Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

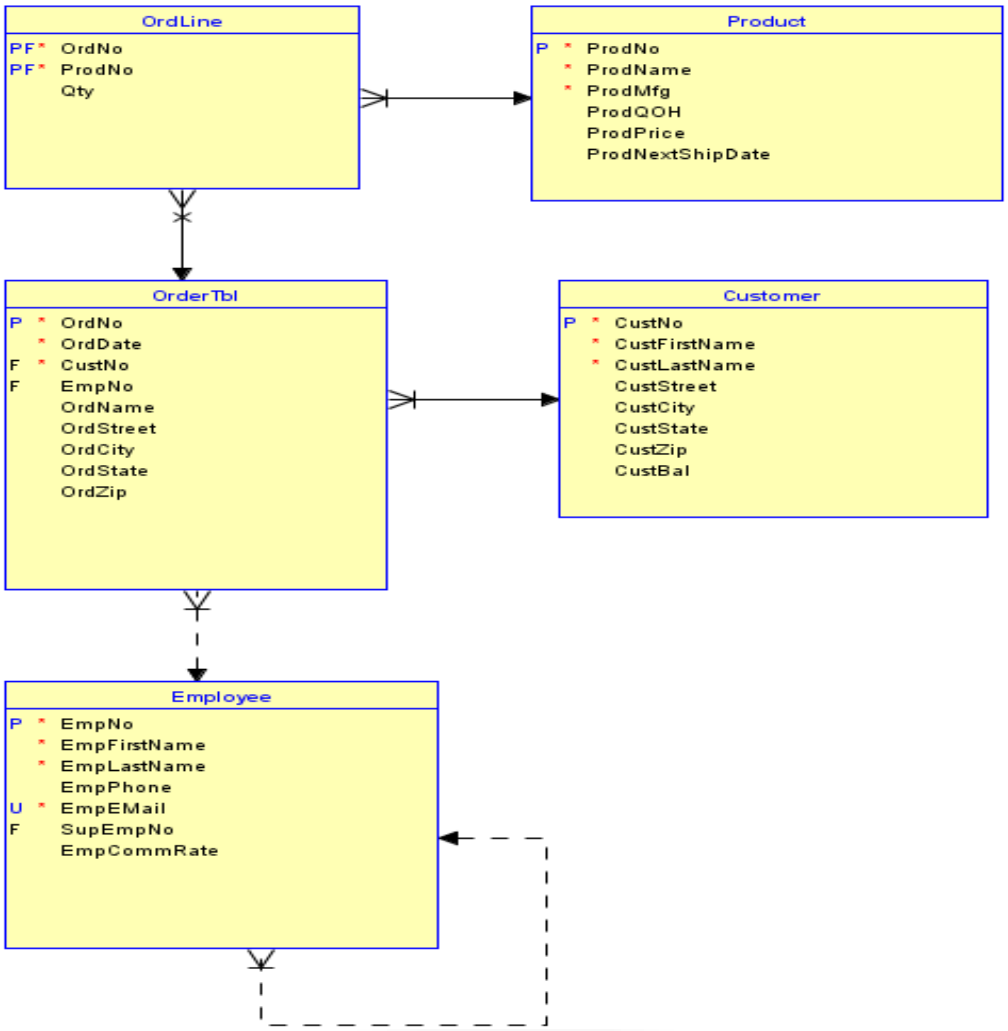
Module 6
SQL for Data Mining Input

Lesson 3: Statement Patterns for Item Sets and Association Rules

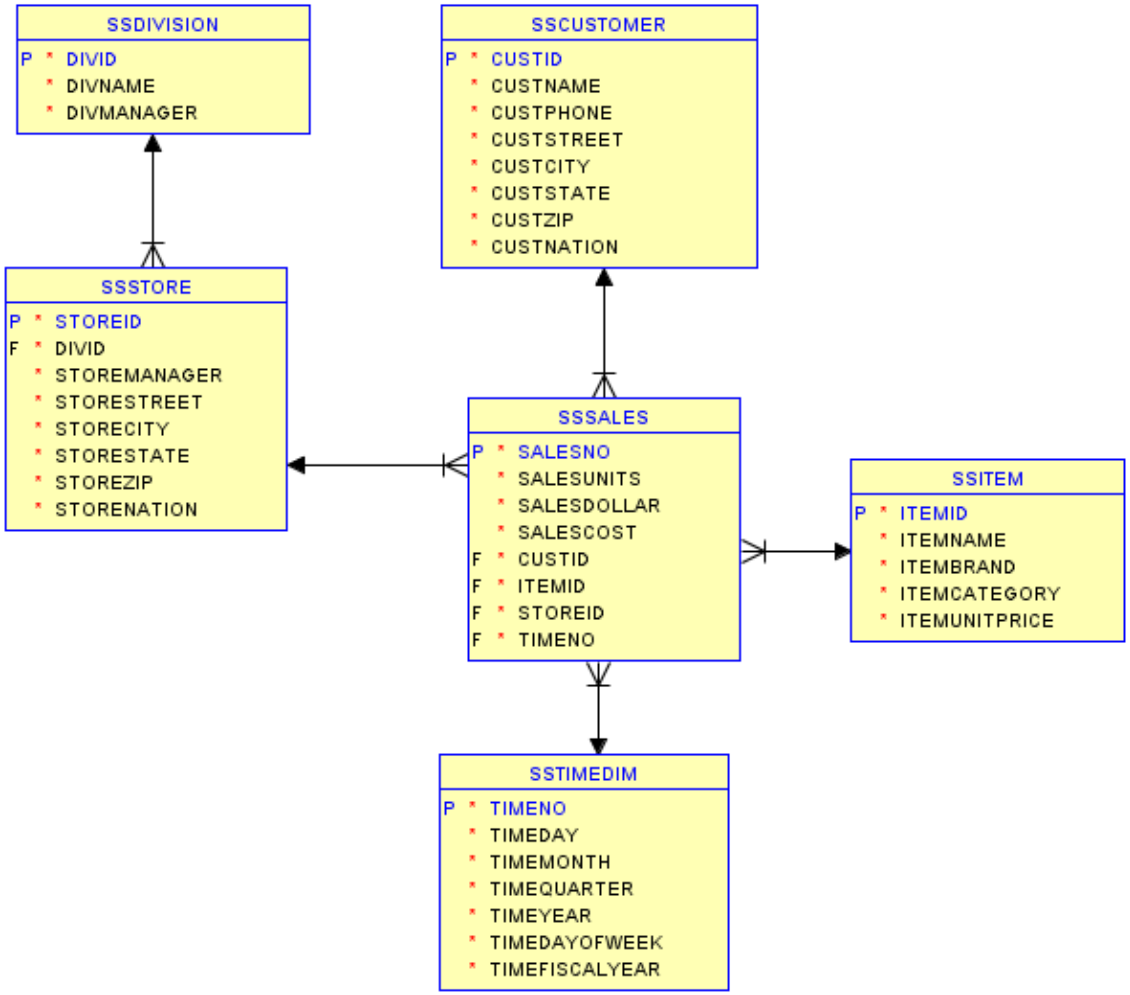
Lesson Objectives

- Apply statement patterns for item sets and association rules
- Write SELECT statements to generate item sets with more than 2 items
- Write SELECT statements to generate association rules with more than 2 items
- Gain insight about atypical join patterns

Order Entry Tables (Operational Database)



Store Sales Tables (Data Warehouse)



Statement Pattern for Item Sets

```
SELECT <BasketId1>, [ <BasketId2>, …    <BasketIdm>, ]
       <ItemId AS Item1>, <ItemId AS Item2>,
       [ <ItemId AS Itemn> ]
 FROM EventTable ET1, EventTable ET2, … [ EventTable ETn ]
 WHERE <BasketJoinConditions>
   AND <ItemSetConditions> ;
-- <BasketJoinConditions>: n-1 sets of self join conditions
-- ETi.BasketIdj = ETi+1.BasketIdj
-- m conditions in each set of join conditions
-- <ItemSetConditions>: ETi.ItemId < ETi+1.ItemId
-- n-1 <ItemSetConditions>
```


