

TP Business Intelligence Project Orion

La société Orion

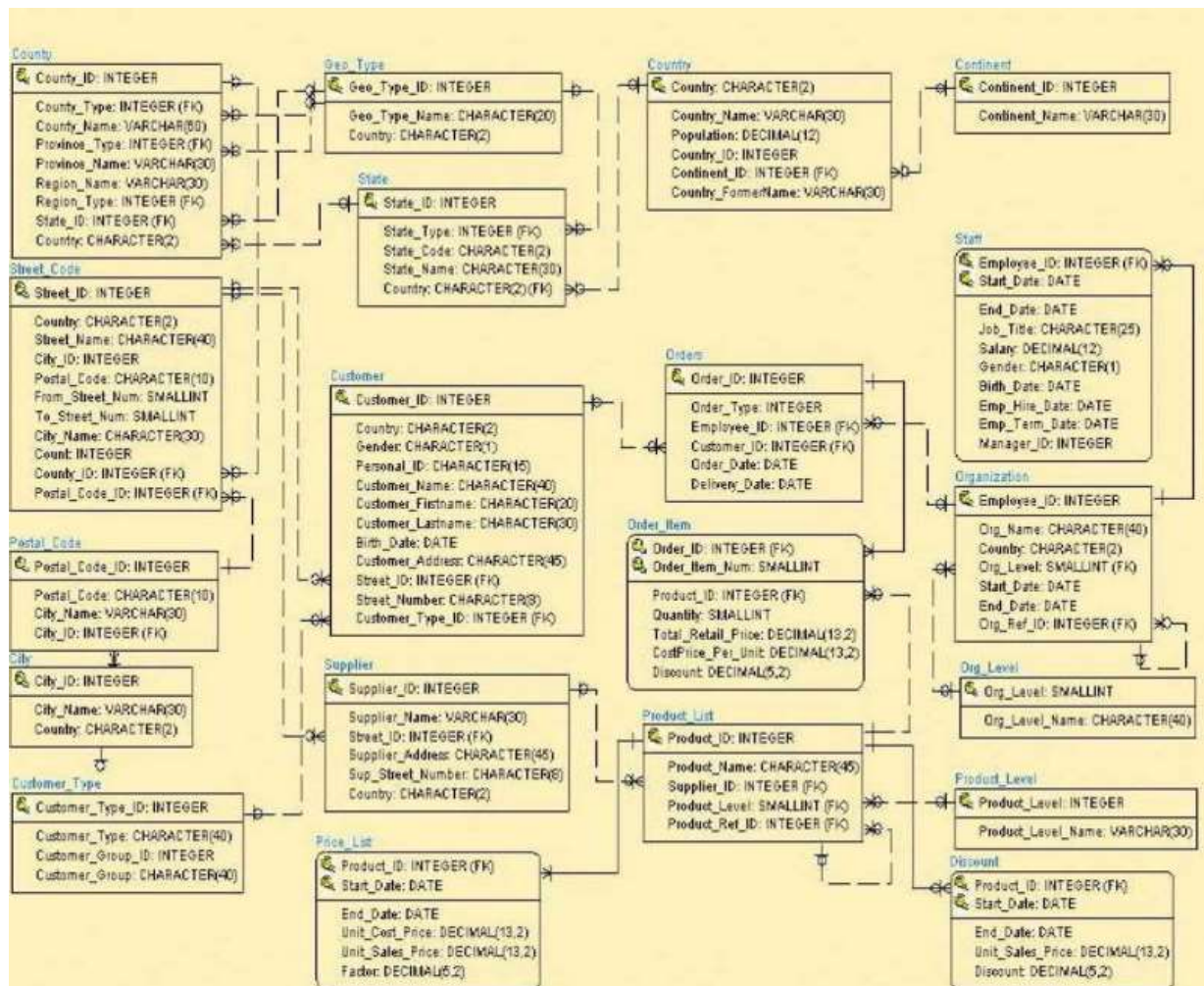
Cette société fictive, présente au niveau mondial, est spécialisée dans la commercialisation d'articles de sport et d'extérieur. Les données disponibles regroupent des informations sur :

- les employés
- les produits
- les clients
- les commandes
- les fournisseurs

La société Orion souhaite améliorer sa performance à l'aide d'un système décisionnel.

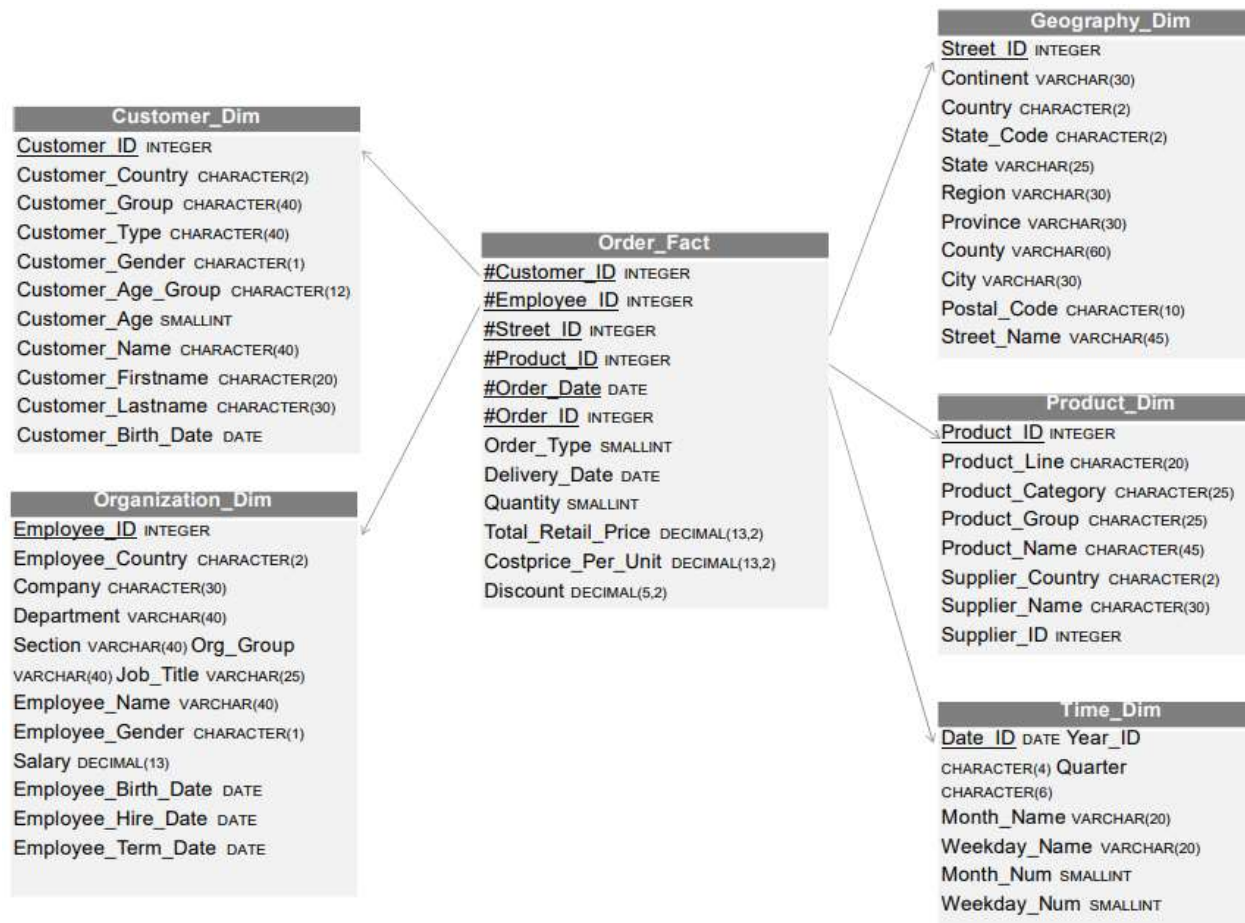
Il faut donc construire un entrepôt de données capable de répondre aux besoins de requête, de reporting, et d'analyses avancées.

