



Professeur : Mr GAOUSSOU CAMARA

Classe : SID

Niveau : Licence 3

Table des Matières

I.	Historique de l'étude.....	2
II.	Analyse et description des sources de données (Sakila,world)	3
III.	Etude comparative de 3 outils de gestion de bases de donnée.....	4
IV.	Etude comparative de 3 outils ETL	5
V.	Etude comparative de 3 outils de Reporting.....	6
VI.	Identification	7
1.	Les indicateurs (mesures) que le magasin souhaite suivre dans cette étude.....	8
2.	Les tables de dimensions et éventuellement leurs hiérarchies	8
3.	La granularité de l'entrepôt de données.....	8
VII.	Le modèle conceptuel de données	9
VIII.	Le modèle physique de données.....	10
IX.	Alimentation Acteur	10
X.	Alimentation Catégorie	12
XI.	Alimentation Temps	13
XII.	Alimentation Géographie	14
XIII.	Alimentation Rentabilité	15
XIV.	Les rapports conçus avec Tableau commentés.....	16
4.	Montant des locations de films par pays	16
5.	Montant des films crédités à chaque acteur dans chaque continent.....	17
6.	Montant des locations de films par semestre	19
XV.	Le tableau de bord.....	20
7.	Lien du tableau de bord	21

I. Historique de l'étude

Un magasin de la place qui a comme activité la location de films souhaite analyser la rentabilité de ces films à partir du jeu de donnée Sakila et World. Pour cela, il se propose de mettre en place un entrepôt de donnée nommé SakilaRentabilité. Cet entrepôt de donnée nous permet de répondre aux questions suivantes :

- Quel est le montant des locations de films généré par chaque catégorie dans chaque pays ?
- Quel est le montant des films crédité à chaque acteur dans chaque continent ?
- Quel est le montant des locations de films généré par semestre ?

II. Analyse et description des sources de données (Sakila,world)

- Sakila est une base de donnée relationnelle .IL est constitué de 16 tables et 7 vues. Ces 16 tables sont : actor,address,category,city,country,customer,film,film_actor,film_category, film_text,inventory,language,payment,rental,staff,store ; et ces 7 vues :actor_info,customer_list,film_list,nicer_but_slower_film_list,sales_by_film__category,sales_by_store,staff_list .Ces tables sont reliées par plusieurs relations entre elles. Exemple les tables film et language sont reliées par la colonne language_id, clé primaire de la table language qui est clé étrangère de la table film. Les tables city et country sont reliées par la colonne country_id, clé primaire de la table country qui est clé étrangère de la table city.
- World est une base de donnée qui comporte 3 tables : city, country,countrylanguage. Ces tables sont reliées par plusieurs relations entre elles.les tables city et countrylanguage sont reliées par la colonne countrycode.Cette dernière est la clé primaire de la table countrylanguage qui est clé étrangère de la table city.
- Les deux bases de données Sakila et World sont reliées par les tables world_country et Sakila_country.Ces deux tables ont des colonnes

semblables : la colonne name de la table world_country et la colonne country de la Sakila_country.

III. Etude comparative de 3 outils de gestion de bases de donnée

Ils existent plusieurs SGBD pour la gestion des bases de données. Parmi lesquelles nous pouvons en citer : MYSQL, Microsoft SQL SERVER BI et ORACLE BI. Le choix de l'un de ces SGBD se fait à travers les critères suivants :

- Taille de la base de données cible
 - Ouverture par rapport aux outils de restitution
 - Capacité de traitement en parallèle
 - Support des technologies en vogue
-
- ❖ MYSQL est un système de gestion de base de données achetée par ORACLE depuis le 21/01/2012. IL est un serveur de base de données relationnelles SQL, développé dans un souci de performances élevées en lecture, ce qui signifie qu'il est davantage orienté vers le service de données déjà en place que vers celui de mises à jours fréquentes et fortement sécurisées. Il est multi-utilisateurs, multi_threadé et fonctionne parfaitement avec PHP. MYSQL est totalement open source et gratuit ,ses performances sont excellentes .
 - ❖ ORACLE est un système de gestion de base de données publiée par Oracle Corporation. Ce dernier est maintenant propriétaire de MYSQL. Les différences entre ORACLE BI et MYSQL sont :
 - MYSQL est gratuit alors qu'ORACLE BI est payante.
 - ORACLE BI prend en charge les bases de données distribuées alors que MYSQL n'y supporte pas.

- ORACLE convient aux déploiements en entreprise, tandis que MYSQL convient aux petites et moyennes entreprises.
- Avec ORACLE vous payer des frais de licence alors que MYSQL est gratuit.

❖ Microsoft SQL Serveur est un système de gestion de base de données relationnelle éditée et commercialisé par Microsoft depuis 1994. IL a plusieurs avantages qui le diffèrent avec les autres :

- Déploiement par un setup, mise en œuvre et administration par des interfaces graphiques intuitives.
- programmable
- Gestion avancée de la sécurité en offrant deux modes d'authentification (Authentification Windows et Authentification SQL Server).
- Prise en compte des spécificités des projets décisionnels.
- Coût relativement moins cher par rapport aux autres SGBD du marché.

IV. Etude comparative de 3 outils ETL

L'objectif des ETL est de produire des données propres, faciles d'accès et qui peuvent être exploitées efficacement par la Business Intelligence. Les données brutes peuvent être extraites de différentes sources, en particulier Base de données existantes. Ils existent plusieurs ETL parmi lesquels : Talend Studio for Data Integration, Pentaho Data Integration et SQL Server Integration Services (SSIS).

