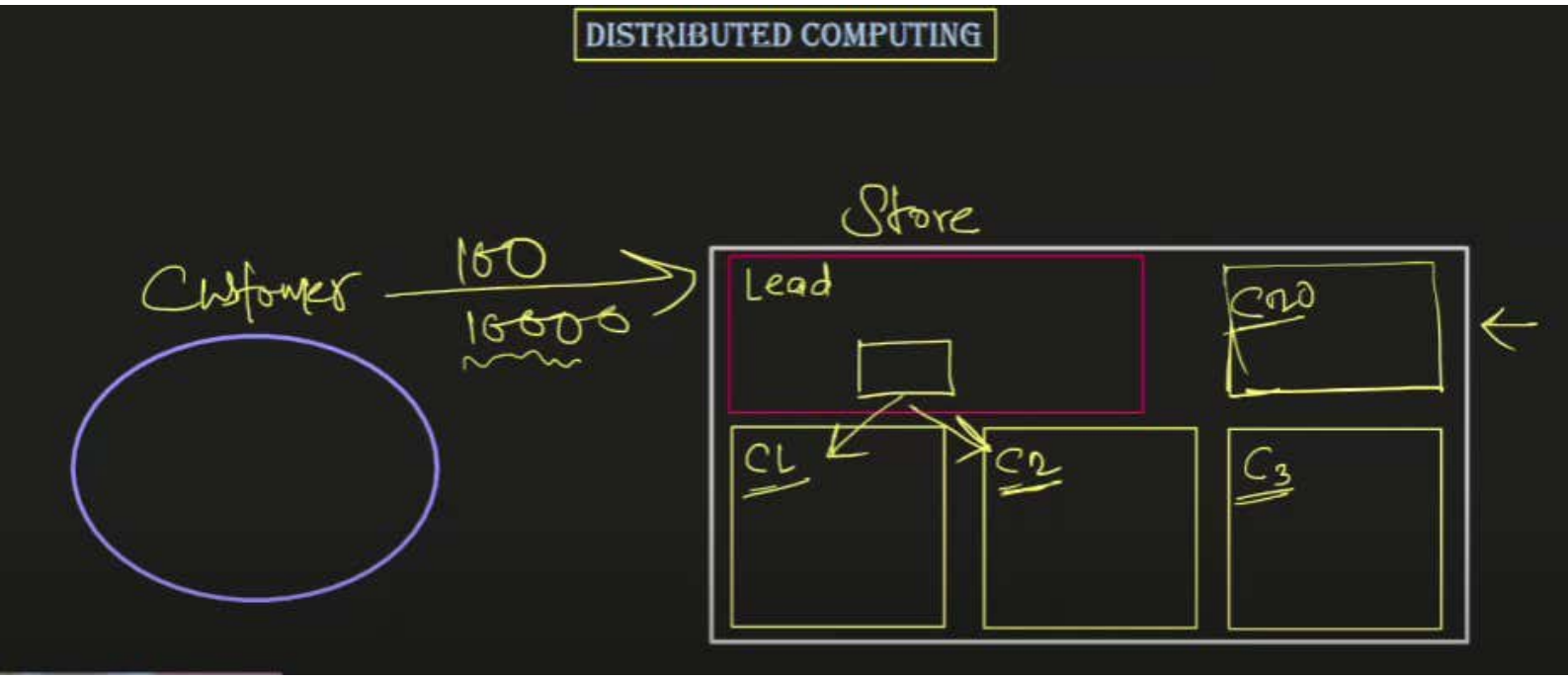
TOPICS TO BE COVERED

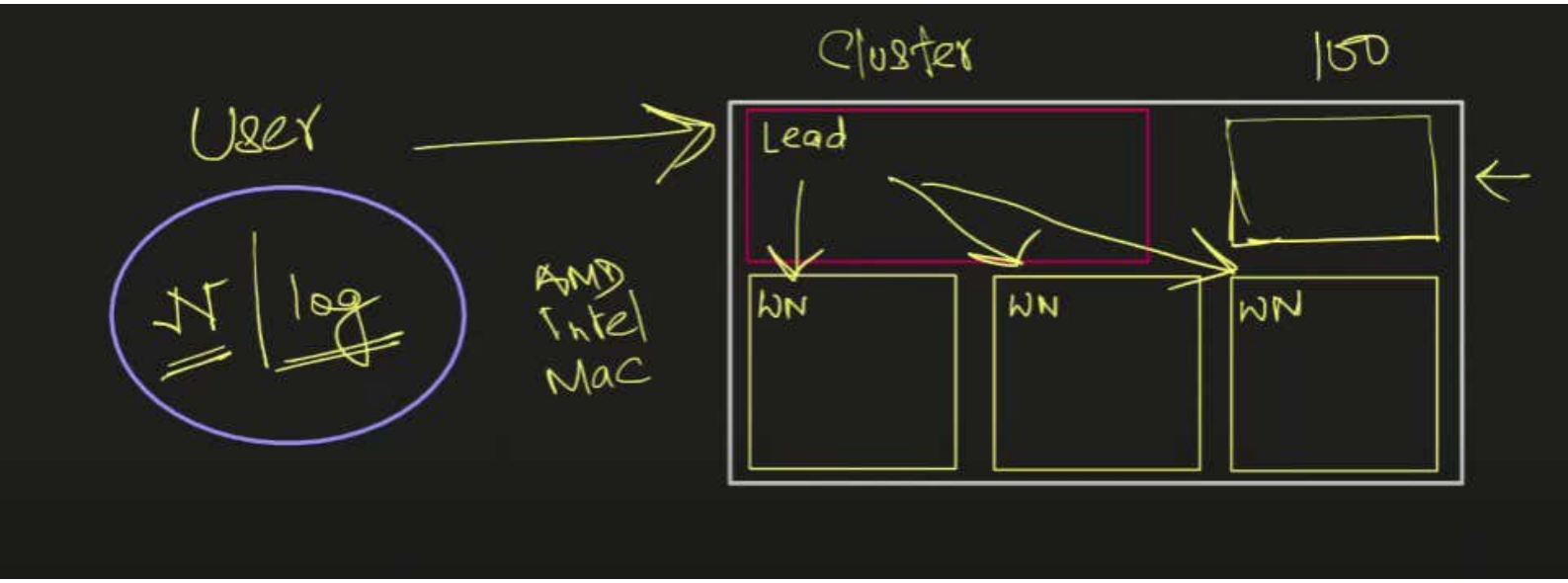
{DVC MLFlow}

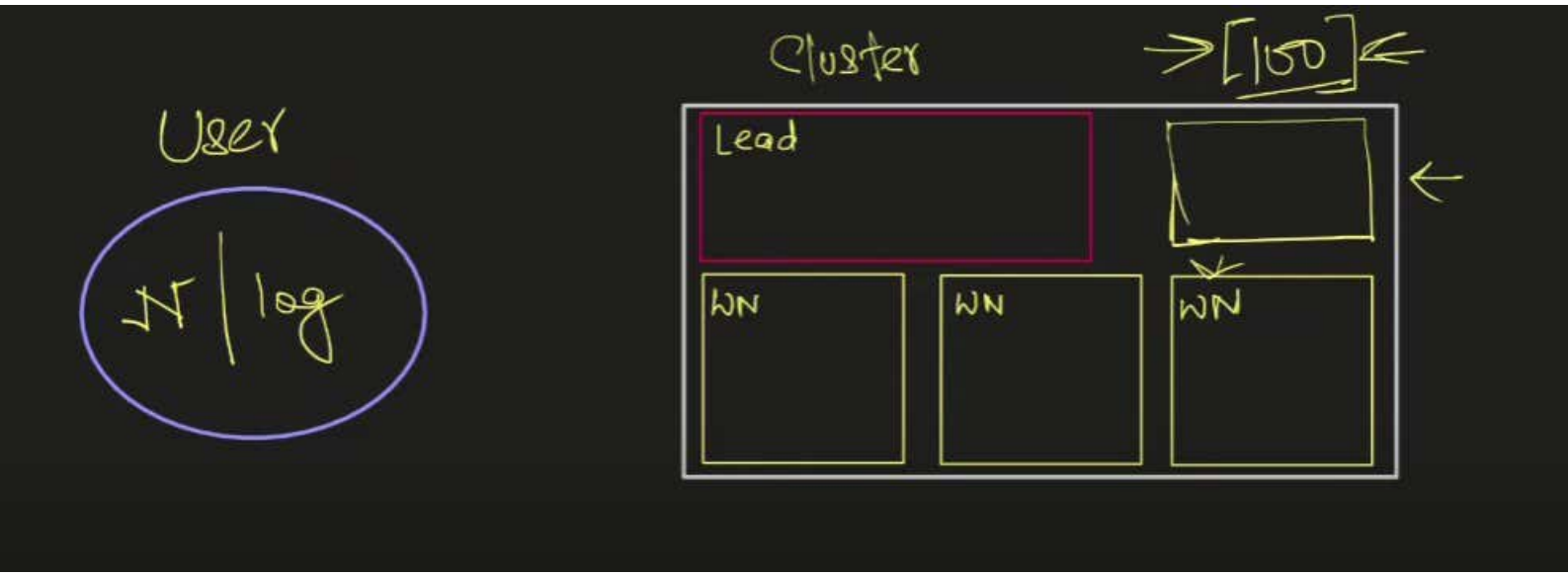
- Our playlist so far. (Pre-requisite)
- Distributed Computing
  - Cluster
  - Lead-Node Server
  - Communication
  - Concurrency [Speed, Fault Tolerance]
  - Same as inside Apache Spark internally (Map Reduce)
- Microservices (How related to Docker & Kubernetes?)
- Kubernetes Internals (next slide)

