

Master 1 ALMA
Université de Nantes
2010-2011

Projet Olap Entrepôt de données

MARGUERITE Alain
RINCE Romain

Université de Nantes
2 rue de la Houssinière,
BP92208, F-44322 Nantes cedex 03, FRANCE

Encadrant : Patricia SERRANO

Table des matières

1	Conception de l'entrepôt	2
1.1	Sujet d'étude	2
1.2	Tables et dimensions	3
2	Consultation de l'entrepôt	4
2.1	Requête 1 : GROUP BY	4
2.2	Requête 2 : GROUP BY CUBE	4
2.3	Requête 3 : Rank	4
2.4	Requête 4 : GROUPING SETS	5
2.5	Requête 5 : GROUPING and GROUP BY ROLLUP	5
2.6	Requête 6 : TOP N	5
2.7	Requête 7 : NTIL	6
2.8	Requête 8 : WINDOW	6
2.9	Requête 9 : PARTITION BY	6

1 Conception de l'entrepôt

1.1 Sujet d'étude

Le domaine étudié dans ce projet est celui d'un réseau de transport en commun d'une grande ville. La mise en place et le maintien d'un tel réseau nécessite la connaissance d'une immense quantité d'informations. Un entrepôt de données OLAP semble tout à fait adapté pour un tel usage puisqu'il permet le stockage et l'accès à l'information d'une très grande base de données. Le choix du schéma en étoile assure une certaine simplicité de création et d'utilisation de la base de données. De plus nos choix de conceptions vont permettre une double utilité à cet entrepôt :

- **Une aide à la décision quant à l'aménagement et le l'entretien du réseau.**

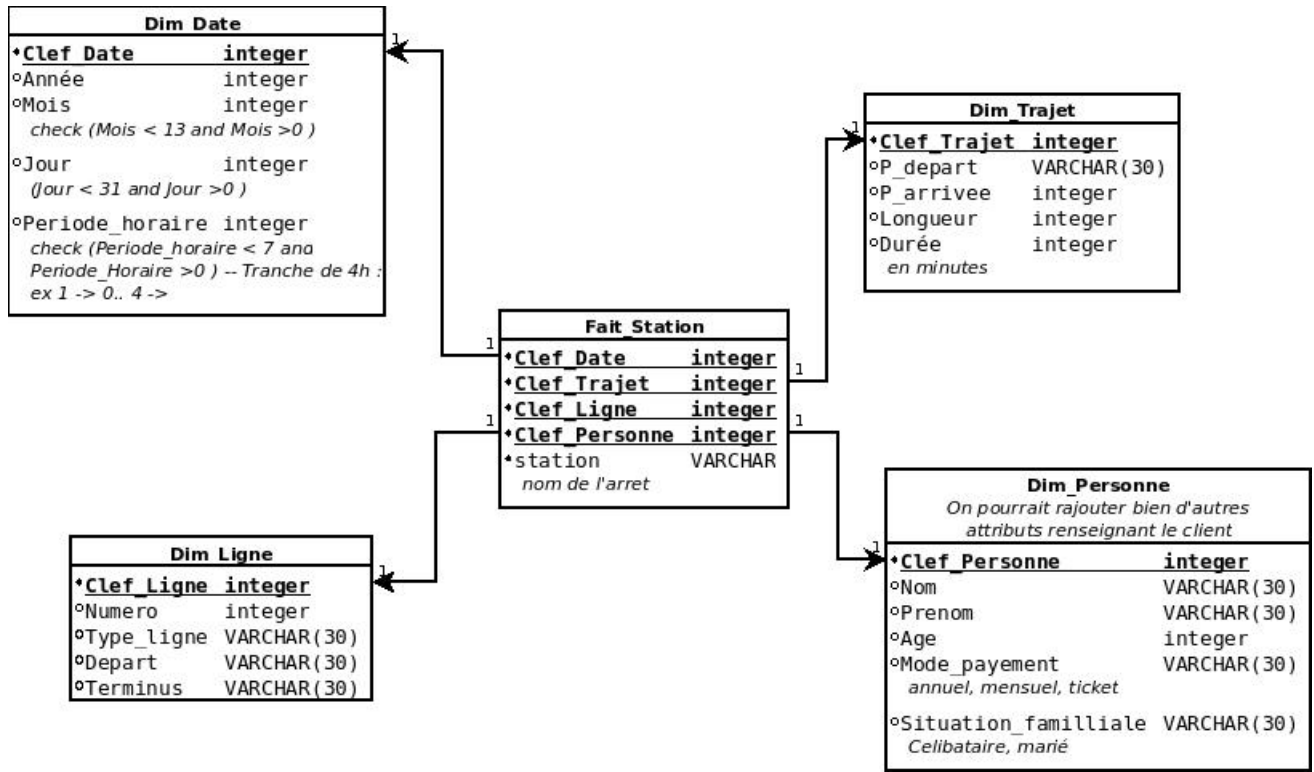
En effet, cet entrepôt sera en mesure de renseigner l'organisme responsable du réseau de nombreux statistiques et d'informations quant à la fréquentation des utilisateurs. Il sera par exemple possible d'avoir une idée de quelles zones requièrent le plus d'aménagements par rapport au taux de fréquentations. L'organisme aura donc toujours connaissance de l'état d'utilisation de son parc de manière globale.