❖ Talend Studio for Data Integration est capable de supporter les projets ETL et ELT. Son interface graphique offre une représentation visuelle des flux de données de l'organisation.

Talend Open Studio for Data Integration est une plateforme d'intégration des données basée sur le langage java. Talend Open Studio permet de répondre à toutes les problématiques liées au traitement des données de manières générales, dans la chaîne décisionnelle. Talend open Studio a plusieurs avantages qui le différencient des autres ETL, parmi lesquels nous pouvons en citer :

- Interface intuitive basée sur Eclipse de java
 - Vue graphique des jobs grâce aux interfaces graphiques élaborées des composants.
 - Talend est compatible quel que soit l'OS car le résultat des développements est en fait un compilé de code java.
 - Talend Open studio est gratuit pour un utilisateur sur le référentiel.
- ❖ Pentaho Data Integration est un logiciel d'ETL qui permet la conception ainsi que l'exécution des opérations de manipulation et de transformations de données très complexes. Pentaho Data Integration a pour principal intérêt de récupérer diverses sources dans divers formats, les traiter, les transformer et former un résultat puis finalement exporter dans le format souhaité vers une destination souhaitée. Il a beaucoup d'avantages qui le différencient des autres en particulier Talend Open Studio :
- Un large appui de point d'émission de données
 - Réutilisation facile des interrogations et des composants de transformation avec les modèles.
 - Possibilité d'avoir en données sources une requête SQL
 - Entièrement intégré avec la suite Open source de Pentaho.

V. Etude comparative de 3 outils de Reporting

Le reporting est un outil servant de support à la prise de décision par les dirigeants en affichant un contenu détaillé de la situation. Il existe plusieurs outils de reporting parmi lesquels Tableau, Power BI, Bime Analytics.

- ✓ Tableau est un outil permettant de réaliser de très belles visualisations de données. Il s'est rapidement imposé sur le marché de la Business Intelligence, et permet de simplifier les données brutes en les

convertissant dans un format compréhensible .Cette solution permet de l'analyse rapide de données, et la génération de visualisations sous forme de tableaux et de feuilles de tableur. Ainsi, les professionnels de tous les départements sont en mesure de comprendre les données et d'interpréter les analyses. En effet, Tableau a plusieurs avantages qui le diffèrent avec les autres outils de reporting.

Il permet de générer des tableaux de bord et des rapports pour prédire les revenus en se basant sur les comportements passés des consommateurs. On peut aussi effectuer des calculs pour transformer les données existantes en fonctions des besoins. L'interface de Tableau utilise un système de glisser-déposer pour permettre de poser des questions aux données. Il suffit d'entrer les types de données dans les axes x et y, et Tableau se chargera de créer instantanément la visualisation.

- ✓ Power BI est un outil de business Intelligence .Il permet de convertir les données en provenance d'une large variété de sources en tableaux interactifs et en rapports BI. Avec Power BI, il est possible de prendre en charge des données en provenance de différentes sources et de générer des visualisations après le processus de nettoyage et d'intégration. Il est plus facile d'intégrer des rapports avec Power BI qu'avec Tableau Desktop. Les points de différence entre Tableau Desktop et Power BI sont nombreux. La possibilité de stocker les données sur un serveur est possible avec Tableau Desktop, ce qui n'est pas le cas avec Power BI.
- ✓ Bime Analytics est un outil de reporting. Ce logiciel a des points forts comme sa capacité de générer des Dashboards avec un large choix de modèles graphiques, ce qui lui permet de s'adapter aux besoins des différents secteurs. Il propose une analyse très poussée aux niveaux des tableaux croisés, un partage en temps réel des données, une intégration sans SQL des données, une capacité d'intégration à différents logiciels pour optimiser la collecte de l'information.

VI. Identification

1. Les indicateurs (mesures) que le magasin souhaite suivre dans cette étude

Le montant de location est l'indicateur que le client souhaite suivre dans cette étude. Les axes d'analyses sont : catégories, pays, acteur, continent, semestre.

