Virtual University	**Assignment No. 03** **Semester: Fall 2019** **CS614: Data Warehousing**	**Total Marks: 15** **Due Date: 27-Jan-2020**

Objective:

The objective of this assignment is to enhance the learning capabilities of the students about:

- Join Techniques
- Hash based join
- Sort-Merge join

Instructions:

Please read the following instructions carefully before submitting assignment:

You need to use MS word document to prepare and submit the assignment on VU-LMS.

It should be clear that your assignment will not get any credit if:

- The assignment is submitted after due date.
- The assignment is not in the required format (.doc or docx)
- The submitted assignment does not open or file is corrupt.
- Assignment is copied(partial or full) from any source (websites, forums, students, etc)

Studen Id: MC18

Assignment

Question

HyperMart is an online shopping store currently acquiring a large number of customers.

Normalized table structures for the shopping store are given below:

Order Table

Order_ID	Order_Date
100	2-Feb-2019
201	2-Feb-2019
300	4-Feb-2019
…. …	

1

Order details Table

M

Detail_ID	Order_ID	Product_ID	Product_Quantity	Sale_Amount
100	100	2	1	100
101	100	3	2	200
200	201	5	1	50
201	201	7	1	90
300	300	15	1	600
301	300	56	1	800
302	300	57	1	850
….	…	…	….	…

Table Information

- Assume 1:12 record count ratio between Order and Order detail for online store's database
- Assume 10,000 orders

Task

For the given relations you are required to calculate the following costs in terms of I/O operations and find out which joining technique is better for given scenario. Tasks you are required to perform are as under;

1. Cost of Sort-Merge Join
2. Cost of Hash-Join
3. On the basis of your calculations, suggest better joining technique between sort-merge and hash join for the given scenario

M log M + N log N + (M+N)

= 10,000 log 10,000 + 120,000 log 120,000 + 10,000 + 120,000
= 10,000 * 4 + 120,000 (5.07918124605) + 10,000 + 120,000
= 779,501.7495

2(M + N)
= 2(10,000 + 120,000)
= 2(130,000)
= 260,000

Matching Pase cost:

M = 10,000 N= 120,000
= 10,000 + 120,000
= 130,000

Cost Of Hash-Join 260,000 + 130,000
= 390,000 I/Os Ans

Best joining techniques is Hash-Join because its cost is 390,000.

Best of Luck!