



UNIVERSITE DE CORSE
Master Informatique
parcours DFS et DE
1^{ère} année
2025-2026

BD partie 3 CH2 – Modélisation Dimensionnelle

Evelyne VITTORI
vittori_e@univ-corse.fr



Plan du cours



CH1 – Principes des Datawarehouse

- Objectifs
- Différences avec une BD
- Architectures DW, DL et DLH



CH2 – Modélisation dimensionnelle

- Concepts de modélisation dimensionnelle
- Schémas en étoile et en flocon



CH3 – Processus ETL

- Définition et rôle d'un processus ETL
- Principaux outils
- Mise en pratique avec PentahoDI

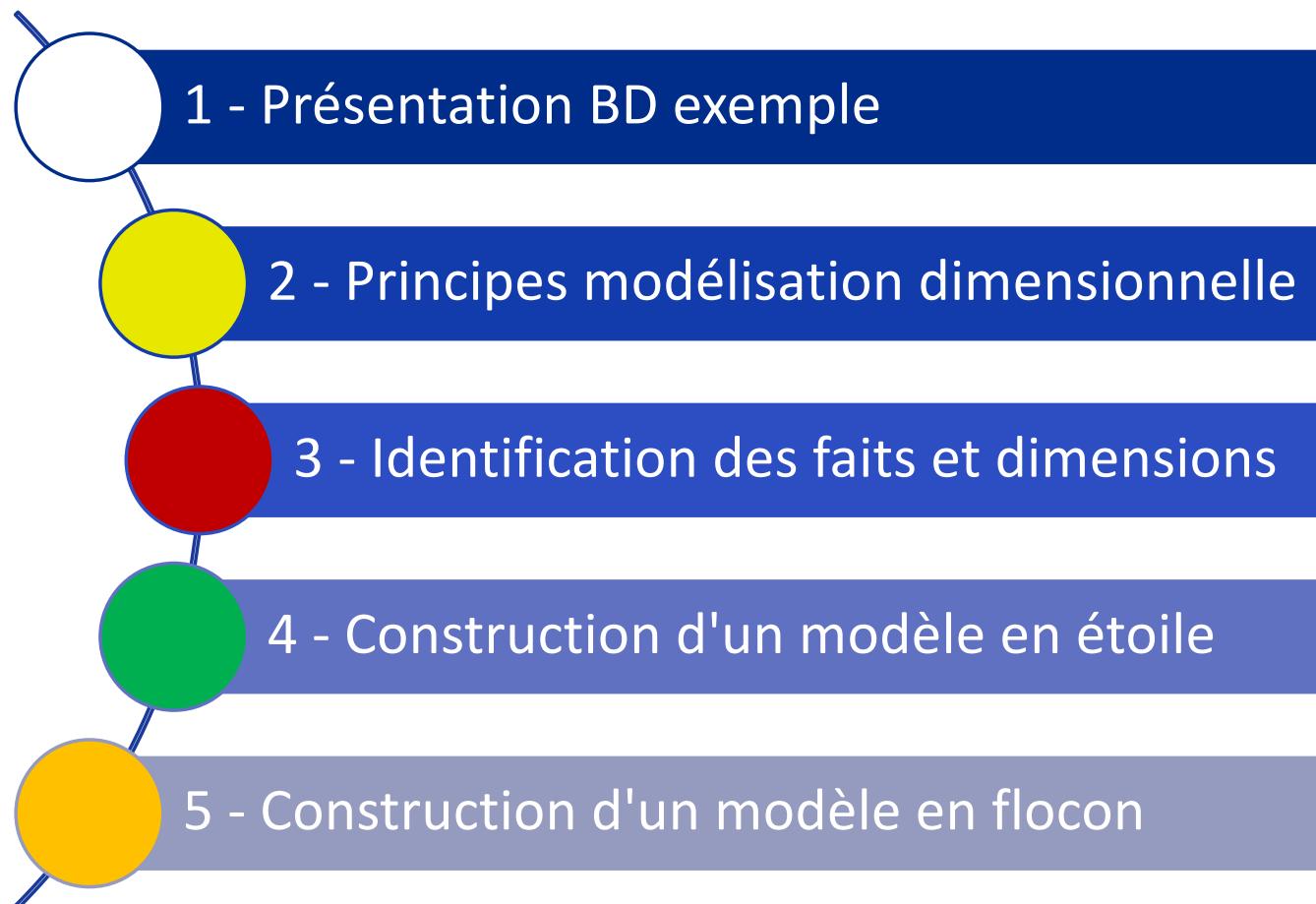


CH4 – Exploitation d'un DW

- Principes OLAP
- Notion d'hypercube OLAP
- Langage MDX
- Mise en pratique avec IcCube



CH2 – Modélisation dimensionnelle





1 - BD Exemple

Skippercruises

Base de données sur les ventes de croisières avec skipper



- **Table des Skippers** : Informations sur les skippers (nom, expérience, nationalité)
- **Table des Bateaux** : Détails des bateaux utilisés (type, capacité, marque)
- **Table des Croisières** : Liste des croisières disponibles (dates, départ, destination, skipper, bateau utilisé, prix total)
- **Table des Clients** : Détails des clients (nom, age, nationalité)
- **Table des Réservations** : Réservations faites par les clients pour des croisières (avec montant des options choisies)
- **Table des Paiements** : Suivi des paiements effectués par les clients par réservations

Exemple BDskipperCruises



paiements
id_paiement serial
id_reservation integer
montant_paiement numeric (10,2)
date_paiement date
mode_paiement character varying(50)

clients
id_client serial
nom character varying(100)
age integer
nationalite character varying(50)

reservations
id_reservation serial
id_client integer
id_croisiere integer
montant_options numeric (10,2)
date_reservation date

croisières
id_croisiere serial
date_depart date
date_arrivee date
id_skipper integer
id_bateau integer
ville_depart character varying(100)
ville_destination character varying(100)
prix_base numeric(10,2)

bateaux
id_bateau serial
nom character varying(100)
type character varying(50)
marque character varying(100)

skippers
id_skipper serial
nom character varying(100)
experience integer
nationalite character varying(50)