Voici le schéma relationnel de la base de données opérationnelle de l'entreprise d'où proviendront les données de l'entrepôt :



Ces tables sont stockées dans la base de données Microsoft Access nommée orion.mdb, hormis la table Staff stockée dans le fichier Microsoft Excel nommé staff.xls.

Voici le schéma en étoile de l'entrepôt de données :



Création des tables de l'entrepôt

```

CREATE TABLE Customer_Dim(
  Customer_ID INTEGER PRIMARY KEY,
  Customer_Country CHARACTER(2),
  Customer_Group CHARACTER(40),
  Customer_Type CHARACTER(40),
  Customer_Gender CHARACTER(1),
  Customer_Age SMALLINT,
  Customer_Name CHARACTER(40),
  Customer_firstname CHARACTER(20),
  Customer_Lastname CHARACTER(30),
  Customer_Group_age CHARACTER(12),
  Customer_Birth_Date DATE
);
  
```

```

CREATE TABLE Organization_Dim (
  Employee_ID INTEGER PRIMARY KEY,
  Employee_Country CHARACTER(2),
  Company CHARACTER(30),
  Department VARCHAR(40),
  Section VARCHAR(40),
  Org_Group VARCHAR(40),
  Job_Title VARCHAR(25),
  Employee_Name VARCHAR(40),
  Employee_Gender CHARACTER(1),
  Salary DECIMAL(13),
  Employee_Birth_Date DATE,
  Employee_Hire_Date DATE,
  Employee_Term_Date DATE
);
  
```

```
CREATE TABLE Geography_Dim (
  Street_ID INT PRIMARY KEY,
  Continent VARCHAR(30),
  Country CHARACTER(2),
  State_Code CHARACTER(2),
  State VARCHAR(25),
  Region VARCHAR(30),
  Province VARCHAR(30),
  County VARCHAR(60),
  City VARCHAR(30),
  Postal_Code CHARACTER(10),

  Street_Name VARCHAR(45)
);
```

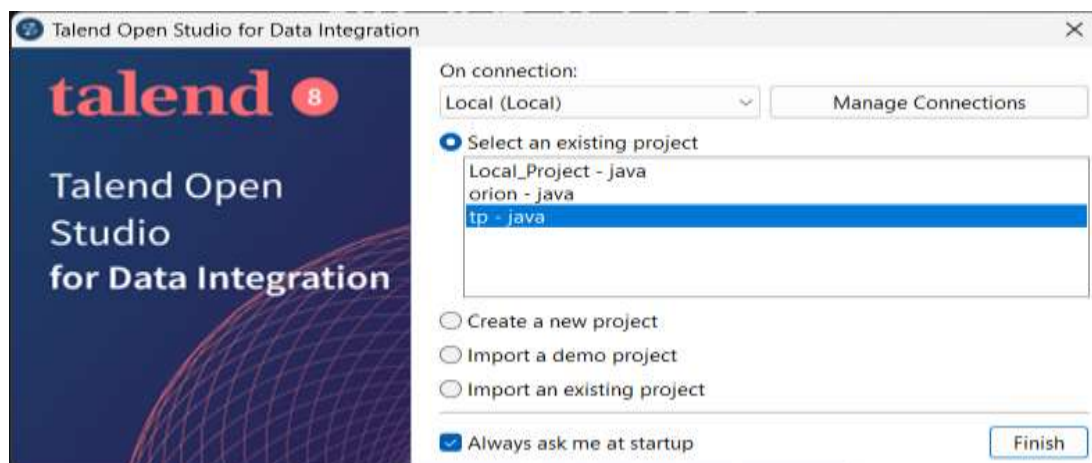
```
CREATE TABLE Product_Dim(
  Product_ID INT PRIMARY KEY,
  Product_Line CHARACTER(20),
  Product_Category CHARACTER(25),
  Product_Group CHARACTER(25),
  Product_Name CHARACTER(45),
  Product_Country CHARACTER(2),
  Supplier_Name CHARACTER(30),
  Supplier_ID INTEGER
);
```

```
CREATE TABLE Time_Dim (
  Date_ID DATE PRIMARY KEY,
  Year_ID CHARACTER(4),
  Quarter CHARACTER(6),
  Month_Name VARCHAR(20),
  Weekday_Name VARCHAR(20),
  Month_Num SMALLINT,
  Weekday_NUM SMALLINT
);
```


```
);
```

