

1. Upload the two files to HDFS (if not uploaded before).

1. Load the files as investors and stock\_prices.

2. Display both files to make sure they are loaded correctly

```
grunt> dump investors

/pig/input/investor.txt"

Output(s):
Successfully stored 27 records (879 bytes) in: "hdfs://ip-172-31-26-120.ec2.internal:8020/tmp/temp1498552
859/tmp-540888138"

2024-03-07 06:54:04,257 INFO input.FileInputFormat: Total input files to process : 1
917055 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to
process : 1
2024-03-07 06:54:04,257 INFO util.MapRedUtil: Total input paths to process : 1
(1,Pippa,Dickens,BAC,240)
(2,Gavin,Thomson,CAH,60)
(3,Brian,Johnston,GE,850)
(4,Jessica,Henderson,MCD,200)
(5,Andrea,Arnold,PFE,130)
(6,Vanessa,Robertson,BAC,275)
(7,Megan,Clark,CAH,65)
(8,Lisa,Butler,GE,800)
(9,Amanda,Piper,MCD,210)
(10,Joe,Chapman,PFE,125)
(11,Pippa,Dickens,CAH,280)
(12,Gavin,Thomson,GE,35)
(13,Brian,Johnston,MCD,810)
(14,Jessica,Henderson,PFE,190)
(15,Andrea,Arnold,BAC,105)
(16,Vanessa,Robertson,CAH,250)
(17,Megan,Clark,GE,75)
(18,Lisa,Butler,MCD,700)
(19,Amanda,Piper,PFE,225)
(20,Joe,Chapman,BAC,115)
(21,Jessica,Henderson,CAH,210)
(22,Andrea,Arnold,GE,65)
(23,Vanessa,Robertson,MCD,875)
(24,Megan,Clark,PFE,220)
(25,Lisa,Butler,BAC,125)
(26,Amanda,Piper,CAH,35)
(27,Joe,Chapman,GE,810)
grunt> □
```

```
(,)  
grunt> dump stockprices[]
```

```
2024-03-07 07:49:26,561 INFO input.FileInputFormat: Total input files to process : 1  
4239359 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2024-03-07 07:49:26,561 INFO util.MapRedUtil: Total input paths to process : 1  
(BAC,25.0)  
(CAH,48.0)  
(GE,12.0)  
(MCD,170.0)  
(PFE,40.0)  
(,)  
grunt> #[]
```

3. Display the structure of relation investors and stock\_prices.

```
grunt> DESCRIBE investors;  
investors: {id: int, name: chararray, lname: chararray, stock: chararray, noStock: int}  
grunt> DESCRIBE stockprices;  
stockprices: {stock: chararray, stockprice: float}  
grunt> #[]
```

4. Perform explain on stock\_prices and observe the logical, physical, and MapReduce execution plans.

```
grunt> explain stockprices  
2024-03-07 07:08:31,729 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled  
is deprecated. Instead, use yarn.system-metrics-publisher.enabled  
1784527 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2024-03-07 07:08:31,729 WARN data.SchemaTupleBackend: SchemaTupleBackend has already been initialized  
1784527 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFilter, PushHubFilter, SplitFilter, StreamTypeCastInserter]}  
2024-03-07 07:08:31,729 INFO optimizer.LogicalPlanOptimizer: {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFilter, PushHubFilter, SplitFilter, StreamTypeCastInserter]}  
#-- New Logical Plan:  
#-- stockprices: (Name: LOStore Schema: stock#60:chararray,stockprice#61:float)  
|--- stockprices: (Name: LOForEach Schema: stock#60:chararray,stockprice#61:float)  
|   |--- (Name: LGenerate[false,false] Schema: stock#60:chararray,stockprice#61:float)ColumnPrune:OutputU  
|   |   |--- (Name: ColumnPrune:InputUids=[60, 61]  
|   |   |   |--- (Name: Cast Type: chararray Uid: 60)  
|   |   |   |--- stock:(Name: Project Type: bytearray Uid: 60 Input: 0 Column: (*))  
|   |   |   |--- (Name: Cast Type: float Uid: 61)  
|   |   |   |--- stockprice:(Name: Project Type: bytearray Uid: 61 Input: 1 Column: (*))  
|   |   |--- (Name: LOInnerLoad[0] Schema: stock#60:bytearray)  
|   |   |--- (Name: LOInnerLoad[1] Schema: stockprice#61:bytearray)  
|--- stockprices: (Name: LLoad Schema: stock#60:bytearray,stockprice#61:bytearray)RequiredFields:null
```

