UNIVERSITÉ DE TECHNOLOGIE DE BELFORT-MONTBÉLIARD

BD51

Business Intelligence & Data Warehouse

Final Exam A13

Redactor : Christian FISCHER

Département Informatique **Automne 2013**

Working conditions :
Documentation alloewed : BD51 Course and Lab Book.
Duration : 2 h

Evaluation : The subject includes 2 separate parts.

Dossier 1 **12 pts**	SQL OLAP Queries 1 : Writing SQL queries 2 : Create a partitionned table	 10 pts 2 pts
Dossier 2 **8 pts**	Design and Modeling of Data Warehouse	2 pts per question

Part 1: Querying the RADAR data warehouse

Global Context

The Ministry of Interior has decided to call "for the Experts at Belfort" for the implementation of decision support and control of speed cameras set up for the management of road safety system.

Information on the performance of radars

Radar Saint-Julien en Genevois on Highway A41 in the direction Annecy to Geneva, located in the department of Haute-Savoie (74) carried 155,304 flashes (or Minutes = *Procés Verbal in french*) since its release . This radar takes 462 flashes per day.

Dashboard example (source Le figaro) for the 2012 performance

21,2 millions de flashs enregistrés en 2012

LES 10 CABINES-RADARS* LES PLUS ACTIVES

	DÉPT.	ROUTE	LIEU ET SENS DE CIRCULATION	VITESSE LIMITE	Flashs/jour
1er	37	A10	À Tours, Bordeaux > Paris	90 Km/h	444
2e	91	A10	À Janvry, Bordeaux > Paris	110	316
3e	84	A7	À Bédarrides, Lyon > Marseille	110	303
4e	73	RN201	À Barberaz, Chambéry > Albertville	90	302
5e	94	A86	PK25, M.-Alfort > Ch.-Le-Roi	90	300
6e	74	A41	PK159, Annecy > Genève	50	286
7e	34	A9	PK91, Nîmes > Béziers	110	280
8e	59	A1	PK206, Lille > Paris	90	275
9e	92	RD131	À Nanterre, Rueil-M. > La G.-Colombes	50	275
10e	59	A1	PK207, Paris > Lille	90	257

*Enregistrés par les cabines-radars ETF et ETFD

Source : Auto Plus

LES 10 DÉPARTEMENTS OÙ LES RADARS FLASHENT LE PLUS

Nombre de flashs totals enregistrés* par an

1er	Rhône	826 520
2e	Nord	821 854
3e	Alpes-Maritimes	758 438
4e	Essonne	715 236
5e	Yvelines	598 078
6e	B.-du-Rhône	528 790
7e	Moselle	517 538
8e	Val-de-Marne	516 895
9e	Paris	497 394
10e	Var	495 046

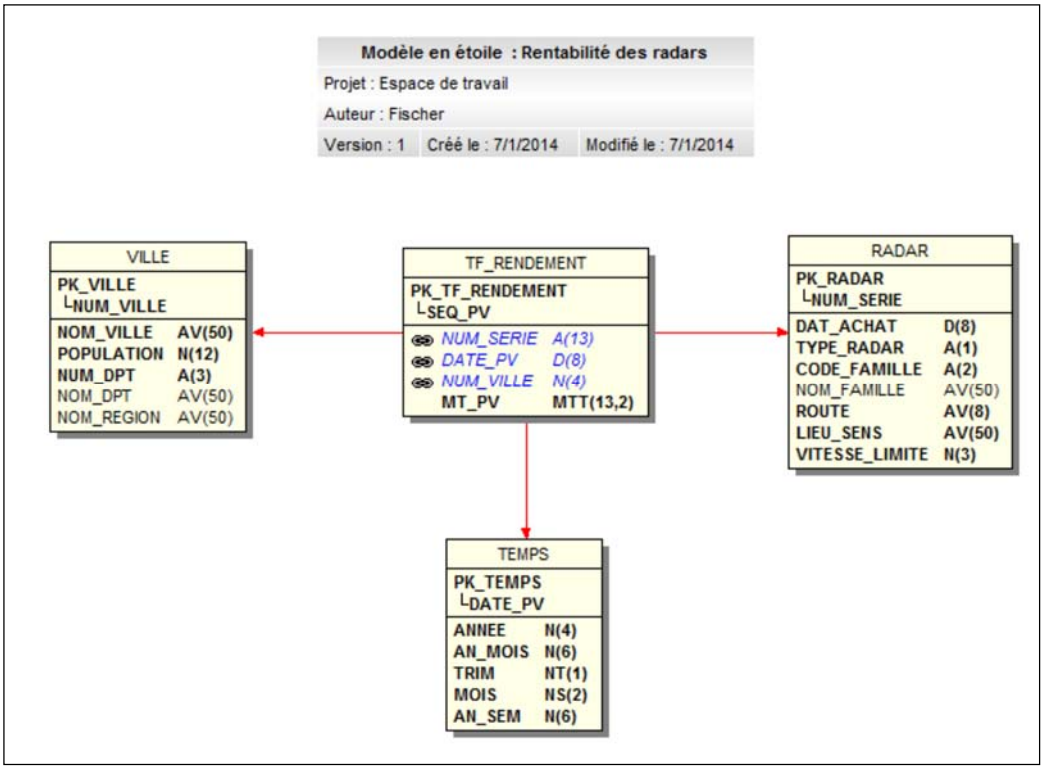
*Enregistrés par tous les types de radars