Exemple BD skipperCruises (1/5)

table Skippers

id_skipper [PK] integer	nom character varying (100)	experience integer	nationalite character varying (50)
1	Jean Dupont	12	France
2	Luc Martin	8	Espagne
3	Pierre Durand	17	Espagne
4	Antoine Leclerc	2	Italie
5	Thomas Garnier	6	Espagne
6	Charles Blanc	18	France
7	Victor Lefevre	16	Espagne
8	Olivier Bernard	34	Espagne
9	Marc Petit	27	Espagne
10	Philippe Robert	29	Espagne



table Bateaux

id_bateau [PK] integer	nom character varying (100)	type character varying (50)	marque character varying (100)	capacite integer
1	Le Grand Bleu	Catamaran	Jeanneau	92
2	Horizon	Yacht	Beneteau	63
3	Étoile du Matin	Yacht	Bavaria	76
4	Vent du Large	Yacht	Beneteau	14
5	Océan Sauvage	Yacht	Bavaria	59

Exemple BD skipperCruises (2/5)



table Croisières

clés étrangères

id_croisiere [PK] integer	date_depart date	date_arrivee date	id_skipper integer	id_bateau integer	ville_depart character varying (100)	ville_destination character varying (100)	prix_base numeric (10,2)
1	2024-01-17	2024-01-19	3	2	Propriano	Propriano	1916.24
2	2022-09-12	2022-09-17	7	2	Nice	Ajaccio	1524.08
3	2023-01-10	2023-01-15	4	2	Cannes	Cannes	2103.87
4	2022-10-04	2022-10-11	5	4	Ajaccio	Propriano	1278.34
5	2023-02-22	2023-03-01	3	3	Antibes	Ajaccio	2724.26
6	2022-10-14	2022-10-15	9	3	Bastia	Propriano	4812.49
7	2022-04-22	2022-04-27	4	3	Propriano	Bastia	1084.83
8	2023-07-06	2023-07-10	10	1	Propriano	Propriano	4523.87
9	2024-10-05	2024-10-11	8	4	Antibes	Cannes	1372.66
10	2024-05-30	2024-06-01	9	3	Calvi	Cannes	2212.08

Exemple BD skipperCruises (3/5)

table Clients

id_client [PK] integer	nom character varying (100)	age integer	nationalite character varying (50)
1	Alice Moreau	75	Espagne
2	Sophie Girard	19	France
3	Marie Leroy	27	Italie
4	Paul Roux	58	Italie
5	Clara Lefebvre	44	Italie
6	Lucas Simon	75	Espagne
7	Emma Michel	66	Espagne
8	Louis Fontaine	75	Espagne
9	Léo Lambert	71	Espagne
10	Julie François	72	France
11	Chloé Dupuis	60	France
12	Nina Meunier	27	France
13	Maxime Richard	26	Italie
14	Camille Henry	28	Italie
15	Sarah David	35	France
16	Elodie Bertrand	20	Espagne
17	Cédric Faure	20	Italie
18	Nicolas Brun	47	France
19	Mathilde Olivier	59	Espagne
20	Thomas Dupuis	37	Italie



Exemple BD skipperCruises (4/5)

clés étrangères

id_reservation [PK] integer	id_client integer	id_croisiere integer	montant_options numeric (10,2)	date_reservation date
1	3	10	475.07	2024-02-06
2	12	6	302.45	2022-03-08
3	11	7	132.08	2022-02-25
4	15	4	751.23	2022-06-16
5	4	2	667.28	2022-05-26
6	20	6	794.68	2022-04-27
7	11	10	971.87	2022-07-15
8	14	7	877.78	2022-03-18
9	16	7	755.65	2022-01-06
10	15	2	424.79	2022-09-03
11	11	5	584.78	2022-10-12
12	17	2	698.24	2022-08-25
13	6	6	932.30	2022-06-21
14	3	4	103.64	2022-10-01
15	5	9	154.41	2022-10-12
16	6	3	576.57	2022-02-01
17	9	3	124.87	2022-03-05
18	15	5	481.69	2022-08-14

table
Reservations

Une réservation correspond à une vente



*Le prix total payé par le client pour une réservation =
prix_base + montant_options*

Exemple BDskipperCruises (5/5)