2. Les tables de dimensions et éventuellement leurs hiérarchies

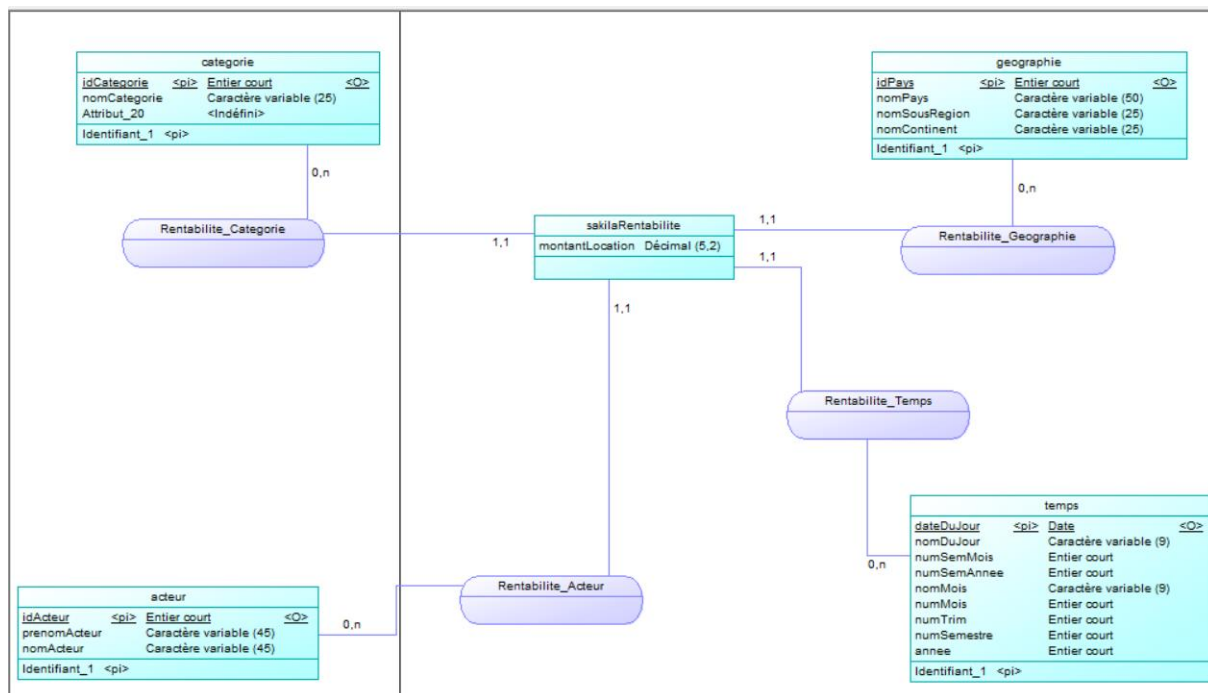
Les tables de dimensions sont :

- Catégorie (idcatégorie, nomcatégorie)
- Acteur (idacteur, prenomacteur, nomacteur)
- Temps (datedujour, nomdujour, numsemmois, nommois, nummois, numtrim, numsemestre, année)
- Géographie (idpays, nompays, numsousrégion, nomcontinent)
 - La table Temps d'écrit des jours de location. On peut déterminer une hiérarchie dans cette table : datedujour-nomdujour-numsemmois-nommois-nummois-numtrim-numsemestre-année.
 - La table Géographie d'écrit des pays ou les locations se sont effectués. On peut identifier une hiérarchie dans cette table : Pays-sousrégion-continent.

3. La granularité de l'entrepôt de données

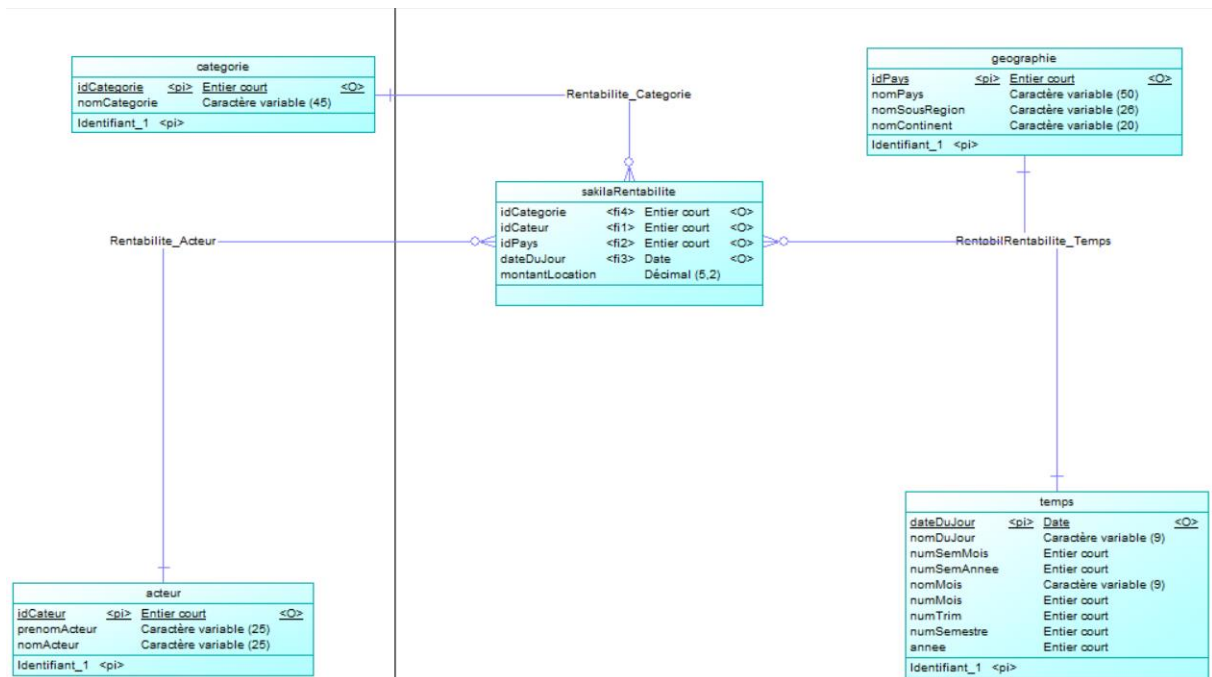
La granularité de l'entrepôt est : Catégorie-Acteur-Jour-Pays.

VII. Le modèle conceptuel de données



A partir de cette modèle conceptuel conçue par **Power AMC**, nous pouvons en tirer l'indicateur que le client souhaite suivre dans cette étude, les tables de dimensions, la table de fait et les différentes relations qui existent entre elles. L'indicateur est le **montant de location**. On a 4 tables de dimensions : catégorie, acteur, temps, géographie et un table de fait SakilaRentabilite. Ils existent quatre associations qui lient la table sakilaRantabilite aux tables de dimensions : rentabilite_categorie, Rentabilite_Géographie, Rentabilite_Temps et Rentabilite_Acteur. Les cardinalités sont de (1,1) au côté de sakilaRentabilite et (0, n) aux côtés des tables de dimensions.

