Des techniques basiques

Les Foreign Data Wrapper
pour faciliter l'interopérabilité des données

Marie-Claude Quidoz (CEFE/CNRS)

22/11/2013

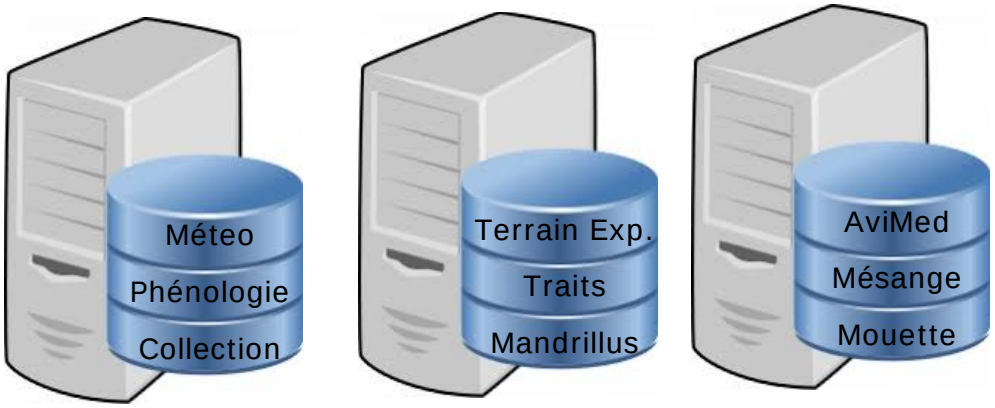
Centre d'Ecologie Fonctionnelle & Evolutive

UMR 5175, Montpellier

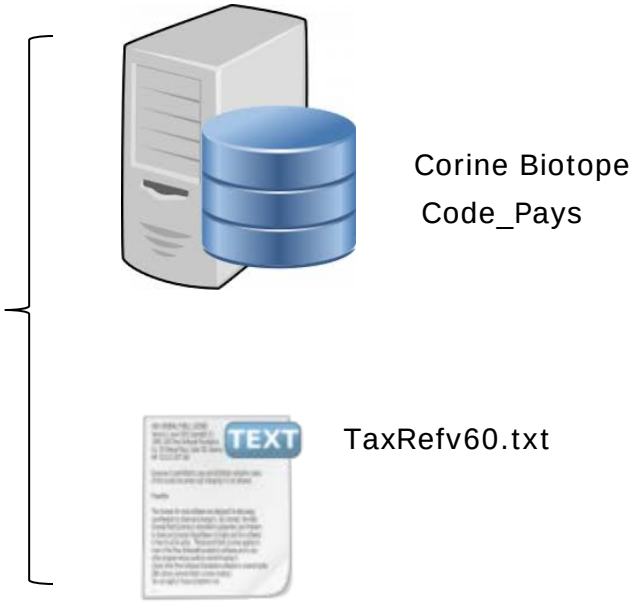
Introduction

- Une réponse à l'interopérabilité ponctuelle
  - Pour accéder à des données distantes
    - Distantes = en dehors de la base de données
    - Données stockées dans le même SGBD, dans un autre SGBD, dans un fichier CSV, …
- Utilisations courantes
  - Utiliser une base de données de référentiels
  - Avoir un seul point d'accès pour son SI
- Prérequis : avoir une connaissances des données à intégrer (syntaxe et structure)