- **Une génération d'informations précieuse d'un point de vue commerciale.**

Toutes les données nécessaires pour établir une stratégie commerciale (profils clients, zone dynamiques ou non ...) pourront être fournies facilement par l'entrepôt présenté dans ce rapport. Cet aspect est non négligeable et permet d'obtenir des informations sur les clients sans les importuner par des enquêtes.

1.2 Tables et dimensions

L'entrepôt sera composé de 4 tables de dimensions et d'une table de faits :



La granularité choisie permet de connaître chaque utilisation d'une station (arrêt de tram ou de bus ...). Ainsi lorsque qu'un utilisateur à une station donnée, monte, sors ou reste dans le véhicule, le fait est enregistré dans la table de faits. La structure de la base de données peut alors fournir un grand nombre d'informations sur ce fait : à quel trajet il appartient, sur quelle ligne ect...

2 Consultation de l'entrepôt

2.1 Requête 1 : GROUP BY

Nombre de stations traversées en 1 journée ;

```
SELECT Station , count(*)
FROM Fait_Station F, Dim_Date DD
WHERE DD.Clef_Date=18 and F.Clef_Date=DD.Clef_Date
GROUP BY Station;
```

2.2 Requête 2 : GROUP BY CUBE

```
SELECT
DL.numero
,DD.Clef_Date
,DP.Nom
,COUNT(Station) AS Nb_stations
FROM Fait_Station F, Dim_Date DD, Dim_Personne DP, Dim_Ligne DL
WHERE F.Clef_Date=DD.Clef_Date and F.Clef_Personne=DP.Clef_Personne and F.Clef_Ligne=DL.
GROUP BY CUBE(DL.numero,DD.Clef_Date,DP.Nom)
ORDER BY DL.Numero,DD.Clef_Date,DP.Nom;
```

2.3 Requête 3 : Rank

Donne les 10 dernières stations traversées

```
SELECT * FROM(
SELECT
```

```

RANK() OVER (ORDER BY DD.Clef_Date DESC) AS ranking,
F.Clef_Date,
F.Clef_Trajet,
F.Clef_Ligne,
DP.Nom,
F.Station
FROM Fait_Station F, Dim_Date DD, Dim_Personne DP, Dim_Ligne DL ,Dim_Trajet DT
WHERE F.Clef_Date=DD.Clef_Date and F.Clef_Personne=DP.Clef_Personne and F.Clef_Ligne=DL.C
)
Where ranking <10;

```

2.4 Requête 4 : GROUPING SETS

Nombre de fois qu'une station est utilisée par date et par ligne, puis les deux réunies (dernière ligne). On obtiendrait la même chose avec un ROLLUP

```

SELECT DD.Clef_Date, DL.Numero as ID_ligne, count(F.Station)
FROM Fait_Station F, Dim_Date DD, Dim_Personne DP, Dim_Ligne DL ,Dim_Trajet DT
WHERE F.Clef_Date=DD.Clef_Date and F.Clef_Personne=DP.Clef_Personne and F.Clef_Ligne=DL.C
GROUP BY GROUPING SETS((DD.Clef_Date,DL.Numero),(DD.Clef_Date),());

```

2.5 Requête 5 : GROUPING and GROUP BY ROLLUP

```

SELECT DD.Clef_Date, DL.Numero, count(F.Station), GROUPING(DD.Clef_Date) as daate, GROUPING
FROM Fait_Station F, Dim_Date DD, Dim_Personne DP, Dim_Ligne DL ,Dim_Trajet DT
WHERE F.Clef_Date=DD.Clef_Date and F.Clef_Personne=DP.Clef_Personne and F.Clef_Ligne=DL.C
GROUP BY ROLLUP (DD.Clef_Date,DL.Numero);

```

2.6 Requête 6 : TOP N

```

SELECT * FROM (
    SELECT F.station,count(station) as nb
FROM Fait_Station F
GROUP BY F.station
ORDER BY nb DESC
)

```

```
WHERE ROWNUM <=3;
```

2.7 Requête 7 : NTIL

```
select DT.P_DEPART , DT.P_ARRIVEE, count(*) as nb, NTILE(3) over (order by count(*) desc)
FROM Fait_Station F ,Dim_Trajet DT
WHERE F.Clef_Trajet=DT.Clef_Trajet
GROUP BY ( DT.P_DEPART , DT.P_ARRIVEE);
```

2.8 Requête 8 : WINDOW

```
SELECT DD.Clef_Date, DD.Annee,DD.Mois,DD.Jour, DP.nom, sum(DT.Duree) as temps,
       sum(sum(DT.Duree)) over (order by DD.Clef_Date rows unbounded preceding)as cumul
FROM Fait_Station F, Dim_Date DD, Dim_Personne DP, Dim_Ligne DL ,Dim_Trajet DT
WHERE DP.Clef_Personne=1 and F.Clef_Date=DD.Clef_Date and F.Clef_Personne=DP.Clef_Personne
GROUP BY (DD.Clef_Date, DD.Annee,DD.Mois,DD.Jour, DP.nom);
```

2.9 Requête 9 : PARTITION BY

```
SELECT DD.Annee,DT.longueur ,sum(DT.Longueur) over ( partition by(DD.Annee)) as total
FROM Fait_Station F, Dim_Date DD,Dim_Trajet DT
WHERE F.Clef_Date=DD.Clef_Date and F.Clef_Trajet=DT.Clef_Trajet;
order by DD.Annee asc
```