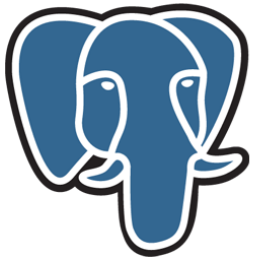


PostgreSQL

PGDay.IT 2011

Monash University Prato Centre
Venerdì 25 Novembre 2011

ETL con Kettle

Giulio Calacoci

Italian PostgreSQL Users Group

www.itpug.org
www.postgresql.org

Chi sono?

- 2ndQuadrant Italia:
  - Business Intelligence
  - Software Development:
    - Java
    - PHP

Sommario

- ETL
  - Concetti di base
  - Esempi di applicazione
- Kettle
  - Componenti
  - Casi d'uso reale

Principali Utilizzatori

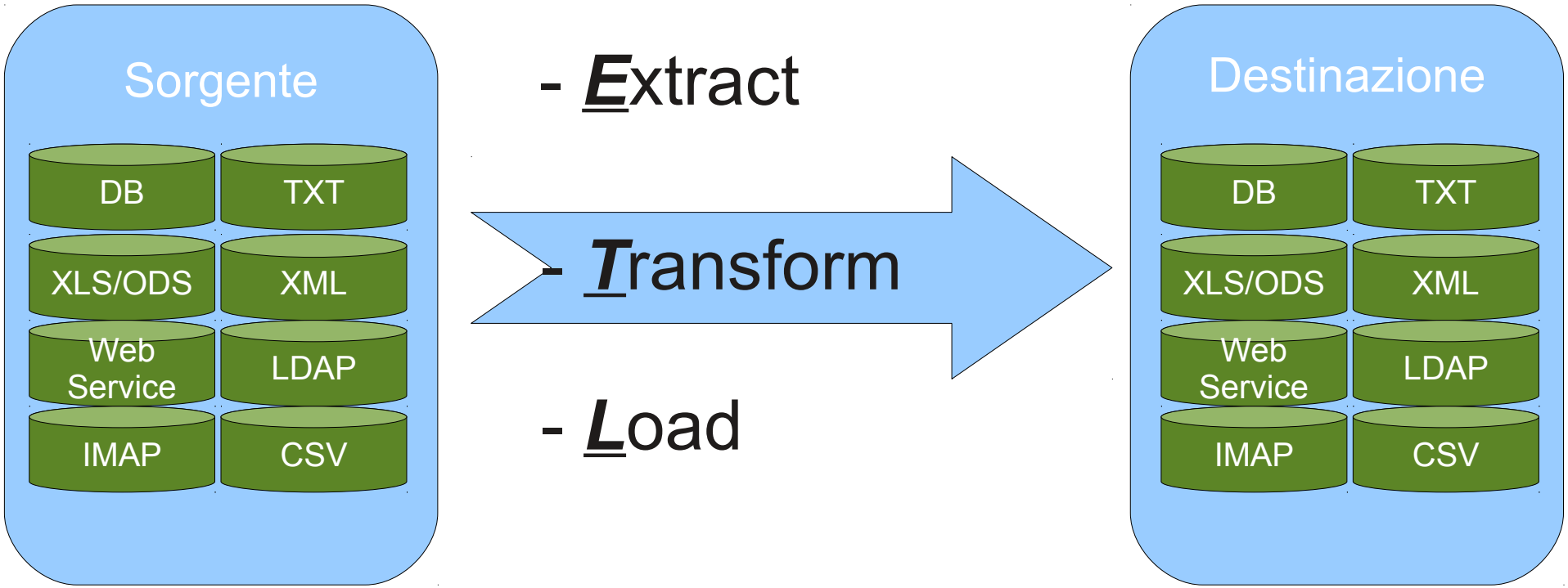
- Developer:
  - Migrazioni di Applicazioni
  - Migrazioni di Database
  - Importazione / Esportazione dati
  - QA
- Business Intelligence Expert:
  - OLTP
  - Fonti di dati ausiliarie
  - WebService
  - QA

