

Data Engineer Roadmap 2024

#	Methods	Tools
1	Operation System	Linux
2	Programming Language	Python-Java
3	Web Framework	Python->FastAPI Java->Spring
4	Version Control	Git
5	Version Control System Hosting	GitLab
6	Advanced SQL Fundamental	CTEs PartionOver Windowing Materialized View
7	Databases	PostgreSQL MongoDB Redis
8	File Formats and Serialization	Parquet Avro
9	Block Storage	Ceph
10	Object Storage	MinIO
11	Query Engine	SparkSQL Trino
12	Pipeline Orchestration	Apache Airflow
13	Data Processing (Stream)	Apache Kafka
14	Data Processing (Stream & Batch)	Apache Spark
15	Data Visualization	PowerBI Metabase
16	Containerization	Docker
17	Container Orchestration	Kubernetes
18	CI/CD	Gitlab CI Jenkins
19	Infrastructure as Code	Ansible
20	Observability (Logging)	Sentry, EFK
21	Observability (Monitoring)	AlertManager Grafana
22	Observability (Tracing)	Joeger