Exemple : Utiliser des référentiels



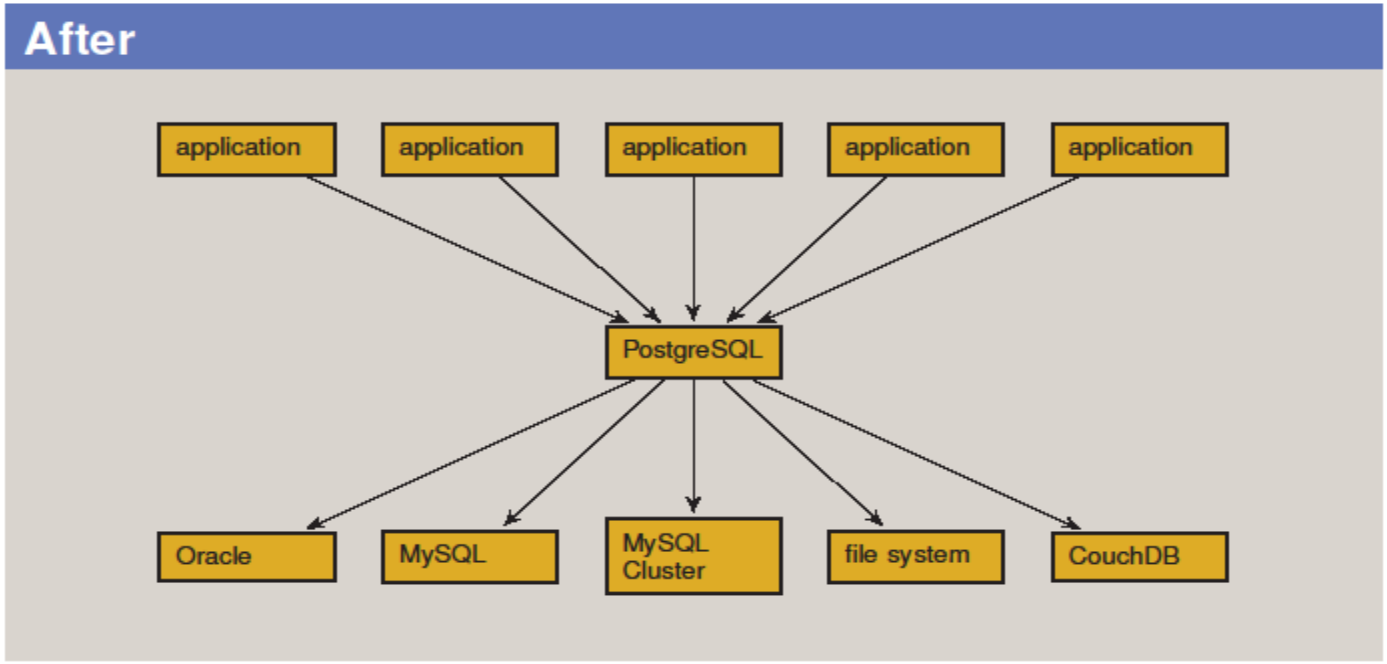
Serveurs de données thématiques



Serveurs de données pour les référentiels

Exemple : Avoir un seul point d'accès

Implications for Application Design and Deployment



Foreign Data Wrapper

- SQL/MED
  - SQL Management of External Data
  - Méthode pour accéder aux données distantes
- Conforme à la norme ISO/IEC 9075-2 (SQL/MED)
- Implémentation
  - Opérationnelle depuis PostgreSQL 9.1
  - Super intéressant depuis PostgreSQL 9.3
- Similitude avec dblink
  - Mais en + simple + performant + standard

Principe

- Définit une table étrangère
  - CREATE FOREIGN TABLE
- Sur un serveur étranger
  - CREATE SERVER
- Accessible depuis un wrapper de données
  - Nom du FOREIGN DATA WRAPPER qui gère le serveur

Exemple avec le FDW PostgreSQL

- CREATE EXTENSION postgres_fdw;

- CREATE SERVER avimed
  FOREIGN DATA WRAPPER postgres_fdw
  OPTIONS (host '193.49.134.127', dbname 'AviMed');

- CREATE FOREIGN TABLE releve(numero_releve integer, point_geographique character(50), date_releve date)
  SERVER avimed
  OPTIONS (schema_name 'terrestre', table_name 'releve_avifaune');

- Utilisation : SELECT * FROM releve;

Testé avec la version 9.3

PostgreSQL

De nombreux FDW

- Pour SGBD SQL :
  - PostgreSQL_fdw, Oracle_fdw, Mysql_fdw
- Pour SGBD NoSQL :
  - CouchDB_fdw, Mongo_fdw
- Pour fichiers :
  - file_fdw, json_fdw
- Autres :
  - www_fdw, ldap_fdw
- http://wiki.postgresql.org/wiki/Foreign_data_wrappers

Exemple avec le FDW www

- CREATE EXTENSION www_fdw;

- CREATE SERVER google_server
  FOREIGN DATA WRAPPER www_fdw
  OPTIONS (uri
  ‘https://ajax.googleapis.com/search/web?v=1.0’);

- CREATE FOREIGN TABLE google_table (title texte, link text,
  snippet text, q text)
  SERVER google_server;

- Utilisation : SELECT * FROM google_table
  WHERE q=‘cat dog’;

