Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

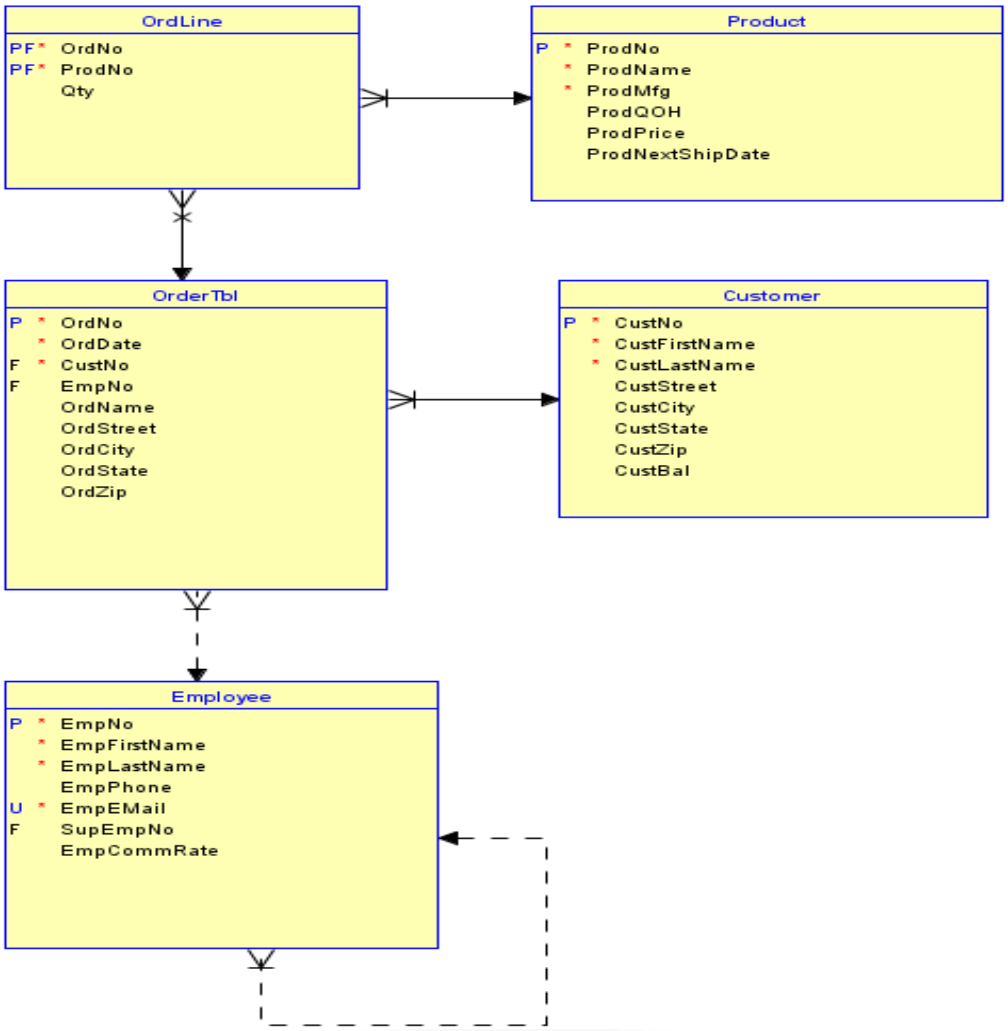
Module 6
SQL for Data Mining Input

Lesson 5: SQL Coding to Generate Matrix Input for Association Rule Mining

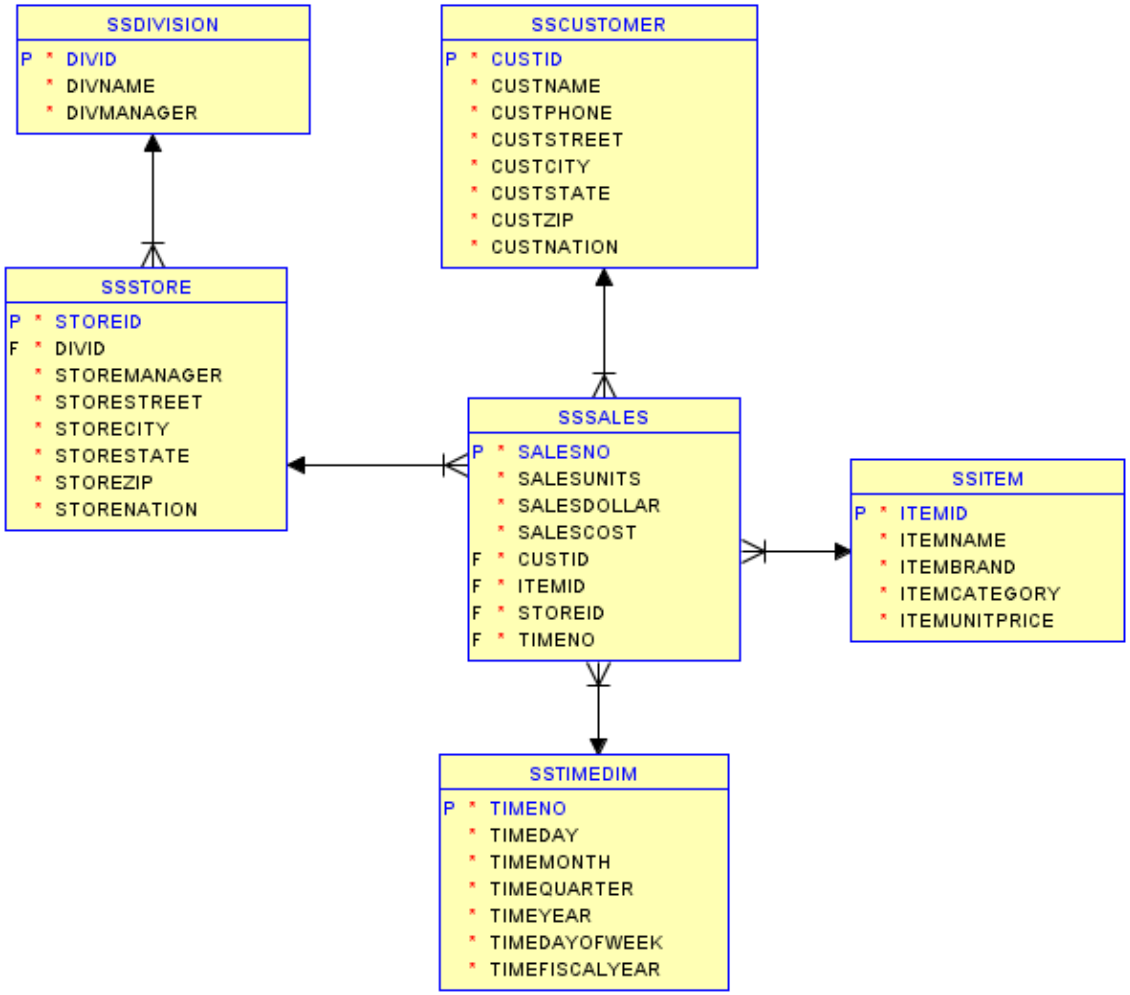
Lesson Objectives

- Apply the statement pattern to develop SELECT statements
- Use the CASE function and subqueries in the SELECT clause when appropriate
- Write SELECT statements to generate cross product results
- Discuss limitations of matrix input

Order Entry Tables (Data Lake)



Store Sales Tables (Data Warehouse)



Input Formats for Association Rule Mining

- Fixed item set size: BasketId, ItemId1, ItemId2, … ItemIdn
- Basket-Item matrix with BasketId values on rows and ItemId values on columns
- Nested representation with BasketId, {list of ItemIds}
- Parent-child relationships in SQL databases

Matrix Input Format

	Products									
Orders	*P0036566*	*P0036577*	*P1114590*	*P1412138*	*P1445671*	*P1556678*	*P3455443*	*P4200344*	*P6677900*	*P9995676*
O1231231	1	0	0	0	1	0	0	0	0	0
O1579999	0	0	0	0	0	1	0	0	1	1
O1615141	1	0	0	0	1	0	0	1	0	0
O1656777	0	0	0	0	1	1	0	0	0	0
O2233457	0	1	0	0	1	0	0	0	0	0
O2334661	1	0	0	1	0	1	0	0	0	0
O3252629	0	0	0	0	0	0	0	1	0	1
O3331222	0	0	0	1	0	1	1	0	0	0
O3377543	0	0	0	0	1	0	0	0	0	1
O4714645	1	0	0	0	0	0	0	0	0	1
O5511365	0	0	0	1	1	1	1	0	1	0
O7847172	0	0	0	0	0	1	0	0	1	0
O7959898	0	0	0	1	0	1	1	0	1	0
O7989497	0	0	1	1	1	0	0	0	0	0
O8979495	0	0	1	1	1	0	0	0	0	0
O9919699	0	1	1	0	0	0	0	1	0	0

New SQL Elements

- **CASE function**
  - Conditional assignment of result values
  - WHEN clause for condition
  - THEN for assignment when true
  - ELSE for assignment when false
