Data Science

1) problem / Business understanding.

2) Data collection and understanding.

3) Data preparation

4) EDA

5) Feature Engineering

6) feature selection / feature reduction

7) feature scaling. (ML) standerscaler
Z score

8) model selection and training

9) validation of the model

10) Model tune (Hyperparameter tuning)

11) Method (Gridsearchcv, Randomisedsearchcv)

12) validation of model

13) model save (Pickle / Joblib)
preprocessed data
(scaler.)

14) containerization (Docker)

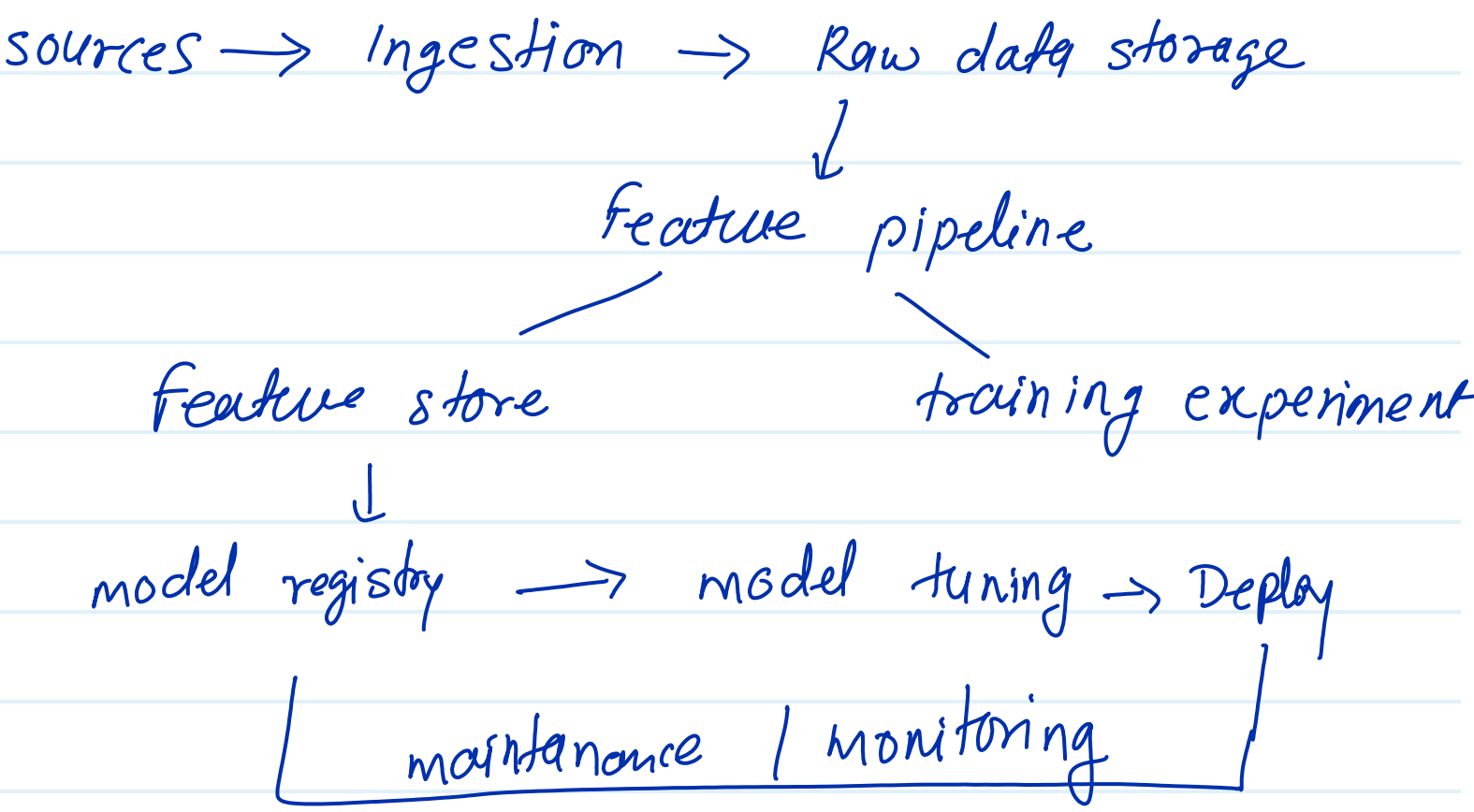
15) Deploy (AWS, Azure, GCP, Heroku)

HLD (High level Design)

1) objective of project

2) major logical component

3) High level Design of project



LLD (Low level Design)

1. Assumptions & versions
2. Data contract & schema
3. ingestion and pipeline specification
4. model training & experiment
5. model Registry & Deployment
6. monitoring, alerts
7. Retraining and automation
8. Governance, auditing
9. Testing strategy

Team

DS1 - problem understanding

DS2 - Data engg & prepcaration lead.

DS3 - EDA & feature engg. & selection

DS4 - model training & evaluation

DS5 - Deployment / monitoring

DS3 - EDA

10:00 login → meeting - Discuss about roles and responsibility on project

30 minutes

setup - System setup, project flow and role, logic designing preparing flow of work.

HLD & LLD Documents - Read

6 hours

2 hours

Day-2 team meeting.

Still data in ingestion process.

according to doc. I am preparing code. - 4-6 hours

meeting →

Day-3

client — company

low budget

high budget

raw analysis

week