

04/02/2025 - 05/02/2025

Forum Beta Février 2025 Présentation DBT

Dossier
Facile

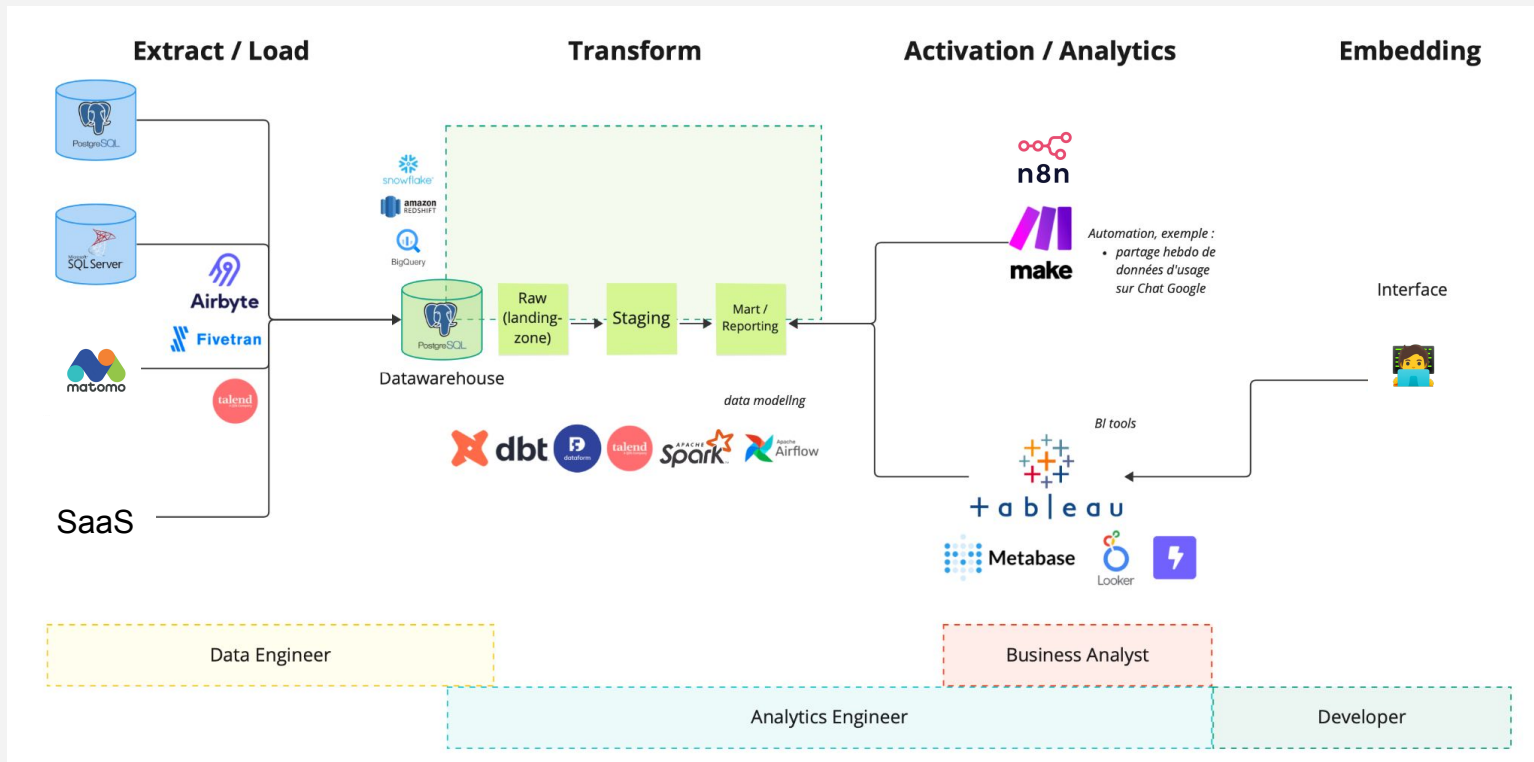
Programme

- Qu'est ce qu'un ETL ?
- Pourquoi utiliser DBT ?
- Concepts clés de DBT
- Exemple d'application avec DossierFacile

DossierFacile

Le dossier de location numérique de l'État

Qu'est-ce qu'un ETL / ELT ?



Pourquoi utiliser un ELT et DBT ?