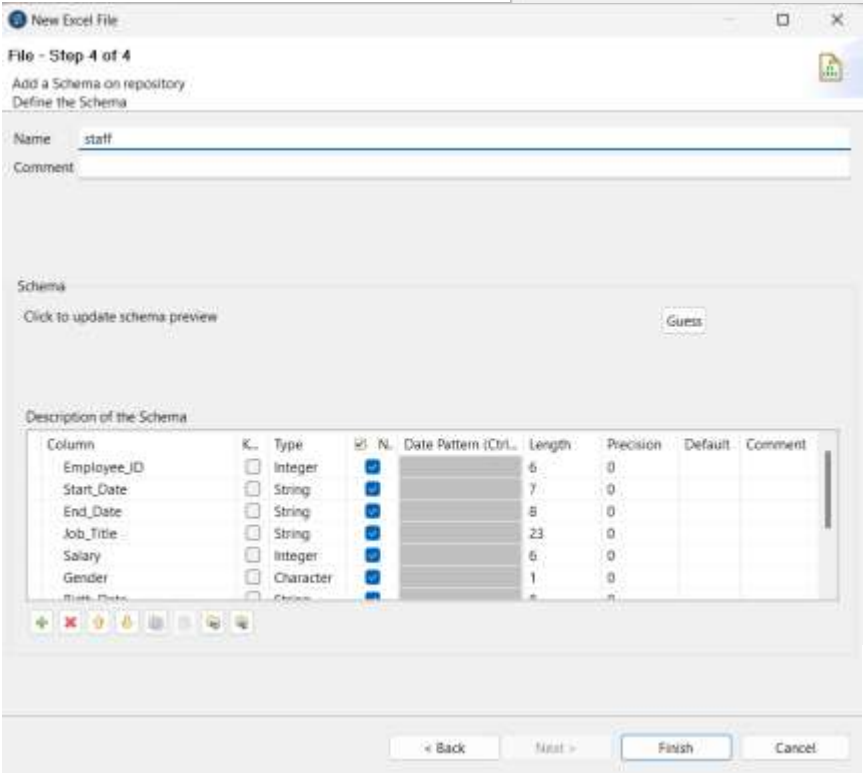
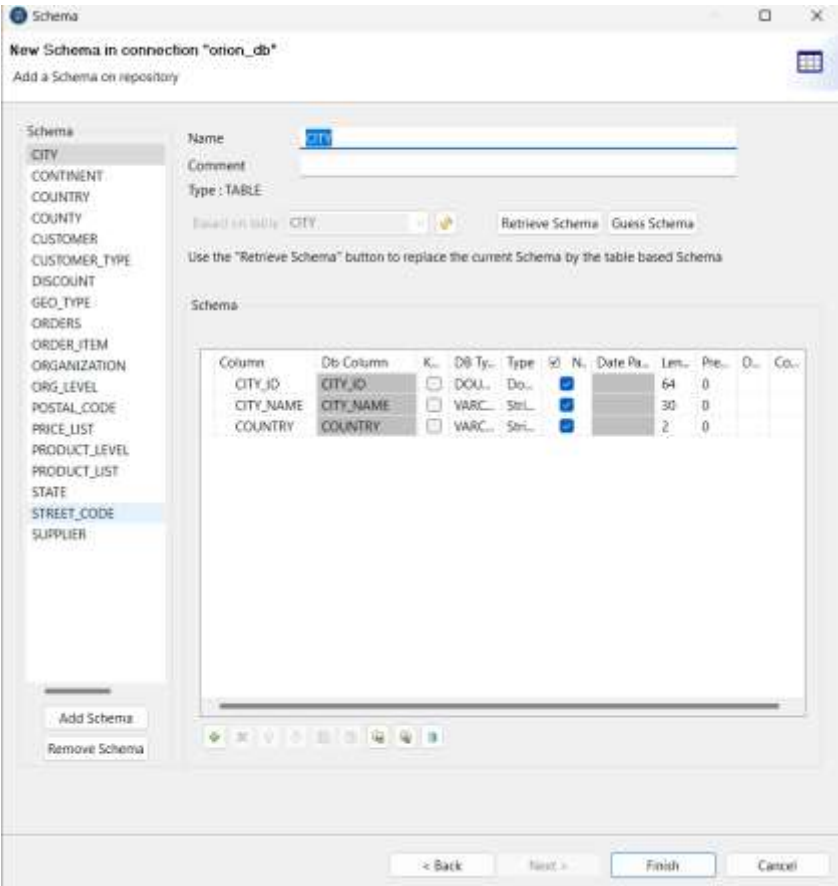
```
CREATE TABLE Order_Fact (
  Customer_ID INTEGER,
  Employee_ID INTEGER,
  Street_ID INT,
  Product_ID INT,
  Order_Date DATE,
  Order_ID INTEGER,
  Order_Type SMALLINT,
  Delivery_Date DATE,
  Quantity SMALLINT,
  Total_Retail_Proce DECIMAL(13,2),
  Costprice_Per_Unit DECIMAL(13,2),
  Discount DECIMAL(5,2),
  PRIMARY KEY
  (Customer_ID,Employee_ID,Street_ID,Product_ID,Order_Date,Order_ID),
  CONSTRAINT fk_cust_id FOREIGN KEY
  (Customer_ID) REFERENCES Customer_Dim
  (Customer_ID),
  CONSTRAINT fk_org_id FOREIGN KEY
  (Employee_ID) REFERENCES
  Organization_Dim (Employee_ID),
  CONSTRAINT fk_geo_id FOREIGN KEY
  (Street_ID) REFERENCES Geography_Dim
  (Street_ID),
  CONSTRAINT fk_prod_id FOREIGN KEY
  (Product_ID) REFERENCES Product_Dim
  (Product_ID),
  CONSTRAINT fk_ord_id FOREIGN KEY
  (Order_Date) REFERENCES Time_Dim
  (Date_ID)
);
```

Ouvrir Talend Open Studio. Créer un nouveau projet nommé orion avec l'option Java.



