Comment on a unifié l'orchestration data pour 80+ engineers

Ismail Mezzour

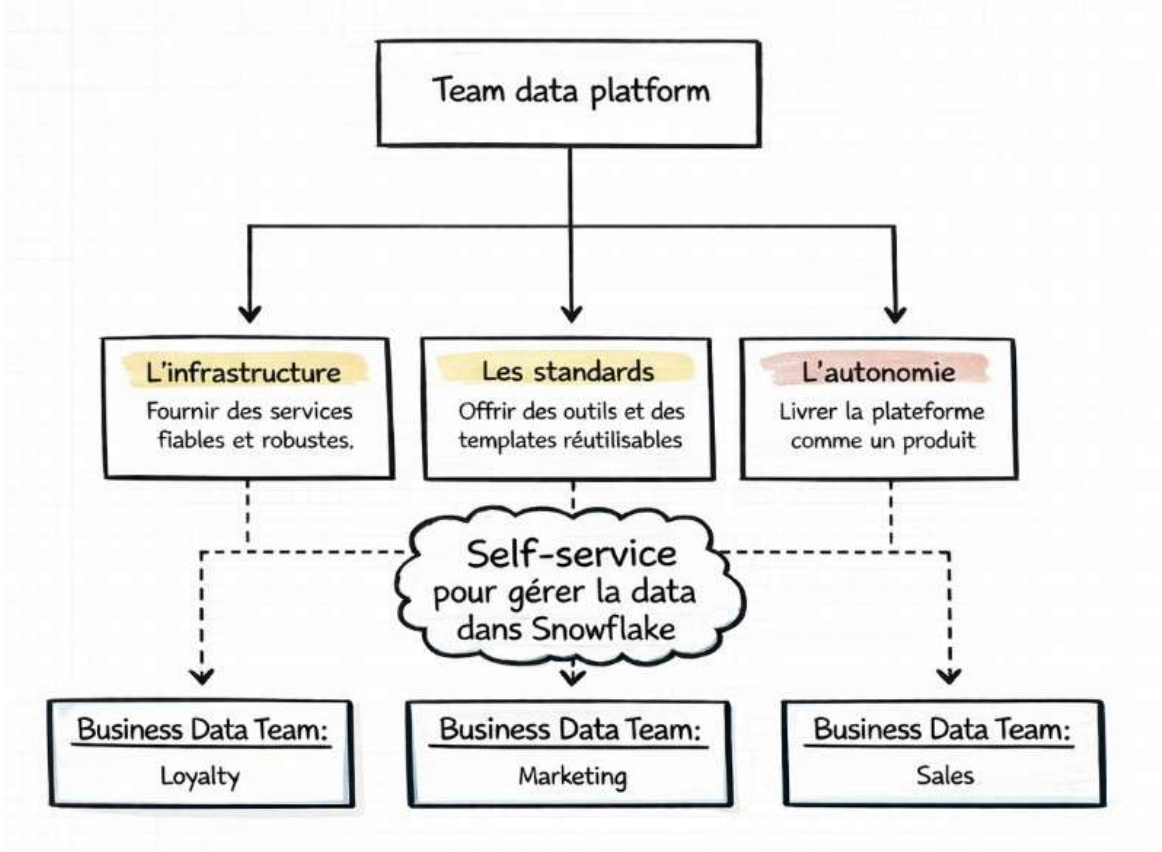
Les plateformes data ne deviennent pas complexes

Elles deviennent pratiquement ingérables.

Et à partir de là… ce n’est plus un système qu’on maîtrise. C’est un système qu’on subit.



La promesse – l’équipe data plateform



En réalité – la data platefome



Loyalty

Talend

Projet hérité, pas le temps de migrer.

Sales

Control-M

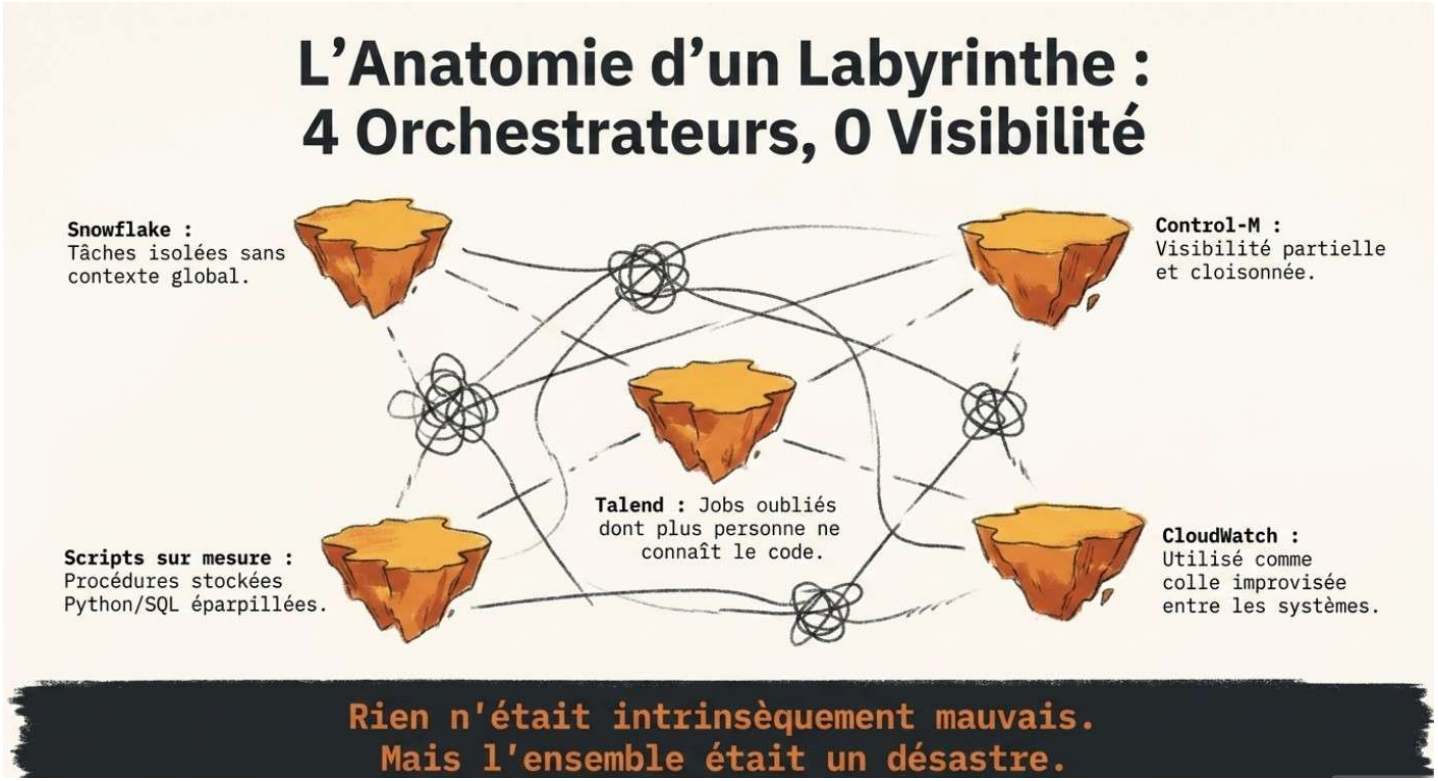
Partagé avec la tech qui l'utilise déjà.