02/04/2013

LE FIGARO.fr

The team responsible for the data warehouse has recently reviewed the design of the current analysis system.
Mr Fischer C. validated the decision model describing the information on the profitability of radar system.

Validate Star Model
Ville = City, Temps = Period, TF_rendement = Performance fact table



This model is managed by an Oracle 11g Enterprise Edition Server with Table Partitioning option.

Additionnal model information
MT_PV = Minutes (Amount_PV) amount calculated from the recorded speed and the speed limit during daily PV loading in the data warehouse.
SEQ_PV = Oracle sequence to identify each PV or flash made by a radar.

Work to do : Writing the above SQL Queries.

Use ANSI joins, table aliases and column aliases.

Query 1 : (0,5 pt)
Calculate the number of flash (or minutes = PV) in the fact table

Query 2 : (0,5 pt)
Calculate the number of flash (or minutes = PV) by year.
Sorting in decreasing the number of PV.

Query 3 : (1 pt)
Calculate the number and the amount of flash (or minutes = PV) for all cities.
Sorting the amount decreasing and increasing on the city in case of equality.
The flash number must be equal to à 0 for the cities without any flash.

Query 4 : (1 pt)
Calculate the amount of flash by department, the total amount by region and the general total
The result is sorted by region and department.

Query 5 : (1 pt)
For the year 2013, calculate the number and amount of flashes generated per month
The result will be sorted by months.

Query 6 : (1 pt)
For the year 2013 flashes, calculate the flash amount by city and family radar.
Add the total amount per family radar for all cities.
Add the total amount.

Query 7 : (1 pt)
For the year 2013 flashes, calculate the amount generated per month, and the amount of cumulative flash (or PV) from January to December.
The result will be sorted by months.

Query 8 : (1,5 pt)
Reproduce the first table (from the Figaro source), representing the 10 most active radar for year 2013.
Display the rank, department number, road, traffic direction, speed and number of flashes (or PV) performed daily.

Query 9 : (1,5 pt)
Reproduce the second table (from the Figaro source), representing the 10 departments with the flashes for year 2013.
Display the rank, the department name, and the number of flashes (or PV).

Query Optimisation

The TF_RENDEMENT table contents data since 27 octobre 2003.
This table must be partitioned by calendar year until the end of 2014.

Work to do : (2 pts)

Write the SQL command to create the partitioning TF_RENDEMENT table using an interval partitioning.

Part 2 : Attendance of the Paris Metro

The Paris Metro is a transit system serving the city of Paris and its suburbs. Operated by the Autonomous Paris Transport Authority (RATP), it has sixteen lines, mainly underground, totaling 219.9 km and serves more than 300 stations.