Spécification des données sources

>  orion_db 0.1



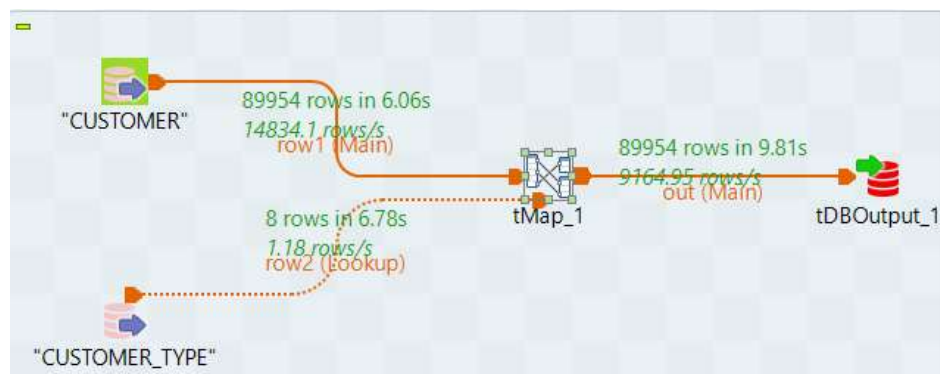
- File Excel
- staff 0.1
 - staff
 - Columns(10)

Spécification des données cibles

The screenshot shows the 'Database Connection' dialog box, Step 2/2, titled 'Update Database Connection - Step 2/2'. It includes a warning: 'You must press the Check Button to check the Database Setting'. The 'DB Type' is set to 'Oracle with SID'. The 'Db Version' is 'Oracle 11 and above'. The 'String of Connection' is 'jdbc:oracle:thin:@OTANA-PC:1521:ORCL'. The 'Login' is 'sys as sysdba', 'Password' is masked with dots, 'Server' is 'OTANA-PC', 'Port' is '1521', 'Sid' is 'ORCL', and 'Schema' is empty. There is an empty field for 'Additional parameters'. A 'Test connection' button is on the right. At the bottom, there are 'Export as context' and 'Revert Context' buttons, a link 'How to install a driver', and navigation buttons '< Back', 'Next >', 'Finish', and 'Cancel'.

Remplissage de la table Customer_Dim

Créer un job nommé Job01_Customer_Dim.



Tmap construction

row1	
Column	
CUSTOMER_ID	
COUNTRY	
GENDER	
PERSONAL_ID	
CUSTOMER_NAME	
CUSTOMER_FIRSTNAME	
CUSTOMER_LASTNAME	
BIRTH_DATE	
CUSTOMER_ADDRESS	
STREET_ID	
STREET_NUMBER	
CUSTOMER_TYPE_ID	
row2	
Expr. key	Column
row1.CUSTOMER_TYPE_ID	CUSTOMER_TYPE_ID
	CUSTOMER_TYPE
	CUSTOMER_GROUP_ID
	CUSTOMER_GROUP

Expression	Type	Variable
Mathematical.INT(TalendDate.formatD...	Double	<input checked="" type="checkbox"/> Age

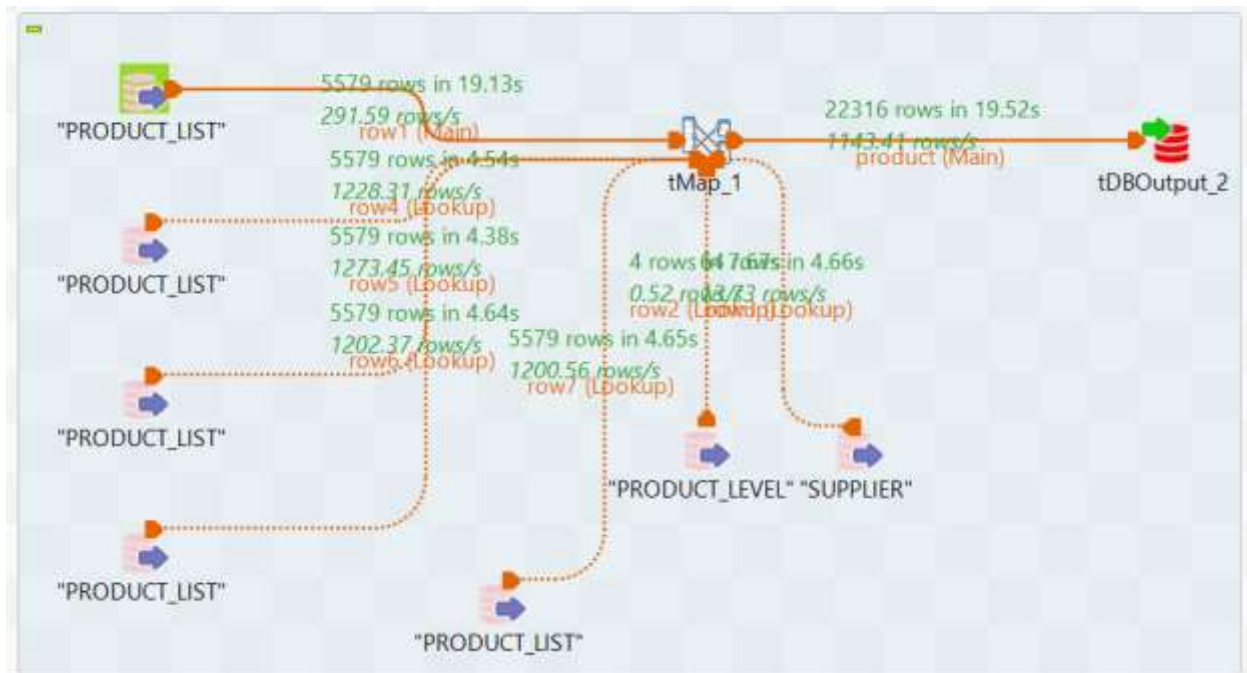
Expression	Column
row1.CUSTOMER_ID	Customer_ID
row1.COUNTRY	Customer_Country
row1.GENDER	Customer_Gender
row1.CUSTOMER_NAME	Customer_Name
row1.CUSTOMER_FIRSTNAME	Customer_firstname
row1.CUSTOMER_LASTNAME	Customer_lastname
row2.CUSTOMER_TYPE	Customer_Type
row2.CUSTOMER_GROUP	Customer_Group
Var.Age	Customer_Age
row1.BIRTH_DATE	Customer_Birth_Date
Var.Age<30?"<30 years": ... Var.Age<46?"30-45 ye...	CUSTOMER_GROUP_AGE

row1	Var	row2
Column	Expression	Column
CUSTOMER_ID	Mathematical.INT(TalendDate.formatD...	Customer_ID
COUNTRY		Customer_Country
GENDER		Customer_Gender
PERSONAL_ID		Customer_Name
CUSTOMER_NAME		Customer_firstname
CUSTOMER_FIRSTNAME		Customer_lastname
CUSTOMER_LASTNAME		Customer_Type
BIRTH_DATE		Customer_Group
CUSTOMER_ADDRESS		Var.Age
STREET_ID		Customer_Birth_Date
STREET_NUMBER		Customer_Age
CUSTOMER_TYPE_ID		CUSTOMER_GROUP_AGE