Statement Pattern for Association Rules

```
SELECT <BasketId1>, [ <BasketId2>, …    <BasketIdm>, ]
       <ItemId AS Item1>, <ItemId AS Item2>,
       [ <ItemId AS Itemn> ]
 FROM EventTable ET1, EventTable ET2, … [ EventTable ETn ]
 WHERE <BasketJoinConditions>
 [ AND <LHSConditions> ]
   AND <RHSConditions> ;
-- <BasketJoinConditions>: n-1 sets of self join conditions
-- ETi.BasketColj = ETi+1.BasketColj
-- <LHSConditions>: ETi.ItemId < ETi+1.ItemId
-- n-2 <LHSConditions>
-- <RHSConditions>: ETi.ItemId <> ETn.ItemId
-- n-1 <RHSConditions>
```


Item Sets of Size 3 (Order Entry Tables)

```
-- Example 1
SELECT OL1.OrdNo, OL1.ProdNo ProdNo1,
       OL2.ProdNo ProdNo2, OL3.ProdNo ProdNo3
 FROM OrdLine OL1, OrdLine OL2, OrdLine OL3
 WHERE OL1.OrdNo = OL2.OrdNo
   AND OL2.OrdNo = OL3.OrdNo
   AND OL1.ProdNo < OL2.ProdNo
   AND OL2.ProdNo < OL3.ProdNo
 ORDER BY OL1.OrdNo, OL1.ProdNo, OL2.ProdNo;
```


Association Rules of Size 3 (Order Entry Tables)

```
-- Example 2
SELECT OL1.OrdNo, OL1.ProdNo ProdNo1,
       OL2.ProdNo ProdNo2, OL3.ProdNo ProdNo3
 FROM OrdLine OL1, OrdLine OL2, OrdLine OL3
 WHERE OL1.OrdNo = OL2.OrdNo
   AND OL2.OrdNo = OL3.OrdNo
   AND OL1.ProdNo < OL2.ProdNo
   AND OL1.ProdNo <> OL3.ProdNo
   AND OL2.ProdNo <> OL3.ProdNo
 ORDER BY OL1.OrdNo, OL1.ProdNo, OL2.ProdNo;
```


Item Sets of Size 3 (Store Sales Tables)

```
-- Example 3
SELECT S1.CustId, S1.TimeNo, S1.StoreId,
       S1.ItemId ItemId1, S2.ItemId ItemId2,
       S3.ItemId ItemId3
 FROM SSSales S1, SSSales S2, SSSales S3
 WHERE S1.CustId = S2.CustId
   AND S2.CustId = S3.CustId
   AND S1.TimeNo = S2.TimeNo
   AND S2.TimeNo = S3.TimeNo
   AND S1.StoreId = S2.StoreId
   AND S2.StoreId = S3.StoreId
   AND S1.ItemId < S2.ItemId
   AND S2.ItemId < S3.ItemId
 ORDER BY S1.CustId, S1.TimeNo, S1.StoreId, S1.ItemId, S2.ItemId;
```


Association Rules of Size 3 (Store Sales Tables)

```
-- Example 4
SELECT S1.CustId, S1.TimeNo, S1.StoreId,
       S1.ItemId ItemId1, S2.ItemId ItemId2,
       S3.ItemId ItemId3
 FROM SSSales S1, SSSales S2, SSSales S3
 WHERE S1.CustId = S2.CustId
   AND S2.CustId = S3.CustId
   AND S1.TimeNo = S2.TimeNo
   AND S2.TimeNo = S3.TimeNo
   AND S1.StoreId = S2.StoreId
   AND S2.StoreId = S3.StoreId
   AND S1.ItemId < S2.ItemId
   AND S1.ItemId <> S3.ItemId
   AND S2.ItemId <> S3.ItemId
 ORDER BY S1.CustId, S1.TimeNo, S1.StoreId, S1.ItemId, S2.ItemId;
```


Summary

- Statement patterns for item sets and association rules
- Join condition differences for data lakes and data warehouses
- SELECT statements to generate item sets and association rules with 3 items
- Specialized SQL skills for data scientist collaboration and atypical join patterns