

1. Contexte & Processus ELT

- **Sources** : CSV d'offres d'emploi & de profils candidats.
- **Stage** : ingestion brute dans des stages Snowflake (format VARIANT/CSV).
- **Transformation** (via dbt) :
 1. **Bronze** (raw → cleaned)
 2. **Silver** (3NF pour matching OLTP)
 3. **Gold** (schéma étoile pour OLAP & reporting)

2. Schéma en 3NF (OLTP – matching)

(voir pdf « schéma 3NF »)

Points clés 3NF

- Tables normalisées, clefs étrangères surfaces.
- Pas de redondance : modifications isolées (ex. type de contrat).
- Optimisé pour les inserts/updates fréquents (matching en temps réel).

3. Schéma étoile (OLAP – reporting)

(voir pdf « schéma étoile »)

Points clés étoile

- **Fait unique** (Faits_Offres) → « mesures » (nb_compétences, salaire, etc.).
- **Dimensions** larges pour slices/dices rapides (date, lieu, compétence...).
- Optimisé pour **reporting** et **visualisations** (PowerBI, Tableau).

4. Choix technologique : Snowflake & dbt

Critère	Pourquoi Snowflake	Pourquoi dbt
Scalabilité	Séparation compute/storage, élasticité instantanée.	Modèle “modular SQL”, exécute chaque modèle indépendamment.
Performance	Micro-partitions, cache résidents.	Tests intégrés, snapshots, materializations (view/table).
Concurrence & coût	Concurrency near-infini, billing à la seconde.	Génération ciblée de tables vues : optimise l’ordre de chargement.
Sécurité & gouvernance	Time Travel, masking policies, RBAC.	Documentation auto (dbt docs), lineage clair.
Écosystème / productivité	Intégration native BI & Python, support semi-structuré.	CI/CD, YAML pour tests, macros Jinja, hooks, snapshots.

En synthèse :

- **Snowflake** fournit un entrepôt cloud-natif, hautement scalable et performant, idéal pour un processus ELT à gros volumes et un usage mixte OLTP/OLAP.
- **dbt** structure et industrialise nos transformations SQL, garantit qualité (tests) et traçabilité (documentation / lineage), facilitant le maintien et l’évolution de nos schémas bronze/silver/gold.