vérifier le résultat du job

Worksheet		Query Builder	
SELECT COUNT(*) FROM Customer_Dim		SELECT * FROM Customer_Dim WHERE SOMME=102	
Query Result		Script Output	
Task completed in 0.059 seconds			
COUNT (*)			
59532			
CUSTOMER_ID	CUSTOMER_GROUP	CUSTOMER_TYPE	CUSTOMER_AGE
1 FR Orion Club Gold members	Orion Club Gold members high activity	M	83 Albert Collet
2 ES Orion Club members	Orion Club members inactive	F	68 Mercedes Martinez
3 IT Orion Club members	Orion Club members high activity	M	53 Pier Egidio Boeris
4 US Orion Club members	Orion Club members low activity	M	53 James Kvarniq
5 US Orion Club Gold members	Orion Club Gold members medium activity	F	48 Sandrine Stephano
6 BE Orion Club members	Orion Club members median activity	M	70 Rent Van Lint
7 ES Orion Club members	Orion Club members high activity	F	48 Julián Escorihuela Monserrate
8 FI Orion Club members	Orion Club members low activity	M	30 Aki Ivonen
9 DE Orion Club Gold members	Orion Club Gold members medium activity	F	51 Cornelia Krah

Remplissage de la table Product_Dim



Tmap :

row1

Column
PRODUCT_ID
PRODUCT_NAME
SUPPLIER_ID
PRODUCT_LEVEL
PRODUCT_REF_ID

row2

Expr. key	Column
	PRODUCT_LEVEL
	PRODUCT_LEVEL_NAME

row3

Expr. key	Column
row1.SUPPLIER_ID	SUPPLIER_ID
	SUPPLIER_NAME
	STREET_ID
	SUPPLIER_ADDRESS
	SUP_STREET_NUMBER
	COUNTRY

row4

Expr. key	Column

product

Expression	Column
row1.PRODUCT_ID	Product_ID
row1.PRODUCT_NAME	Product_Name
row4.PRODUCT_NAME	Product_Line
row5.PRODUCT_NAME	Product_Category
row6.PRODUCT_NAME	Product_Group
row7.PRODUCT_NAME	Product_Country
row3.SUPPLIER_NAME	Supplier_Name
row3.SUPPLIER_ID	Supplier_ID

Column	K.	Type	id	NL	Date Pattern (Ctrl+...)	Length	Precision	Default	Comment
PRODUCT_ID		Double				64	0		
PRODUCT_NAME		String				45	0		
SUPPLIER_ID		Double				64	0		
PRODUCT_LEVEL		Double				64	0		
PRODUCT_REF_ID		Double				64	0		

product

Column	K.	Type	id	NL	Date Pattern (Ctrl+...)	Length	Precision	Default	Comment
Product_ID		Double				64	0		
Product_Name		String				45	0		
Product_Line		String				45	0		
Product_Category		String				45	0		
Product_Group		String				45	0		
Product_Country		String				45	0		

Le resultat :

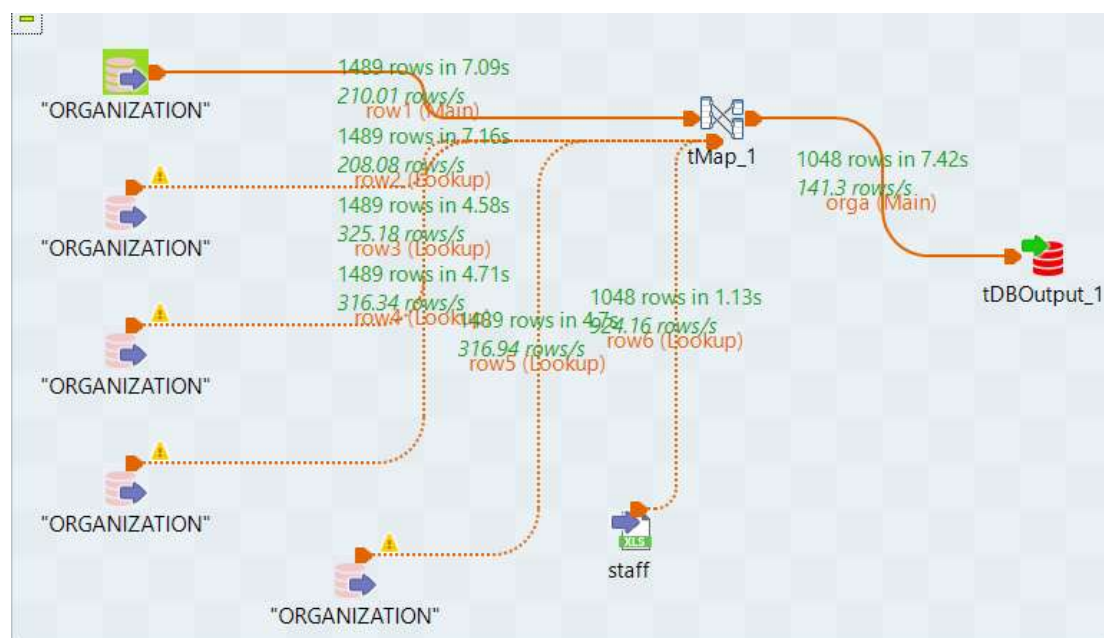
COUNT (*)

