

Apache BEAM

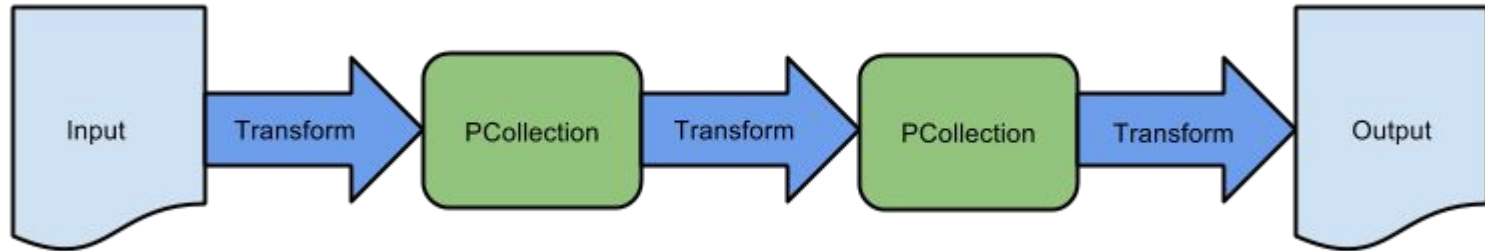


Vue global

- Open-source
- Utilisé pour définir des pipelines

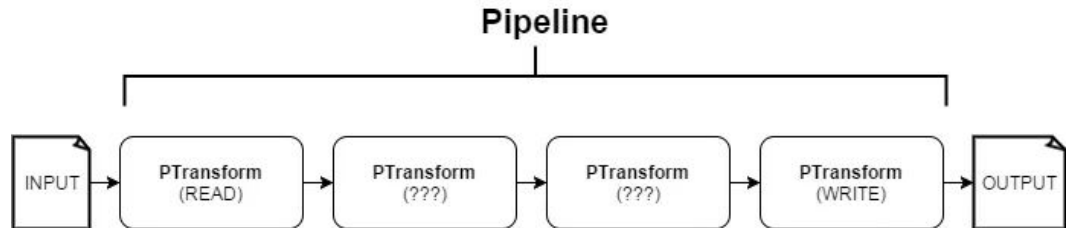


Configuration du pipeline

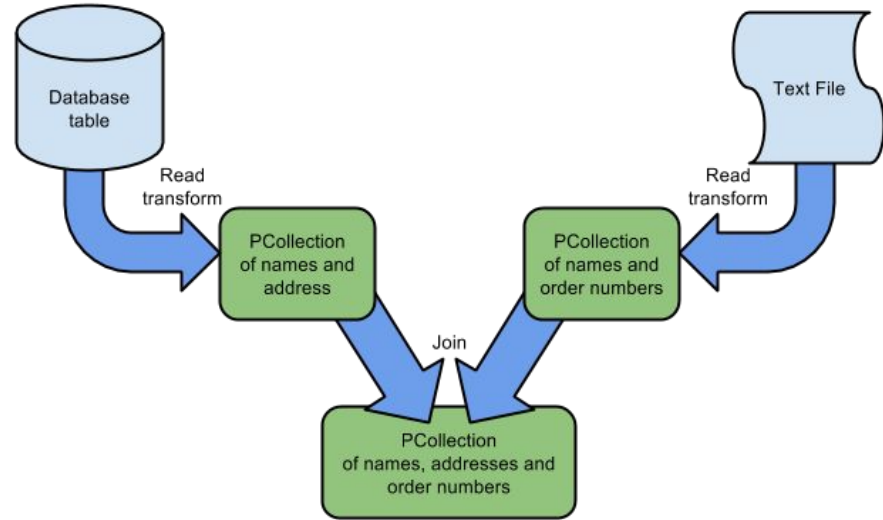
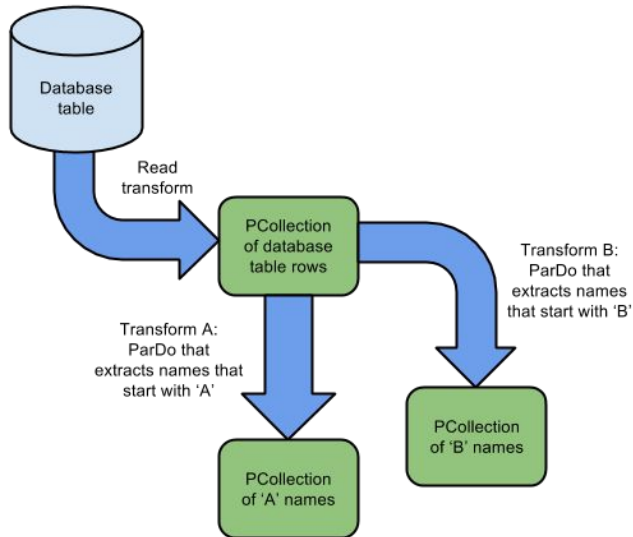