- **Subqueries in SELECT**
  - IN operator with multiple columns
  - Independent of outer query
  - Inside WHEN condition

SQL for Cross Product Results (Order Entry Tables)

```
-- Example 1
-- Cross product of OrderTbl and Product tables
-- Independent subquery inside the CASE function
SELECT OrdNo, ProdNo,
 CASE WHEN (OrdNo, ProdNo) IN
  ( SELECT OrdNo, ProdNo
      FROM OrdLine )
   THEN 1 ELSE 0 END AS Basket
 FROM OrderTbl, Product
-- Optional ORDER BY clause for convenient ordering
 ORDER BY OrdNo, ProdNo;
```

SQL for Cross Product Results (Store Sales Tables)

```
-- Example 2
-- Cross product on 4 tables
-- Independent subquery inside the CASE function
SELECT CustId, TimeNo, StoreId, ItemId,
 CASE WHEN (CustId, TimeNo, StoreId, ItemId ) IN
  ( SELECT CustId, TimeNo, StoreId, ItemId
     FROM SSSales )
  THEN 1 ELSE 0 END AS Basket
 FROM SSCustomer, SSTimeDim, SSStore, SSItem
-- Optional ORDER BY clause
 ORDER BY CustId, TimeNo, StoreId, ItemId;
```

Statement Pattern for Cross Product Results

```
SELECT <BasketIdColList>, ItemIdCol,
       CASE WHEN ( <BasketIdColList>, ItemIdCol ) IN
       ( SELECT <BasketIdColList>
          FROM ItemTable )
       THEN 1 ELSE 0
 FROM  <BasketTableList>, ItemTable
 WHERE [ ( <BasketIdColList> ) IN
  ( SELECT <BasketIdColList>
     FROM ItemTable
     GROUP BY <BasketIdColList>
     HAVING COUNT(*) > 1 ) ]
 [ AND <TableConditionList> ] ;
```

Summary

- Matrix format for some association rule mining algorithms
- SQL elements for conditional assignment of column values and independent related subqueries
- Example statements and statement pattern for cross product results