VIII. Le modèle physique de données



Ce graphique nous montre le modèle physique de donnée qui a été tiré à partir du modèle conceptuel de donnée. La spécificité de ce modèle est la migration des clés primaires des tables de dimensions vers la table de mesure (sakilaRentabilite). Ce modèle physique nous a permis de choisir le système de gestion de base de données MYSQL pour la continuité du travail.

IX. Alimentation Acteur

Job alimentationacteur 0.1 x Job alimentationcategorie 0.1

200 rows in 0,24s
816,33 rows/s
row1 (Main)

tMap_1

200 rows in 1,26s
158,73 rows/s
sortie1 (Main)

"acteur"

Designer Code

Job(alimentationacteur 0.1) Contexts(alimentationacteur) Composant Exécuter (Job alimentationacteur)

Job alimentationacteur

Exécution simple

Exécution en mode Debug

Paramètres avancés

Cible d'exécution

Exécution pour la mémoire

Exécution

Exécuter Arrêter Effacer

Démarrage du Job alimentationacteur à 17:26 05/11/2022.

[statistics] connecting to socket on port 3482

[statistics] connected

[statistics] disconnected

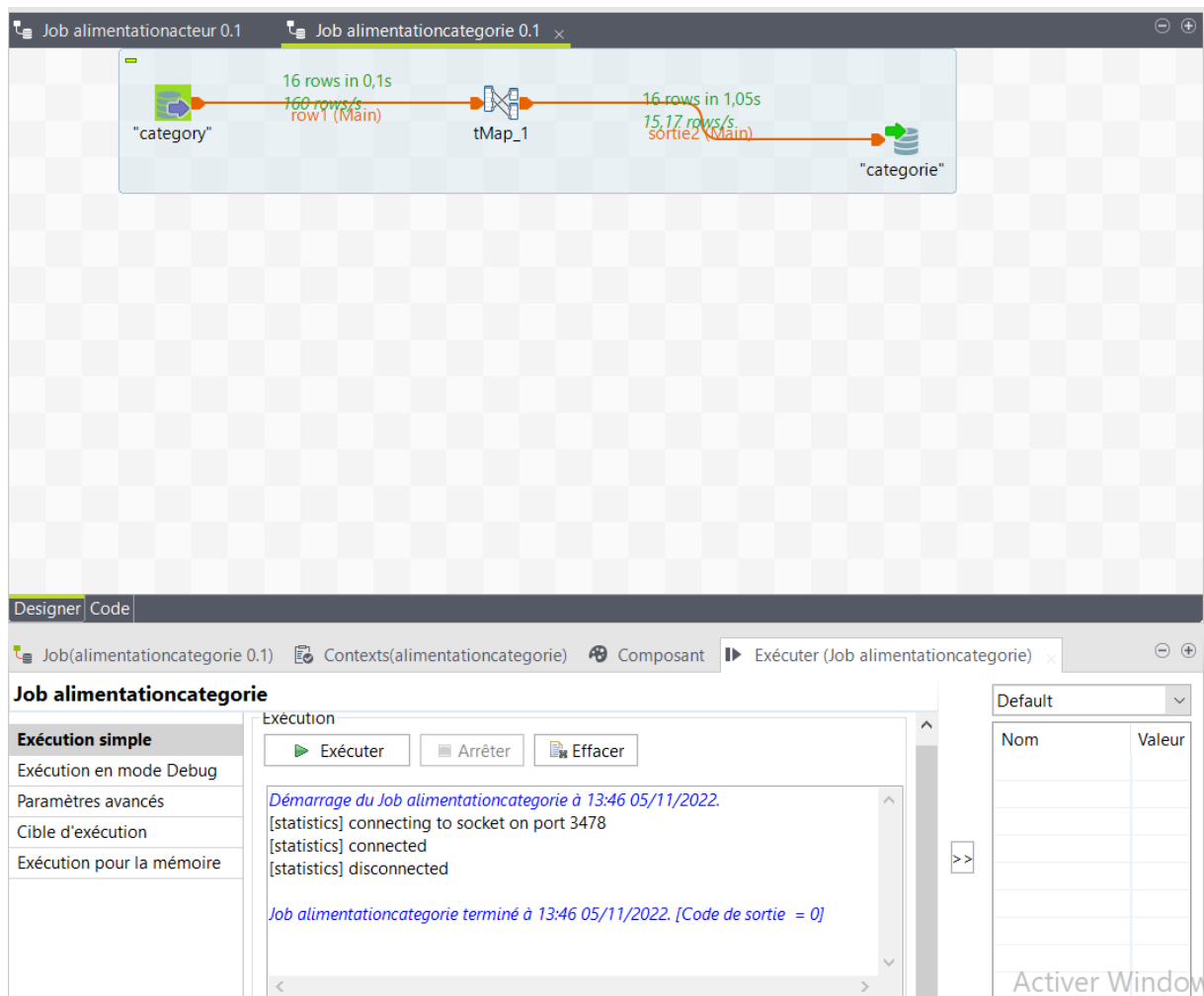
Job alimentationacteur terminé à 17:26 05/11/2022. [Code de sortie = 0]

Default

Nom	Valeur

Ce graphique nous montre le job de l'alimentation de la table acteur. Nous alimentons notre table acteur à partir de la table actor. Nous sélectionnons trois colonnes sur les quatre parce-que la colonne last_update n'a pas trop d'impact. La table actor contenait 200 acteurs et tous ses acteurs ont été extraire et alimenter à la table acteur.