Configuration du pipeline

- Pipeline : Encapsule le processus
- PCollection : Les données à l'intérieur du pipeline
- PTransform : Une étape dans le pipeline



Variantes pipeline





SDKs

- Modèle commun à tous les langages
- Utilisation similaire pour toutes les données



Pipeline runners

- Processus Backend
- What / Where / When / How



Google
Cloud
Dataflow



Apache
Samza



Apache
Gearpump



Apache
Flink



Apache
Apex



Apache
Spark



Pourquoi Beam ?

- Différents runners utilisables
- Même modèle pour les données bornées et non
- Fonctionnalités avancées



DÉMONSTRATION



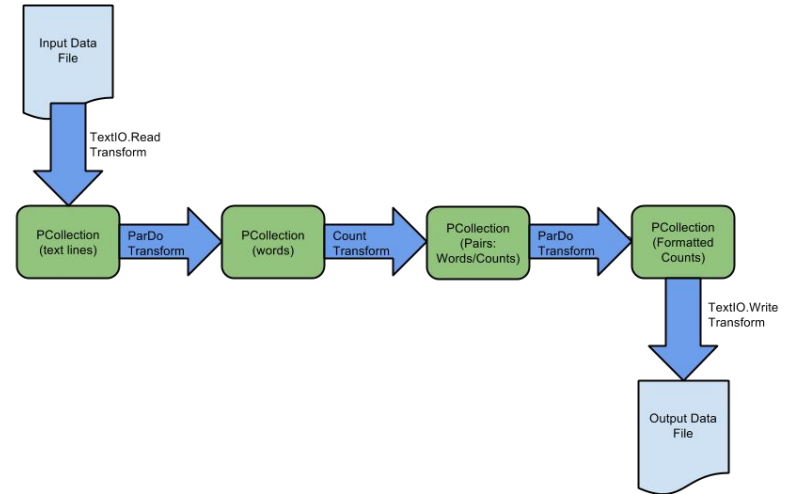
Installation (Java)

- Prérequis :
 - Java 7 (ou +)
 - maven

```
➤ mvn archetype:generate \  
  -DarchetypeGroupId=org.apache.beam \  
  -DarchetypeArtifactId=beam-sdks-java-maven-archetypes-examples \  
  -DarchetypeVersion=2.9.0 \  
  -DgroupId=org.example \  
  -DartifactId=word-count-beam \  
  -Dversion="0.1" \  
  -Dpackage=org.apache.beam.examples \  
  -DinteractiveMode=false
```

Exécution

```
~ mvn compile exec:java -Dexec.mainClass=org.apache.beam.examples.MinimalWordCount
```





**Merci de votre attention !
Place aux questions**



Sources

<https://beam.apache.org/documentation/programming-guide/>

<https://beam.apache.org/documentation/pipelines/design-your-pipeline/>

<https://meritis.fr/bigdata/apache-beam-projet-dunification-prometteur/>

<https://www.slideshare.net/JeanBaptisteOnofr/introduction-to-apache-beam>

<https://github.com/apache/beam>

<https://beam.apache.org/get-started/wordcount-example/#windowedwordcount-example>