Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

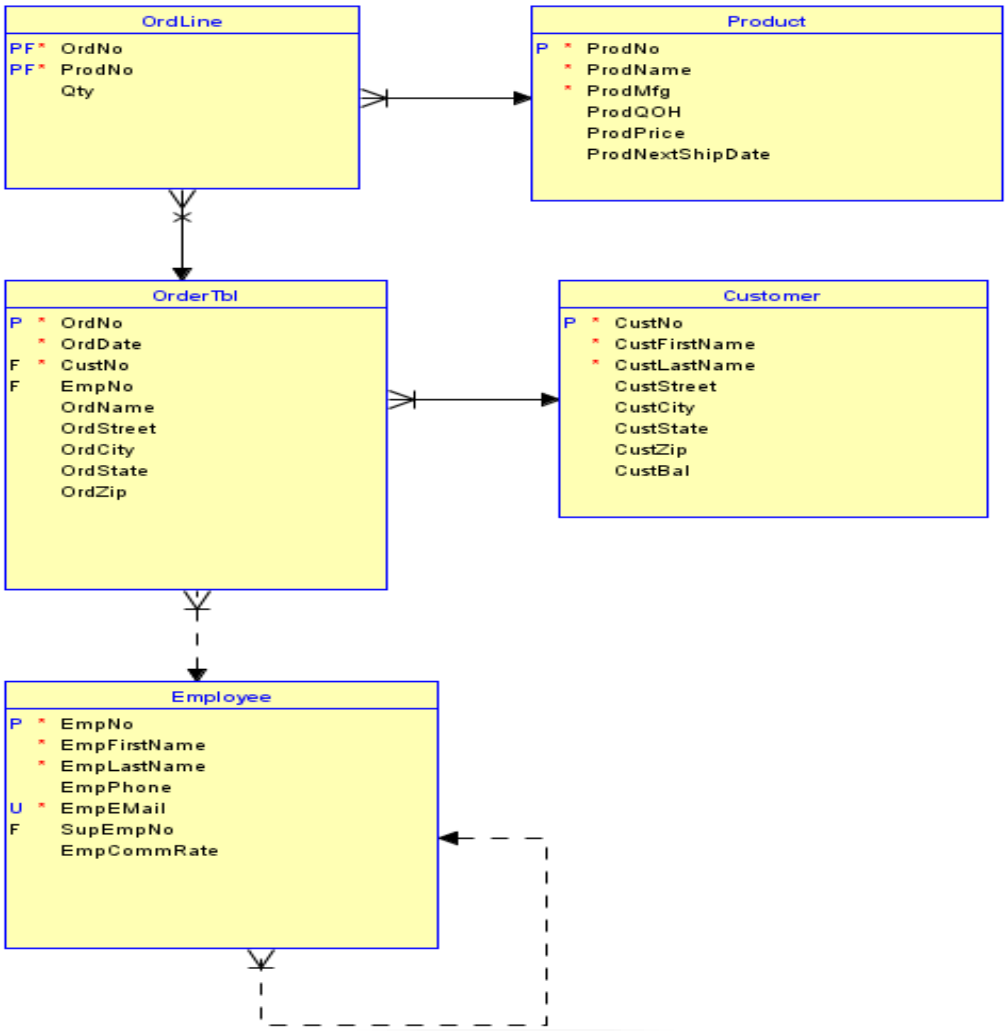
Module 6
SQL for Data Mining Input

Lesson 2: SQL Coding for Simple Item Sets and Association Rules

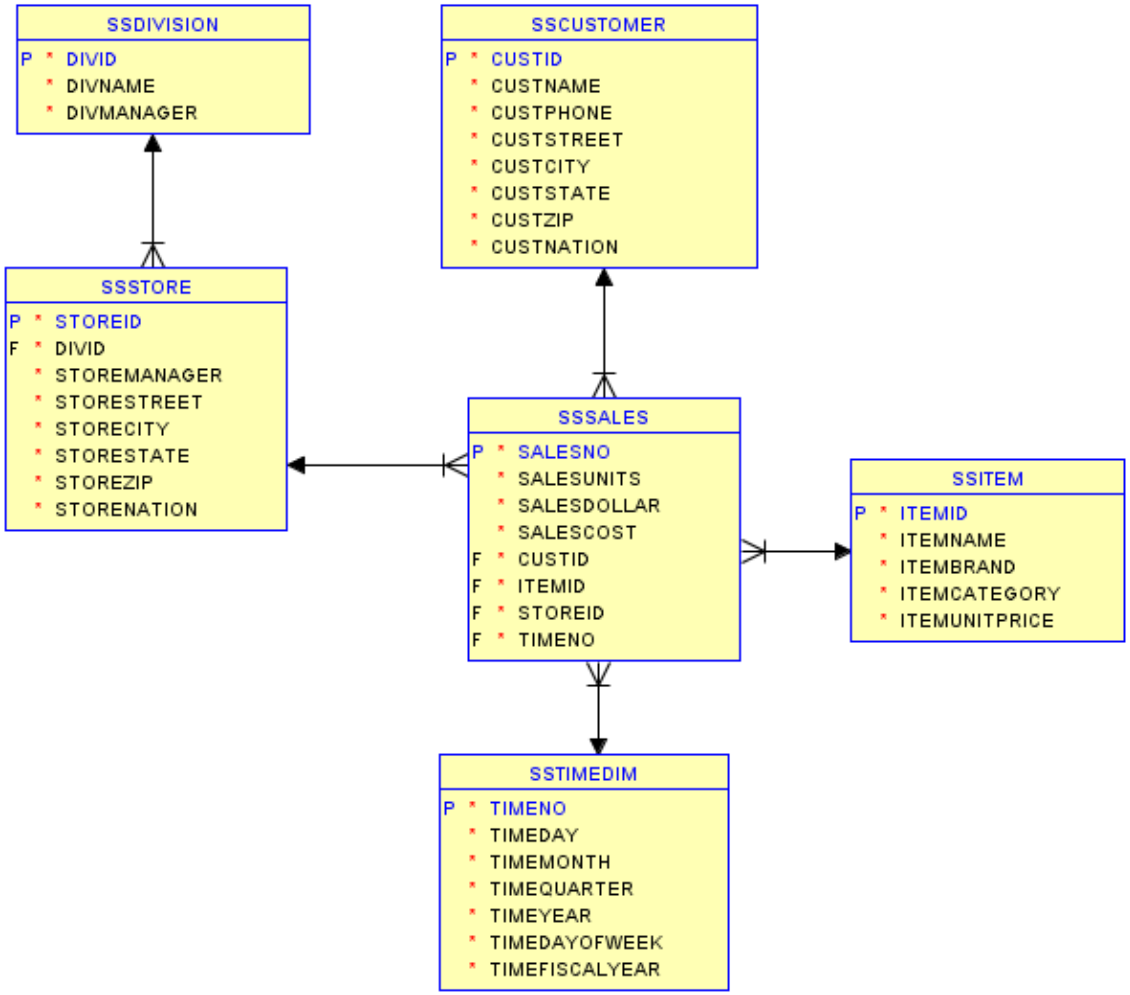
Lesson Objectives

- Explain concepts for query formulation: basket identification, item sets, and association rules
- Write SELECT statements to generate item sets with 2 items
- Write SELECT statements to generate association rules with 2 items

Order Entry Tables (Operational Database)



Store Sales Tables (Data Warehouse)



Item Sets versus Association Rules

Item set

- Combination (ordering not important)
- Used by association rule mining algorithms
- Examples: (I1, I2), (I1, I3), (I2, I3), (I1, I2, I3)

Association rule

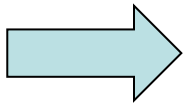
- Left hand side (LHS) with one or more items
- Right hand side (RHS) with a single item
- Ordering between LHS and RHS
- Examples: (I1->I2), (I2->I1), (I1->I3), (I3->I1), (I2->I3), (I3->I2), (I1,I2->I3), (I2,I3->I1), (I1,I3->I2)

Item Sets of Size 2 (Order Entry Tables)