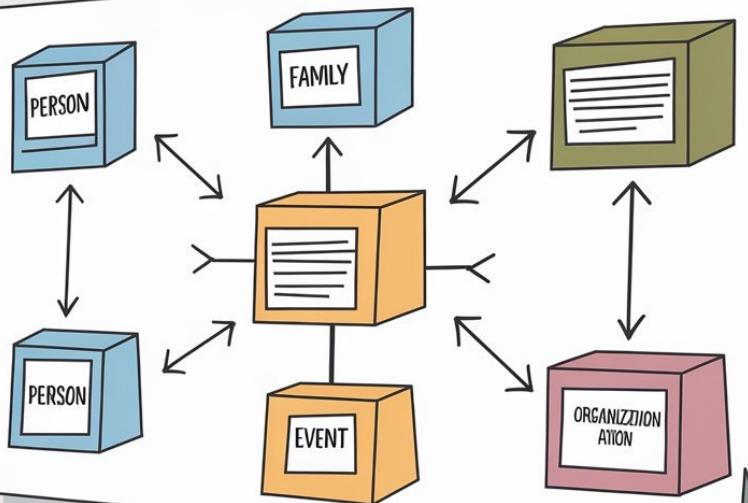
table
Paiements

Plusieurs
paiements
possibles pour
une réservation

clé étrangère

<code>id_paiement</code> [PK] integer	<code>id_reservation</code> integer	<code>montant_paiement</code> numeric (10,2)	<code>date_paiement</code> date	<code>mode_paiement</code> character varying (50)
1	2	4331.81	2022-10-09	Carte
2	6	5565.06	2022-06-18	Carte
3	10	1583.44	2022-09-17	Carte
4	24	922.06	2023-02-28	Virement
5	4	221.63	2022-09-03	Chèque
6	5	700.08	2022-08-25	Chèque
7	15	802.97	2024-09-02	Carte
8	23	665.74	2022-10-11	Virement
9	6	5136.54	2022-07-19	Chèque
10	30	2418.75	2022-04-20	Chèque
11	2	3493.02	2022-05-13	Carte
12	16	1670.47	2022-02-07	Virement
13	15	1205.66	2023-01-20	Virement
14	25	292.91	2024-08-28	Virement
15	27	1270.05	2022-04-16	Carte
16	16	1024.60	2022-07-03	Chèque
17	26	2656.42	2023-07-27	Virement
18	5	1095.32	2022-08-27	Chèque
19	8	1442.86	2022-03-22	Virement
20	24	1868.90	2023-01-22	Carte