4499

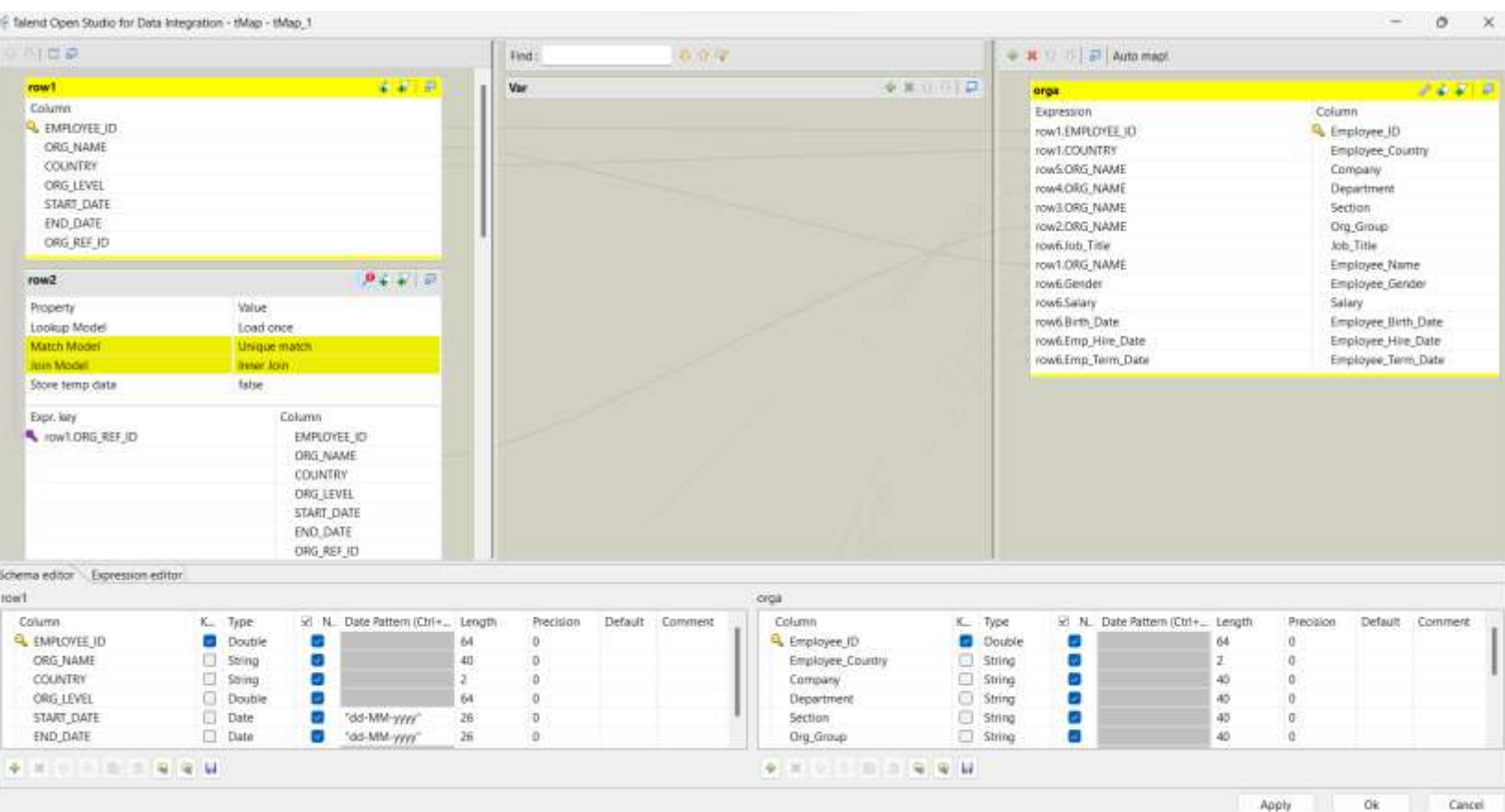
PRODUCT_ID	PRODUCT_LINE	PRODUCT_CATEGORY	PRODUCT_GROUP	PRODUCT_NAME	PR_SUPPLIER_NAME	SUPPLIER_ID
2.1000E+11				Children		
2.1010E+11	Children			Children Outdoors		
2.1010E+11	Children Outdoors	Children		Outdoor things, Kids		
2.1010E+11	Outdoor things, Kids	Children Outdoors	Children	Boy's and Girl's Xri Pants with Braces	Scandinavian Clothing A/S	50
2.1010E+11	Outdoor things, Kids	Children Outdoors	Children	Children's Jacket	Luna sastretería S.A.	4742
2.1010E+11	Outdoor things, Kids	Children Outdoors	Children	Children's Jacket Sidney	Scandinavian Clothing A/S	50
2.1010E+11	Outdoor things, Kids	Children Outdoors	Children	Children's Rain Set	Scandinavian Clothing A/S	50
2.1010E+11	Outdoor things, Kids	Children Outdoors	Children	Children's Rain Suit	Scandinavian Clothing A/S	50
2.1010E+11	Outdoor things, Kids	Children Outdoors	Children	Rain Suit for Children	Scandinavian Clothing A/S	50

9 rows selected.

Remplissage de la table Organization_Dim



Tmap



le resultat :

Worksheet

Query Builder

```
SELECT COUNT(*) FROM organization_dim;
SELECT * FROM organization_dim WHERE ROWNUM<10;
```

Script Output

Query Result

Task completed in 0.077 seconds

COUNT(*)

1048

EMPLOYEE_ID	ORG COMPANY	DEPARTMENT	SECTION	ORG_GROUP	JOB_TITLE	EMPLOYEE_NAME
120101	AU Orion Australia	Sales Management	Sales Management	Sales Management	Director	Patrick Lu
120102	AU Orion Australia	Sales Management	Sales Management	Sales Management	Sales Manager	Tom Zhou
120103	AU Orion Australia	Sales Management	Sales Management	Sales Management	Sales Manager	Wilson Daves
120104	AU Orion Australia	Administration	Administration	Administration	Administration Manager	Karen Billington
120105	AU Orion Australia	Administration	Administration	Administration	Secretary I	Liz Povey
120106	AU Orion Australia	Administration	Administration	Administration	Office Assistant II	John Wornsey
120107	AU Orion Australia	Administration	Administration	Shipping Charges	Office Assistant III	Sherie Sheedy
120108	AU Orion Australia	Administration	Goods Entrance	Stock Admin	Warehouse Assistant II	Gladys Gromek
120109	AU Orion Australia	Administration	Goods Entrance	Stock Admin	Warehouse Assistant I	Gabriele Baker

9 rows selected.