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

OrdLine

OrdNo	ProdNo	Qty
O1	P1	1
O1	P2	5



Result

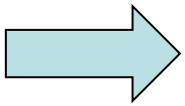
OrdNo	ProdNo1	ProdNo2
O1	P1	P2

Association Rules of Size 2 (Order Entry Tables)

```
-- Example 2 generates double rows of Example 1.
SELECT OL1.OrdNo, OL1.ProdNo ProdNoLHS,
       OL2.ProdNo ProdNoRHS
 FROM OrdLine OL1, OrdLine OL2
 WHERE OL1.OrdNo = OL2.OrdNo
   AND OL1.ProdNo <> OL2.ProdNo
 ORDER BY OL1.OrdNo, OL1.ProdNo;
```

OrdLine

OrdNo	ProdNo	Qty
O1	P1	1
O1	P2	5



Result

OrdNo	ProdLHS	ProdRHS
O1	P1	P2
O1	P2	P1

Item Sets of Size 2 (Store Sales Tables)

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


Association Rules of Size 2 (Store Sales Tables)

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


Summary

- Examples using a data lake and data warehouse
- Usually, 1 basket identification column for a data lake but more than 1 column for a data warehouse
- No ordering for item sets
- Ordering between LHS and RHS in association rules
- SELECT statements with < condition for item sets and <> condition for association rules