



































 Query: 'Derive the total sales of A4-type laptop personal computers that were sold in the U.S.A. as the Summer Campaign during August, 2019.' Many Join operations between the Fact Table and the Dimension Tables are required. 	SELECT SUM(Sales) FROM Fact_Table WHERE Product_ID IN (SELECT Product_ID FROM Product_Table WHERE Product_Category = 'A4 Laptop') AND Shop_ID IN (SELECT Shop_ID FROM Shop_Table WHERE Country = 'U.S.A.') AND Promotion_ID IN (SELECT Promotion_ID FROM Promotion_Table WHERE Promotion_Category = 'Summer Campaign') AND Time_ID IN (SELECT Time_ID FROM Time_Table WHERE The_Year = 2019 AND The_Month = 'Aug.')	
2020/7/2 Adv	vance Data Engineering (©H.Yokota)	73









- Query Example 1 SELECT Product_ID, Shop_ID, SUM(Sales) FROM Fact_Table GROUP BY Product_ID, Shop_ID
- Query Example 2 SELECT Product_ID, SUM(Sales) FROM Fact_Table GROUP BY Product ID
- The result of Query Example 1 can be used for calculating Query Example 2. (Product_ID, Shop_ID) ≥ Product_ID

2020/7/2

Advance Data Engineering (©H.Yokota)

77







Smallest Parent Example								
					Product ID	Time ID	SUM(Sales)	
Product ID	Sh	op ID	SUM(Sales)	P001	'18 3Q	5	
P001	Sh	op-A	1	0	P002	'18 3Q	20	
P002	Sh	op-A	3	5	P001	'18 4Q	15	
P001	Sh	op-B	2	0	P002	'18 4Q	40	
P002	Sh	op-B	5	0	P001	'19 1Q	10	
L,			P002	'19 1Q	25			
Produ			Sales)		Product ID	Promotion ID	SUM(Sales)	
PO	P001 30							
20 P002 85 pe D		e Data Eng	ineering (©H.Yokota)	:	81			























Question (3-3)							
 Trace the results of using the Apriori algorithm on the store example with the following conditions respectively. 1. minimum support 40% and confidence 70%. 2. minimum support 30% and confidence 60% Show the candidate and large itemsets for each database scan, and indicate the derived association rules. 							
	Customer	Items					
	C1	Apple, Donut, Eggs					
	C2	Apple, Donut,					
	C3	Apple, Beer, Chips					
	C4	Beer, Chips					
	C5	Chips, Eggs					
2020/7/2	C6	Apple, Beer, Chips	93				