Remplissage de la table Time_Dim

Dans cette table, il faut rentrer toutes les dates du 01/01/1998 au 31/12/2002. Avec un programme pl/sql

```
DECLARE
vQuarter CHARACTER(6);
vMonth_Name VARCHAR(20);
vWeekday_Name VARCHAR(20);
vMonth_Num SMALLINT;
vWeekday_Num SMALLINT;
vDate_ID DATE := TO_DATE('01-01-1998', 'DD-MM-YYYY');
```

```

BEGIN
  WHILE vDate_ID < TO_DATE('31-12-2002', 'DD-MM-YYYY') LOOP
    vQuarter := TO_CHAR(vDate_ID, 'YYYY') || 'Q' || TO_CHAR(vDate_ID, 'Q');
    vMonth_Num := TO_NUMBER(TO_CHAR(vDate_ID, 'MM'));
    vMonth_Name := TO_CHAR(vDate_ID, 'Month');
    vWeekday_Num := TO_NUMBER(TO_CHAR(vDate_ID, 'D'));
    vWeekday_Name := TO_CHAR(vDate_ID, 'Day');
    INSERT INTO Time_Dim (Date_ID, Year_ID, Quarter, Month_Name,
Weekday_Name, Month_Num, Weekday_Num)
    VALUES (vDate_ID, TO_CHAR(vDate_ID, 'YYYY'), vQuarter, vMonth_Name,
vWeekday_Name, vMonth_Num, vWeekday_Num);
    vDate_ID := vDate_ID + 1; -- increment the date by one day
  END LOOP;
  COMMIT; -- commit the changes
  DBMS_OUTPUT.PUT_LINE('Time_Dim table populated successfully.');
```

-- display a message

```

EXCEPTION
  WHEN OTHERS THEN
    DBMS_OUTPUT.PUT_LINE('Error: ' || SQLCODE || ' - ' || SQLERRM); --
display an error message
END;
/
```

Resultat :






Worksheet

Query Builder

```
SELECT COUNT(*) FROM time_dim;  
SELECT * FROM time_dim WHERE ROWNUM<10;
```

Script Output x

Query Result x

     Task completed in 0.042 seconds

COUNT (*)

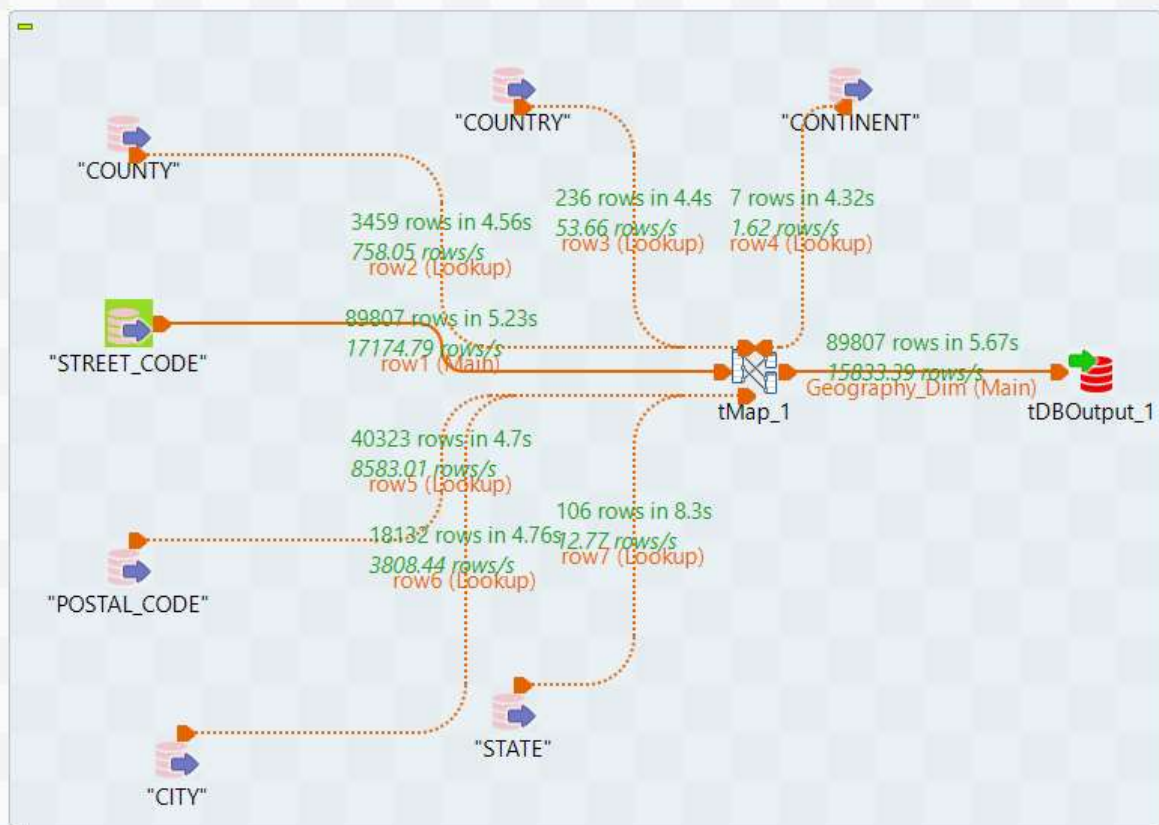
1825

DATE_ID	YEAR	QUARTE	MONTH_NAME	WEEKDAY_NAME	MONTH_NUM	WEEKDAY_NUM

01-JAN-98	1998	1998Q1	January	Thursday	1	5
02-JAN-98	1998	1998Q1	January	Friday	1	6
03-JAN-98	1998	1998Q1	January	Saturday	1	7
04-JAN-98	1998	1998Q1	January	Sunday	1	1
05-JAN-98	1998	1998Q1	January	Monday	1	2
06-JAN-98	1998	1998Q1	January	Tuesday	1	3
07-JAN-98	1998	1998Q1	January	Wednesday	1	4
08-JAN-98	1998	1998Q1	January	Thursday	1	5
09-JAN-98	1998	1998Q1	January	Friday	1	6

9 rows selected.

Remplissage de la table Geography_Dim



Talend Open Studio for Data Integration - tMap - tMap_1

Find:

Var:

Auto map

Geography_Dim

Expression	Column
row1.STREET_ID	Street_ID
row4.CONTINENT_NAME	Continent
row3.COUNTRY_NAME	Country
row7.STATE_CODE	State_Code
row7.STATE_NAME	State
row2.REGION_NAME	Region
row2.PROVINCE_NAME	Province
row2.COUNTY_NAME	County
row6.CITY_NAME	City
row5.POSTAL_CODE	Postal_Code
row1.STREET_NAME	Street_Name

Schema editor | Expression editor

row1

Column	K..	Type	N..	Date Patter...	Length	Precisi...	Def...	Com...
STREET_ID	<input checked="" type="checkbox"/>	Double	<input checked="" type="checkbox"/>		64	0		
COUNTRY	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		2	0		
STREET_NAME	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		40	0		
CITY_ID	<input type="checkbox"/>	Double	<input checked="" type="checkbox"/>		64	0		
POSTAL_CODE	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		10	0		

