* Load without alias

truck\_events = LOAD 'truck\_event\_text\_partition.csv' USING PigStorage(',');

DESCRIBE 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;

* 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;

* SORT data

ordered\_data = ORDER drivers BY name asc;

DUMP ordered\_data;

* FILTER data

filtered\_events = FILTER truck\_events BY NOT (eventType MATCHES 'Normal');

grouped\_events = GROUP filtered\_events BY driverId;

DESCRIBE grouped\_events;

DUMP grouped\_events;