# IST769 Homework Submission Template

## Basic Information

Your Name: Wanyue Xiao

SUID: 720633297

Your Email: xwanyue@syr.edu

Date Due: April 9, 2020

Homework #: HW09

## Instructions

For each answer, please include your answer as text, and any screenshot(s) which demonstrate your answer was executed. Most importantly, make sure to include evidence your answer is correct. This will most likely be a screenshot. If you had issues, problems, or had to make assumptions include them in your answer.

## Your Answers:

1. **Scenario:** Given that Cassandra is good at storing time series data, this scenario is to create a table which records the transactions among users and vendors.

**Rules:**

1. Each user and vendor will have a unique User\_email or Vendor\_name.
2. Each product has a unique Product\_name.
3. Each vendor can provide a same type of products but the Product\_name should be different. For example, vendor A and vendor B can both sell fruit apple but with different Product\_name.
4. **I assumed that each user will not purchase a same certain product in a certain store within one day. For example, if the user ‘A’ purchase ‘Apple’ in ‘Tops’ on ‘April 9, 2020’, then A will not purchase that product at that day again!**
5. Each transaction must contain the unique user\_email, the product that purchased by user, the vendor who supplies the product, the quantity of product, the total price, and the purchase date.

**Variables:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **Example** |
| User\_email | text | JamesSmith@gmail.com |
| Product\_name | text | TOPS001 - Sweet Potato |
| Vendor\_name | text | Tops |
| Product\_qty | Int | 5 |
| Order\_price | Int | 5 |
| Promo | text | Yes |
| Purchase\_date | date | April 8, 2020 |

In this case, I assumed that the database system separate node based on user\_emial, which ensure each node only contains transaction records of a certain user.

**partition key: user\_emial**

The cluster key could be the combination of vendor\_name, product\_name and purchase date, which ensure the uniqueness of record.

**cluster key: vendor\_name, product\_name, purchase\_date**

1. Create table user\_transaction(

User\_email text,

Product\_name text,

Vendor\_name text,

Product\_qty int,

Order\_price int,

Promo text,

Purchase\_date date,

Primary((user\_email), vendor\_name, product\_name, purchase\_date)

)

**The code of creating table:**



**Describe user\_transaction;**

A screenshot of a cell phone

Description automatically generated

1. **The code of inserting data into the user\_transaction table:**

A picture containing computer

Description automatically generatedA screenshot of a cell phone

Description automatically generated

**The result:**

A screenshot of a cell phone

Description automatically generated

1. **SELECT \* FROM user\_transaction WHERE vendor\_name = ‘Tops’;**

Try to retrieve records from table by filtering the vendor\_name column. This will report an error **since vendor\_name is not a partition key but a cluster key.**



**CREATE MATERIALIZED VIEW user\_transaction\_by\_vendor AS SELECT \* FROM user\_transaction WHERE product\_name IS NOT NULL AND purchase\_date IS NOT NULL AND user\_emial IS NOT NULL and vendor\_name IS NOT NULL PRIMARY KEY (vendor\_name, user\_email, product\_name, purchase\_date);**

Then we can create a materialized view which **sets ‘vendor\_name’ as a primary key / partition key.**  


**SELECT \* FROM user\_transaction\_by\_vendor WHERE vendor\_name = ‘Tops’;**

Finally we can get the desired output.

A close up of a screen

Description automatically generated

**DESCRIBE user\_transaction;**

A screenshot of a cell phone

Description automatically generated

1. **SELECT \* FROM user\_transaction WHERE promo = ‘Yes’;**

Try to retrieve records from table by filtering the promo column. This will report an error **since vendor\_name is not a key.**



**CREATE INDEX ix\_user\_transaction\_promo ON user\_transaction (promo);**

Then we can create a index to re-order the table by using promo.

**SELECT \* FROM user\_transaction WHERE promo = ‘Yes’;**

Finally we can get the result.

A screenshot of a computer

Description automatically generated

**DESCRIBE user\_transaction;**

A screenshot of a cell phone

Description automatically generated

## Student Reflection:

To achieve the highest grade on the assignment you must be as descriptive and personal as possible with your reflection. Ask yourself the following questions.

1. What new information have I learned from completing this assignment?  
   **Cassandra, the structure of Cassandra, and the difference between partition key and cluster key.**
2. What were my challenges / roadblocks I encountered on the way to completing it?  
   **Try to get up early and pay attention to the lecture. Try to understand the exact different between partition key and cluster key. Take the last two question as an example, I am not sure that am I set a new/different partition key by using the cluster key to create a materialized view or am I just re-arrange the original table and store the result to a view?**
3. To be better prepared to attempt this assignment I should \_\_\_\_\_\_\_ ?

**Review before class.**

1. If I could make this assignment better, I would \_\_\_\_\_?

**Try to focus!!!**

1. Rate your comfort level with completing this assignment:

**1**  
1 ==> I can do this on my own and explain how to do it.

2 ==> I can do this on my own without any help.

3 ==> I can do this with help or guidance from others. If you choose this level, please list those who helped you.

4 ==> I don't understand this at all yet and need extra help. If you choose this, please try to articulate that which you do not understand.