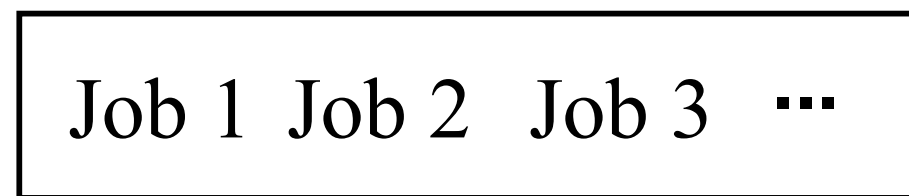


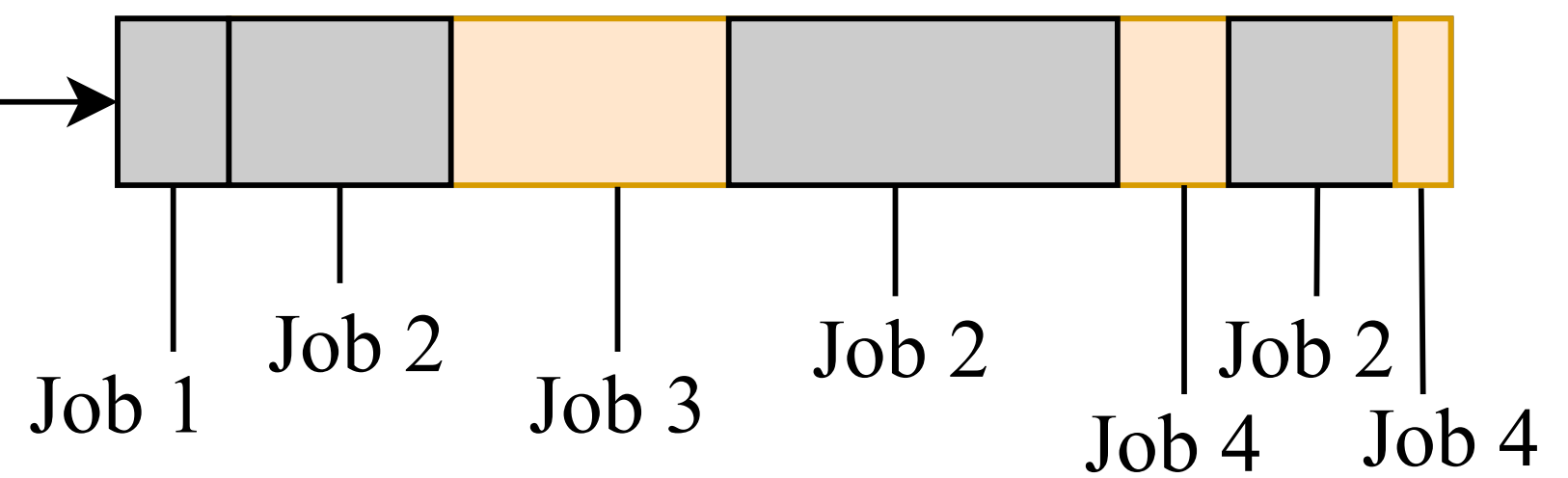
Trade-off line

Latency Memory Accuracy ...

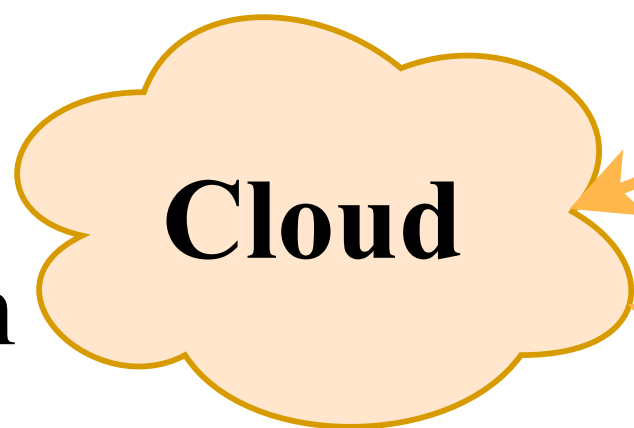
Runtime



Job queue



Pretrainig,
optimization



Caching,
coordinating

Edge

Lightweight
input collection

Cache management

