1. DSA ⇒ Basic to intermediate

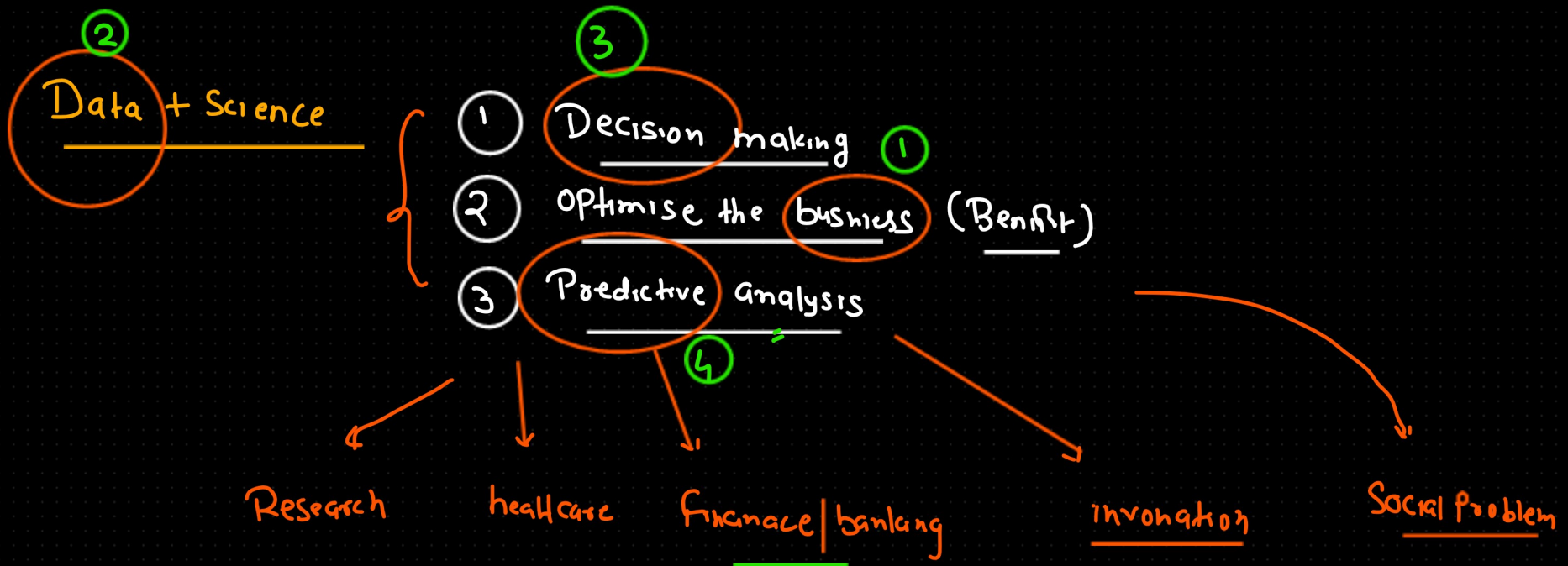
= Python

2. Python
3. Stats
4. ML ①②③
5. ML Related Project
6. Project Related Question
7. MLOPS
8. HLD, LLD, DPR

=

8. DL Interview
9. CV Interview ques
10. NLP interview ques
11. DL Based Proj.
12. Resume

1. Impact of DS in today's world
2. AI vs DS vs ML vs DL
3. Generic Architecture of DS Project-



AI, DS, ML, DL

AI ⇒ (artifical intelligence) ⇒ Product-

VA ⇒ Alexa, google assistant, Siri

RS ⇒ NetFlix, Spotify, Amazon, OTT platform

Chatbot ⇒ ChatGPT, Bard ai

autonomous car ⇒ self Drive car, tesla

AI ⇒ Artifical intelligence

↳ creativity =

↳ imagination =

↳ love | hate (emotion)