5. Group investors by their stock symbols that they have purchased and display the results.

```
grunt> aggStocks = group investors by stock;  
grunt> dump aggStocks
```

```

Input(s):
Successfully read 27 records (748 bytes) from: "hdfs://ip-172-31-26-120.ec2.internal:8020/user/hadoop/HW5
/pig/input/investor.txt"

Output(s):
Successfully stored 5 records (857 bytes) in: "hdfs://ip-172-31-26-120.ec2.internal:8020/tmp/temp14985528
59/tmp729236169"

2024-03-07 07:13:09,685 INFO input.FileInputFormat: Total input files to process : 1
2062483 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths t
o process : 1
2024-03-07 07:13:09,685 INFO util.MapRedUtil: Total input paths to process : 1
(GE,{(12,Gavin,Thomson,GE,35),(8,Lisa,Butler,GE,800),(3,Brian,Johnston,GE,850),(17,Megan,Clark,GE,75),(27
,Joe,Chapman,GE,810),(22,Andrea,Arnold,GE,65)})
(BAC,{(1,Pippa,Dickens,BAC,240),(6,Vanessa,Robertson,BAC,275),(15,Andrea,Arnold,BAC,105),(20,Joe,Chapman,
BAC,115),(25,Lisa,Butler,BAC,125)})
(CAH,{(7,Megan,Clark,CAH,65),(16,Vanessa,Robertson,CAH,250),(26,Amanda,Piper,CAH,35),(21,Jessica,Henderso
n,CAH,210),(2,Gavin,Thomson,CAH,60),(11,Pippa,Dickens,CAH,280)})
(MCD,{(18,Lisa,Butler,MCD,700),(9,Amanda,Piper,MCD,210),(13,Brian,Johnston,MCD,810),(23,Vanessa,Robertson
,MCD,875),(4,Jessica,Henderson,MCD,200)})
(PFE,{(19,Amanda,Piper,PFE,225),(24,Megan,Clark,PFE,220),(10,Joe,Chapman,PFE,125),(5,Andrea,Arnold,PFE,13
0),(14,Jessica,Henderson,PFE,190)})
grunt>
grunt> []

```

## 6. Combine both relations with a union command and display the results.

```

grunt> combined = union investors,stockprices;
grunt> dump combined

```

```

2024-03-07 07:17:10,148 INFO util.MapRedUtil: Total input paths to process : 2
(BAC,25.0)
(CAH,48.0)
(GE,12.0)
(MCD,170.0)
(PFE,40.0)
(,)
(1,Pippa,Dickens,BAC,240)
(2,Gavin,Thomson,CAH,60)
(3,Brian,Johnston,GE,850)
(4,Jessica,Henderson,MCD,200)
(5,Andrea,Arnold,PFE,130)
(6,Vanessa,Robertson,BAC,275)
(7,Megan,Clark,CAH,65)
(8,Lisa,Butler,GE,800)
(9,Amanda,Piper,MCD,210)
(10,Joe,Chapman,PFE,125)
(11,Pippa,Dickens,CAH,280)
(12,Gavin,Thomson,GE,35)
(13,Brian,Johnston,MCD,810)
(14,Jessica,Henderson,PFE,190)
(15,Andrea,Arnold,BAC,105)
(16,Vanessa,Robertson,CAH,250)
(17,Megan,Clark,GE,75)
(18,Lisa,Butler,MCD,700)
(19,Amanda,Piper,PFE,225)
(20,Joe,Chapman,BAC,115)
(21,Jessica,Henderson,CAH,210)
(22,Andrea,Arnold,GE,65)
(23,Vanessa,Robertson,MCD,875)
(24,Megan,Clark,PFE,220)
(25,Lisa,Butler,BAC,125)
(26,Amanda,Piper,CAH,35)
(27,Joe,Chapman,GE,810)
grunt> []

```