ETL

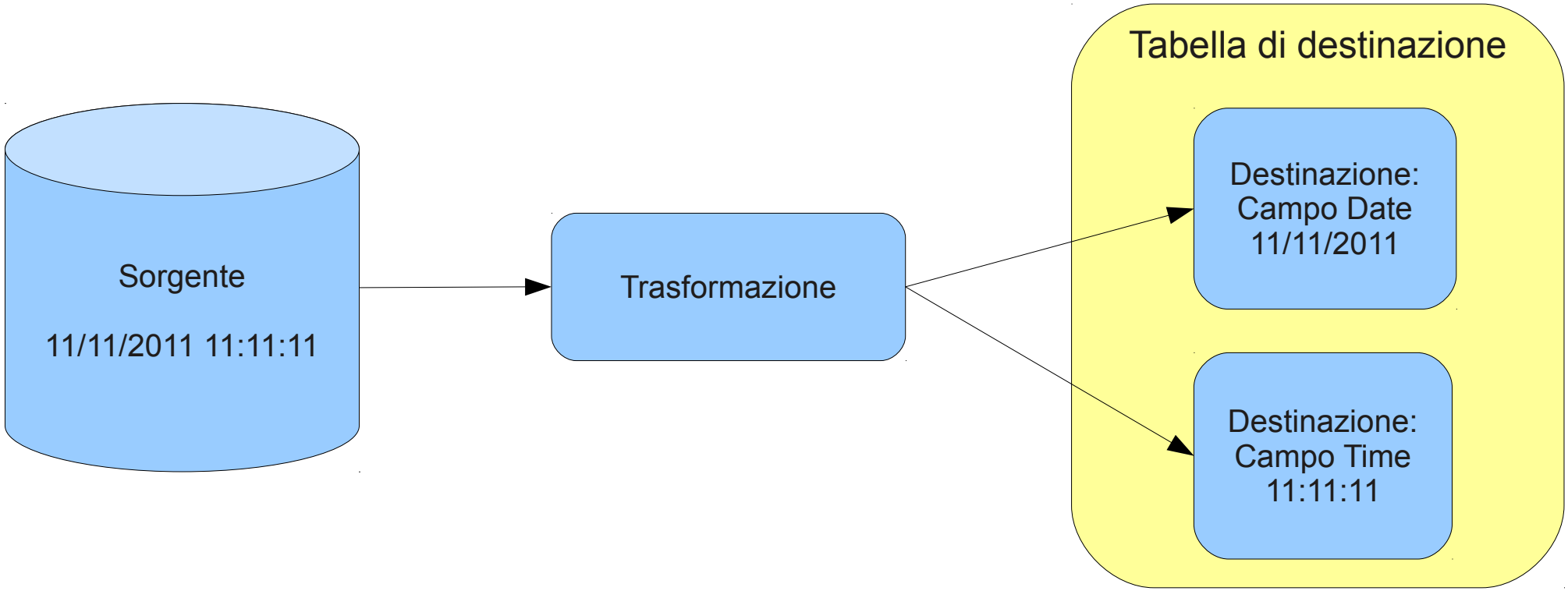
Chi non ha mai scritto un programma del genere?

```php
<?php
function csv_file_to_mysql_table($source_file, $target_table, $max_line_length=10000) {
    if (($handle = fopen("$source_file", "r")) !== FALSE) {
        $columns = fgetcsv($handle, $max_line_length, ",");
        foreach ($columns as &$column) {
            $column = str_replace(".","",$column);
        }
        $insert_query_prefix = "INSERT INTO $target_table (".join(",",$columns).")\nVALUES";
        while (($data = fgetcsv($handle, $max_line_length, ",")) !== FALSE) {
            while (count($data)<count($columns))
                array_push($data, NULL);
            $query = "$insert_query_prefix (".join(",",quote_all_array($data)).");";
            mysql_query($query);
        }
        fclose($handle);
    }
}
```