1950-60

= Symbolic AI

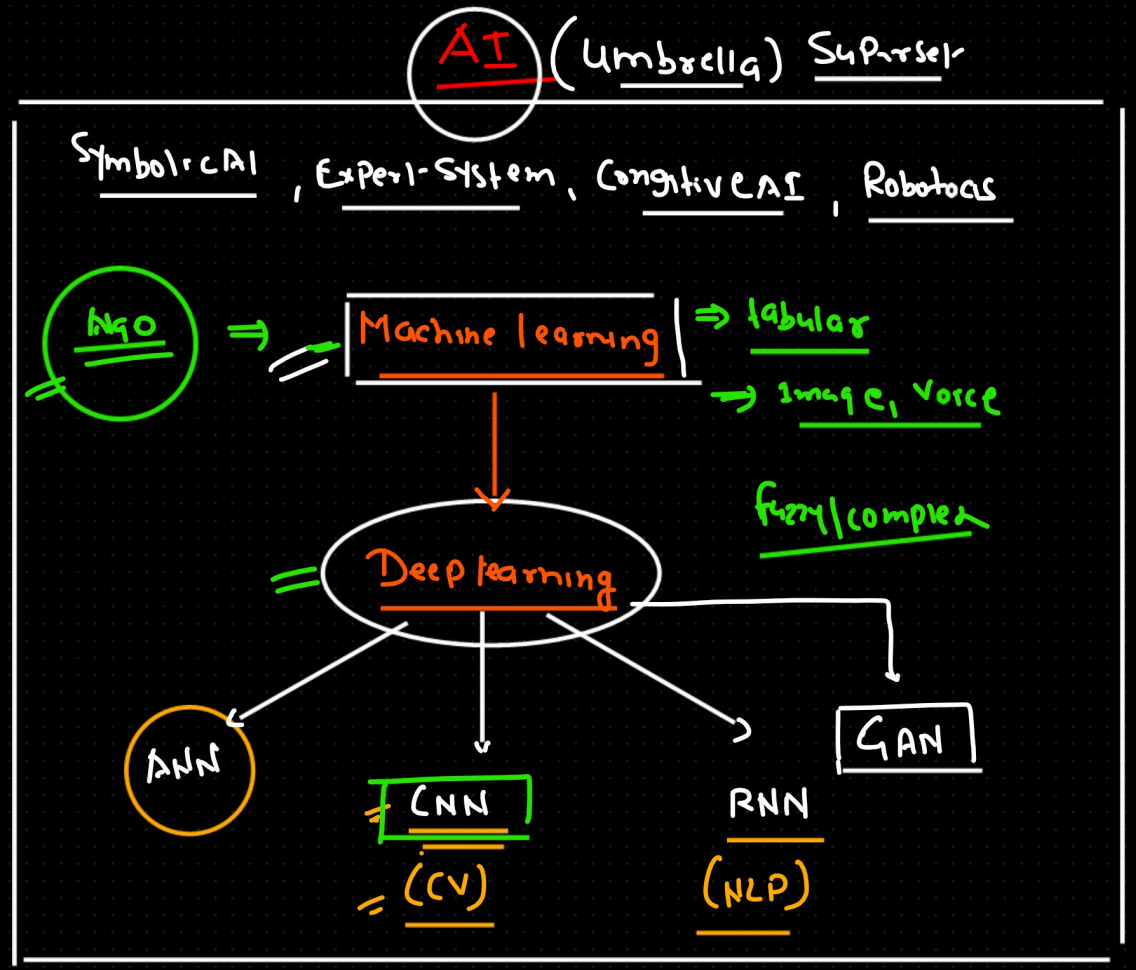
= If-else ⇒ Rule Based AI

= Expert System ⇒ Assistant

= Knowledge base

Machine learning

= Pattern recognition

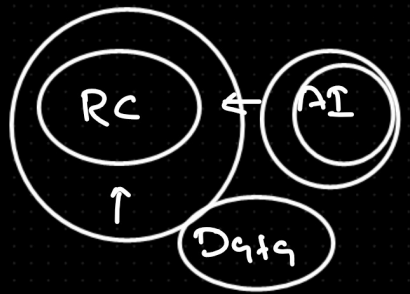


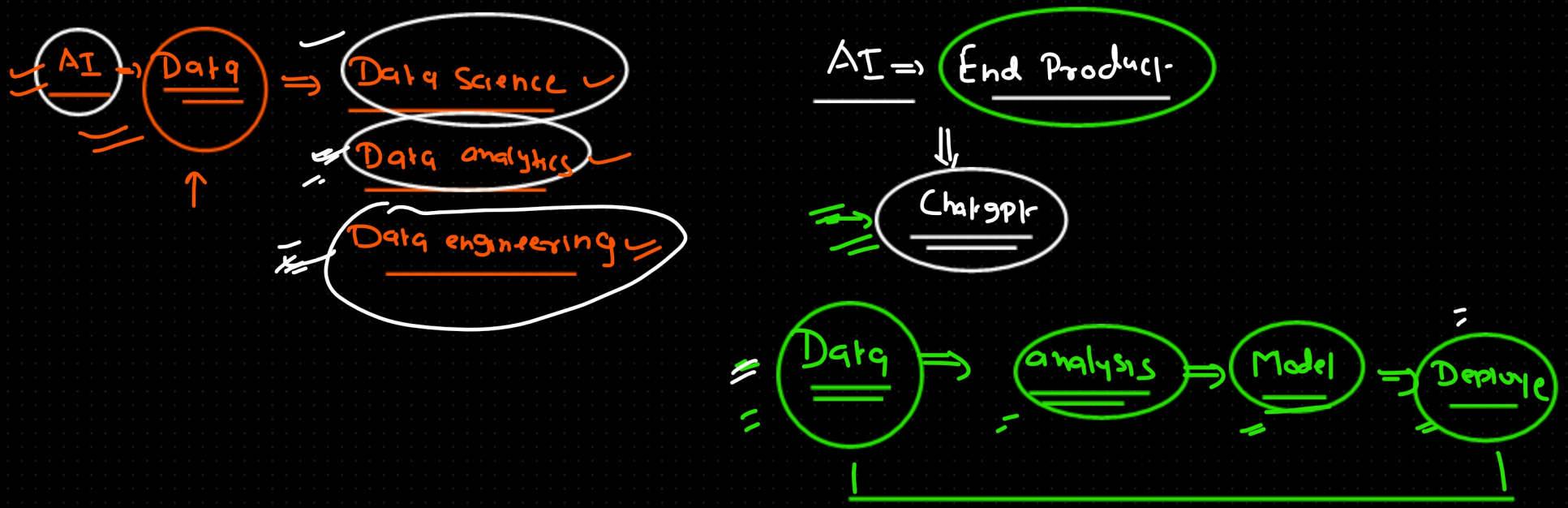
Chatgpt =>