Marketing

Snowflake Tasks

Tasks & procédures stockées en SQL.

La découverte du labyrinthe

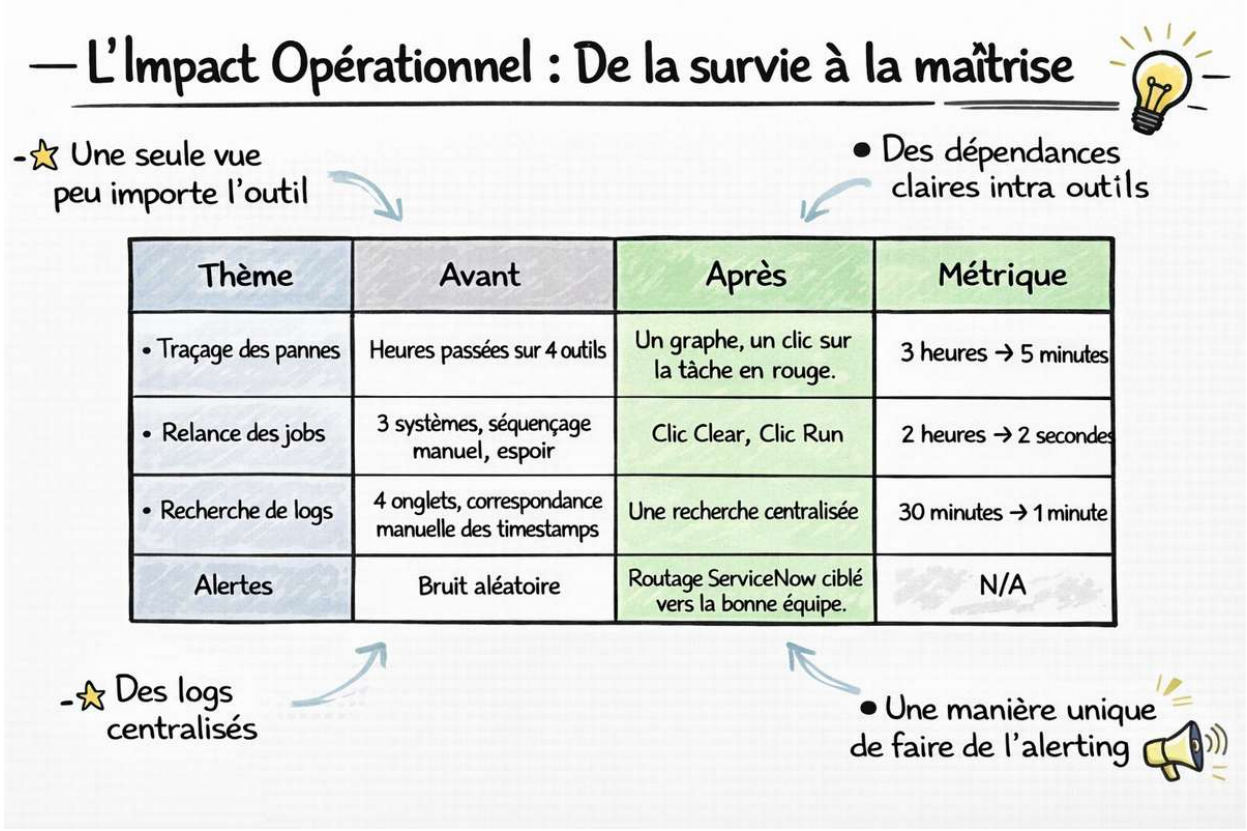


Et à ce moment-là… on commence à voir apparaître **deux types de problèmes**.