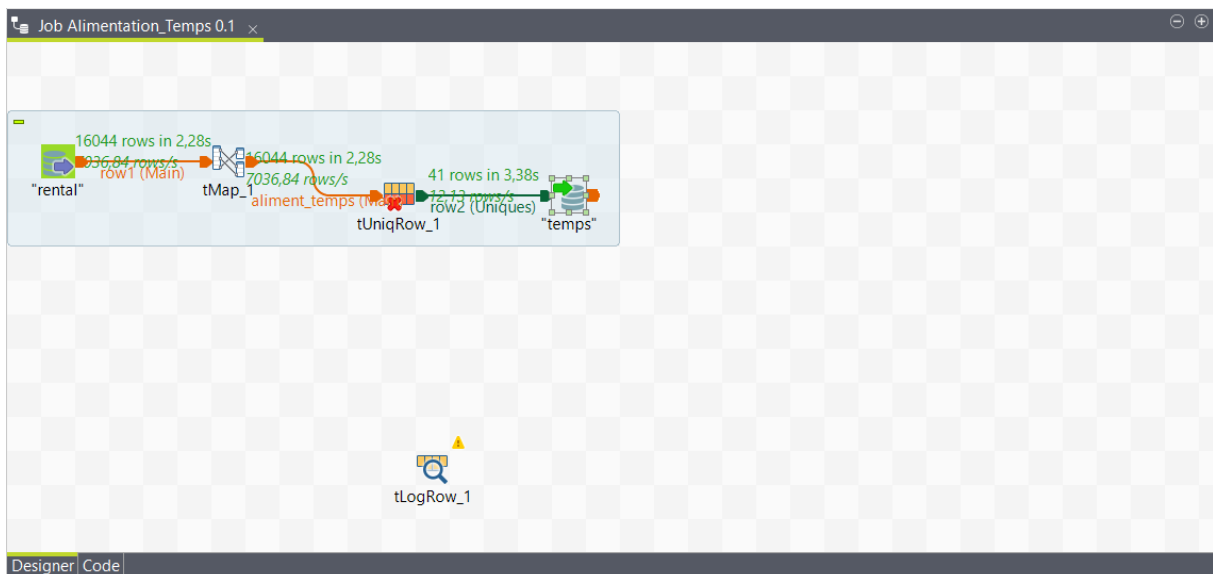
X. Alimentation Catégorie



Ce graphique nous montre le job de l'alimentation de la table catégorie. Nous alimentons notre table catégorie à partir de la table category. Nous sélectionnons deux colonnes sur les quatre parce-que la colonne last_update

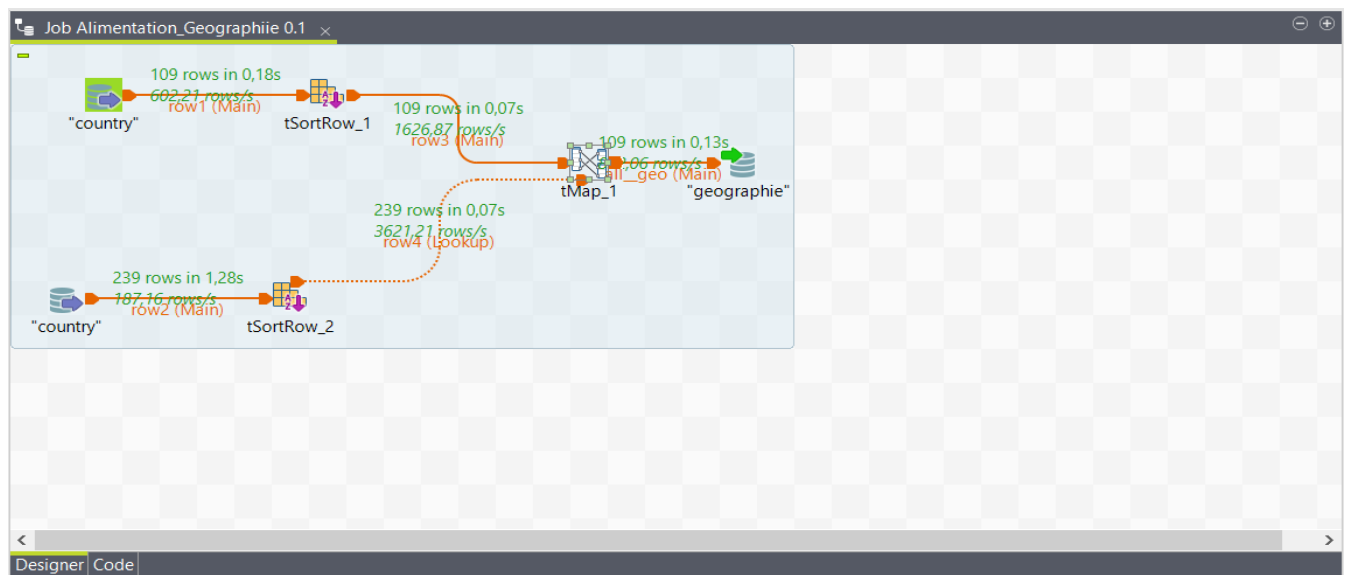
n'a pas trop d'impact. La table catégorie contenait 16 catégories et toutes ses catégories ont été extraites et alimentées à la table catégorie.