DIMENSIONAL MODELING

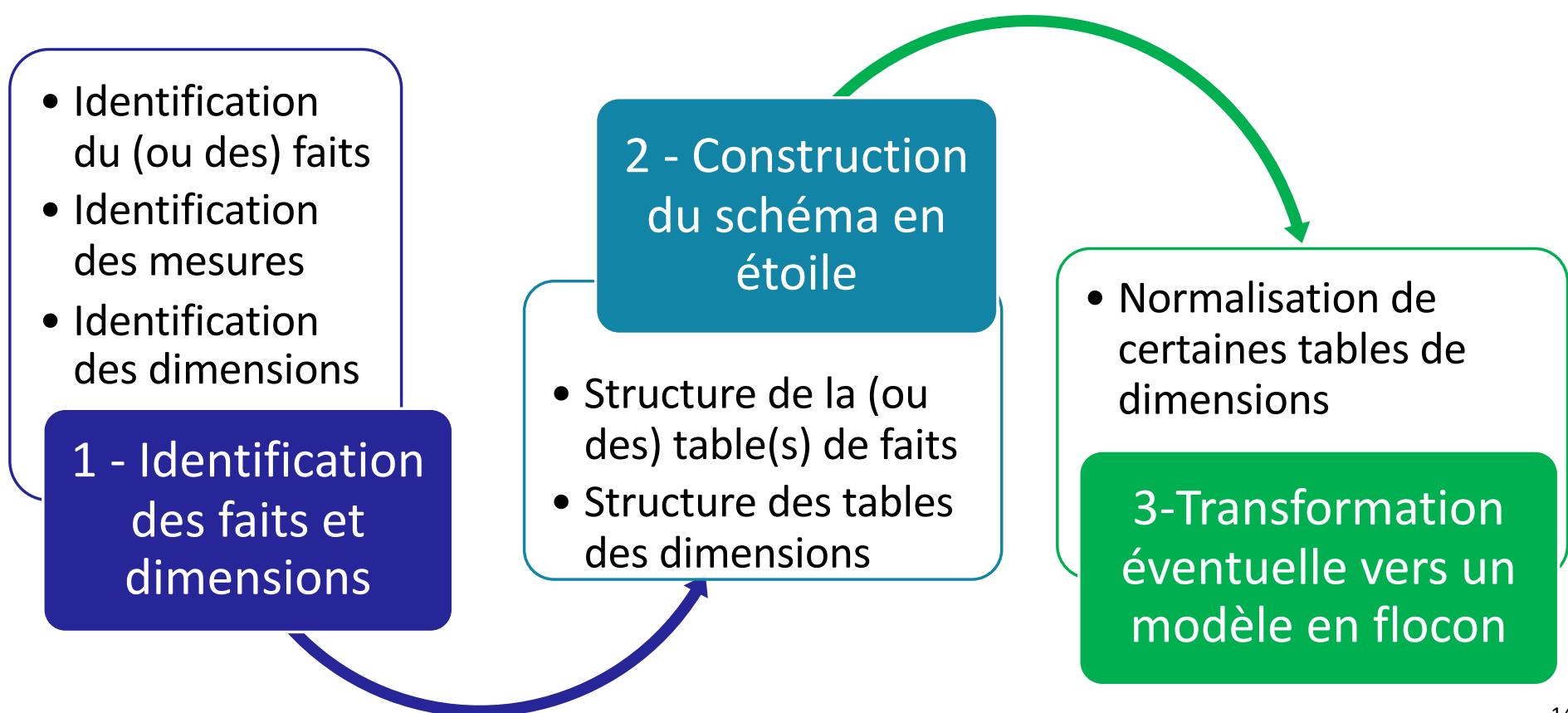


2 - Principes de modélisation dimensionnelle

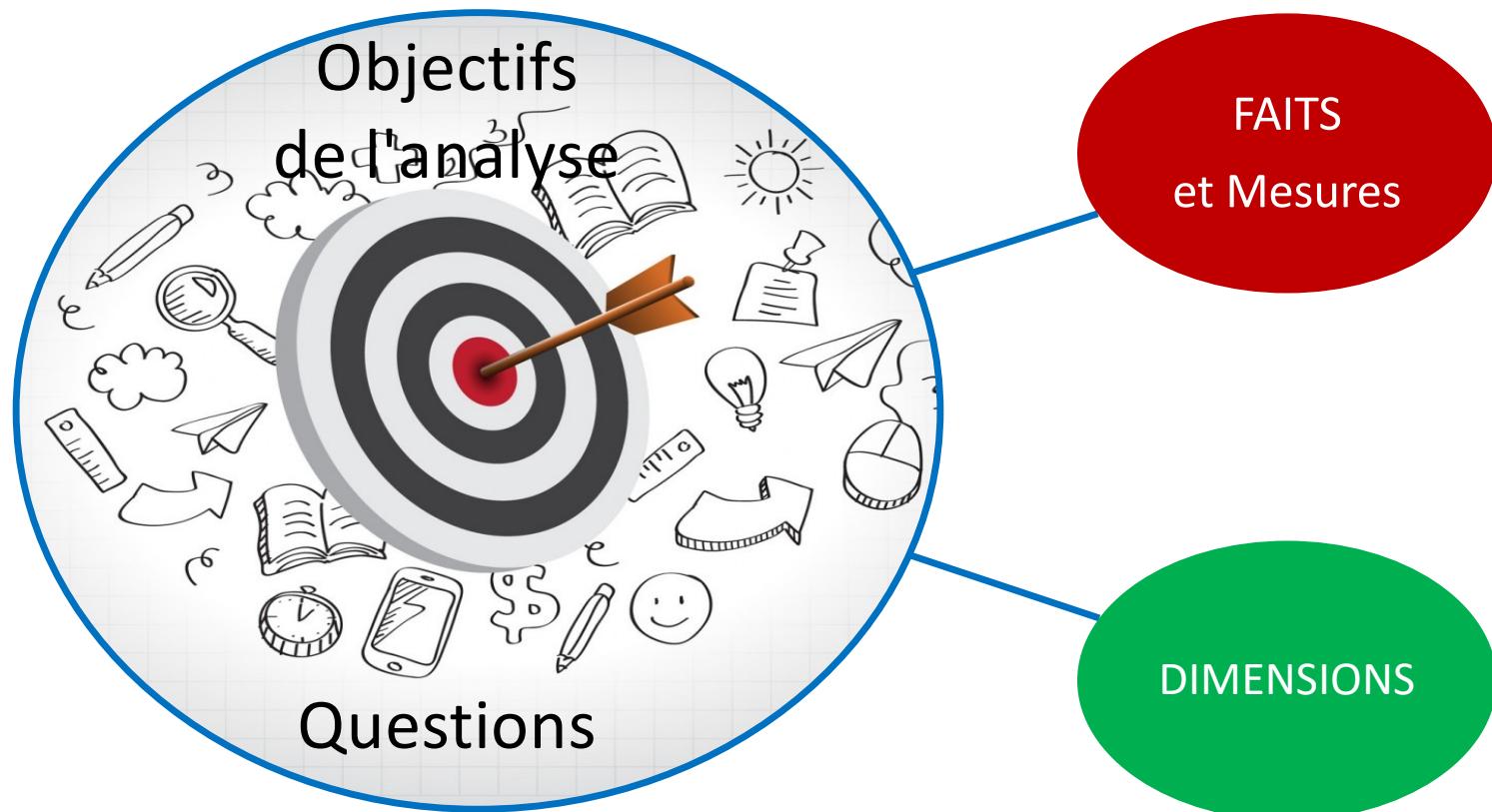
Qu'est-ce que la modélisation dimensionnelle?

- **Modélisation basée sur un objectif précis**
 - Organiser les données pour répondre à un objectif d'analyse spécifique.
- Les données sont structurées afin de mettre en évidence :
 - le sujet analysé **le (ou les) faits**
 - les différentes perspectives de l'analyse **les dimensions**
- Plusieurs approches possibles pour la construction du schéma
 - Modèle en étoile ou en flocon

Démarche de modélisation dimensionnelle



3 - Identification des faits et dimensions



Objectifs BD skippercruises



Comprendre les performances commerciales des croisières (analyse des revenus générés)

- Répondre à des questions précises
 - "Quel skipper génère le plus de revenus ?"
 - "Quel type et quelle marque de bateau génère le revenu le plus important?"
 - "Les bateaux de plus grandes capacités génèrent-ils plus de revenus ?"
 - "Quels mois ou trimestres sont les plus rentables ?"
 - "Les ventes augmentent-elles ou diminuent-elles selon la saison ?"
 - "Quelles destinations attirent le plus de clients et génèrent le plus de revenus ?"

Une réservation correspond à une vente

Identifier le (ou les faits)

- Qu'est-ce qu'un Fait ?

Un fait désigne un événement ou un processus que l'on désire mesurer et analyser

(ex. : mouvement de stock, consultation médicale, ventes)

Exemple BD skippercruises



Le fait principal est la **vente de croisières**

Identifier les mesures des faits



■ Qu'est-ce qu'une Mesure ?

Une mesure est une **valeur quantitative** (aggrégeable) décrivant un fait.

Exemple BD skippercruises

Le fait *vente de croisières* peut être décrit par 2 mesures :



- ✓ **Revenu par croisière** = montant total généré par les ventes (réservations) de chaque croisière.
- ✓ **Nombre de réservations par croisière** = nombre total de réservations associées à chaque croisière.