Le problème visible

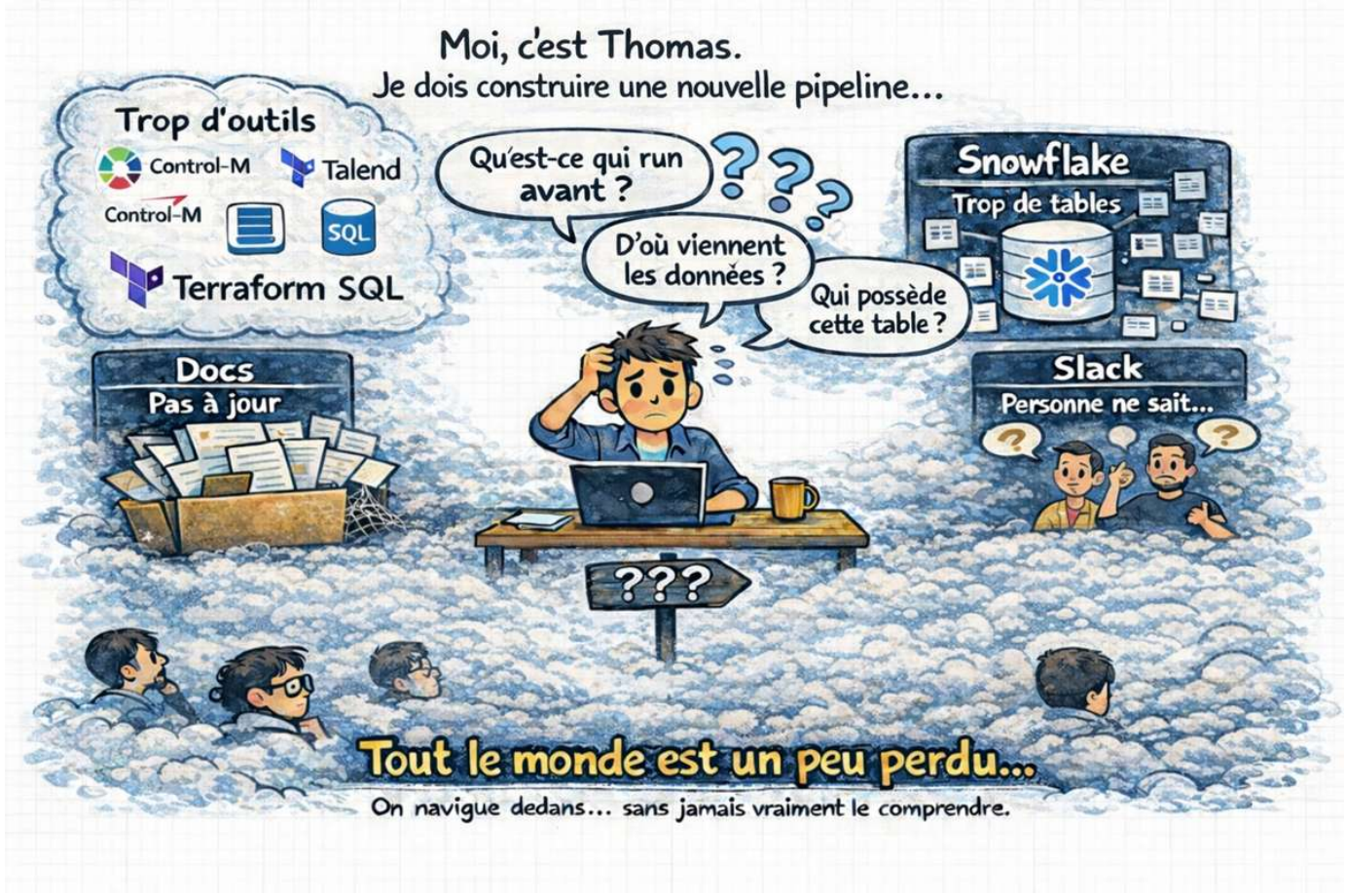


L'impact opérationnel

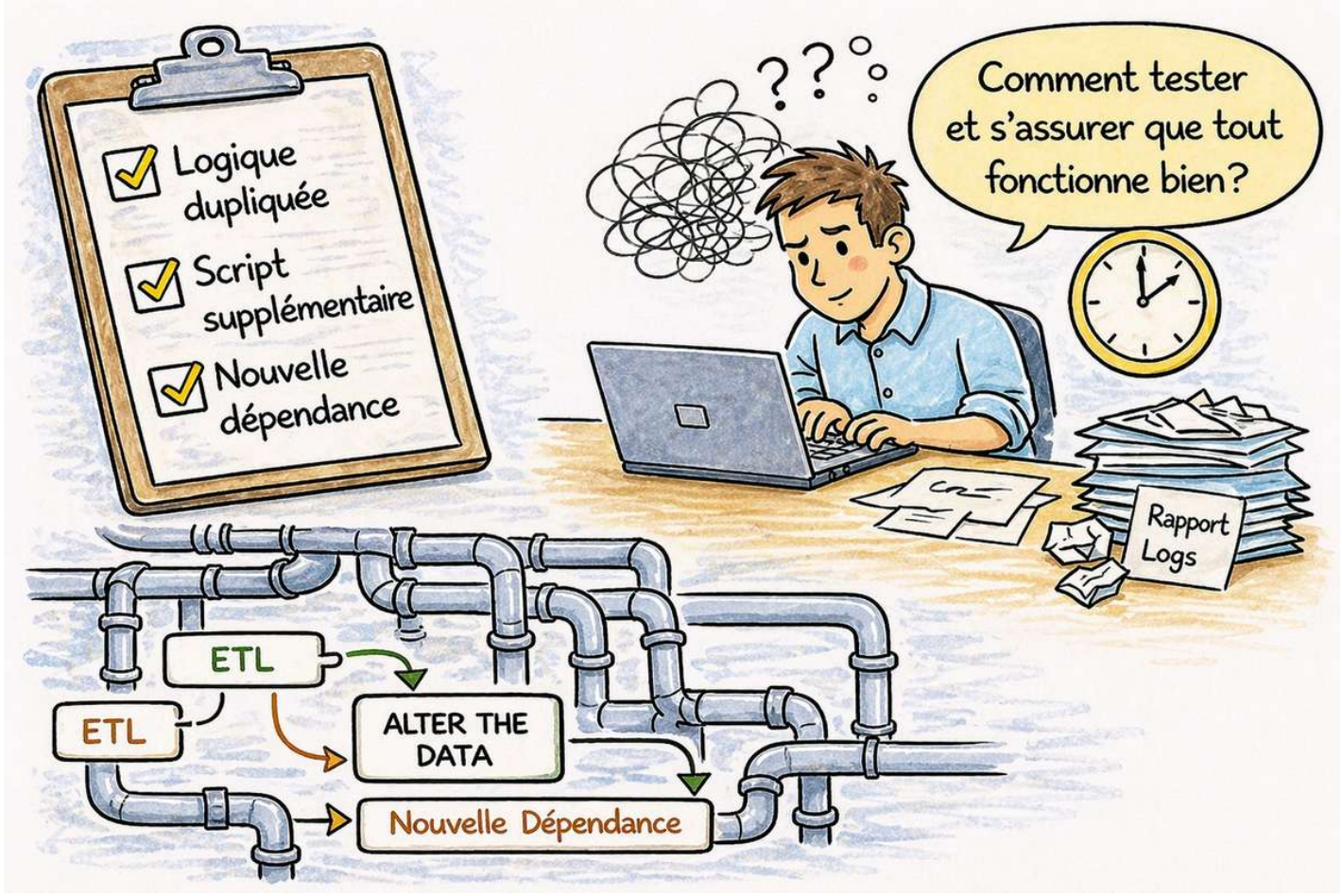


Thème	Avant	Après	Métrique
• Traçage des pannes	Heures passées sur 4 outils	Un graphe, un clic sur la tâche en rouge.	3 heures → 5 minutes
• Relance des jobs	3 systèmes, séquençage manuel, espoir	Clic Clear, Clic Run	2 heures → 2 secondes