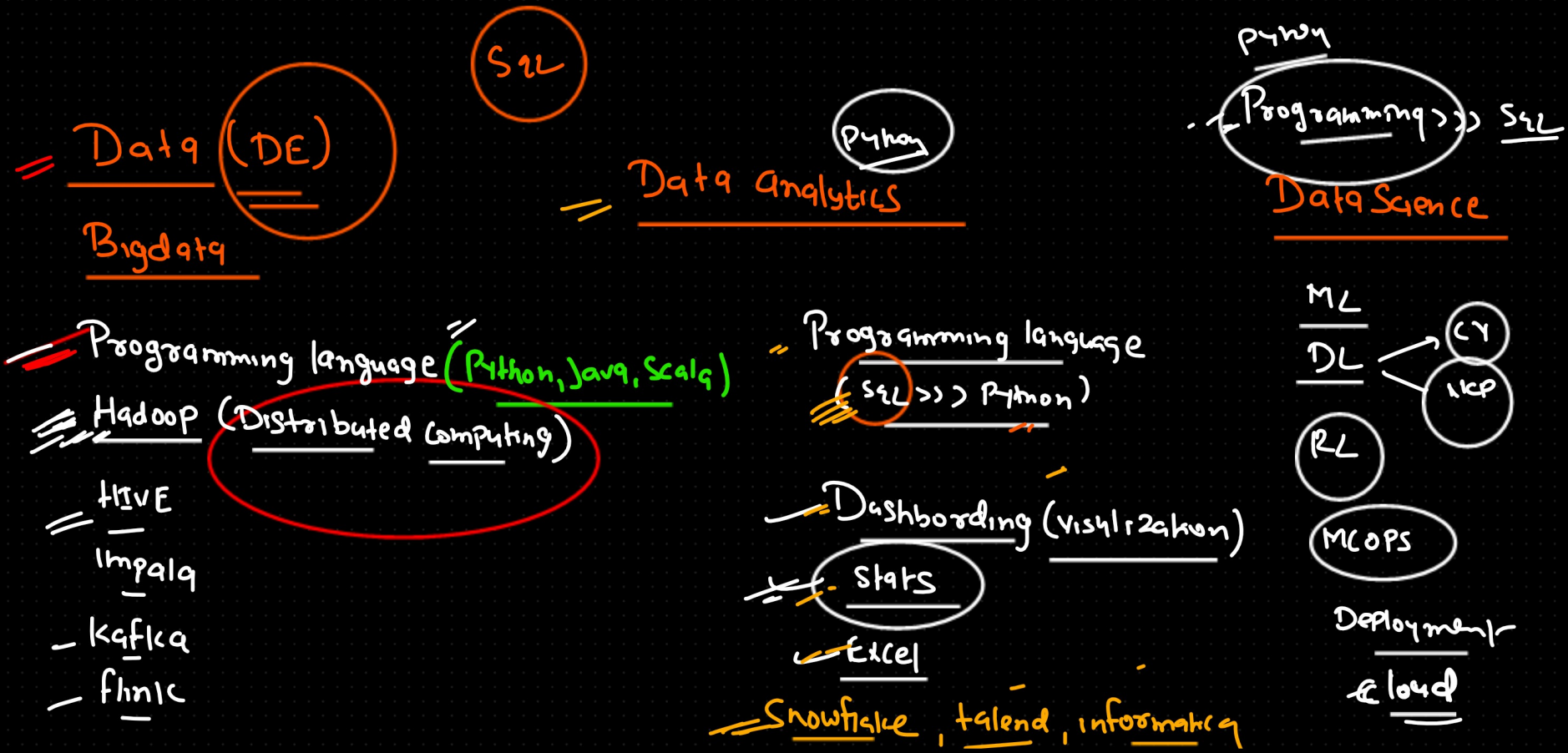
SVM IDT

{ Objective fn
loss
Optimizer }

Data
- cleaning
- Analysis
- stats
- Model
- evaluation




AI
Data
Data Science
Data analytics
Data engineering
AI => End Product
Chatgpt
Data
analysis
Model
Deploye


SQL
Data (DE)
Bigdata
Programming language (Python, Java, Scala)
Hadoop (Distributed Computing)
HIVE
Impala
kafka
flink
Python
Data analytics
Programming language
(SQL >>> Python)
Dashboarding (visualization)
Stats
Excel
Snowflake, talend, informatica
Python
Programming >>> SQL
Data Science
ML
DL
CV
NLP
RL
MLOPS
Deployment
cloud