Identifier les mesures des faits



- Les mesures peuvent être décrites de manière détaillée dans un **dictionnaire de mesures**
- *Exemple BD skippercruises*



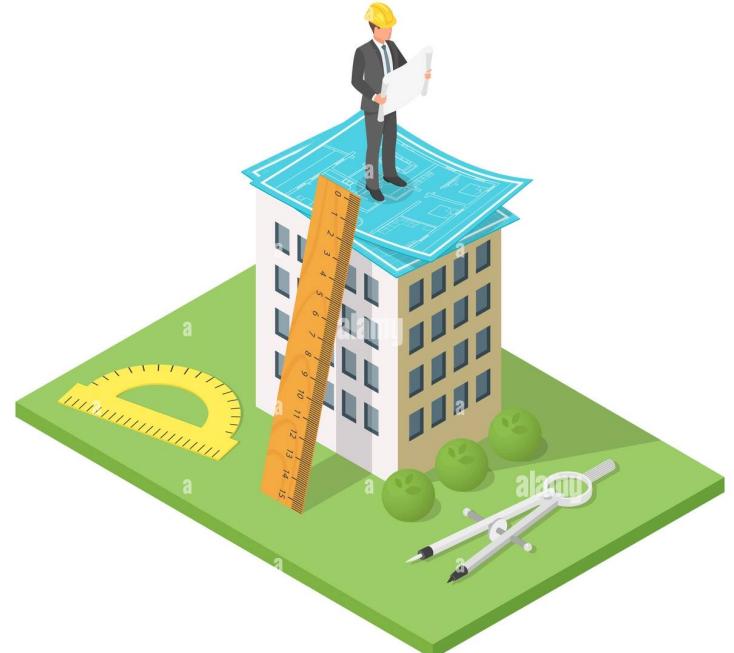
Libellé	Désignation	Type	Formule de calcul
revenu_total	montant total généré par les ventes d'une croisière	réel	Somme des montants des réservations d'une croisière calculés par la somme : prix_base + montant_option
nombre_reservations	nombre total de réservations pour une croisière	entier	comptage des réservations pour une croisière

Identifier les dimensions

- Qu'est-ce qu'une Dimension ?

Une **dimension** fournit un contexte autour des faits et permet de les analyser sous différents angles
(ex. : temps, produit, client)

Identifiées à partir des questions posées par l'analyse



Identifier les dimensions

Exemple BD skippercruises

Objectifs de l'analyse

1. "Quel **skipper** génère le plus de revenus ?"
2. "Quels **types et marques de bateau** génèrent le plus de ventes ?"
3. "Les bateaux de plus grande **capacité** génèrent-ils plus de revenus ?"
4. "Quels **mois ou trimestres** sont les plus rentables ?"
5. "Les ventes augmentent-elles ou diminuent-elles selon la **saison** ?"
6. "Quelles **destinations** génèrent le plus de revenus ?"

Dimensions

- **Skipper** : Nom
- **Bateau** : type, marque, capacité
- **Date** : année, mois, trimestre, saison
- **Destination** : ville_destination



4 – Construction d'un modèle en étoile

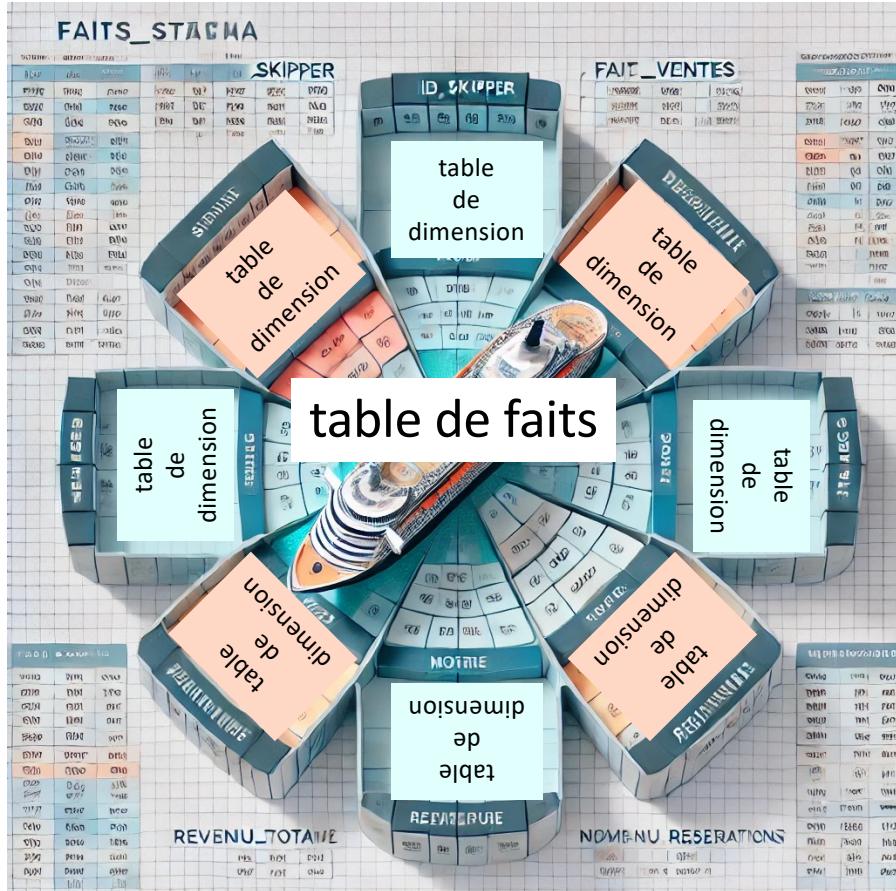
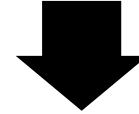


Table de faits

- Mesures



Tables des dimensions

- Attributs bruts et transformés
- Hiérarchies
- Attributs faibles

La table de faits



- Attributs de la table de faits
 - Clés étrangères vers les dimensions
 - Mesures associées au fait décrit

faits_ventes
id_vente serial
id_skipper integer
id_bateau integer
id_date integer
id_destination integer
revenu_total numeric(10,2)
nombre_reservations integer

Différents types de table de Faits

- Tables de faits **transactionnelles**
 - Capturent des événements spécifiques
(ex. : une vente ou une réservation)
- Tables de faits de **snapshot périodique**
 - Capturent des mesures à un moment donné, souvent à des intervalles réguliers
(ex. : l'inventaire à la fin de chaque mois, les réservations à une date spécifique).
- Tables de faits **cumulatives**
 - Capturent un flux d'événements
(ex. : nombre total de produits vendus jusqu'à une certaine date).