• Recherche de logs	4 onglets, correspondance manuelle des timestamps	Une recherche centralisée	30 minutes → 1 minute
Alertes	Bruit aléatoire	Routage ServiceNow ciblé vers la bonne équipe.	N/A

Le problème invisible



Solution de Thomas



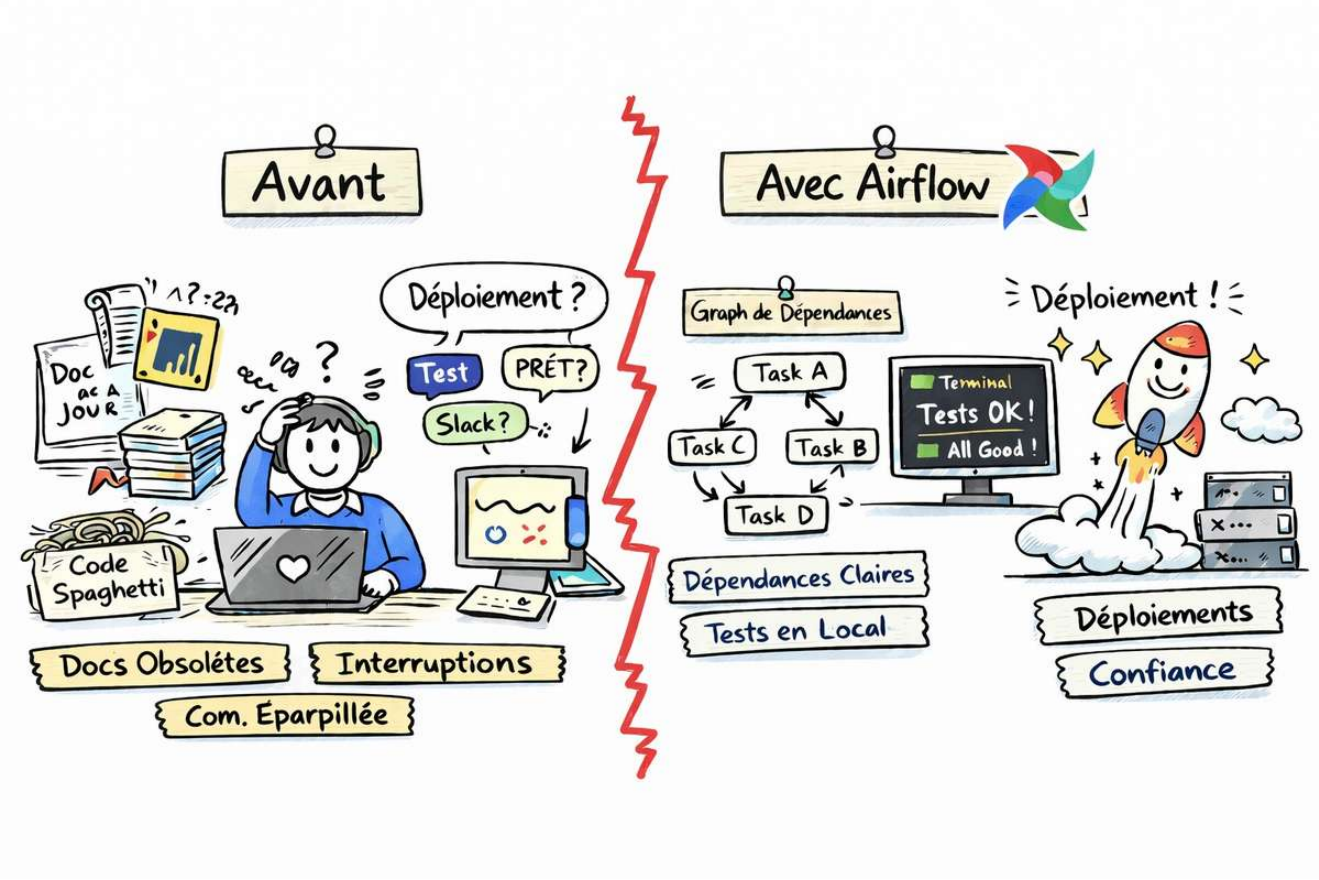
Et au bout d'un moment…

Le problème n'est plus quand ça casse.

