6-Month Learning Roadmap: Cloud + Data Engineering Path

Goal:

Become a Cloud/Data Engineer with a pathway into AI, Cloud Architecture, or Technical Leadership.

Month 1: Foundation Setup

Core Skills:

- Learn Python:

- Python for Everybody Coursera (Free Audit): https://www.coursera.org/specializations/python

- Linux & Bash:

- LinuxCommand.org: http://linuxcommand.org/

- Learn Shell: https://www.learnshell.org/

- Git & GitHub:

- Introduction to Git and GitHub Coursera (Free Audit): https://www.coursera.org/learn/introduction-git-github

Tools to Install:

- Python 3.x

- VS Code or PyCharm

- Git & GitHub account

Month 2: Cloud Computing Essentials

Choose a Cloud Platform: AWS (preferred), or Azure

Learn Cloud Basics:

- AWS Cloud Practitioner Essentials AWS SkillBuilder (Free): https://explore.skillbuilder.aws/learn/course/external/view/elearning/134/aws-cloud-practitioner-esse

ntials

Start Core AWS Services:

- EC2, S3, IAM, Lambda, RDS, VPC (use AWS Free Tier)

Certification Goal:

- AWS Certified Solutions Architect Associate (SAA-C03): https://aws.amazon.com/certification/certified-solutions-architect-associate/

- FreeCodeCamp AWS SAA Course on YouTube: https://www.youtube.com/watch?v=Ia-UEYYR44s

Month 3: Data Engineering Skills

Core Topics:

- SQL:

- LeetCode SQL: https://leetcode.com/problemset/database/

- DataLemur SQL Interview Practice: https://datalemur.com/

- ETL Pipelines:

- Data Engineering Zoomcamp DataTalksClub: https://github.com/DataTalksClub/data-engineering-zoomcamp

- Tools:

- Apache Airflow: https://airflow.apache.org/

- Apache Spark: https://spark.apache.org/

Project:

- Create a data pipeline (CSV -> Clean -> Load into S3 or RDS using Python + Airflow)

Month 4: Infrastructure & DevOps

Learn DevOps Basics:

- Docker Getting Started Guide: https://docs.docker.com/get-started/

- Kubernetes Basics FreeCodeCamp: https://www.youtube.com/watch?v=KzvWxjZ0LR4

- Terraform for Beginners FreeCodeCamp: https://www.youtube.com/watch?v=SLB_c_ayRMo

- CI/CD: GitHub Actions Docs: https://docs.github.com/en/actions

Project:

- Containerize a data pipeline and deploy on AWS (EC2, S3, Airflow)

Month 5: Real-World Projects + Certification

Capstone Project Ideas:

- Weather/Stock Data Pipeline (scheduled via Airflow)

- Dashboard with Streamlit: https://streamlit.io/ or Grafana: https://grafana.com/

Certifications:

- AWS Solutions Architect Associate exam

- Optional: Databricks Data Engineer Associate: https://www.databricks.com/learn/certification/data-engineer-associate

Month 6: Expansion (Choose Your Path)

If leaning ML/AI:

- Learn ML with Scikit-learn Documentation: https://scikit-learn.org/stable/user_guide.html

- Deploy a simple ML model using Flask & AWS Lambda: https://www.youtube.com/watch?v=UbCWoMf80PY

If leaning Cloud Architecture:

- Learn multi-region design, monitoring (CloudWatch, Prometheus)

- Study AWS Well-Architected Framework: https://aws.amazon.com/architecture/well-architected/

Bonus Tips

- Maintain a GitHub portfolio with READMEs and screenshots
- Post learning updates weekly on LinkedIn
- Join communities: r/dataengineering: https://www.reddit.com/r/dataengineering/, AWS Discord: https://discord.gg/aws

Resources Summary

- Coursera: https://www.coursera.org
- AWS SkillBuilder: https://skillbuilder.aws
- DataTalksClub Zoomcamp: https://github.com/DataTalksClub/data-engineering-zoomcamp
- FreeCodeCamp: https://www.youtube.com/c/Freecodecamp
- GitHub: https://github.com
- Apache Airflow: https://airflow.apache.org
- Apache Spark: https://spark.apache.org
- Streamlit: https://streamlit.io
- Grafana: https://grafana.com