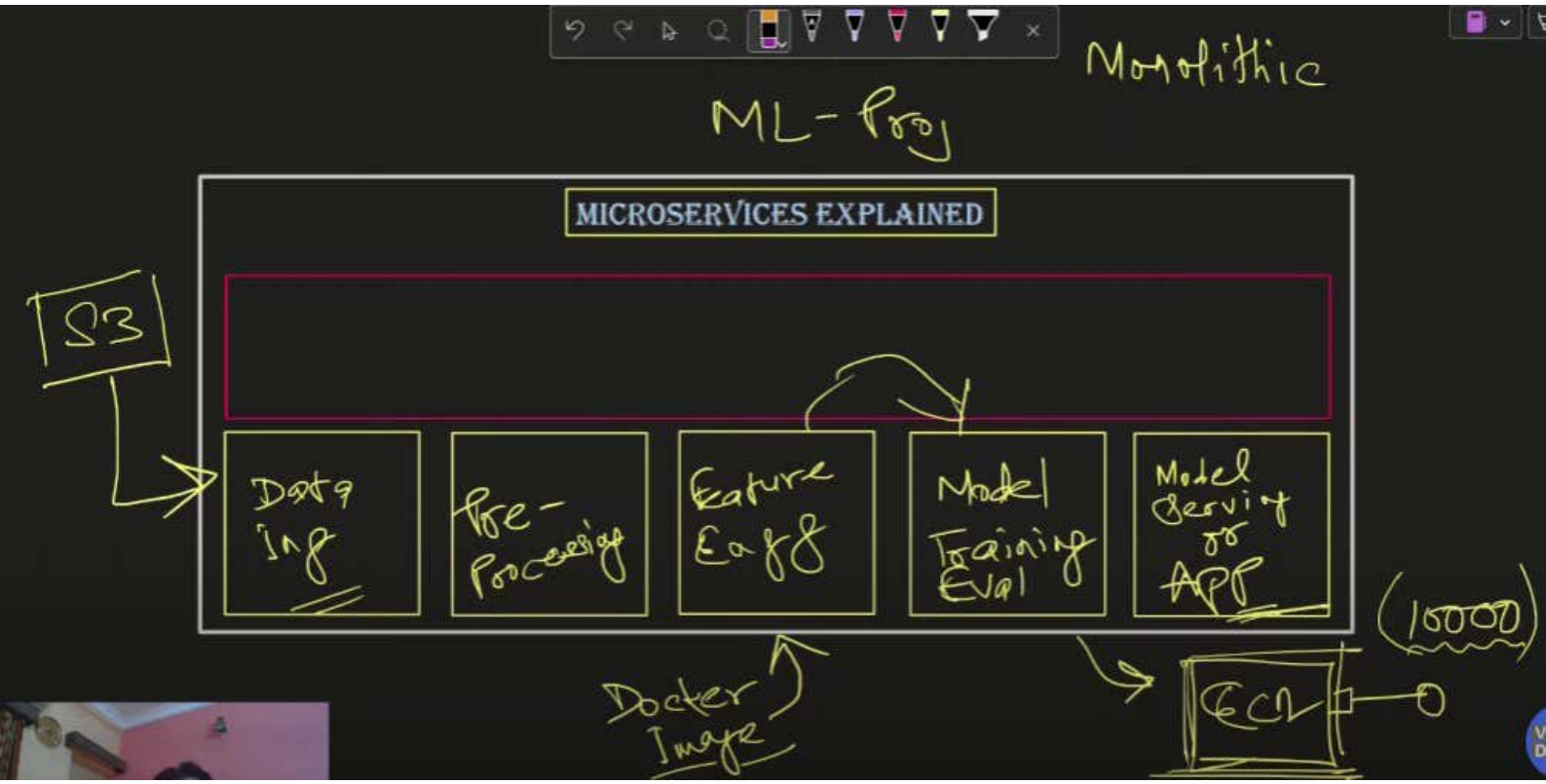
KUBERNETES INTERNALS

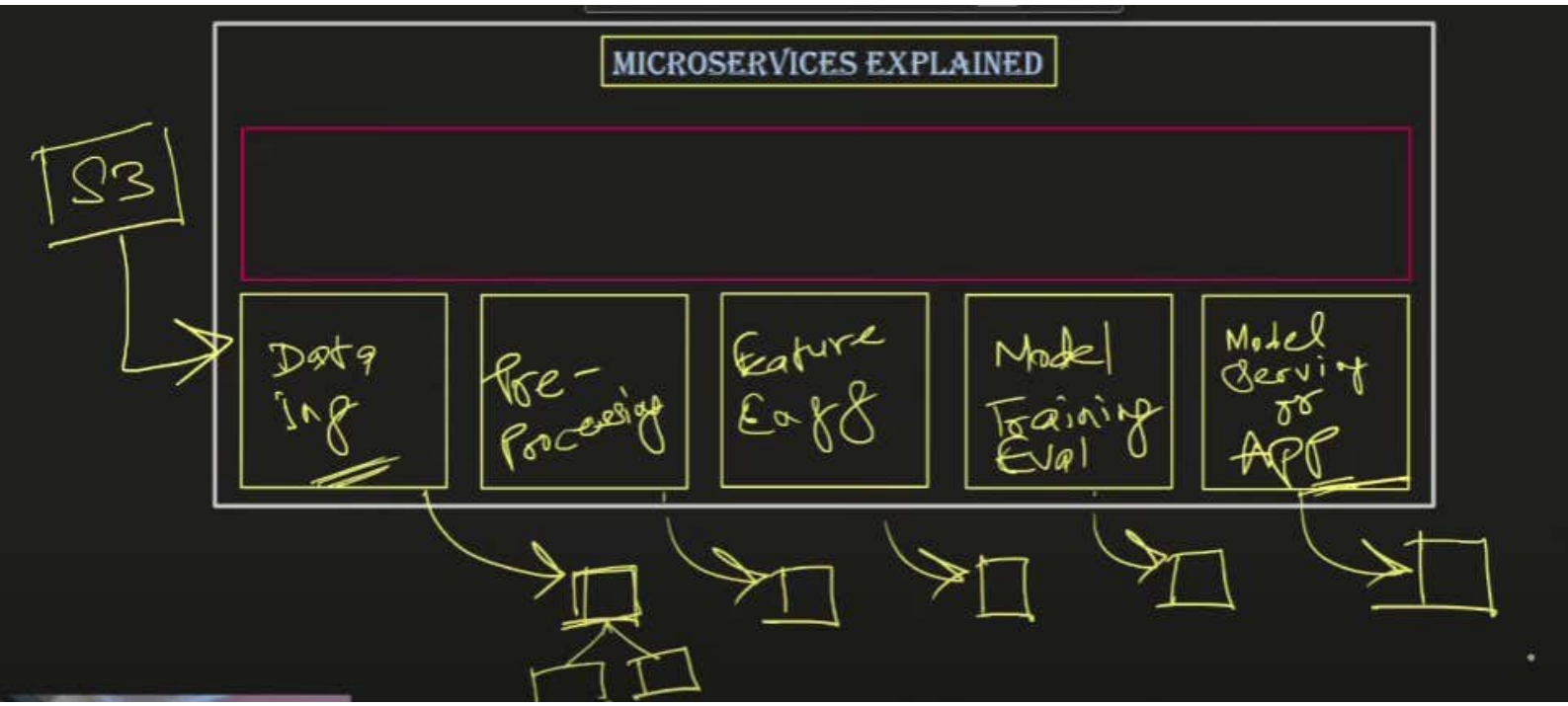
- Master Node (Control Plane)
  - Resource Manager
  - Communication (User to API Server)
  - Database (etcd)
- Worker Node
  - Kubelet
  - Kube-Proxy
  - Pods
  - SharedDB (volumes)
  - Kube-manifest (yaml)
  - Service
  - Namespace
  - Scheduler
  - Replica Set
- VS Code - Kubernetes Extension