Conclusion

- SELECT, INSERT, UPDATE et DELETE
- Un compagnon qui peut être intéressant
  - CREATE MATERIALIZED VIEW
    v_releve AS SELECT * FROM releve;
- Attendre encore un peu pour une utilisation sur un serveur en production
- PGAdmin – pas encore adapté
- Possibilité d’écrire ses propres wrapper

Références

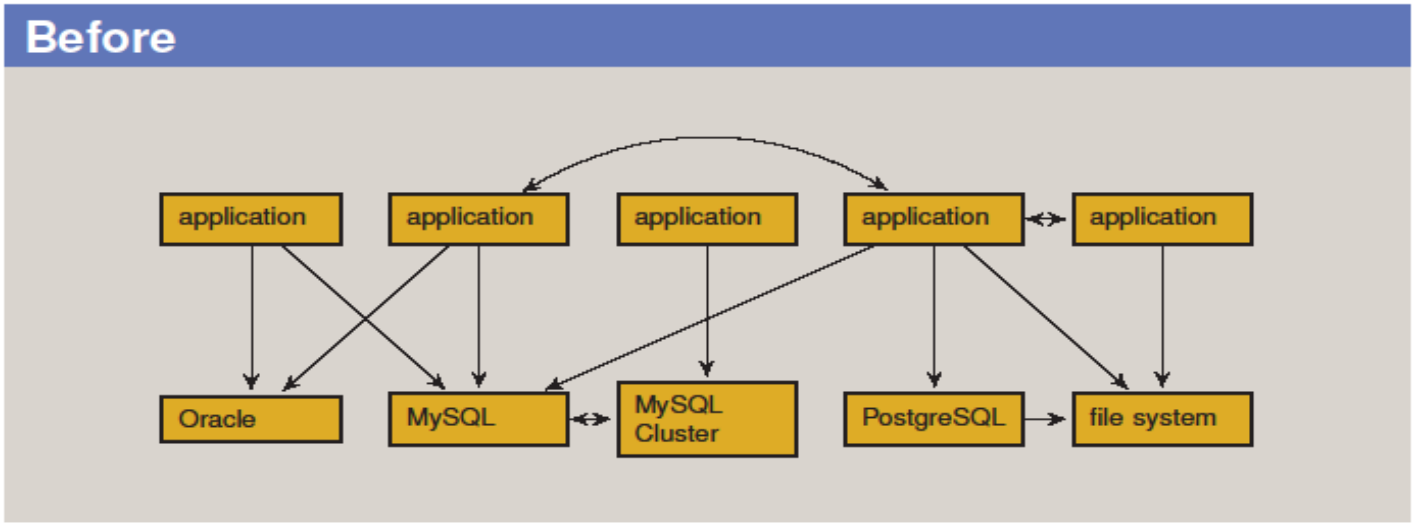
- http://si.cenlr.org/13-09-2013/postgresql-fdw-foreign-data-wrapper-taxref-inpn
- http://wiki.postgresql.org/wiki/Foreign_data_wrappers
- http://fr.slideshare.net/petereisentraut/sqlmed-doping-for-postgresql
- http://wiki.hsr.ch/Datenbanken/files/SQLMED_and_More_Schwendener_Paper.pdf

Compléments

- Plusieurs points d’accès à son SI
- Technologies spécifiques
- ETL

Exemple : Avoir un seul point d'accès

Implications for Application Design and Deployment



SQL/MED Doping for PostgreSQL, Peter Eisentraut, PGCon 2009

Technologies spécifiques

- Quelques solutions
  - Linked Servers by Microsoft
  - DBLink Oracle
  - dblink PostgreSQL

- Utilisé généralement pour les SELECT mais possible aussi pour INSERT, DELETE, UPDATE

Exemple de requête SELECT

```
CREATE OR REPLACE VIEW v_rbdd AS
 SELECT *
 FROM dblink('dbname=Referentiel host=193.49.134.127
                    user=quidoz password=xxxxx',
        'SELECT "LB_NOM","RANG","CD_NOM"
         FROM taxrefv60
         WHERE "RANG" = ''KD''')
 AS t1(lb_nom text, rang text, cd_nom integer);
```

```
SELECT * FROM v_rbdd;
```