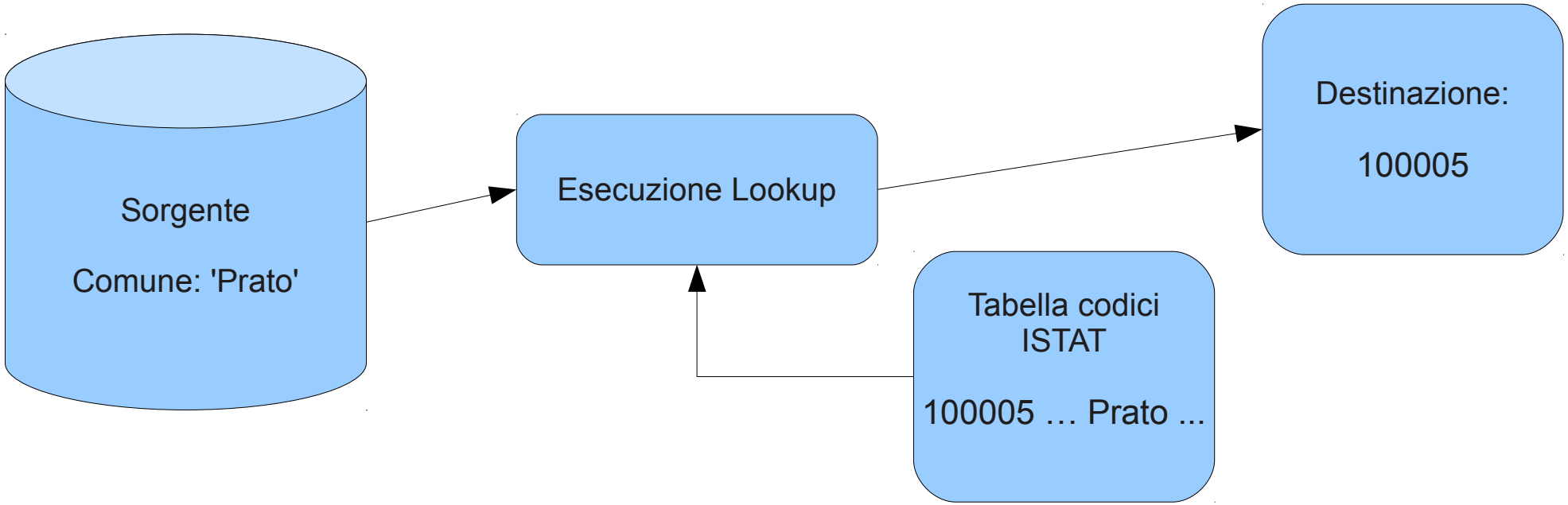
ETL


Sorgente
DB
TXT
XLS/ODS
XML
Web Service
LDAP
IMAP
CSV
- Extract
- Transform
- Load
Destinazione
DB
TXT
XLS/ODS
XML
Web Service
LDAP
IMAP
CSV

Esempio: Manipolazione dati



Esempio: Importazione con Lookup



Kettle

- Open Source
  - GNU GPL
  - Pentaho Data Integration
- Scritto in Java
- OLTP (PostgreSQL)
- Warehouse (PostgreSQL o Greenplum)