XI. Alimentation Temps



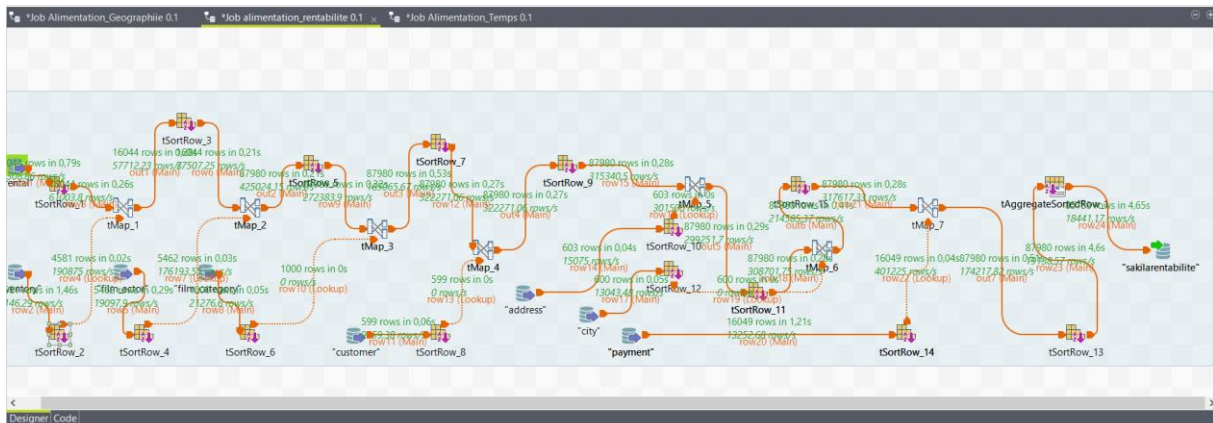
Ce graphique nous montre le job de l'alimentation de la table temps. Nous alimentons notre table temps à partir de la table rental. La table rental contenait 16044 lignes avec l'intégration d'un tUniqRow on parvient à charger 41 lignes dans la table temps. La table rental contenait 7 colonnes, mais on a utilisé que les colonnes rental_id et rental qui sont indispensables à cette étude.

XII. Alimentation Géographie



Ce graphique nous montre le job de l'alimentation de la table géographie. Nous alimentons notre table géographie à partir des tables country de sakila et country de world. La table country_sakila contenait 109 lignes et country_world 239 lignes. Grace à une jointure entre les deux tables, on parvient à alimenter 109 lignes dans la table géographie.