Le problème… c'est tout le temps qu'on perd quand ça ne casse pas.



L'impact opérationnel d'Airflow



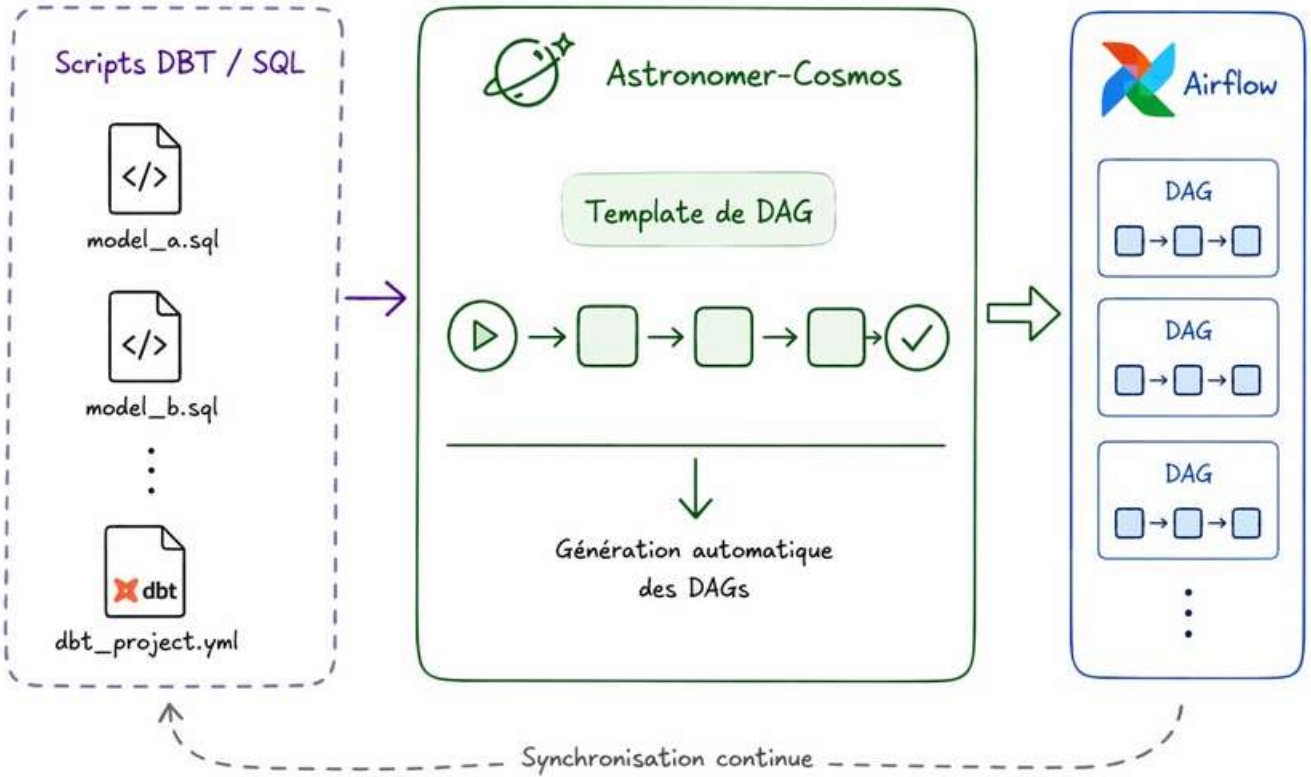
Mais est-ce qu'on s'est arrêtés là ?

Ajouter Airflow sans guidelines, sans templates, sans standards… c'est recréer le chaos. Juste ailleurs.

Qu'est-ce qui nous garantit qu'on ne va pas reproduire exactement les mêmes erreurs ?