- Spark
- Nifi
- Airflow
- Oozie

- MongoDB
- Cassandra
- HBASE

ETL { Extract, transform, load }

Data Pipeline

Storage, Processing, Management

SQL

DE ⇒ SQL >>> Programming

↑
ETL
Analytics

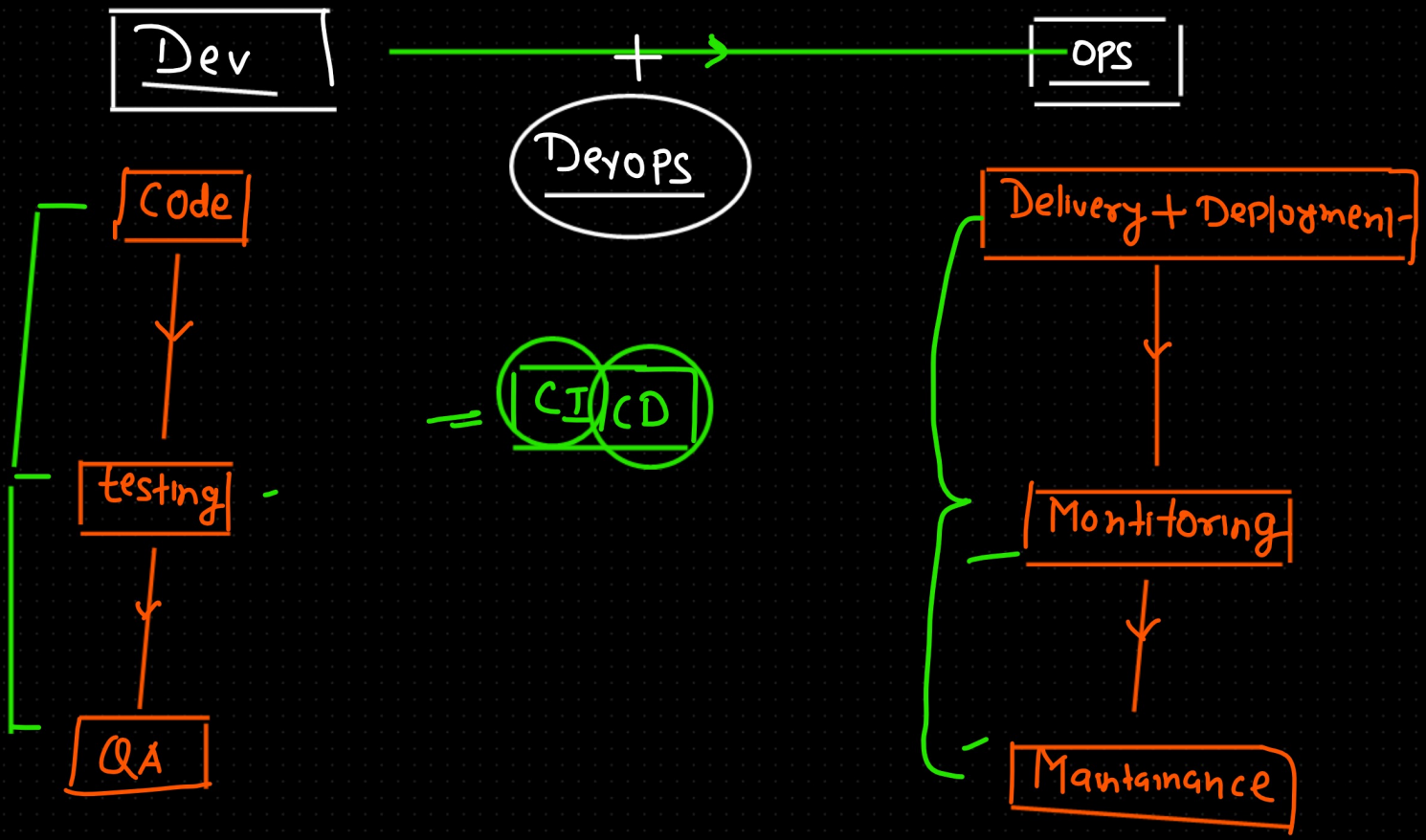
Stats

Basic knowledge ⇒ Dist com.

