- Load without alias

```
truck_events = LOAD 'truck_event_text_partition.csv' USING PigStorage(',');
```

- DESCRIBE a Relation
```
DESCRIBE truck_events;
```

- DUMP data on console
```
DUMP truck_events;
```

- Load with alias

```
truck_events = LOAD 'truck_event_text_partition.csv' USING PigStorage(',')
AS (driverId:int, truckId:int, eventTime:chararray, eventType:chararray,
longitude:double, latitude:double, eventKey:chararray, correlationId:long,
driverName:chararray, routeId:long, routeName:chararray,
eventDate:chararray);
```

```
DESCRIBE truck_events;
```

- Take sample records

```
truck_events_subset = LIMIT  truck_events 100;
```

```
dump truck_event_subset;
```

- Choose specific columns

```
specific_columns = FOREACH truck_events_subset GENERATE driverId,
eventTime, eventType;
```

```
DESCRIBE specific_columns;
```

```
DUMP specific_columns;
```

- STORE output

```
STORE specific_columns INTO 'output_directory' USING PigStorage(',');
```

- JOIN two datasets

```
truck_events = LOAD 'truck_event_text_partition.csv' USING PigStorage(',')
AS (driverId:int, truckId:int, eventTime:chararray,
eventType:chararray, longitude:double, latitude:double,
eventKey:chararray, correlationId:long, driverName:chararray,
routeId:long,routeName:chararray,eventDate:chararray);

drivers =  LOAD 'drivers.csv' USING PigStorage(',')
AS (driverId:int, name:chararray, ssn:chararray,
location:chararray, certified:chararray, wage_plan:chararray);

join_data = JOIN  truck_events BY (driverId), drivers BY (driverId);

DESCRIBE join_data;
DUMP join_data;
```

- SORT

```
ordered_data = ORDER drivers BY name asc;
DUMP ordered_data;
```

- FILTER
```
filtered_events = FILTER truck_events BY NOT (eventType MATCHES
'Normal');
DUMP filtered_events;
```

- SPLIT FILTER
```
SPLIT filtered_events INTO normal_events if eventType == 'Normal',
others if eventType != 'Normal';
dump normal_events;
dump others;
```

- GROUP
```
grouped_events = GROUP filtered_events BY driverId;
DESCRIBE grouped_events;
DUMP grouped_events;
```