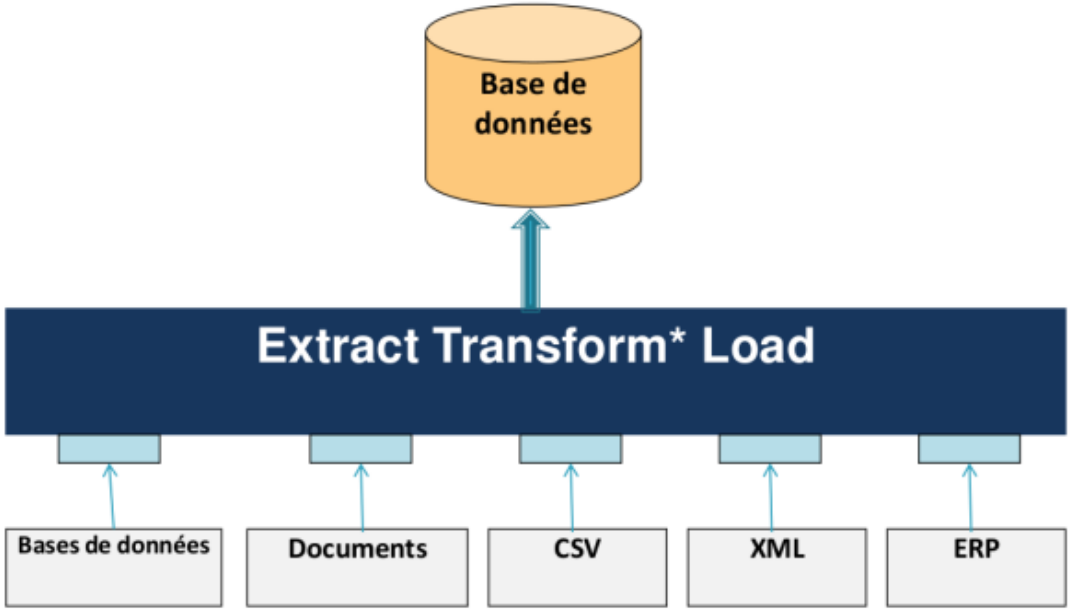
lb_nom	rang	cd_nom
Animalia	KD	183716
Bacteria	KD	187502
Monera	KD	187501
Chromista	KD	436946
Stramenopila	KD	524110
Fungi	KD	187496
Plantae	KD	187079
Protozoa	KD	187545
Protoctista	KD	187538

Testé avec la version 9.2

Bilan de DBlink (avec PostgreSQL)

- Pas très souple mais ca marche
  - Il faut définir la structure de la sortie
- dblink_connect et dblink_disconnect
- Moins riche que dans d'autres SGBD
- Pour en savoir plus
  - http://www.postgresonline.com/journal/archives/44-Using-DbLink-to-access-other-PostgreSQL-Databases-and-Servers.html

