



Module 6

SQL for Data Mining Input

Lesson 6: SQL Coding to Generate Nested Input
for Association Rule Mining



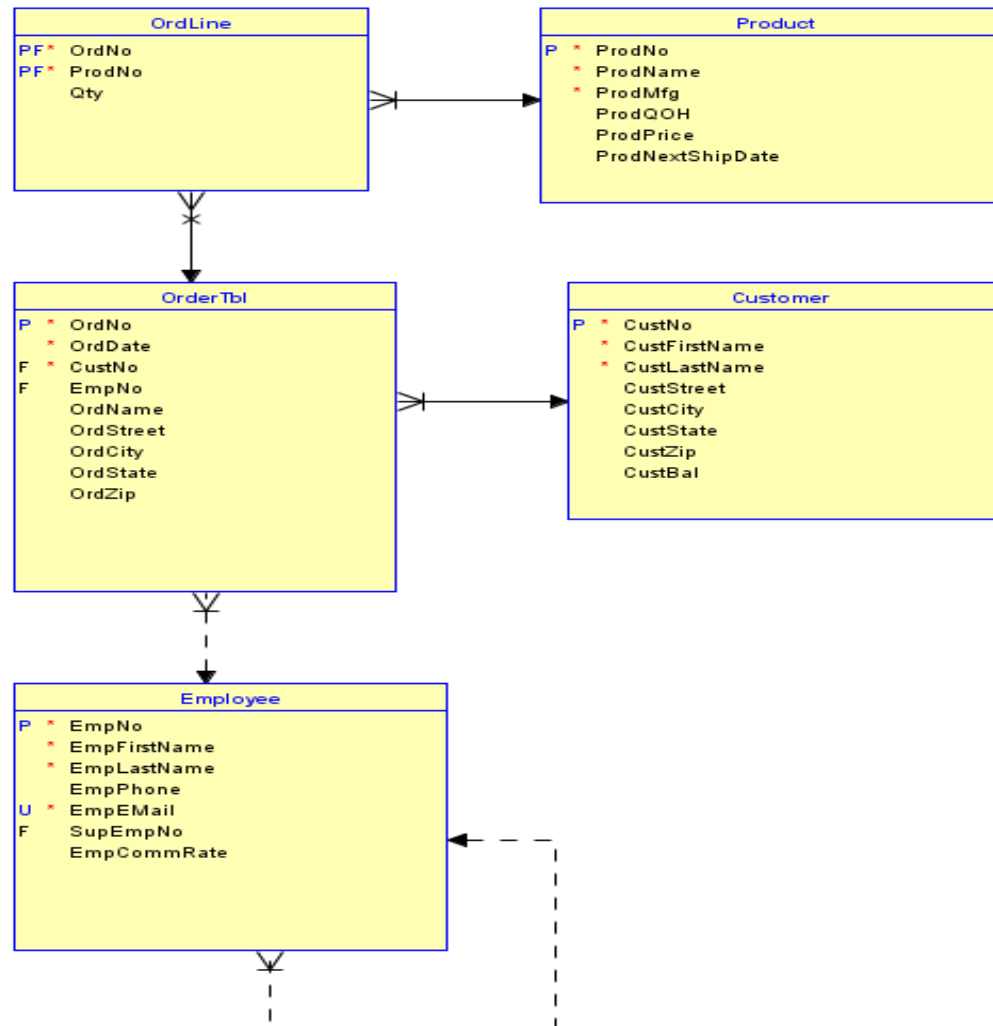
Lesson Objectives

Explain aggregate functions for array results in Oracle and PostgreSQL

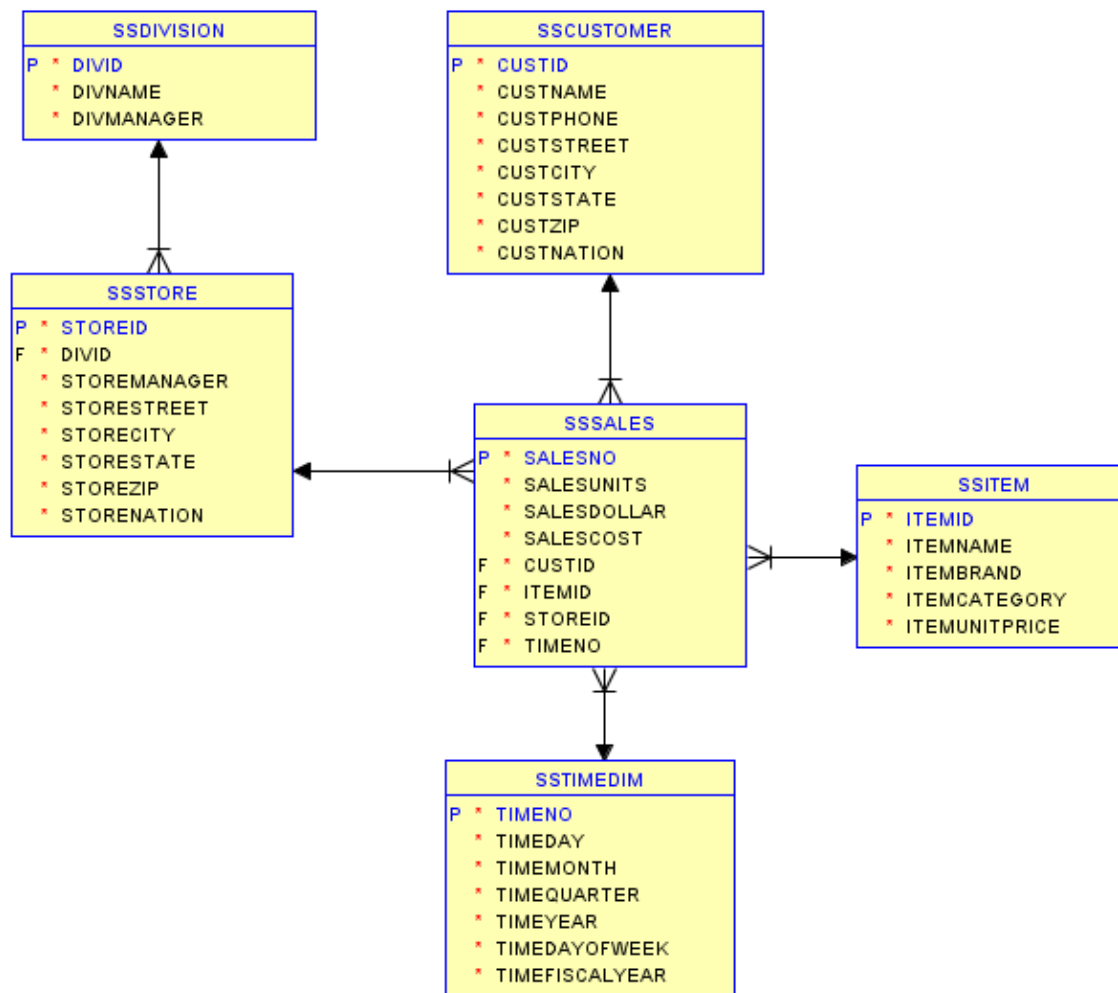
Write SELECT statements to generate array results in Oracle and PostgreSQL