Lines table

Ligne	Parcours	Mise en service	Longueur en km[8] (*en surface*)	Nombre de stations	Matériel	Voitures par rame	Rames (en pointe)
1	La Défense ↔ Château de Vincennes	1900	16,6 (*0,6*)	25	MP 05	6	45
2	Porte Dauphine ↔ Nation	1900	12,3 (*2,2*)	25	MF 01	5	37
3	Pont de Levallois ↔ Gallieni	1904	11,7	25	MF 67	5	40
3bis	Gambetta ↔ Porte des Lilas	1921	1,3	4	MF 67	3	4
4	Porte de Clignancourt ↔ Mairie de Montrouge	1908	12,1	27	MP 89 CC	6	40
5	Bobigny - Pablo Picasso ↔ Place d'Italie	1906	14,6 (*2,5*)	22	MF 01	5	45
6	Charles de Gaulle - Étoile ↔ Nation	1907	13,7 (*6,1*)	28	MP 73	5	37
7	La Courneuve - 8 mai 1945 ↔ Villejuif - Louis Aragon / Mairie d'Ivry	1910	18,6	38	MF 77	5	60
7bis	Louis Blanc ↔ Pré Saint-Gervais	1911	3,1	8	MF 88	3	6
8	Balard ↔ Pointe du Lac	1913	23,4 (*4,1*)	38	MF 77	5	50
9	Pont de Sèvres ↔ Mairie de Montreuil	1922	19,6	37	MF 67 MF 01 (en test)	5	57
10	Boulogne - Pont de Saint-Cloud ↔ Gare d'Austerlitz	1913	11,7	23	MF 67	5	22
11	Châtelet ↔ Mairie des Lilas	1935	6,3	13	MP 59 MP 73	4	20
12	Front Populaire ↔ Mairie d'Issy	1910	15,3	29	MF 67	5	37
13	Asnières Gennevilliers Les Courtilles / Saint-Denis - Université ↔ Châtillon - Montrouge	1911	24,3 (*2,4*)	32	MF 77	5	50
14	Saint-Lazare ↔ Olympiades	1998	9,2	9	MP 89 CA	6 (à terme : 8)	18

Under the Open Data program, the commercial service RATP offers free following information:

1. Dataset: Geographical Positions stations RATP network

This file contains the geographic location of stations and stopping points of the RATP network in Ile-de-France.

Details of the data

Column	Description
A	Station ID / Stop point ID
B	Geographical coordinates (X) in WGS84 format
C	Coordonnée géographique (Y) au format WGS84
D	Geographical coordinate (Y) in WGS84 format
E	Common (and district) of station / stop point
F	Connected network (metro/rer/bus)

Complement:

- When it comes to a station with multiple access, the geographical position is filled at the barycenter access the corresponding station.
- The geographical coordinates are in WGS 84 format.

Sample Excel data integration

IdStation_PtArret	Coordonnée géographique (X)	Coordonnée géographique (Y)	NomStation_PtArrêt	Ville	Réseau
1975	2.33871281165883	48.8844176451841	Abbesses	PARIS-18EME	metro
1981	2.32674567371924	48.828398514348	Alésia	PARIS-14EME	metro
1978	2.3949898158233	48.8561744489676	Alexandre-Dumas	PARIS-11EME	metro
1980	2.30088336611718	48.8646460306735	Alma-Marceau	PARIS-16EME	metro
1982	2.285373878396	48.8924723754181	Anatole-France	LEVALLOIS-PERRET	metro
11037	2.30131372490699	48.7547197227554	Antony	ANTONY	metro
1984	2.34397282520569	48.8831322902739	Anvers	PARIS-18EME	metro
1986	2.2893216639617	48.875533092935	Argentine	PARIS-17EME	metro
1988	2.35637778086357	48.8652929436681	Arts-et-Métiers	PARIS-03EME	metro
225437	2.28407447167184	48.9301634125573	Asnieres-Gennevilliers Les Courtilles	ASNIERES-SUR-SEINE	metro
1989	2.32063223779825	48.8608996856827	Assemblée Nationale	PARIS-07EME	metro
1992	2.3920219501423	48.9038299162483	Aubervilliers Pantin (4 Chemins)	AUBERVILLIERS	metro
1994	2.29493807524416	48.8469642745973	Avenue Emile-Zola	PARIS-15EME	metro
1995	2.39798013009568	48.8509782113622	Avron	PARIS-11EME	metro
1997	2.27821832700082	48.8362754634973	Balard	PARIS-15EME	metro
1999	2.34973719426844	48.8836207315381	Barbès-Rochechouart	PARIS-09EME	metro