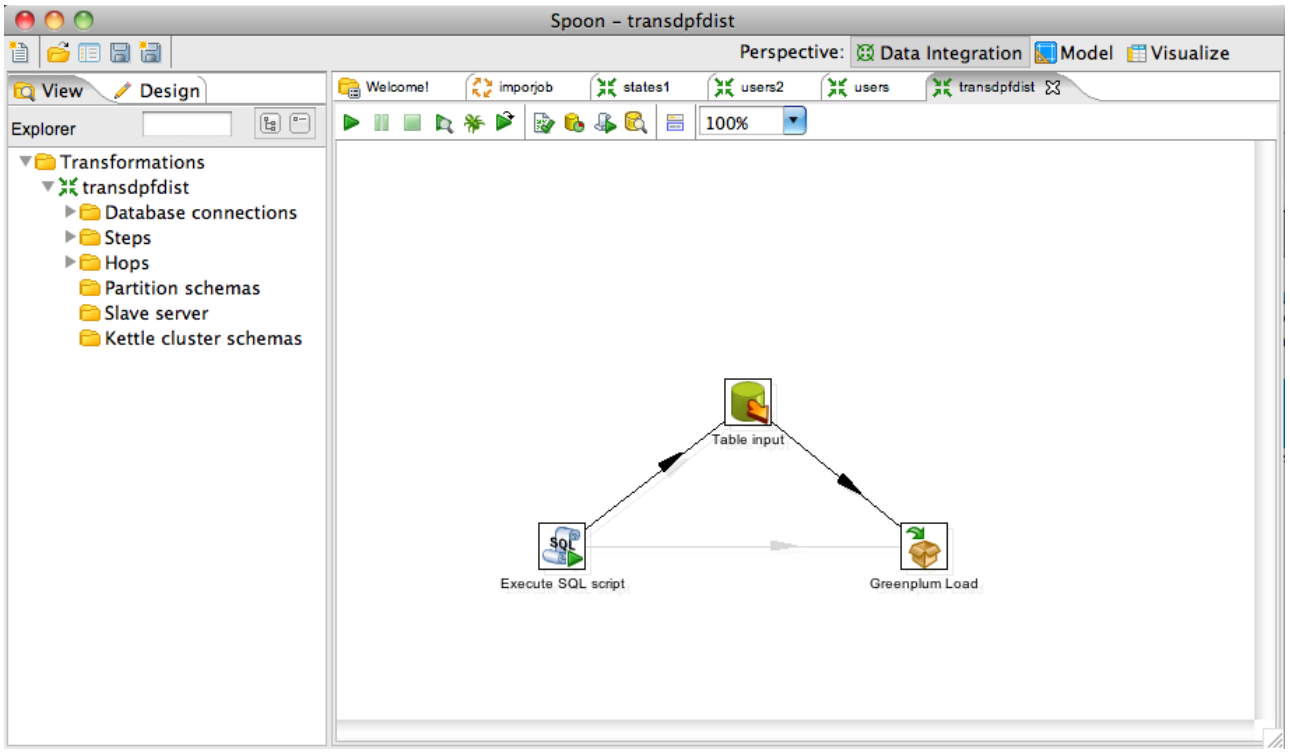
pentaho
POWERFUL ANALYTICS MADE EASY™

Composto da :

- SPOON
- PAN
- KITCHEN

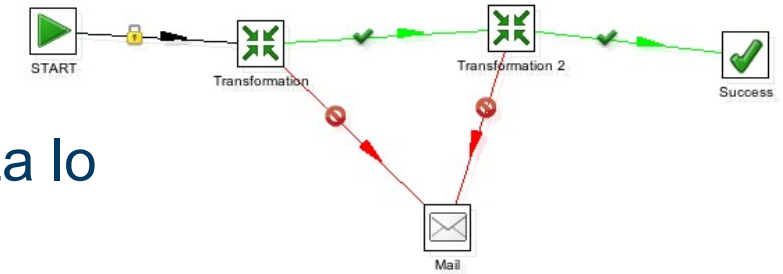
SPOON

- Interfaccia grafica
- Modellazione
- Flusso dati
- Niente codice generato
- Job
- Trasformazioni



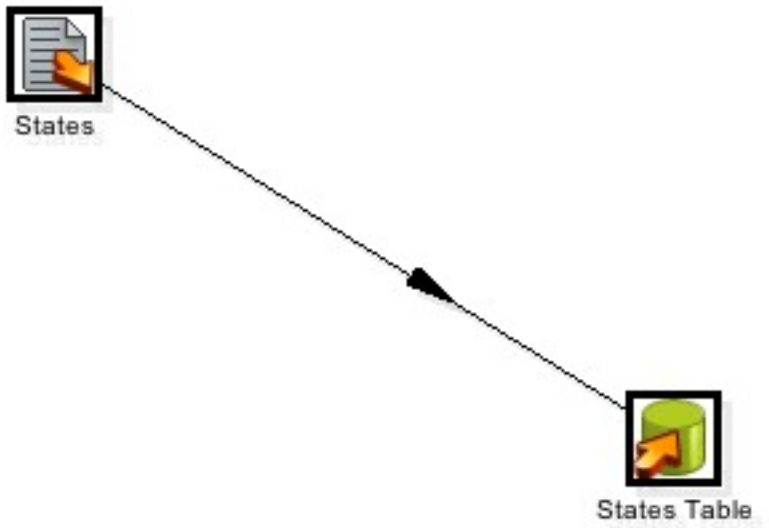
Job

- Serie di operazioni
- Insieme di componenti generici per l'esecuzione di un lavoro
- Esecuzione in ordine
- Il risultato dell'esecuzione influenza lo stato successivo
- Hop