Les tables de dimensions

- Une table de dimension contient des données descriptives décrivant un axe de contexte des faits analysés.
- **Clé primaire**
 - Identifiant unique de la dimension : clé de substitution (ex. : id_skipper, id_date).
- **Attributs descriptifs (membres)**
 - Attributs nécessaires aux analyses
 - Issus directement des tables sources ou transformés

Exemples de tables de dimension

DWCruisesEtoile

dimension_bateau
id_bateau
nom
type
capacite
marque

dimension_date
id_date
jour
mois
trimestre
annee
saison

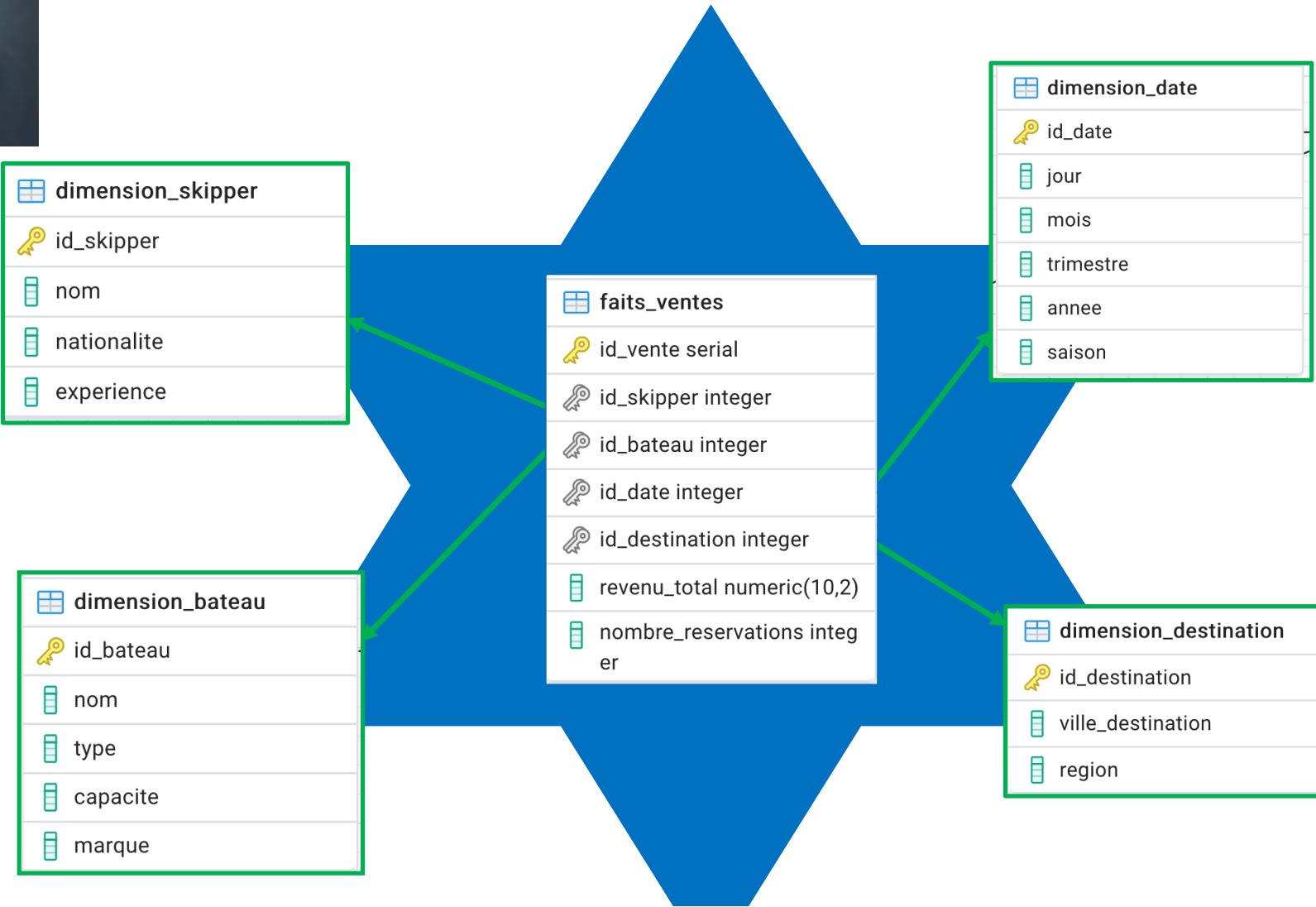
dimension_skipper
id_skipper
nom
nationalite
experience

dimension_destination
id_destination
ville_destination
region





Modèle en étoile





Attributs bruts ou transformés

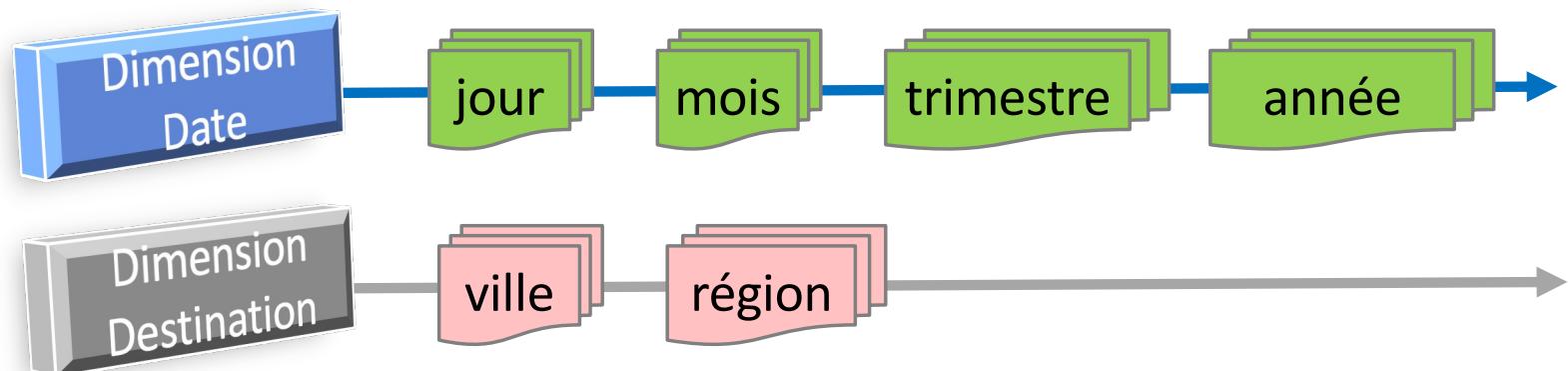
- Attributs **bruts** : Importés directement à partir des tables de la base source.
 - Informations descriptives telles qu'elles sont stockées dans la base de données opérationnelle.
- Attributs **transformés** : Résultent de règles de gestion, de calculs ou d'algorithmes de transformation