XIII. Alimentation Rentabilité

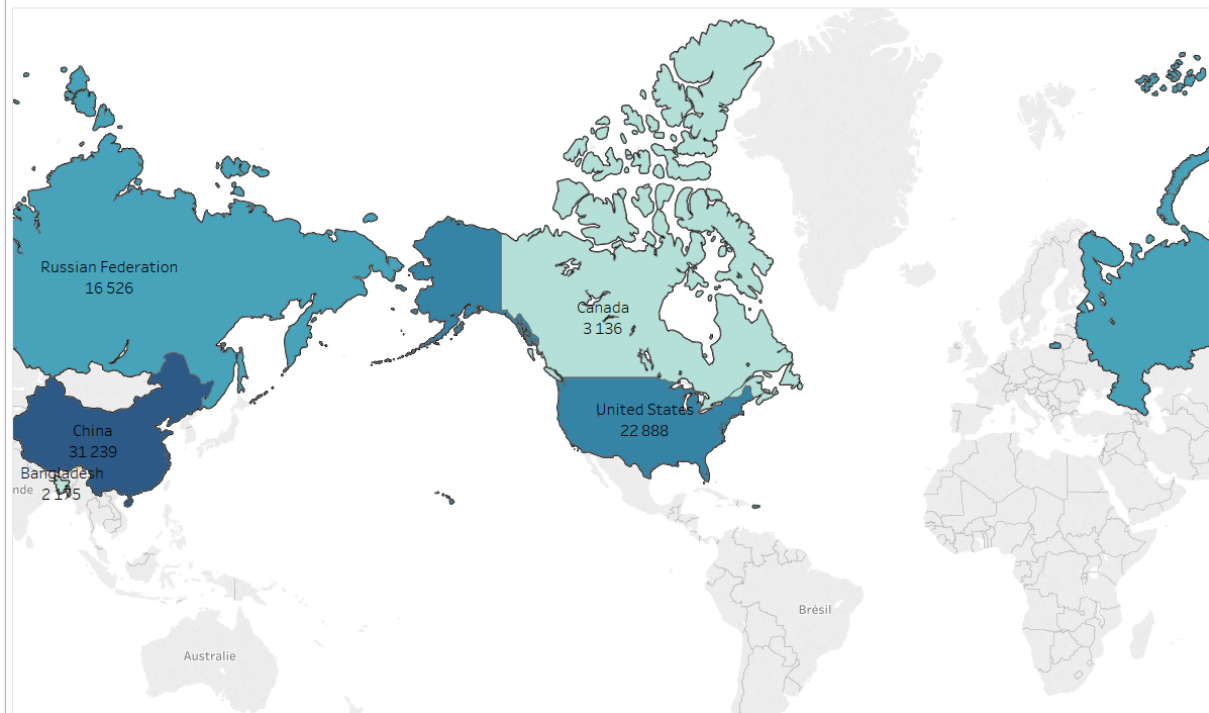


Ce graphique nous montre le job de l'alimentation de la table sakilarentabilite. L'alimentation de cette table a été faite grâce à une utilisation de plusieurs tables et des fonctions Talend Open Studio. On a commencé à faire une jointure entre les tables rental et inventory de grâce la clé de jointure inventory_id de inventory qui est migré vers la table rental comme clé étrangère. Ensuite on a fait une jointure entre la sortie 1 et la table film_actor grâce à la colonne film_id qui est utilisé comme clé de jointure entre les deux tables. La même colonne film_id qu'on avait utilisé pour faire la jointure précédente est réutilisée pour faire la jointure entre la sortie 2 et film_category. On a fait la jointure entre la sortie 3 et la table customer grâce à la colonne customer_id de customer. La colonne address_id est utilisée pour faire la jointure entre la table address_id et la sortie 4. On a fait la jointure entre la table city et la sortie 5 grâce à la colonne city_id de city qui constitue la clé de jointure. Enfin par une jointure de la table payment et la sortie 6 et l'utilisation de la fonction tAggregateSortedRow on parvient à obtenir une alimentation complète de la sakilarentabilite.

XIV. Les rapports conçus avec Tableau commentés

4. Montant des locations de films par pays

montant locations de films par catégorie et par pays

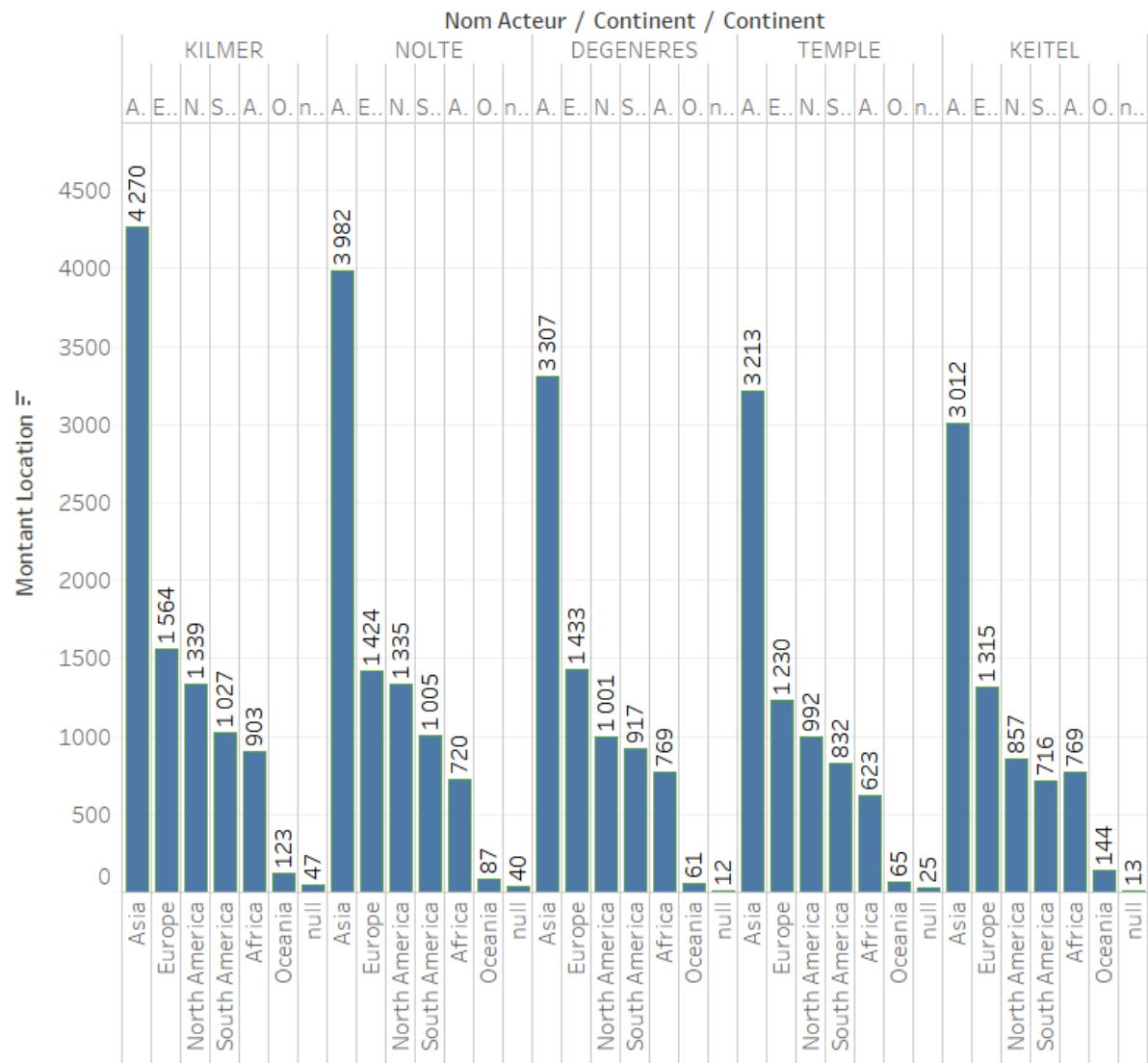


Titre : Montant des locations de films par pays

Ce graphe nous montre une aperçue sur le montant des locations de films dans chaque pays. Nous constatons que le montant des films loués est plus élevé en chine, aux Etat Unis et en Russie. Avec la chine en tête qui englobe le plus grand montant des locations.