Pros

ETL

- Rendre la donnée et le travail sur les indicateurs accessibles à des **personnes non techniques** (Activation / BI)
- **Supprimer les dépendances entre l'applicatif et la BI** (pas de risque de regressions, ressources matérielles différentes).
- **Accélérer les itérations** et de définition d'indicateurs (avec une CI / CD) - moins d'adhérences dev / data
- Stratégie technique permettant de se connecter facilement à d'**autres sources de données** (CRM, SaaS etc)
- Stratégie technique permettant dans le futur le développement de use cases autour de l'**IA** (Data Science)

DBT

- **Robustesse** d'un produit éprouvé par la communauté open- source
- Profiter des fonctionnalités de **supervision, d'orchestration et d'alerting** intégrés (DBTCloud)
- Pré- calculer nos indicateurs afin de réduire la latence
- Construire un modèle de donnée générique à travers les transformations
- Implémenter des tests unitaires sur nos transformations afin de s'assurer de l'absence de régressions

Cons

- Logique de calcul par batch (ex: fréquence 1h)
- Les indicateurs ne seront pas disponibles en temps réel
- Il s'agit donc principalement de présenter la donnée
- Réplication de la donnée, surcoût à prévoir côté architecture

Pourquoi utiliser DBT ?



Devs

- Suivi total des modifications via **Git**
- **≠ Environnement** dev, staging, prod (target)
- CI / CD et **Orchestration** (DBTCloud)
- **Test** unitaire / Test e2e / Lint (SQLFluff)
- **Alerting**

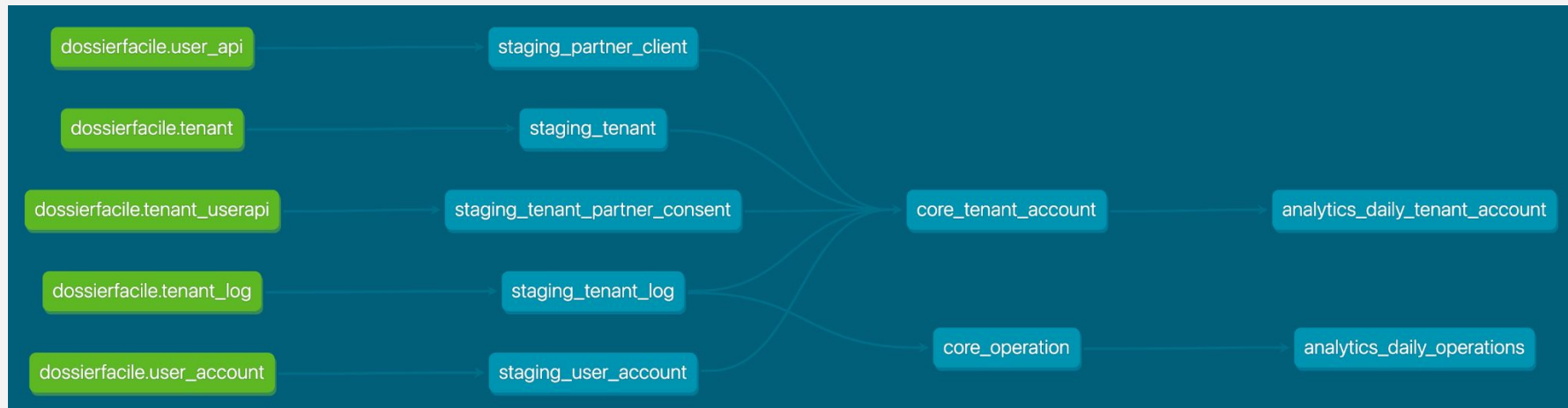
Concepts clés de DBT

- **Sources** : Connexion aux bases de données brutes
- **Models** : Requête SQL de transformation de la donnée

```
select
  CAST(id as INTEGER) as id
  , CAST(tenant_type as VARCHAR) as tenant_type
  , CAST(apartment_sharing_id as INTEGER) as rental_application_id
  , CAST(LEFT(zip_code, 5) as BIGINT) as zip_code
  , CAST(status as VARCHAR) as tenant_status
  , CAST(last_update_date as TIMESTAMP) as updated_at
  , CAST(operator_date_time as TIMESTAMP) as last_operation_at
  , CAST(operator_comment as VARCHAR) as operator_comment
  , CAST(clarification as VARCHAR) as tenant_comment
from {{ source('dossierfacile', 'tenant') }}
```

Concepts clés de DBT

- **Macros** : Fonctions SQL réutilisables
- **Seeds** : Données statiques utilisées dans les transformations
- **Tests** : Vérification automatique de l'intégrité des données
- **Lineage** : Graphe de dépendances entre modèles





Exemple d'application avec DossierFacile

Repo Git : <https://github.com/MTES-MCT/dossierfacile-data>

Metabase : <https://metabase.dossierfacile.fr/dashboard/159-pirate-metrics-v2>