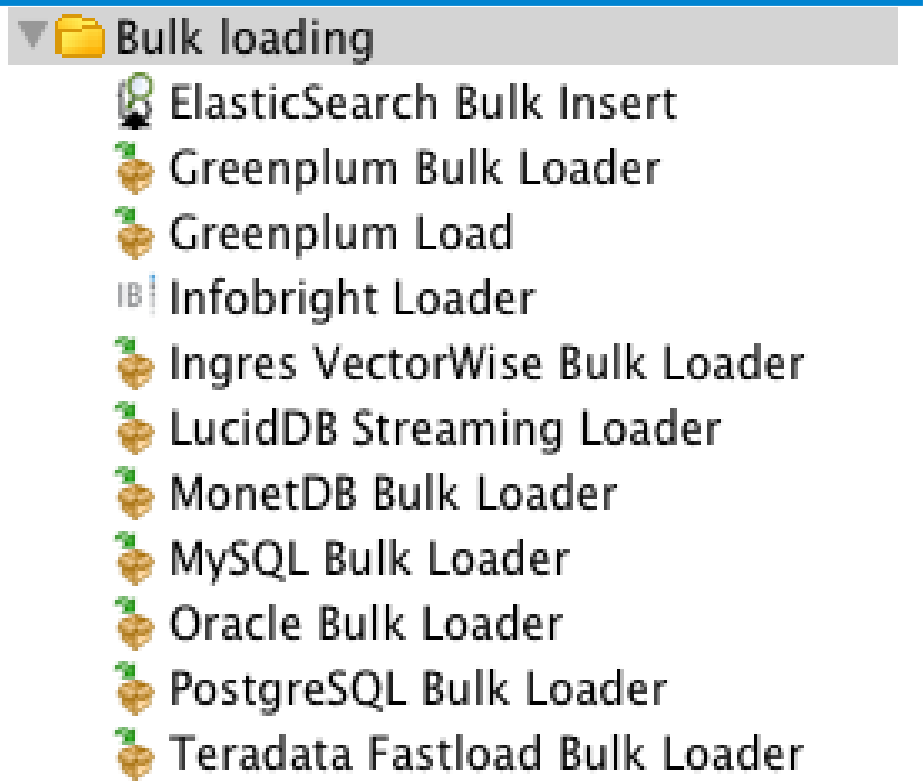
Trasformazione

- Sottoinsieme delle operazioni che compongono un job
- Possibilità di esecuzione in parallelo
- Riusabile
- Insieme di componenti specifici per l'esecuzione delle azioni



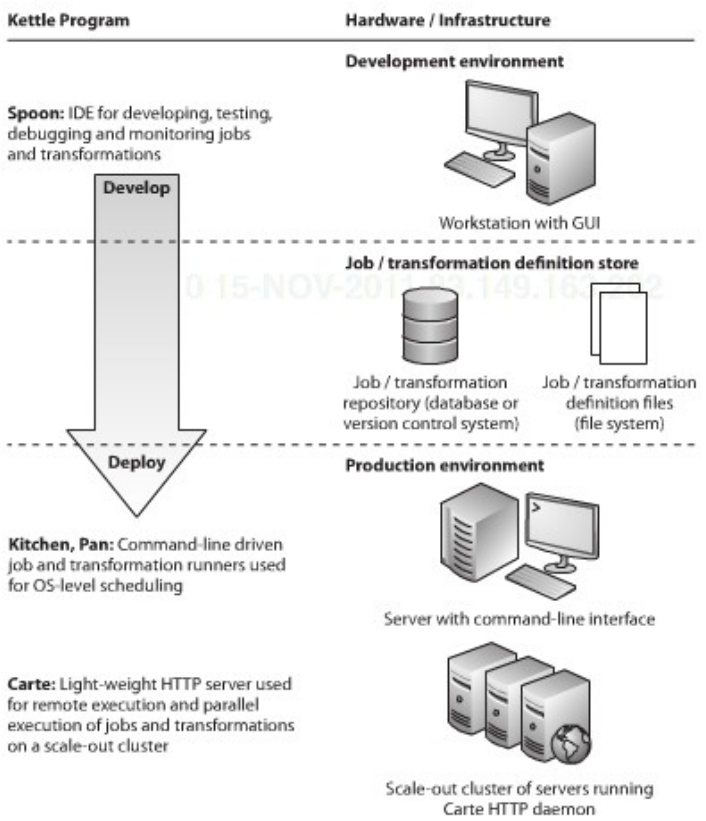
Componenti

- Organizzati in cartelle
- Rappresentano lo “step” di una trasformazione
- Eseguono operazioni specifiche
- Facilmente configurabili