5. Montant des films crédités à chaque acteur dans chaque continent

montant des films crédité à chaque acteur dans chaque continent

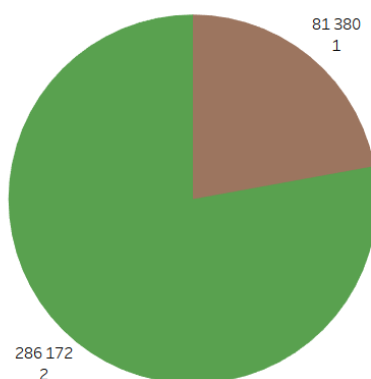


Titre : Montant des films crédités à chaque acteur dans chaque continent

Ce graphe nous montre le montant des films crédités à chaque acteur dans chaque pays .Nous constatons que l'acteur KILMER a le plus grand montant qui lui est crédité dans la continent asiatique que dans les autres continents suivie de NOLTE et DEGENERES .Et s'est aussi KILMER est l'acteur qui a le plus grand montant des films qui lui est crédité .

6. Montant des locations de films par semestre

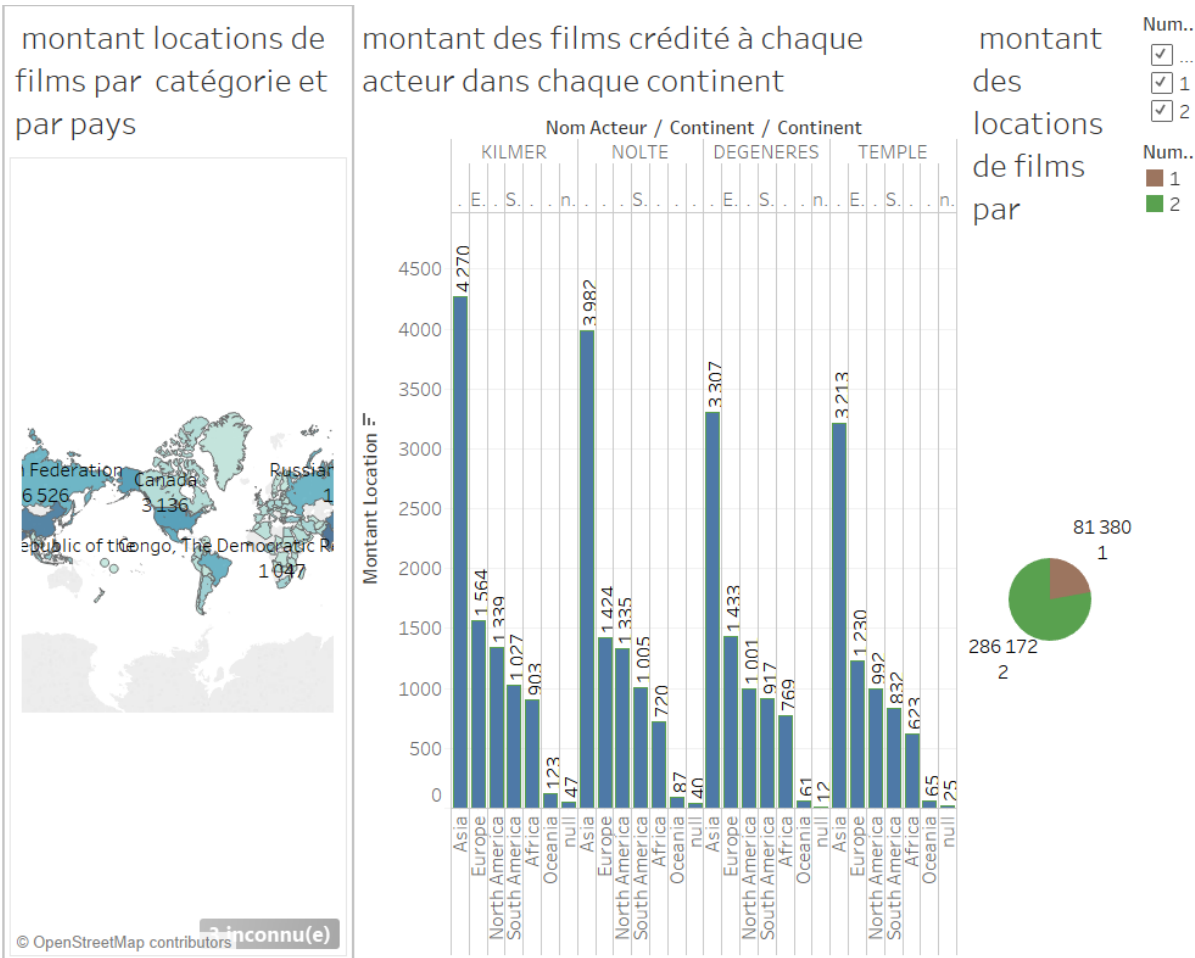
montant des locations de films par semestre



Titre : Montant des locations de films par semestre

Cette graphique en secteur nous renseigne sur le montant des locations de films par semestre. Nous constatons que, le plus grand des montants de locations se fait au niveau du semestre 2 avec un montant de 286172 contre 81380 pour le semestre 1.

XV. Le tableau de bord



Titre : Tableau de bord

Ce graphique nous permet de répondre les questions majeures de cette étude. Il a été conçu à partir de trois rapports. Ce tableau de bord, nous permet de connaître le montant de location des films par catégorie et par pays, le

montant des films crédité à chaque acteur dans chaque continent, le montant des locations de films par semestre.

7. Lien du tableau de bord

[Partagé avec moi - Tableau Online](#)