

Etude de Cas

Projet Composants Hadoop

Vous êtes sollicités par le DSI d'une société de distribution de produits (le choix d'un autre domaine métier est possible) qui souhaite mettre en place une Data Platform Globale à l'ensemble des domaines métier de l'entreprise (Finance, RH, Marketing, distribution, autres...) . Il souhaite que cette plateforme soit opérationnelle à partir de Juin de 2020.

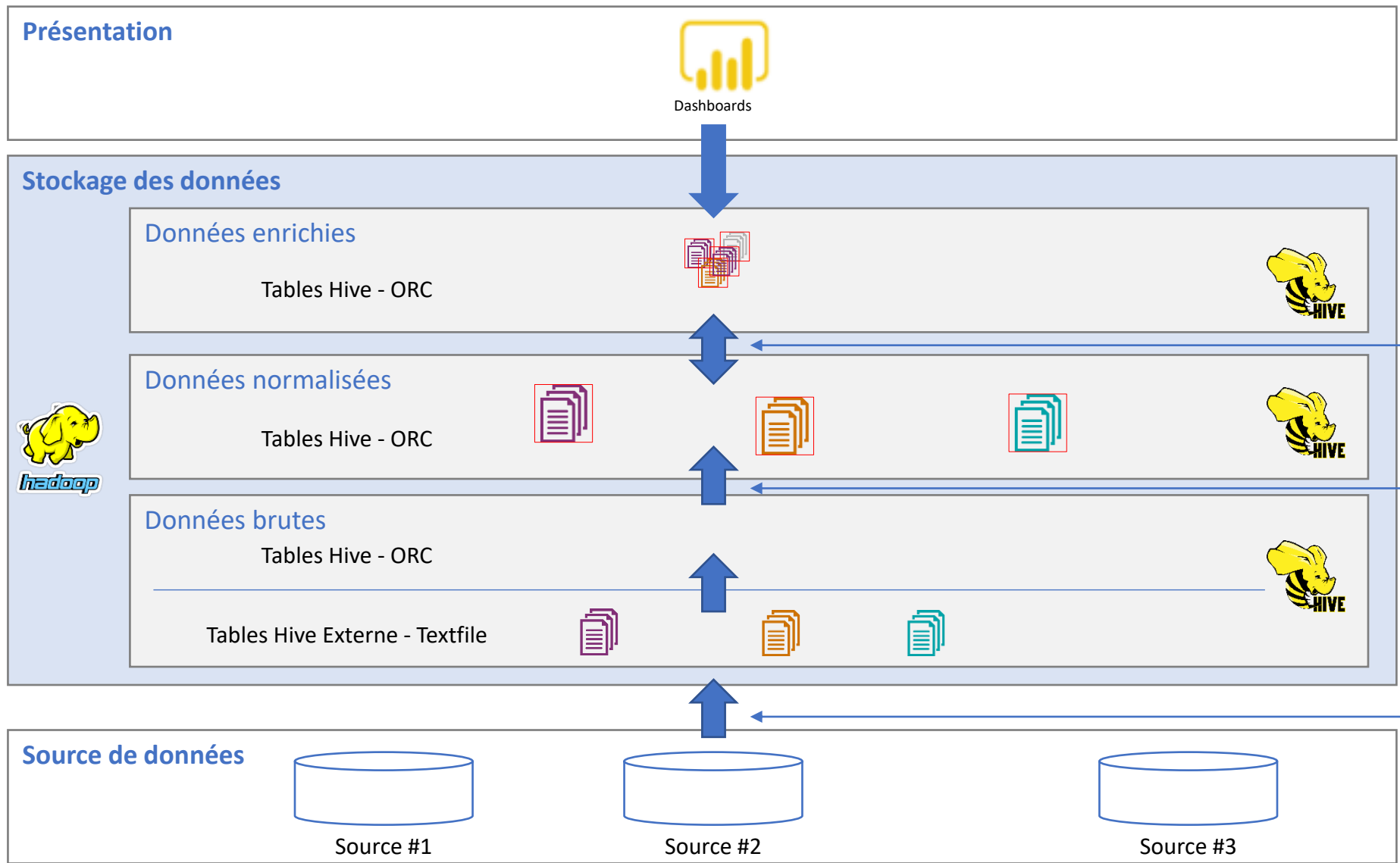
La data platform servira de support aux nouveaux Use Case d'avance Analytics (IA) et Analytics classique (BI).

Les attentes du DSI

Identifiez un cas d'usage métier que vous allez implémenter en réalisant les opérations suivantes:

1. Ingestion de 3 sources de données (Une source étant une base de données)
 1. Ingestion en mode batch avec la solution SQOOP
2. Normaliser l'ensemble des sources de données (alignement des formats de dates, devises et autres)
3. Faire un enrichissement des données en croisant les données de plusieurs sources de données dans un nouveau schéma
4. Réaliser un Dashboard connecté aux sources de données enrichies
5. Merci de respecter les différents étages du diagramme ci-dessous.

Organisation du stockage des données sur HDFS



L'utilisation de la
plateforme Cloudera
Quick start est
recommandée

Transformation & enrichissement

Logos for data processing and transformation tools:

- jupyter
- python
- APACHE spark
- ANACONDA
- Java

Ingestion

Logo for data ingestion:

- oozie