

# BD51

## Business Intelligence & Data Warehouse

### Final examination A14

Writer : Christian FISCHER

Conditions of realization:

Authorized documents: BD51 Course and TP Supports

Duration: 2h

Notation: The subject understands 2 independent parts

Part 1	SQL OLAP Queries for the SNCF	
<b>12 pts</b>	To write 8 SQL queries	12 pts
Part 2	Study and Modeling the ESN Data Warehouse	
<b>8 pts</b>	Star model Management of the calendar Software architecture	4 pts 2 pts 2 pts

### Part 1 : Interrogation of the data warehouse on the trains frequentation in Paris region

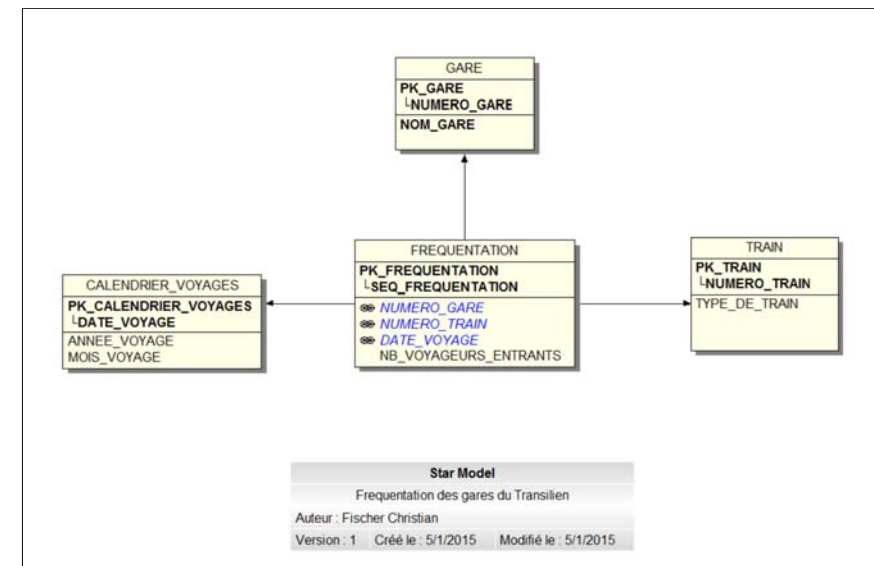
*Presentation of the context*

For the SNCF, the valorization of the data is a lever of innovation for more informed, fluid and personalized mobility. Already invested in the Open DATA the SNCF proposes today a challenge of data science whose objective is to build a predictive model of the frequentation of the train stations.

The team charged of the data warehouse recently examined the current systems design of analysis.

Mr. Fischer has validated the decisional model describing the information system relating to the frequentation of the stations.

Star model validated



Vocabulary :

Gare = Train Station

Fréquentation = Frequentation

This model is managed by a base Oracle 11G Enterprise Edition with the partitioning option.

SEQ\_FREQUENTATION = Oracle sequence to identify each frequentation per day, station and train.

**Work to achieve :**

Write the following SQL queries.

Use ANSI joins, alias of tables and alias of columns.

**Query 1 :** (0,5 pt)

Calculate the number of travellers entering (*nombre de voyageurs entrants*) per year for 2014 and 2015

**Query 2 :** (1 pt)

Calculate the number of travellers entering per year and month for the year 2014

**Query 3 :** (1 pt)

Calculate the number of travellers entering per year and month for the year 2014 and 2015. Include the annual total and the grand total in the result using the Grouping Sets clause.

**Query 4 :** (1,5 pt)

Calculate the number of travellers per station for the year 2014.

The result will have to comprise the name of the station, the number of travellers and the ranking of each station according to its frequentation.

**Exemple de résultat :**

	Gare	Nb_entrants	Rang
1	LYCEE HENRI SELLIER	32116112474	1
2	LA REMISE A JORELLE	32116054413	2
3	NEMOURS SAINT-PIERRE	32007110740	3
4	MORET VENEUX LES SABLONS	32006436220	4
5	LE MEE	32006400085	5
6	PARIS GARE DE LYON	32006362802	6
7	VILLENEUVE TRIAGE	32006265035	7
8	MOULIN GALANT	32006116845	8
9	LE BRAS DE FER (EVRY GENOPOLE)	32006113925	9
10	SOUPPES CHATEAUN LONDON	32005220074	10
11	EVRY COURCOURONNES (CENTRE)	32004908546	11
12	BOIGNEVILLE	32004822878	12
13	VILLENEUVE SAINT-GEORGES	32004779861	13
14	BAGNEAUX SUR LOING	32004734470	14
15	MONTIGNY SUR LOING	32004705998	15

**Query 5 :** (1,5 pt)

Calculate the number of travellers for the year 2014 and 2015 and per type of train.

The result will have to comprise the year, the type of train, the number of entering and the ranking according to the frequentation

The result will be sorted on the number of entering.

**Exemple de résultat :**

	Année	TYPE_DE_TRAIN	nb_entrants	rang
1	2015	RER	2115982652996	1
2	2015	Train SNCF Banlieue	2081854303667	2
3	2015	Tramway	2081854303667	2
4	2015	TGV Duplex	2081854303667	2
5	2015	RER Rénové	2081854303667	2
6	2014	RER	2064373319844	6
7	2014	Tramway	2031077369289	7
8	2014	Train SNCF Banlieue	2031077369289	7
9	2014	RER Rénové	2031077369289	7
10	2014	TGV Duplex	2031077369289	7
11	2015	TGV	2013597591230	11
12	2014	TGV	1964485454721	12

**Query 6 :** (1,5 pt)

Calculate the number of travellers for the year 2014 per type of train.