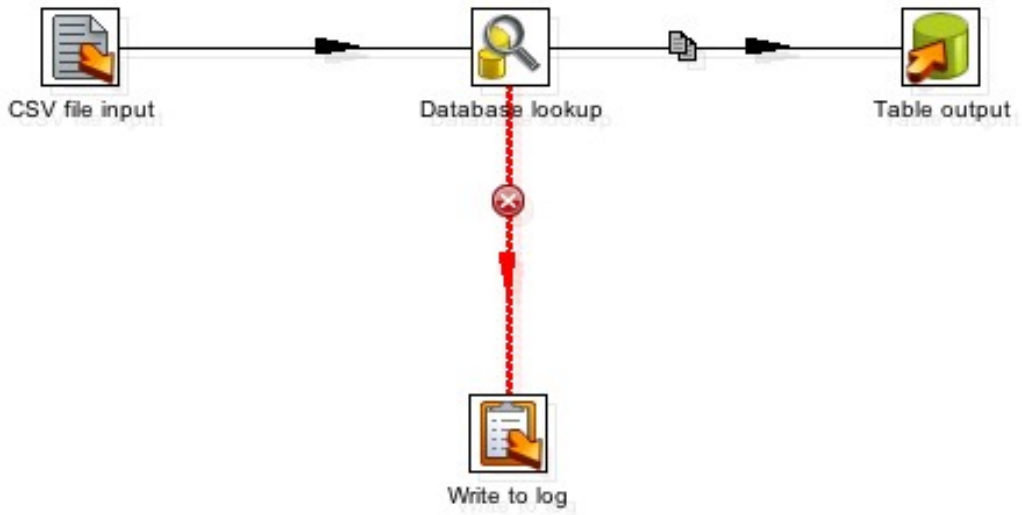


Automazione

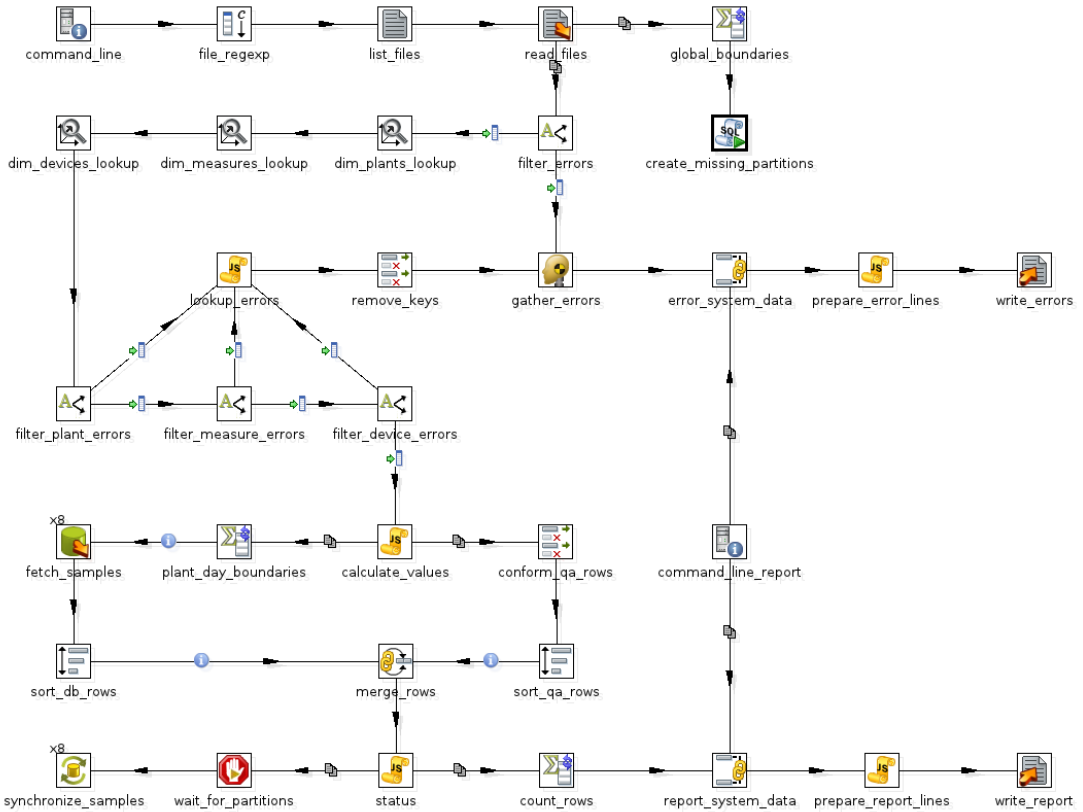
- Fase immediatamente successiva al design
- Eseguibile su server tramite componenti della suite dedicati: Kitchen/Pan
- Esecuzione dei job via cron



Esempi di casi reali: Lookup



Esempio di caso reale: QA Complesso



Conclusioni

- Figura professionale BI Expert
- Estensibile
- Semplificazione operazioni complesse
- Portabile
- Scalabile
- Cache interna

Domande?

- E-Mail: giulio.calacoci@2ndquadrant.it
- URL: www.2ndquadrant.it
- Blog : blog.2ndquadrant.com

Licenza Creative Commons

Attribuzione

Non commerciale

Condividi allo stesso modo

2.5 Italia

http://creativecommons.org/licenses/by-nc-sa/2.5/it/