Des templates de DAGs pour les scripts DBT/SQL avec Astronomer-Cosmos



1. Génération auto des DAGs

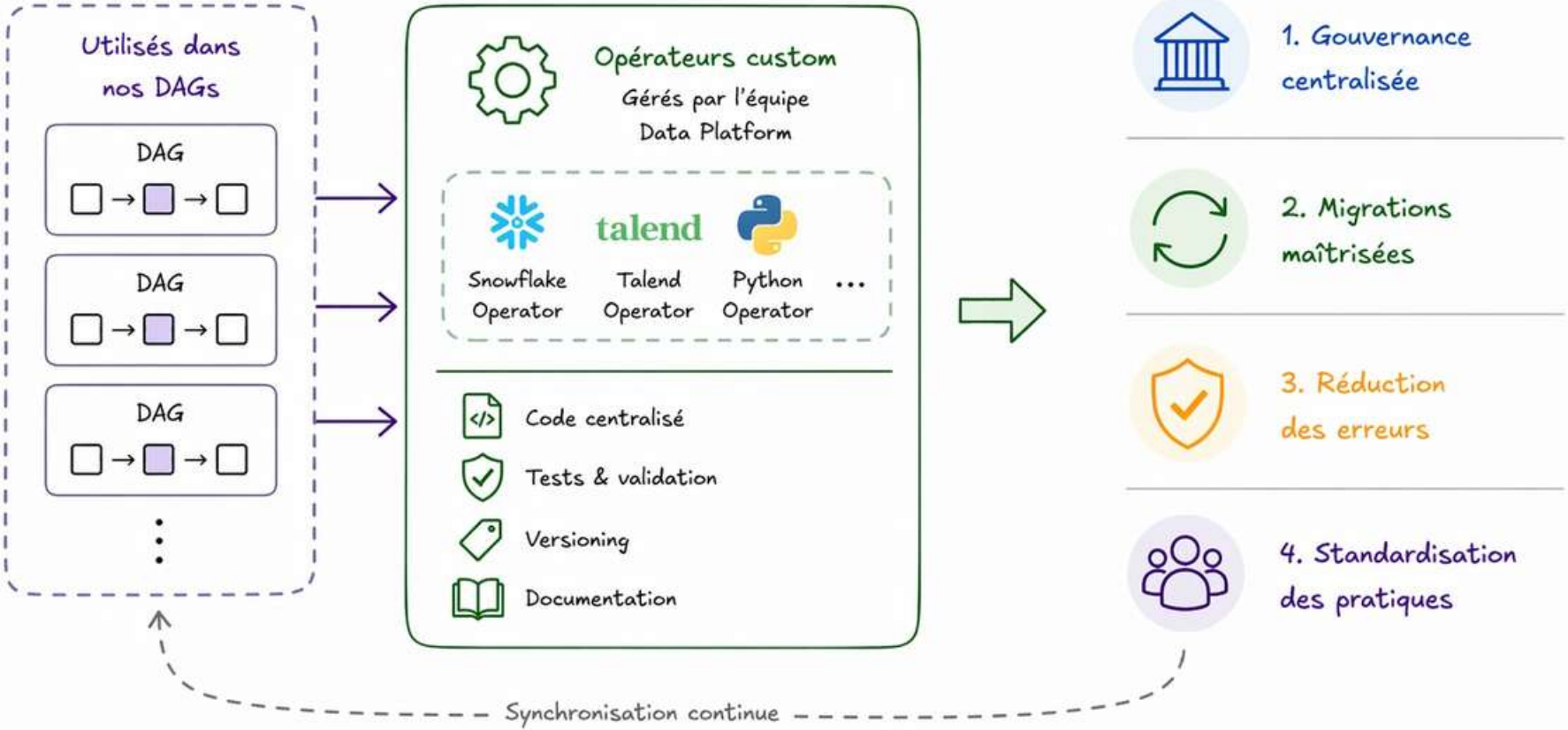
2. Abstraction de l'orchestration

3. Standardisation

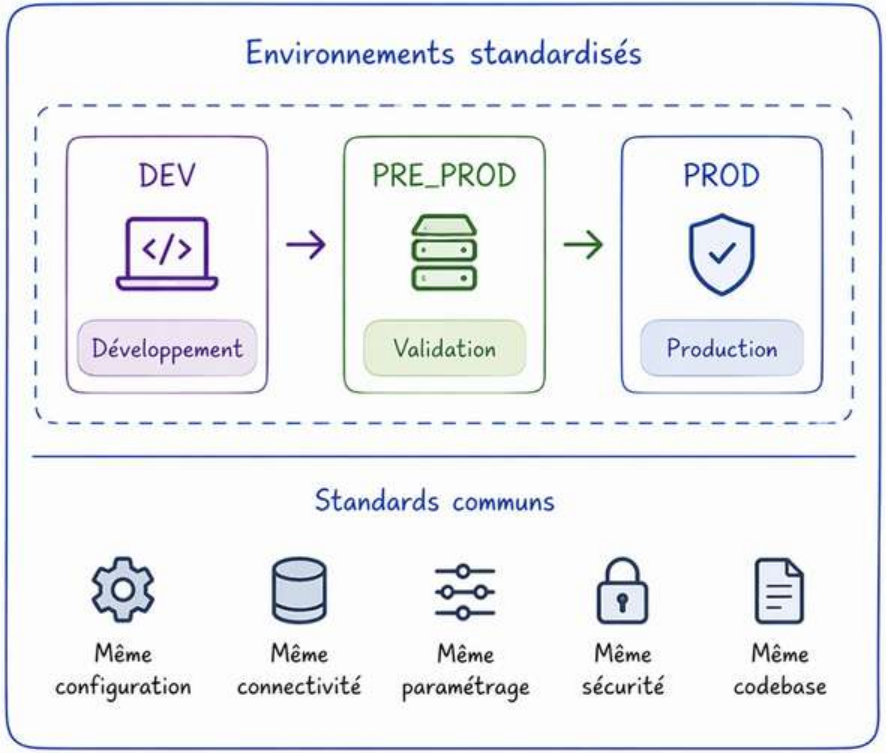