The result will have to comprise the year, the type of train, the number of entering and the ranking according to the frequentation (with and without equality)

The result will be sorted on the number of entering.

**Exemple de résultat :**

	TYPE_DE_TRAIN	nb_entrants	rang	drang
1	RER	2064373319844	1	1
2	RER Rénové	2031077369289	2	2
3	TGV Duplex	2031077369289	2	2
4	Train SNCF Banlieue	2031077369289	2	2
5	Tramway	2031077369289	2	2
6	TGV	1964485454721	6	3

**Query 7 : (2 pt)**

Step 1:

Calculate the number of travellers for the year 2014 per month.

The result will have to comprise the month and the number of entering.

Step 2 :

From the preceding query add a cumulative running sum over 6 months and an average over 6 months.

Example of result:

*Note: in the extract the frequentations are identical for certain months*

	mois	nb_entrants	cumul_6mois	moy_6mois
1	1	1032186659922	1032186659922	1032186659922
2	2	932298794799	1964485454721	982242727360
3	3	1032186659922	2996672114643	998890704881
4	4	998890709367	3995562824010	998890706002
5	5	1032186659922	5027749483932	1005549896786
6	6	998890709367	6026640193299	1004440032216
7	7	1032186659922	6026640193299	1004440032216
8	8	1032186659922	6126528058422	1021088009737
9	9	998890709367	6093232107867	1015538684644
10	10	1032186659922	6126528058422	1021088009737
11	11	998890709367	6093232107867	1015538684644
12	12	1032186659922	6126528058422	1021088009737

**Query 8 : (3 pt)**

Step 1 :

Calculate the number of travellers per year of voyage and type of train for the years 2014 and 2015

Only the trains of the type TGV and the RER must be included in the result.

The result is sorted per Year and Type of train

Example of result:

	ANNEE_VOYA	TYPE_DE_TR	Nb_entrants
1	2014	RER	2064373319844
2	2014	TGV	1964485454721
3	2014	TGV Duplex	2031077369289
4	2015	RER	2115982652996
5	2015	TGV	2013597591230
6	2015	TGV Duplex	2081854303667

Step 2 :

Use of the operator PIVOT to obtain only one line per year with in columns the number of travellers per type of train.

The total of the travellers by type of train TGV and the grand total must be calculated.

Example of result:

	Année	TGV Simple	TGV Duplex	Total TGV	RER	Total
1	2014	1964485454721	2031077369289	3995562824010	2064373319844	6059936143854
2	2015	2013597591230	2081854303667	4095451894897	2115982652996	6211434547893

## Part 2: Modeling of ESN

### Presentation of the context

Conscious of the importance of modifying our nutritional practices and of protecting our environment, company ESN developed the ensign Espace Natural Health which offers a broad product range resulting from the organic farming, labellized and controlled by approved organizations.

The sales network of the shops Espace Natural Health offers food substances of everyday consumption certified, but also of the food supplements.

These products, known as “nutraceutic”, a beneficial physiological effect or guard against the chronic diseases has. They are manufactured starting from food substances and are marketed in the form of tablet, of powder, potion or other medicinal forms.

In addition, ESN offers natural éco-product and biological cosmetic, upkeeping products for the house (without chemical products) (cream of day and night, shower gel, shampoo) many such as crockery compostable, layers for biodegradable with 100% and not bleached babies.

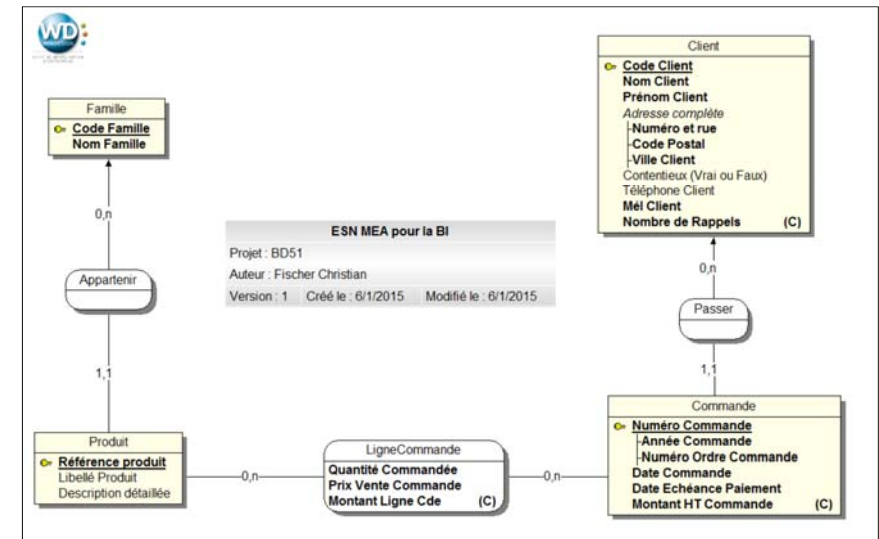
Encouraged by the rise of the market of nutraceutic, the company decided to make evolve its information system to accompany the development by its ensign Espace Natural Health.

The company wishes today to adapt its [www.espace-sante-nature.com](http://www.espace-sante-nature.com) site in order to market its new products nutraceutic. The development of a dynamic site is based on an three thirds (3 tiers) architecture. This architecture requires the access to the database which manages stocks and the marketing of the products.

The installation of reporting tools will have to make it possible to follow the sales per day (in J+1 the data of sales of the site went up in a data warehouse).

The follow-up of the sales will be carried out per day, week, month, quarter and year. The promotional periods must also be managed.

## Model entity association of the current database



### Work to achieve :

1. From the MEA of the current base, to provide a star model used by the decisional application.
2. Define the structure of the sales calendar table and explain the principle of its automatic construction.
3. Suggest and describe a software solution to implement the decisional reporting solution.