaStore: 0: get_unique_constraints : tbl=hive.employee_json.employees
2024-09-11 23:45:04,331 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_unique_constraints : tbl=hive.employee_json.employees
2024-09-11 23:45:04,334 INFO metastore.HiveMetaStore: 0: get_foreign_keys : parentdb=null parenttbl=null foreigndb=employee_json foreigntbl=employees
2024-09-11 23:45:04,337 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_foreign_keys : parentdb=null parenttbl=null foreigndb=employee_json foreignt
bl=employees
2024-09-11 23:45:04,390 INFO metastore.HiveMetaStore: 0: get_databases: @hive#
2024-09-11 23:45:04,390 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_databases: @hive#
2024-09-11 23:45:04,393 INFO metastore.HiveMetaStore: 0: get_materialized_views_for_rewriting: db=@hive#default
2024-09-11 23:45:04,393 INFO HiveMetaStore.audit: ugi=prath    ip=unknown-ip-addr    cmd=get_materialized_views_for_rewriting: db=@hive#default
```

```
2024-09-11 23:45:04,523 INFO exec.SelectOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_SEL_1:10,
2024-09-11 23:45:04,524 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_3:10,
1       John Doe
2       Jane Smith
3       Alice Johnson
4       Bob Brown
5       Charlie Davis
6       Eve White
7       Frank Black
8       Grace Green
9       Henry Gold
10      Isabel Blue
Time taken: 0.295 seconds, Fetched: 10 row(s)
2024-09-11 23:45:04,561 INFO CliDriver: Time taken: 0.295 seconds, Fetched: 10 row(s)
2024-09-11 23:45:04,562 INFO conf.HiveConf: Using the default value passed in for log id: 15916fd5-0cc5-48bf-9cb8-a338d3b8713c
2024-09-11 23:45:04,568 INFO session.SessionState: Resetting thread name to  main
hive>
```

AGGREGATION:

select age,max(salary) as MAX_SALARY from employees group by age;

```
2024-09-15 12:26:22,175 INFO exec.ListSinkOperator: RECORDS_OUT_OPERATOR_LIST_SINK_10:10, RECORDS_OUT_INTERMEDIATE:0,
22      48000.0
25      60000.0
27      52000.0
28      55000.0
29      59000.0
30      50000.0
32      65000.0
33      73000.0
35      70000.0
40      80000.0
Time taken: 38.333 seconds, Fetched: 10 row(s)
```

LIMIT:

select * from employees limit 4;

```
1       John Doe        30      50000.0
2       Jane Smith      25      60000.0
3       Alice Johnson   28      55000.0
4       Bob Brown       35      70000.0
```

COUNT:

select count(*) from employees;

```
10
Time taken: 30.682 seconds, Fetched: 1 row(s)
```

SORT:

select * from employees order by salary desc;

```
5       Charlie Davis   40      80000.0
10      Isabel Blue     33      73000.0
4       Bob Brown       35      70000.0
7       Frank Black     32      65000.0
2       Jane Smith      25      60000.0
9       Henry Gold      29      59000.0
3       Alice Johnson   28      55000.0
8       Grace Green     27      52000.0
1       John Doe        30      50000.0
6       Eve White       22      48000.0
Time taken: 33.027 seconds, Fetched: 10 row(s)
```

SKIP:

select * from (select *, row_number() over() as row_num from employees) temp where row_num>3;

```
7        Frank Black     32       65000.0 4
6        Eve White       22       48000.0 5
5        Charlie Davis   40       80000.0 6
4        Bob Brown       35       70000.0 7
3        Alice Johnson   28       55000.0 8
2        Jane Smith      25       60000.0 9
1        John Doe        30       50000.0 10
Time taken: 33.718 seconds, Fetched: 7 row(s)
```