Geography_Dim

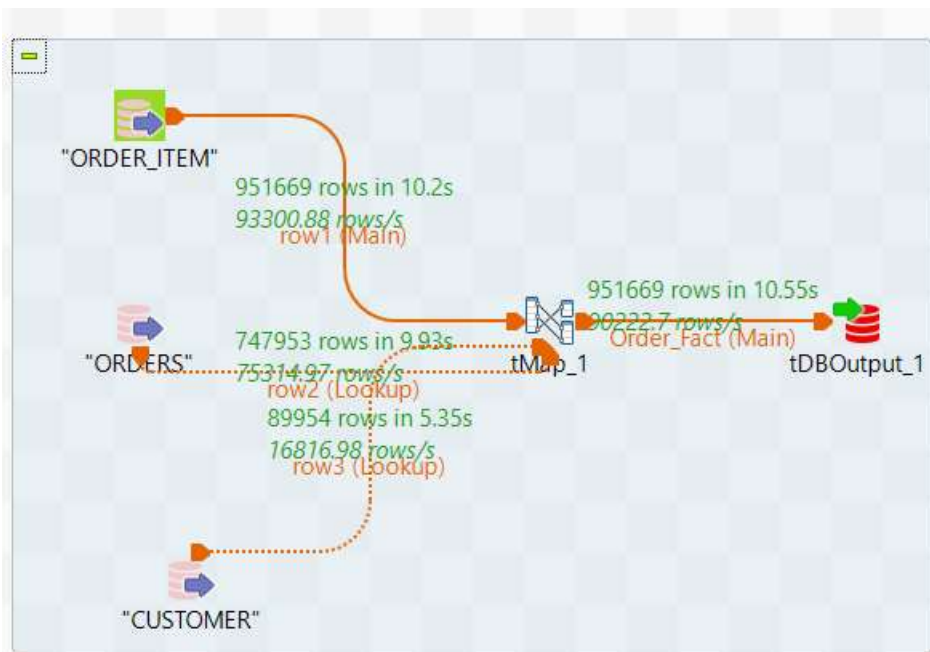
Column	K..	Type	N..	Date Patter...	Length	Precisi...	Def...	Com...
Street_ID	<input checked="" type="checkbox"/>	Double	<input checked="" type="checkbox"/>		64	0		
Continent	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		30	0		
Country	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		30	0		
State_Code	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		2	0		
State	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		30	0		

Apply Ok Cancel

Resultat

STREET_ID	CONTINENT	CD ST STATE	REGION	PROVINCE	COUNTY
400100001					
400100002					
400100003					
400100004					
400100005					
400100006					
400100007					
400100008					
400100009					

Remplissage de la table Order_Fact



Tmap :

Talend Open Studio for Data Integration - tMap - tMap_1

Find:

Var

Auto map!

Expression	Column
row2.CUSTOMER_ID	CUSTOMER_ID
row2.EMPLOYEE_ID	EMPLOYEE_ID
row3.STREET_ID	STREET_ID
row1.PRODUCT_ID	PRODUCT_ID
row2.ORDER_DATE	ORDER_DATE
row2.ORDER_ID	ORDER_ID
row2.ORDER_TYPE	ORDER_TYPE
row2.DELIVERY_DATE	DELIVERY_DATE
row1.QUANTITY	QUANTITY
row1.TOTAL_RETAIL_PRICE	TOTAL_RETAIL_PRO...
row1.COSTPRICE_PER_UNIT	COSTPRICE_PER_U...
row1.DISCOUNT	DISCOUNT

Schema editor | Expression editor

row1

Column	K...	Type	N.	Date Patter...	Length	Precisi...	Def.	Com...
ORDER_ID		Double			64	0		
ORDER_ITEM_NU...		Double			64	0		
PRODUCT_ID		Double			64	0		

Order_Fact

Column	K...	Type	N.	Date Patter...	Length	Precisi...	Def.	Com...
CUSTOMER_ID		Double			64	0		
EMPLOYEE_ID		Double			64	0		
STREET_ID		Double			64	0		

Apply Ok Cancel

Resultat

SQL Worksheet History											
0.053 seconds											
Worksheet Query Builder											
<pre>SELECT COUNT(*) FROM order_fact; SELECT * FROM order_fact;</pre>											
Script Output x Query Result x											
Task completed in 0.053 seconds											
COUNT (*)											
73											
CUSTOMER_ID	EMPLOYEE_ID	STREET_ID	PRODUCT_ID	ORDER_DAT	ORDER_ID	ORDER_TYPE	DELIVERY	QUANTITY	TOTAL_RETAIL_PROCE	COSTPRICE_PER_UNIT	DISCOUNT
63337	120553	6300100213	2.4050E+11	13-APR-98	1230563241	1	13-APR-98	3	98.1	15.7	
70700	120553	6300101268	2.4050E+11	13-APR-98	1230563254	1	13-APR-98	1	67.3	28	
13803	121075	9260101488	2.3010E+11	04-MAY-98	1230678379	1	04-MAY-98	2	346	72.75	
317	120146	1600101981	2.3010E+11	08-JUL-98	1231029300	1	15-JUL-98	1	140.3	58.4	
9187	120127	1600102727	2.2010E+11	08-JUL-98	1231029314	1	08-JUL-98	1	238.1	113.4	
61830	121073	9260124824	2.3010E+11	28-NOV-98	1231828073	1	28-NOV-98	1	96.7	48.45	
61830	121073	9260124824	2.3010E+11	28-NOV-98	1231828073	1	28-NOV-98	1	252	103.95	
28936	120623	4800100365	2.3010E+11	19-DEC-98	1231940165	1	19-DEC-98	2	266.4	60.6	
65715	121023	9260100978	2.4010E+11	09-JAN-99	1232071101	1	14-JAN-99	1	219.9	109.9	
15375	120145	1600102596	2.3010E+11	10-JAN-99	1232071155	1	10-JAN-99	2	25.4	6.1	
15375	120145	1600102596	2.3010E+11	10-JAN-99	1232071155	1	10-JAN-99	4	6.8	.8	

Lancement des jobs

Dans une étude réelle, les données sources évoluent en permanence. Les jobs doivent donc être planifiés régulièrement. Le lancement des jobs pourra se faire par exemple toutes les nuits pour prendre en compte les données modifiées pendant la journée. La planification des jobs peut se faire grâce au planificateur de tâches du système d'exploitation.