Order Entry Tables (Operational Database)





Data Warehouse Diagram



Nested Input Format

- No size restrictions unlike matrix format
- Generate a nested array
- Different aggregate functions in Oracle and PostgreSQL

	ordno [PK] character (8) 	array character[] (8) 
1	01116324	{P1445671}
2	01231231	{P0036566,P1445671}
3	01241518	{P0036577}
4	01455122	{P4200344}
5	01579999	{P1556678,P6677900,P9995676}
6	01615141	{P0036566,P1445671,P4200344}
7	01656777	{P1445671,P1556678}
8	02233457	{P0036577,P1445671}
9	02334661	{P0036566,P1412138,P1556678}
10	03252629	{P4200344,P9995676}



Array Result using PostgreSQL I

```
-- Example 1 with the Order Entry Tables
SELECT OrdNo, ARRAY_AGG(ProdNo
    ORDER BY ProdNo) AS ProdNoList
FROM OrdLine
GROUP BY OrdNo
-- Optional ORDER BY clause
ORDER BY OrdNo;
```



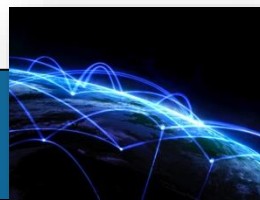
Array Result using PostgreSQL II

```
-- Example 2 with the Store Sales Tables
SELECT S1.CustId, S1.TimeNo, S1.StoreId,
       ARRAY_AGG( DISTINCT ItemId ORDER BY ItemId )
       AS ItemIdList
FROM SSSales S1
GROUP BY S1.CustId, S1.TimeNo, S1.StoreId
-- Optional ORDER BY clause
ORDER BY S1.CustId, S1.TimeNo, S1.StoreId;
```



Extended Array Result using PostgreSQL

```
-- Example 3 with order entry tables
-- HAVING clause to remove orders with only 1 product.
SELECT OrdNo, ARRAY_AGG(ProdNo ORDER BY ProdNo)
        AS ProdNoList
FROM OrdLine
GROUP BY OrdNo
HAVING COUNT(*) > 1
ORDER BY OrdNo;
```



Array Results using Oracle I

```
-- Example 4 using the Order Entry Tables
SELECT OrdNo, LISTAGG(ProdNo, ', ' )
      WITHIN GROUP (ORDER BY ProdNo)
      AS ProdNoList
FROM OrdLine
GROUP BY OrdNo
HAVING COUNT(*) > 1
ORDER BY OrdNo;
```

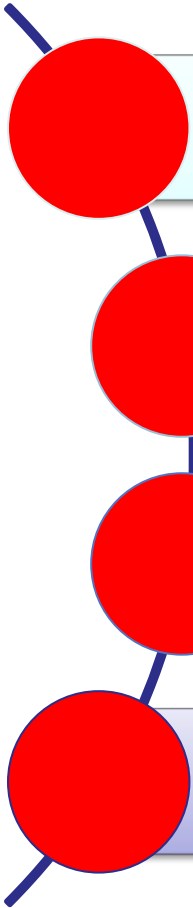


Array Results using Oracle II

```
-- Example 5 using the Store Sales Tables
SELECT S1.CustId, S1.TimeNo, S1.StoreId,
       LISTAGG( DISTINCT ItemId, ', ' )
       WITHIN GROUP (ORDER BY ItemId)
       AS ItemIdList
FROM SSSales S1
GROUP BY S1.CustId, S1.TimeNo, S1.StoreId
HAVING COUNT(DISTINCT ItemId) > 1
ORDER BY S1.CustId, S1.TimeNo, S1.StoreId;
```



Summary



Nested format with basket identifier and array of item identifiers

ARRAY_AGG function for array results in PostgreSQL

LISTAGG function for array results in Oracle SQL

Example statements using both functions