4. Facilité de migration

5. Moins de code custom

Mise en place d'opérateurs qu'on va gérer nous de notre côté



Avoir des environnements standardisés dev, pre_prod, pro



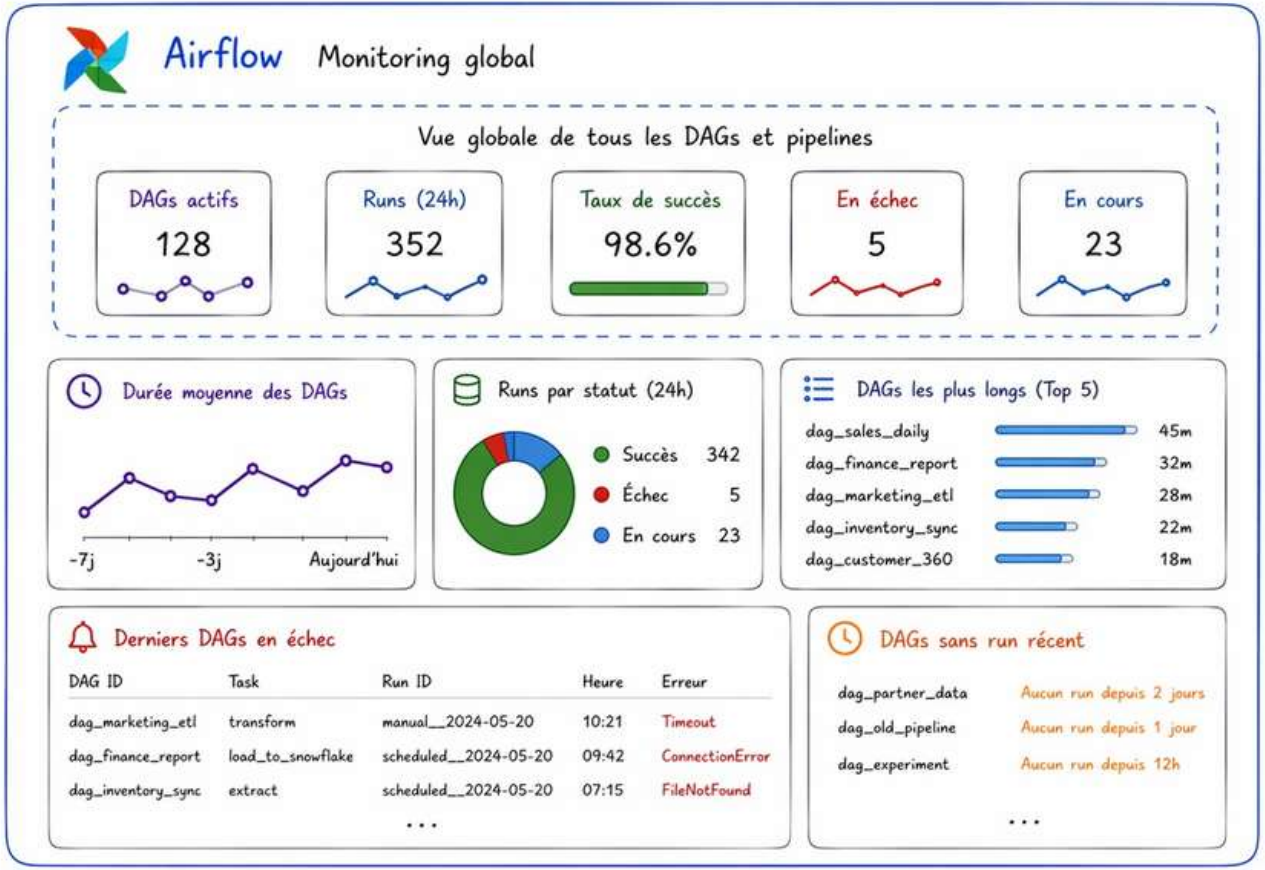
1. Comportement cohérent
→ même exécution

