Programme Data Engineering

|  |  |  |  |
| --- | --- | --- | --- |
| Semaine | Technologie | Objectif | Ajouté / Modifié |
| 1-2 | Docker, Kubernetes, CI/CD, Bash | Conteneurisation, déploiement automatisé + Scripts Bash pour automatiser les tâches et gérer les conteneurs. | ✅ GitHub Actions, Terraform, Bash |
| 3-4 | Apache Airflow, Airbyte, dbt | Orchestration et ETL/ELT moderne | ✅ Pipeline complet Airbyte → Airflow → dbt |
| 7-8 | Pandas + SQL, Data Quality, BigQuery | Manipulation et validation des données | ✅ Great Expectations, GCP Dataflow |
| 9-10 | Kafka, Pub/Sub | Streaming en temps réel | ✅ Pub/Sub pour cloud-native |
| 11-12 | Spark & PySpark | Traitement Big Data sur Cloud | ✅ Spark sur Dataproc |
| 13-14 | Spark avancé, Cloud Storage | Optimisation & stockage Cloud | ✅ Remplacement Hadoop par S3/GCS |
| 15-16 | Approfondissement dbt/Spark & Tests | Transformation avancée ou Big Data | ✅ Choix flexible |
| 17-18 | Projet final | Déploiement complet | ✅ GCP, CI/CD, Terraform |