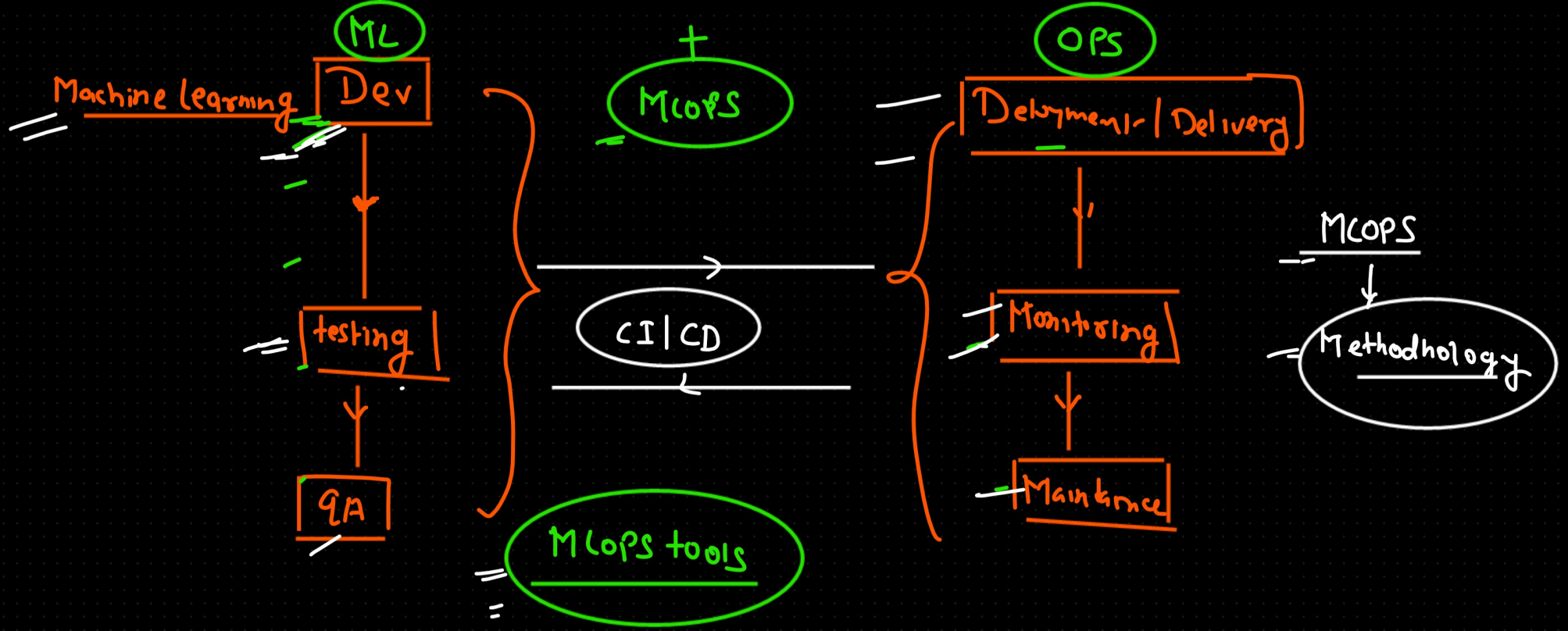
- Nosql
- SQL
- kafka
- Spark

- Aws | gcp | Azure

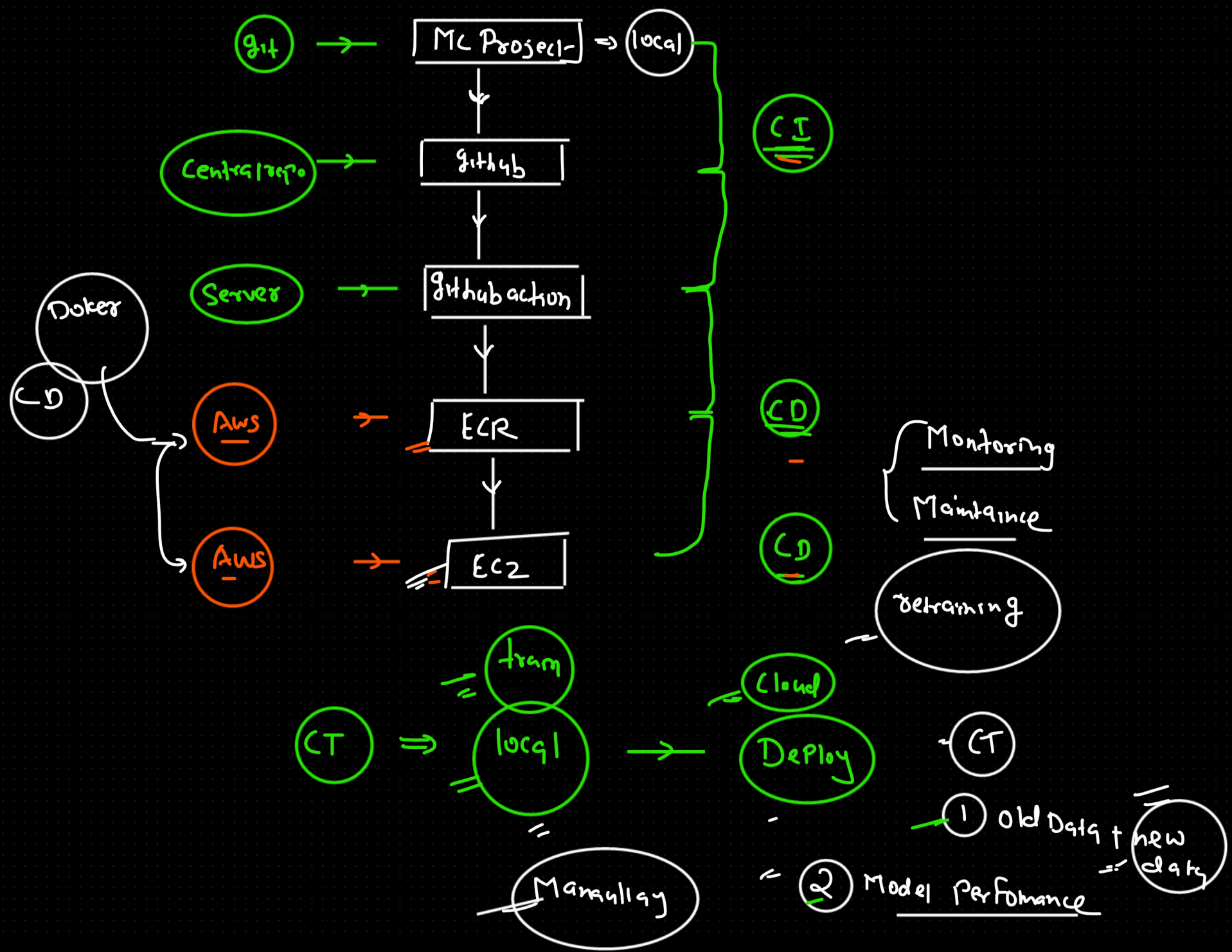
Generic Architecture of DS Project-

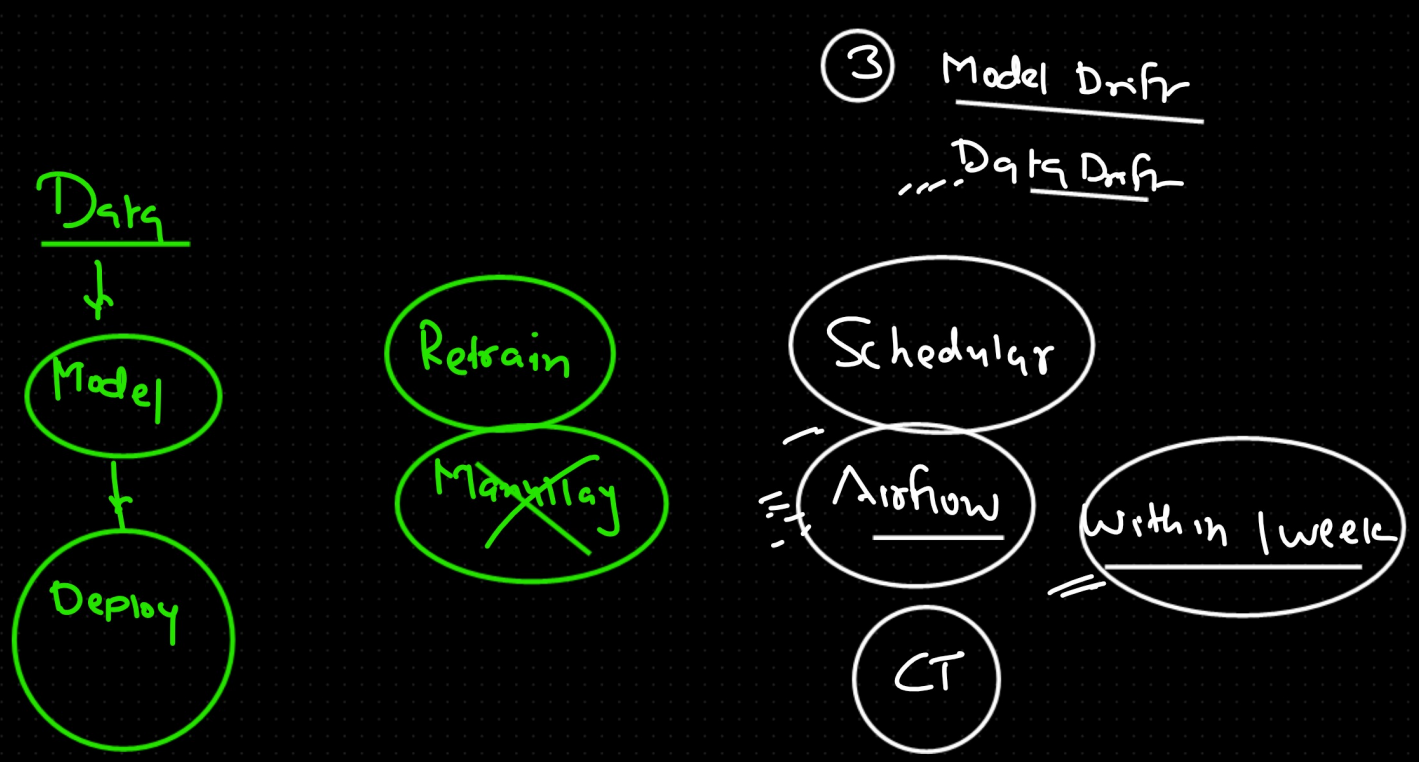


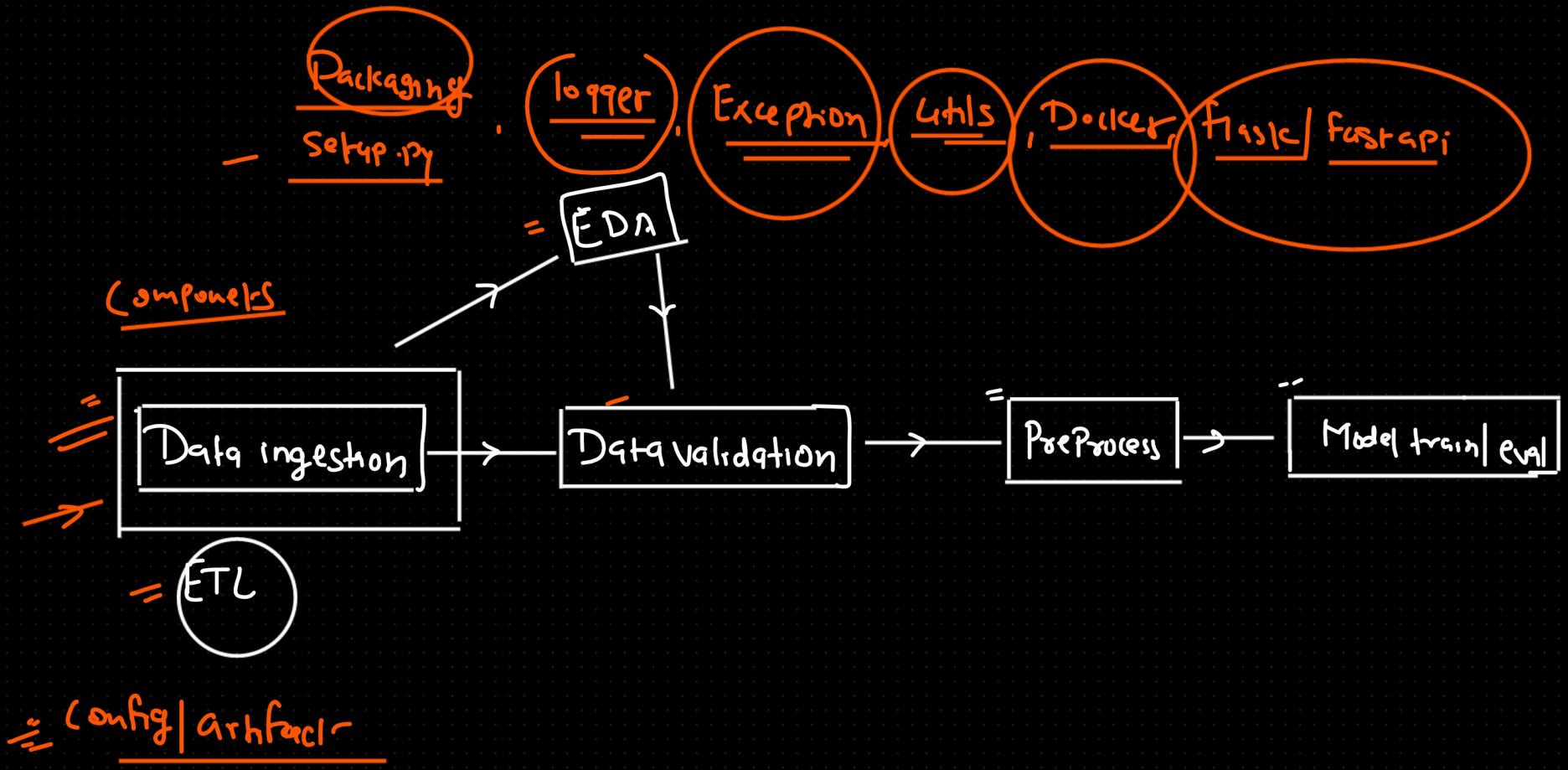
1. Waterfall
2. Agile Methodology



- git, Docker, github action, grafana, kubernets, Selenium, Packaging


git
MC Project-
local
CI
Centralrepo
github
Server
github action
Docker
CD
AWS
ECR
CD
AWS
EC2
Monitoring
Maintaince
CD
Retraining
train
Cloud
CT
local
Deploy
CT
1 Old Data + new data
Manually
2 Model Performance


3 Model Drift
Data Drift
Data
Model
Deploy
Retrain
Manually
Scheduler
Airflow
Within 1 week
CT


Packaging
setup.py
logger
Exception
utils
Docker
flask/ fastapi
EDA
components
Data ingestion
Data validation
PreProcess
Model train/ eval
ETL
config/ artifact