Attributs et niveaux hiérarchiques

- Les attributs d'une dimension peuvent être organisés suivant des hiérarchies
- Chaque attribut a un niveau de hiérarchie (ou de granularité) spécifique



Paramètres et Attributs faibles



- Les paramètres sont des attributs qui déterminent un niveau de granularité
- Les attributs dits **faibles** sont des attributs qui dépendent fortement de paramètres
 - ils ne peuvent pas être compris de manière autonome
- *Exemples*
 - **region** est un attribut faible dans la dimension destination
 - Il n'existe que par rapport au paramètre `ville_destination`. Si `ville_destination` change, alors `region` changera également.
 - **saison** est un attribut faible dans la dimension date
 - Il n'existe que par rapport au paramètre `mois`.

Résumé modèle en étoile

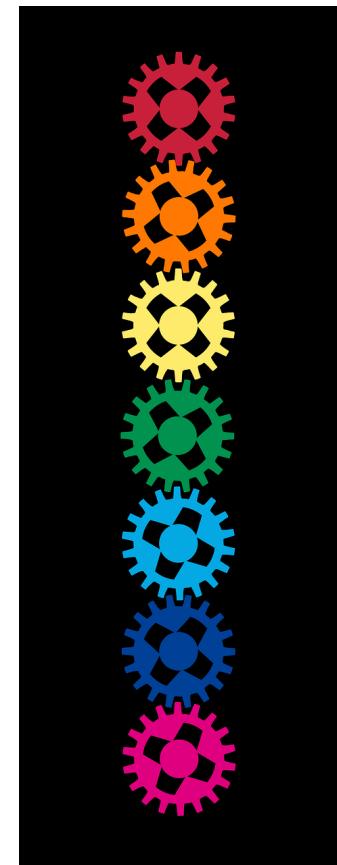
Etapes de conception

1. Construire la table de faits

- décrire les mesures
- Identifier les dimensions avec le niveau de granularité le plus bas

2. Construire les tables des dimensions

- décrire les attributs ou membres
 - bruts ou transformés (formule calcul)
 - niveau hiérarchique
 - faible ou non



Résumé modèle en étoile

Points clés

Le plus utilisé

Fait est au centre avec dimensions autour

Tables de dimensions non liées entre elles

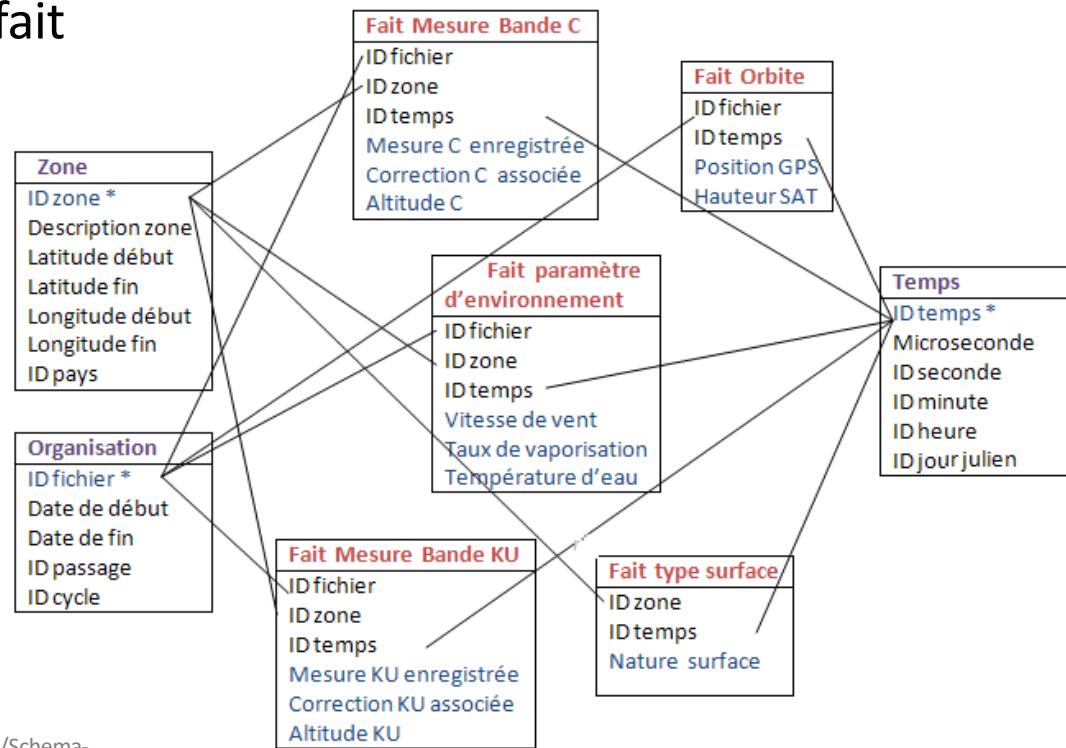
Avantages

- **Simplification** des requêtes
 - Toutes les informations relatives à une dimension sont stockées dans une seule table.
 - Diminution du nombre de jointures
- **Performances** optimales
 - Requêtes OLAP (analyses décisionnelles)

Notion de constellation

- Fusion de plusieurs modèles en étoile qui utilisent des dimensions communes
 - Plusieurs tables de fait

Exemple de Schéma en constellation d'un DW de données d'altimétrie spatiale



source : https://www.researchgate.net/figure/Schema-en-constellation-du-DW-de-donnees-daltimetrie-spatiale-apres-lunification-des_fig1_323826694

A vous de jouer (1) ...