2. Réduction des bugs
→ moins d'écarts liés à la configuration

3. Déploiements fiables
→ ce qui est validé en preprod fonctionne en prod

4. Migrations sécurisées
→ changements testés dans un environnement identique

Observabilité avancée - Airflow



1. Monitoring global

Monitoring centralisé
→ vue globale de tous les DAGs et pipelines

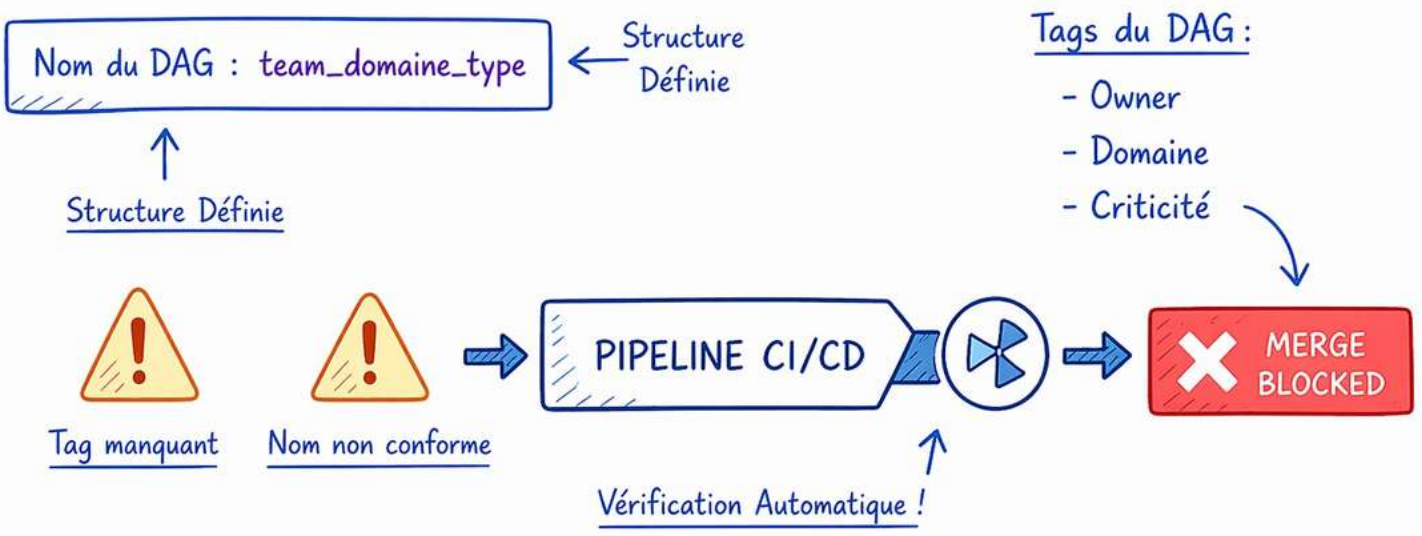
2. Métriques clés

→ durée, volumes, coûts (ex: Snowflake), taux de succès

3. Détection d'anomalies

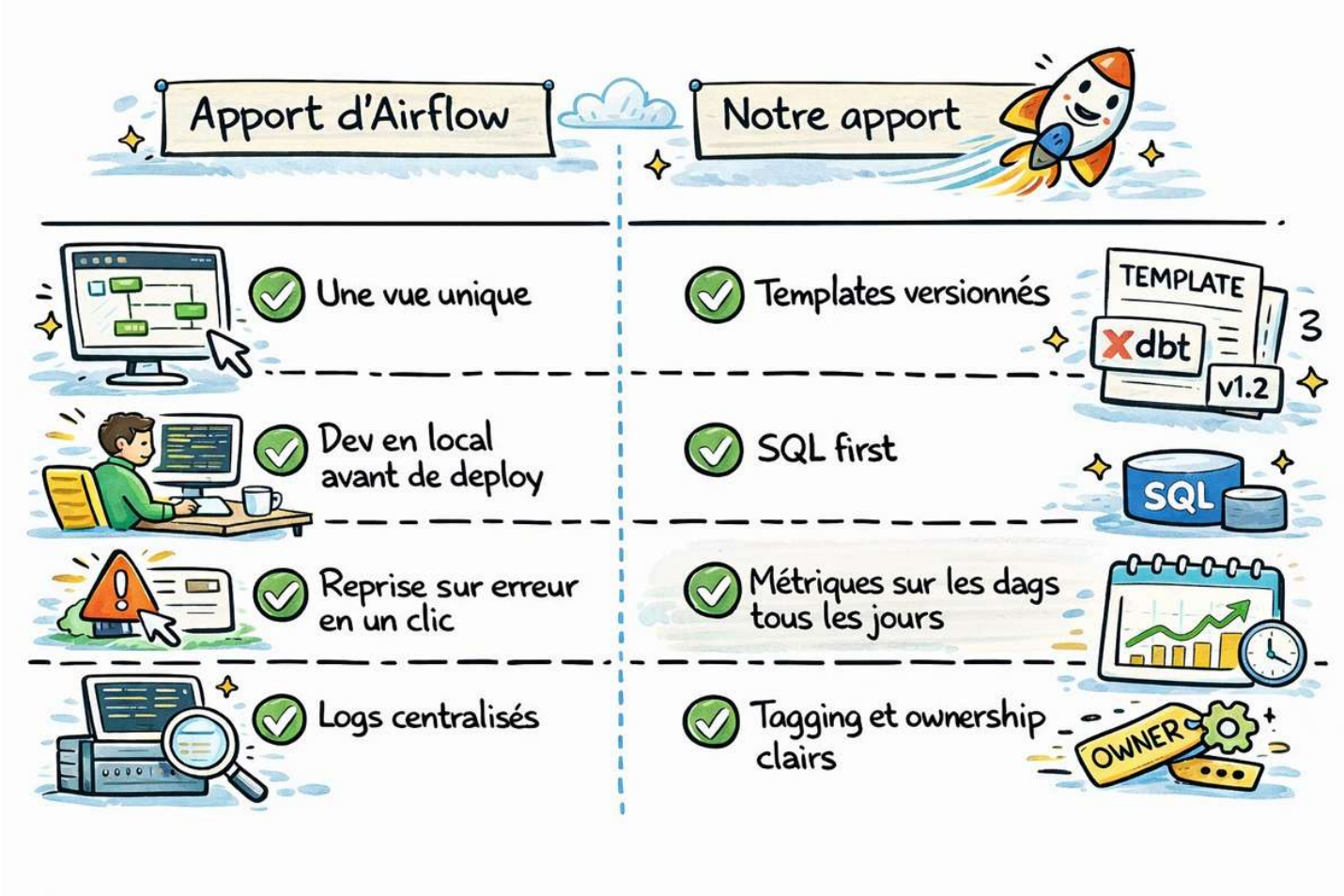
→ alertes sur comportements anormaux (pas juste failures)

Conventions de Nommage & Tagging des DAGs



Parce qu'on n'a pas juste changé d'orchestrateur.
On a construit ce qui empêche le chaos de revenir.

L'impact opérationnel



La Plateforme Aujourd'hui : La phase de mise à l'échelle

200+ DAGs

Production

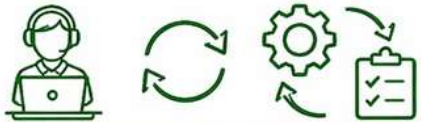
(+ croissance continue)

Expansion des cas d'usage



- Nouveaux besoins
- Intégrations
- Ingestion & Serveurs

Boucle d'amélioration



- 80+ ingénieurs
- Points de friction
- Standards (Cosmos)

Adoption fluide



- Onboarding structuré
- Standards imposés
- Dès le 1er jour

Un nouvel ingénieur pourrait-il comprendre l'ensemble de votre plateforme en une seule journée ?