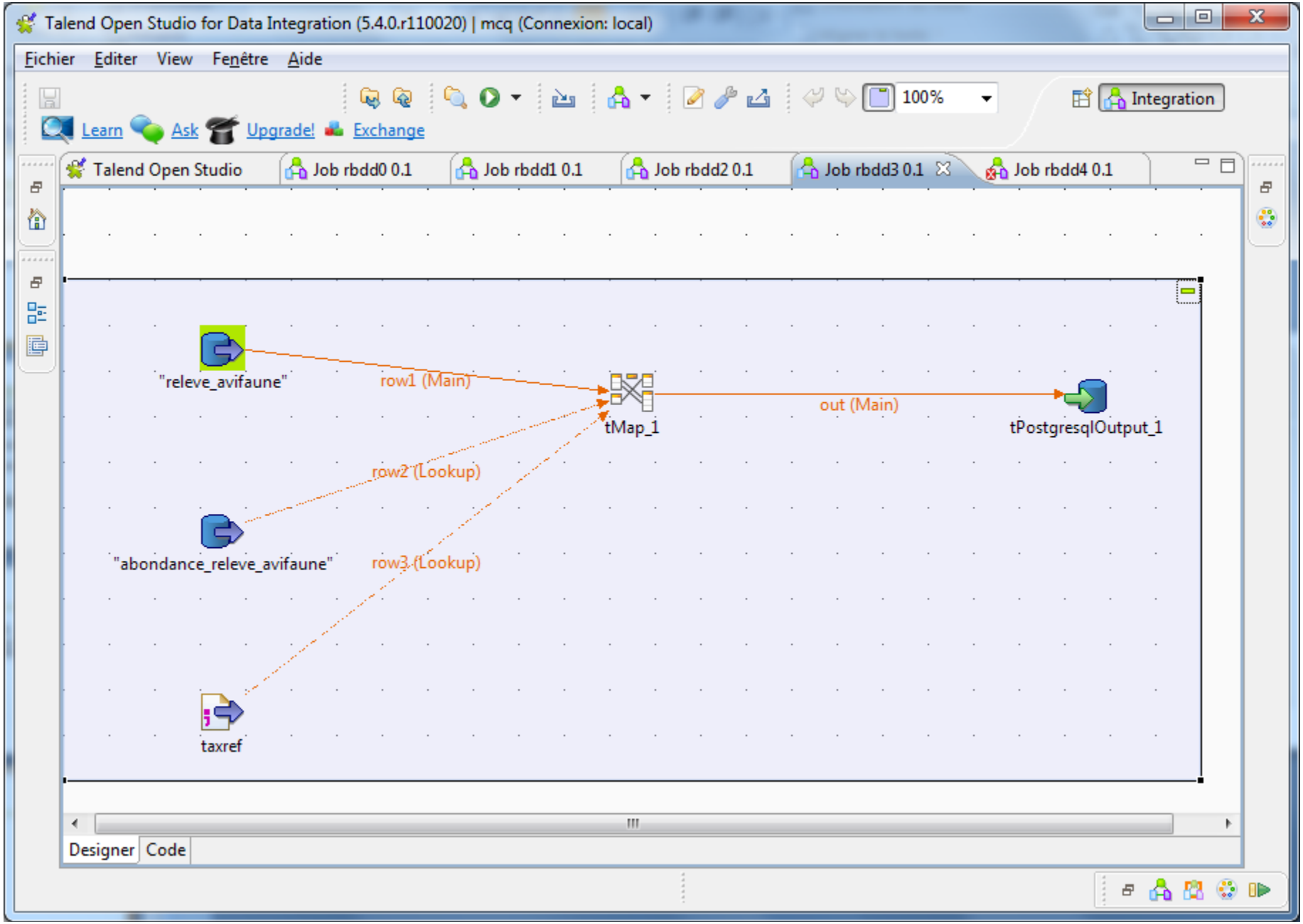
Extract Transform Load



*Conversion de type, correction, filtre, agrégation, formatage

http://www-igm.univ-mlv.fr/~dr/XPOSE2010/Extract_Transform_Load/etl.php

- De nombreux de connecteurs SGBD
  - Talend : SQLite, PostreSQL, MySQL, NoSQL, …


Talend Open Studio for Data Integration (5.4.0.r110020) | mcq (Connexion: local)
Fichier
Editer
View
Fenêtre
Aide
100%
Integration
Learn
Ask
Upgrade!
Exchange
Talend Open Studio
Job rbdd0 0.1
Job rbdd1 0.1
Job rbdd2 0.1
Job rbdd3 0.1
Job rbdd4 0.1
"releve_avifaune"
row1 (Main)
tMap_1
out (Main)
tPostgresqlOutput_1
row2 (Lookup)
"abondance_releve_avifaune"
row3 (Lookup)
taxref
Designer
Code

Talend Open Studio for Data Integration (5.4.0.r110020) | mcq (Connexion: local)

Fichier Editer View Fenêtre Aide

Learn Ask Upgrade! Exchange

Integration

Talend Open Studio | Job rbdd0 0.1 | Job rbdd1 0.1 | Job rbdd2 0.1 | Job rbdd3 0.1 | Job rbdd4 0.1

```
2
3  right (c) 2006-2013, Talend Inc.
4
5   source code has been automatically generated by_Talend Open Studio for Data Integration
6  censed under the Apache License, Version 2.0 (the "License");
7  may not use this file except in compliance with the License.
8  may obtain a copy of the License at
9  ://www.apache.org/licenses/LICENSE-2.0
10
11 ss required by applicable law or agreed to in writing, software
12 ributed under the License is distributed on an "AS IS" BASIS,
13 OUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 the License for the specific language governing permissions and
15 tations under the License.
16
17  mcq.rbdd3_0_1;
18
19 routines.Mathematical;
20 routines.DataOperation;
21 routines.Relational;
22 routines.TalendDate;
23 routines.TalendDataGenerator;
24 routines.Numeric;
25 routines.TalendString;
26 routines.StringHandling;
27 routines.system.*;
28 routines.system.api.*;
29 java.text.ParseException;
30 java.text.SimpleDateFormat;
31 java.util.Date;
32 java.util.List;
33 java.math.BigDecimal;
```

Designer | Code