Conception DWCrusesPerf

- Concevez un modèle en étoile pour permettre l'analyse des performances des croisières en termes de nombre de passagers et de durée des croisières.

Questions auxquelles devra répondre l'analyse :

- Quelle est la durée moyenne pour l'ensemble des croisières ?
- Combien de passagers ont été transportés par chaque bateau ?
- Quels sont les deux bateaux qui ont transporté le plus de passagers?
- Quelle a été la durée moyenne des croisières pour chaque bateau ?
- Quelle est la durée moyenne des croisières gérées par chaque skipper ?
- Quel est le nombre total de passagers embarqués par chaque skipper ?
- Quel est le nombre total de passagers embarqués par niveau d'expérience des skippers ?
 - $0 < \text{expérience} < 10$: débutant - $10 \leq \text{expérience} < 20$: intermédiaire - $20 \leq \text{expérience}$: expert
- Quelle est le nombre de passager par skipper pour chaque bateau?



A vous de jouer (2) ...

Conception DWCrusesPerf

- Définissez le script SQL de création du schéma relationnel de votre datawarehouse sous PostgreSQL.



- Vous pouvez utiliser un assistant ...



- Attention à la conformité des types par rapport à la BD source (skippercruises)

5 – Modèle en flocon



Modèle en étoile

Inconvénients

- **Redondance** dans certaines tables de dimensions (tables dénormalisées)



table
Dimension_
Date

id_date [PK] integer	jour integer	mois integer	trimestre integer	annee integer	saison character varying (20)
1	4	10	4	2022	Automne
2	5	10	4	2024	Automne
3	6	7	3	2023	Été
4	10	1	1	2023	Hiver
5	12	9	3	2022	Automne
6	14	10	4	2022	Automne
7	17	1	1	2024	Hiver
8	22	2	1	2023	Hiver
9	22	4	2	2022	Printemps
10	30	5	2	2024	Printemps

- Pourquoi peut-on dire que cette table n'est pas normalisée?

Qu'est-ce qu'un modèle en flocon?

- Variante du modèle en étoile avec des tables de dimensions normalisées

Création de tables de sous-dimensions supplémentaires

Points clés

- Fait toujours au centre et dimensions en sous-hierarchies
- Un seul niveau hiérarchique par « table » de dimension
- La « table » de dimension de niveau hiérarchique le plus bas est reliée au fait (elle a la granularité la plus fine)



Modèle en flocon



Modèle en flocon

Avantages

- Elimination des redondances
- Optimisation du stockage
 - Intéressant sur des très grands DW

Inconvénients

- Complexification des requêtes
 - Augmentation du nombre de jointures
- Requêtes moins performantes
- Navigation difficile

A vous de jouer (3) ...



Identifiez les éventuelles tables de dimension du modèle en étoile obtenu à l'issue de l'exercice précédent pouvant être normalisées afin de concevoir un modèle en flocon.



Modèle en étoile vs Modèle en flocon

Critères	Modèle en étoile	Modèle en flocon
Structure des dimensions	Toutes les dimensions sont dénormalisées.	Les dimensions sont normalisées en plusieurs niveaux.
Redondance	Il y a de la redondance dans les tables de dimensions.	La redondance est minimisée en décomposant les dimensions.
Simplicité des requêtes	Les requêtes sont simples, avec peu ou pas de jointures entre les tables de dimensions.	Les requêtes sont plus complexes en raison des multiples jointures nécessaires entre les tables de dimensions normalisées.
Performance des requêtes	Meilleure pour les petites bases de données car il y a moins de jointures.	Moins performante pour les petites requêtes à cause des nombreuses jointures.
Maintenance des données	Plus difficile à maintenir car les informations redondantes doivent être mises à jour dans plusieurs endroits.	Plus facile à maintenir car les données sont centralisées dans des tables normalisées, réduisant les mises à jour multiples.
Espace de stockage	Nécessite plus d'espace de stockage à cause de la redondance des données.	Nécessite moins d'espace de stockage car les données sont stockées de manière centralisée et normalisée.
Granularité des dimensions	Les dimensions sont moins flexibles pour des analyses détaillées (puisque tout est dénormalisé).	Plus flexible pour l'analyse détaillée car la granularité des dimensions est explicitement définie dans des tables hiérarchisées.
Utilisation pratique	Utilisé lorsque les performances de requêtes sont prioritaires et que la base est relativement simple.	Utilisé lorsque la maintenance et la réduction de la redondance sont prioritaires, ou dans les grandes bases de données.
Exemple d'application	Idéal pour des bases de données décisionnelles simples avec peu de relations hiérarchiques (par exemple, analyse des ventes simples).	Idéal pour des bases plus complexes avec des hiérarchies profondes (par exemple, analyses par année, mois, trimestre, région).

Question

Pourquoi la normalisation
n'est-elle pas indispensable
dans le cadre d'une
modélisation dimensionnelle?

liens

- <https://slideplayer.fr/slide/1159105/>
- <https://fr.slideshare.net/slideshow/cours-data-warehouse/239246576>