DISTRIBUTED COMPUTING
Store
Customer
100
10000
Lead
C20
C1
C2
C3

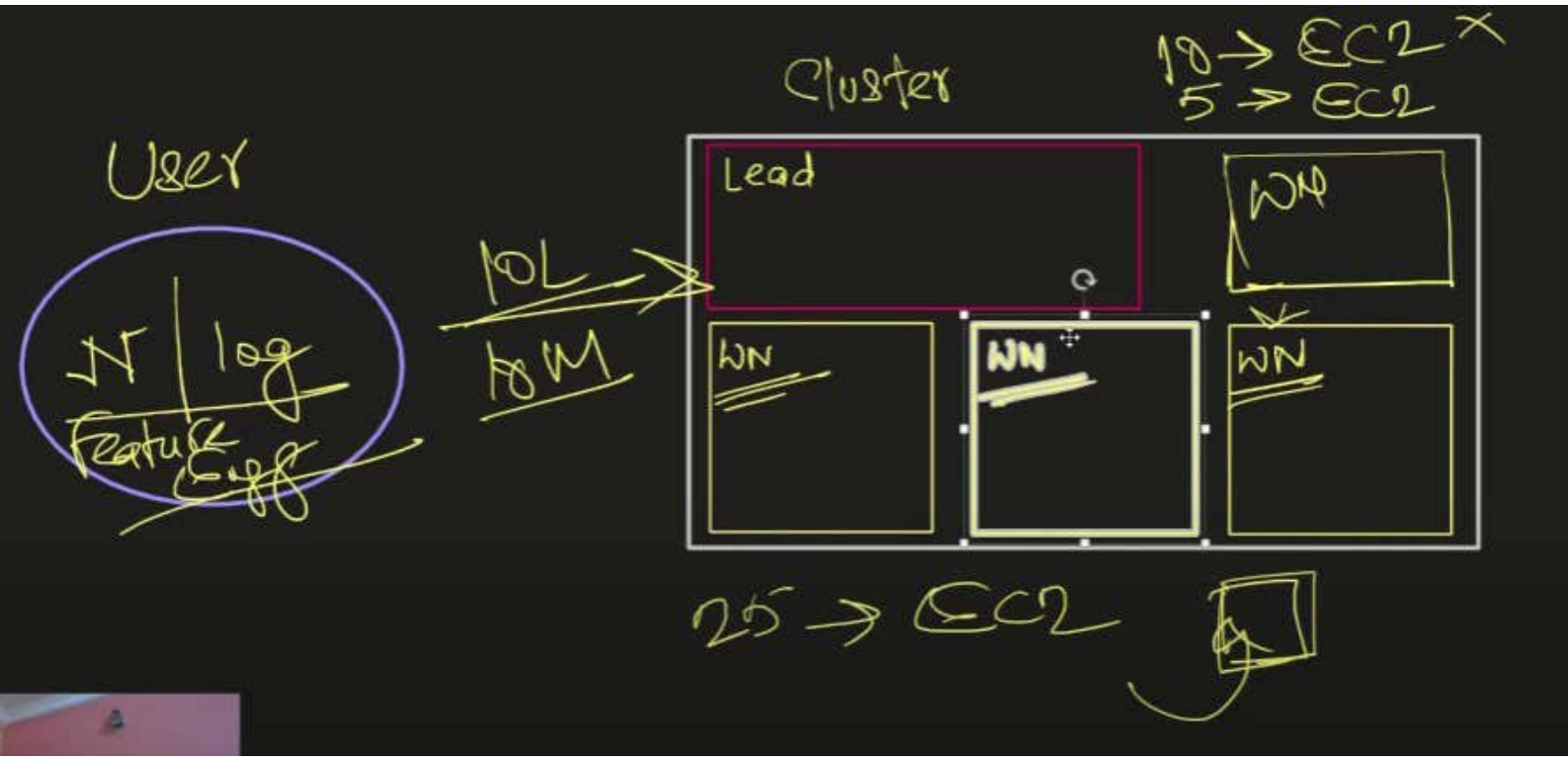

Cluster
User
Lead
WN
WN
WN
AMD
Intel
Mac

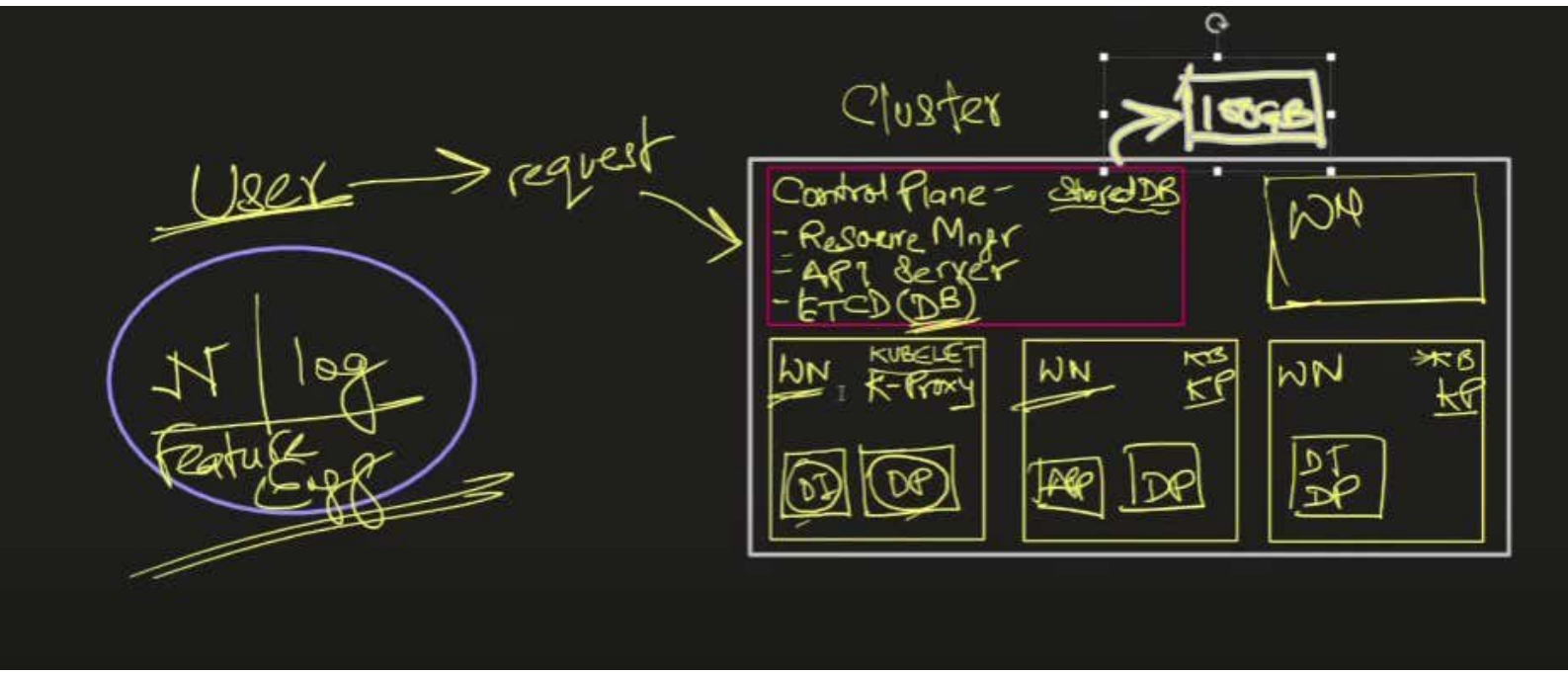

User
Cluster
Lead
WN
WN
WN

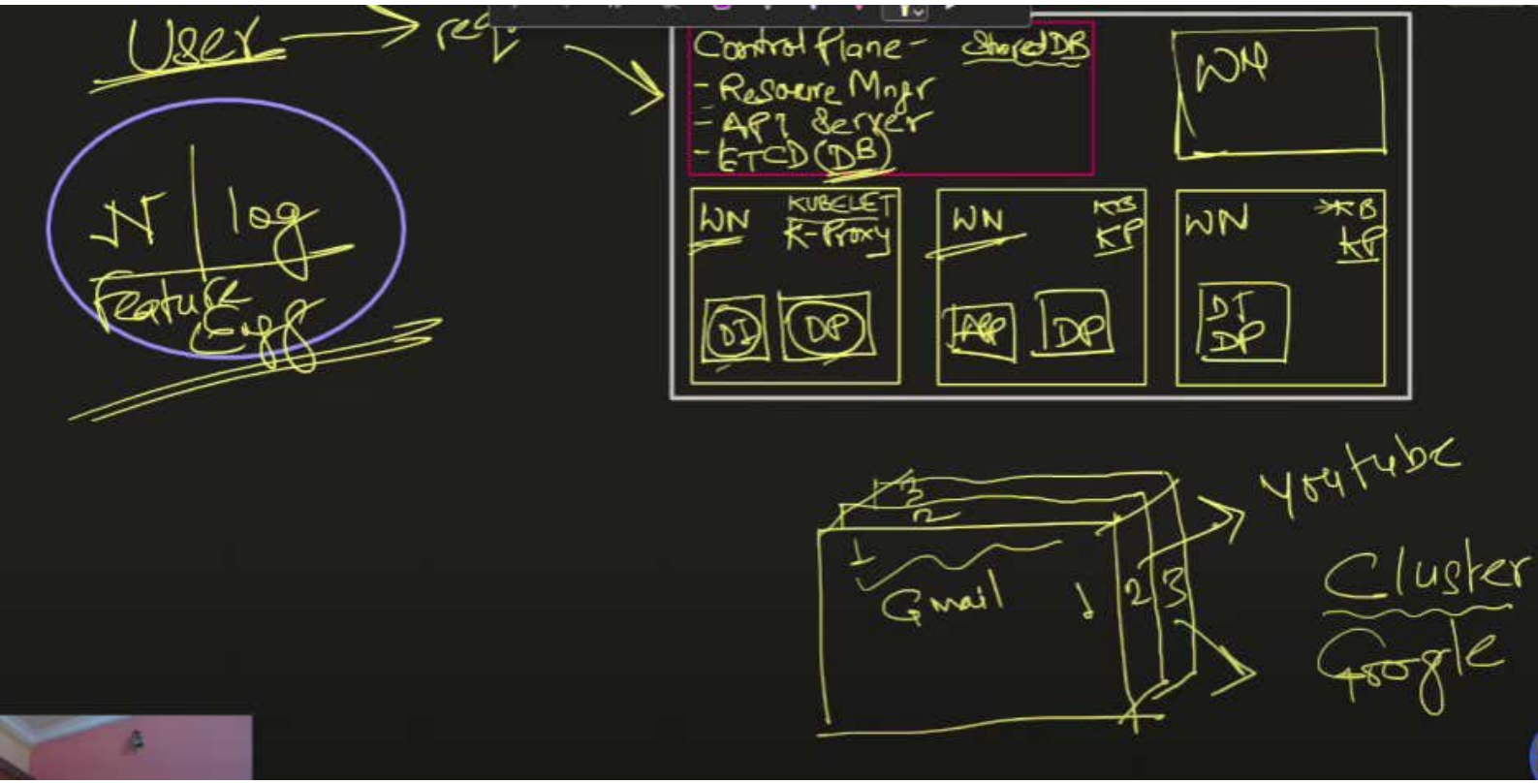

Monolithic
ML-Proj
MICROSERVICES EXPLAINED
S3
Data ing
Pre-Processing
Feature Engg
Model Training Eval
Model Serving or App
(10000)
Docker Image
EC2


MICROSERVICES EXPLAINED
S3
Data Ing
Pre-Processing
Feature Eng
Model Training Eval
Model Serving or APP

Zomato App / Vehicle Insurance
[Resturant Recc]
[Food Recc]
[Coupon]
[Delivery Agent]
[Order tracking]
[ML Project]


Cluster
10 → EC2
5 → EC2
User
Lead
WN
WN
WN
WN
25 → EC2


User
request
Cluster
log
Control Plane -
- Resource Mngr
- API Server
- ETCD (DB)
Shared DB
WN
WN
KUBELET
K-Proxy
WN
WN
DP
DP


User
Control Plane -
- Resource Mngr
- API Server
- ETCD (DB)
Shared DB
WN
KUBELET
K-Proxy
DP
log
Gmail
Youtube
Cluster
Google