

Data Engineer - Modele de fiche de poste

Raison d'être

Concevoir, industrialiser et opérer des pipelines de données fiables, scalables et sécurisés au service des produits/uses cases (BI, Data Science, IA Générative), en garantissant coûts maîtrisés et qualité bout-en-bout.

Missions (focus)

- Build: modélisation (ELT/ETL), ingestion batch/stream, orchestration, data quality by design, optimisation performance/coûts (FinOps).
- Run: supervision des jobs, SLO/SLA, remédiation incidents, capacity planning, continuité de service.
- Gouvernance & Sécurité: application des politiques d'accès (RBAC/ABAC), conformité (logs, lineage, PII), contribution aux glossaires/dictionnaires.
- Acculturation: bonnes pratiques d'usage de la plateforme (patterns d'ingestion, schémas, conventions de nommage).

Activités clés & Livrables

- Développement de pipelines (code repo, CI/CD), plans d'orchestration, modèles de données (ex. lakehouse, warehouse), tests de données (unit/integration/expectations), data contracts.
- Playbooks d'exploitation (SOP), dashboards de monitoring (latence, volumétrie, coût), documentations (README, runbooks).

KPIs d'évaluation

Taux de succès des jobs, latence E2E, coût par pipeline/use case, couverture de tests de données, dette technique (lint/scan), MTTR/MTBF.

Compétences

- Tech: SQL avancé, Python/Scala, ELT (dbt), streaming (Kafka/Event Hubs), orchestration (Airflow/Azure Data Factory/Fabric pipelines), Databricks/Spark, stockage (Delta/Parquet), CI/CD (GitHub Actions/Azure DevOps), IaC (Terraform/Bicep), observabilité.
- Soft: culture produit & valeur, communication claire, sens du service, rigueur.

Stack de référence (indicative)

Azure (Fabric, Synapse, Databricks, Data Factory, Event Hubs), Snowflake, dbt, Great Expectations, MLflow, Terraform, GitHub Actions. (Stacks et parcours de formation cohérents avec notre catalogue Data.IA.)

Interactions (RACI)

- Responsible: Data Engineer pour pipelines & run.
- Accountable: Data Architect/Tech Lead sur standards ; PO Data sur priorisation.
- Consulted: Data Scientist/BI Dev, Data Manager/DQ Manager.
- Informed: Sécurité/RSSI, Product Owner.

Profil & expérience

3–8 ans, expérience cloud, industrialisation, optimisation coûts/perf.

Exemples de résultats

Réduction de 30% des coûts de traitement via pruning & autoscaling ;
amélioration x3 de la latence d'ingestion.

Anti-patterns

Absence de data tests, automatisation prématurée.

D-F