…

Displaying metro and railway station:

IdStation_PtArret	Coordonnée géographique (X)	Coordonnée géographique (Y)	NomStation_PtArrêt	Ville	Réseau
1838	2.36431304821499	48.8431055835176	Gare d'Austerlitz	PARIS-13EME	metro
1841	2.35791867724751	48.8760613055252	Gare de l'Est (Verdun)	PARIS-10EME	metro
1839	2.37406590751466	48.8447045450658	Gare de Lyon	PARIS-12EME	metro
1842	2.35470307836603	48.8799654432891	Gare du Nord	PARIS-10EME	metro

2. Dataset: Positions of stop kpoints on a line

IdStation_PtArret	Ligne desservant arrêt X (Y)	Réseau
1862	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1866	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1632	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
2018	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1770	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1908	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1838	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1874	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1911	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
2019	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1643	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1942	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
54244	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1836	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1994	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1812	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1950	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1717	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1834	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1631	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1837	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1746	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1952	10 (Gare d'Austerlitz/Boulogne - Pont de Saint Cloud)	metro
1679	11 (Châtelet/Mairie des Lilas)	metro
55329	11 (Châtelet/Mairie des Lilas)	metro
1964	11 (Châtelet/Mairie des Lilas)	metro
1855	11 (Châtelet/Mairie des Lilas)	metro
1745	11 (Châtelet/Mairie des Lilas)	metro

3. Dataset : stations attendance

Details: Data file in Excel format attendance metro lines for 2011 (extract limited to 20 stations of the busiest subway).

All metro stations are present in the file.

Trafic des entrants (*) par station pour l'ensemble de l'année 2011

Rang	Réseau	Station	Trafic	Ligne de correspondances					Ville	Arrondissement pour Paris
1	Métro	GARE DU NORD	48 146 629	4	5				Paris	10
2	Métro	SAINT-LAZARE	46 790 941	3	9	12	13	14	Paris	08
3	Métro	GARE DE LYON	34 523 049	1	14	A			Paris	12
4	Métro	MONTPARNASSE-BIENVENUE	31 152 275	4	6	12	13		Paris	14
5	Métro	GARE DE L'EST	19 671 320	4	5	7			Paris	10
6	Métro	REPUBLIQUE	17 095 806	3	5	8	9	11	Paris	11
7	Métro	BIBLIOTHEQUE	15 826 727	14					Paris	13
8	Métro	CHATELET	14 440 964	1	4	7	11	14	Paris	01
9	Métro	LA DEFENSE	13 853 216	1					Puteaux	
10	Métro	LES HALLES	13 113 834	4					Paris	01
11	Métro	HOTEL DE VILLE	12 760 823	1	11				Paris	04
12	Métro	FRANKLIN D. ROOSEVELT	12 640 577	1	9				Paris	08
13	Métro	BASTILLE	12 517 181	1	5	8			Paris	12
14	Métro	OPERA	12 389 715	3	7	8			Paris	09
15	Métro	PLACE D'ITALIE	12 168 442	5	6	7			Paris	13
16	Métro	BELLEVILLE	11 794 952	2	11				Paris	10
17	Métro	PORTE D'ORLEANS	9 598 097	4					Paris	14
18	Métro	PALAIS-ROYAL	9 559 535	1	7				Paris	01
19	Métro	NATION	9 481 963	1	2	6	9		Paris	12
20	Métro	BARBES-ROCHECHOUART	9 323 474	2	4				Paris	18

The **Trafic Column** corresponds to the traffic entering in a station traffic: number of passengers validating a ticket in the station concerned.

Travelers' correspondence on the RATP network are not counted.

From this information you must represent attendance per station as follows:

Example for the entire network:

Example for lines serving the Gare du Nord station.

Work to do :

1. Analyze the consistency and quality of data available RATP through different files and identify problems to be solved in order to establish the mapping attendance.

2. Explain the procedure to create a data warehouse to easily obtain the expected results.

In this perspective an annual attendance loading (ETL) stations the past year (2013 in this case) must be made at the beginning of each year (2014).

Propose a solution to perform this operation.

3. Provide a star model to manage attendance per station / and lines.

4. Propose a software solution to represent the mapping of the attendance of the Paris Metro.