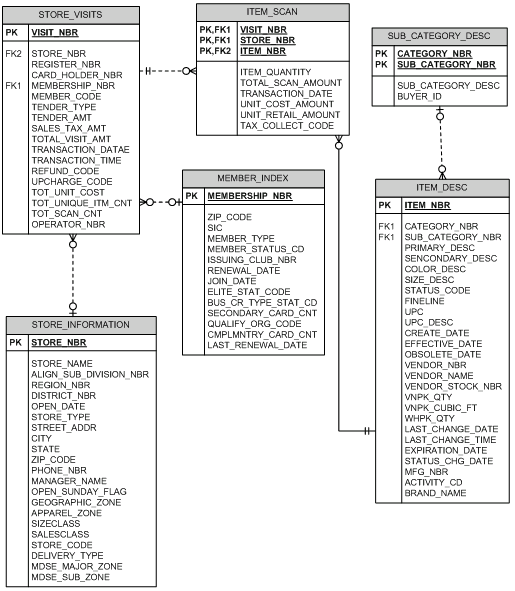
Name: Thien Nguyen & Shen Yang

Assignment 3 Due: April 11, 2014 (11pm)

This assignment will give you some experience with association mining on a large data warehouse. You will use Teradata's Warehouse Miner within a remote desktop environment. You will use data from the Sam's Club database that contains over 1 million records. The database schema diagram follows.



**Note:** In the STORE\_VISITS table, the column TRANSACTION\_DATAE should be TRANSACTION\_DATE.

Group members will work together to perform a marketbasket analysis on a subset of the Sam's Club database. The goal of the analysis is to give us insight into the relationships between items purchased together in a sales transaction (market basket).

To accomplish this, we will use Teradata's Warehouse Miner. We will need to perform the following tasks:

* install the remote desktop for the Teradata Warehouse Miner (see Appendix A)
* build our analytical data set (see Appendix B)
* run the association mining algorithm on our data set (see Appendix A)

For the second task, we want to build our market basket data using items that have been purchased in the month of January 2000 from stores with Store\_nbr of 22, 24 and 26. We will use the UPC\_DESC to identify the items. In our analysis we want to display the description of the item, i.e., the UPC\_DESC, so that will be needed in our market basket data. It may be useful to define this table as a view so that we don't exceed allowable storage for our tables. The resulting table (view) should contain 41,174,279 rows.

SQL:

DROP VIEW "ES86673"."store24";

CREATE VIEW "ES86673"."store24" AS (

SELECT

"\_twmads0"."UPC\_Desc"

,"\_twmads1"."transaction\_Date"

,"\_twmads1"."Store\_Nbr"

,"\_twmads1"."Visit\_Nbr"

FROM "ua\_samsclub"."item\_desc" AS "\_twmads0"

LEFT OUTER JOIN "ua\_samsclub"."item\_scan" AS "\_twmads1"

ON "\_twmads0"."Item\_Nbr" = "\_twmads1"."Item\_Nbr"

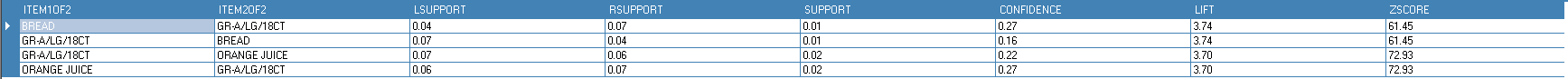
WHERE extract(year from transaction\_date) = 2000 and extract(month from transaction\_date) = 1 and store\_nbr = 22 or store\_nbr = 24 or store\_nbr = 26

);

COMMENT ON "ES86673"."store24" 'TWM : ADS : Project 1';

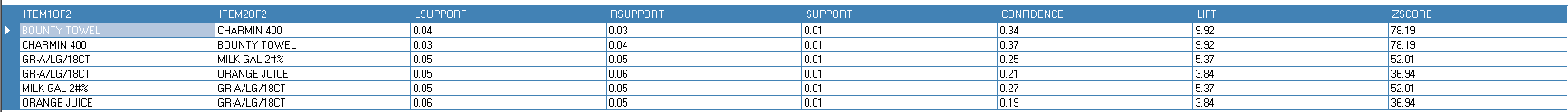
For the third task, we will use the default parameter settings for support and confidence. We will produce pair wise associations between items that have a support of at least 0.01. That is, the associations should hold in at least 1% of the transactions (or visits). We will run 4 different analyses as described below:

1. Do association mining, consider the stores: 22, 24 and 26 for month of January

**Number of Visit\_Nbr:** 167,779

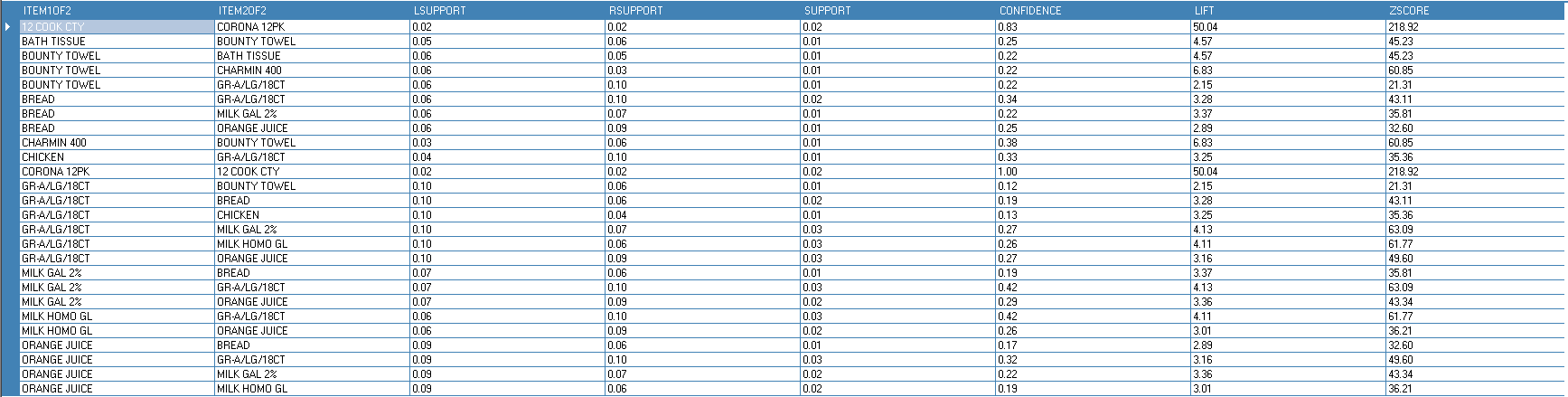
1. Do association mining, consider the store 22 for month of January

**Number of Visit\_Nbr:** 60,834



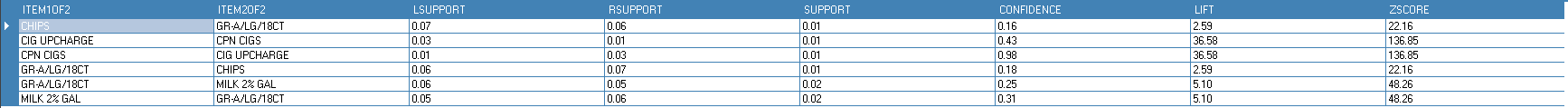
1. Do association mining, consider the store 24 for month of January

**Number of Visit\_Nbr:** 60,043



1. Do association mining, consider the store 26 for month of January

**Number of Visit\_Nbr:** 46,902



This result of each analysis should be stored as a word document and submitted by the due date for each group. Also for each analysis, include the number of transactions that are used in each. You may also be requested to demo your assignment.