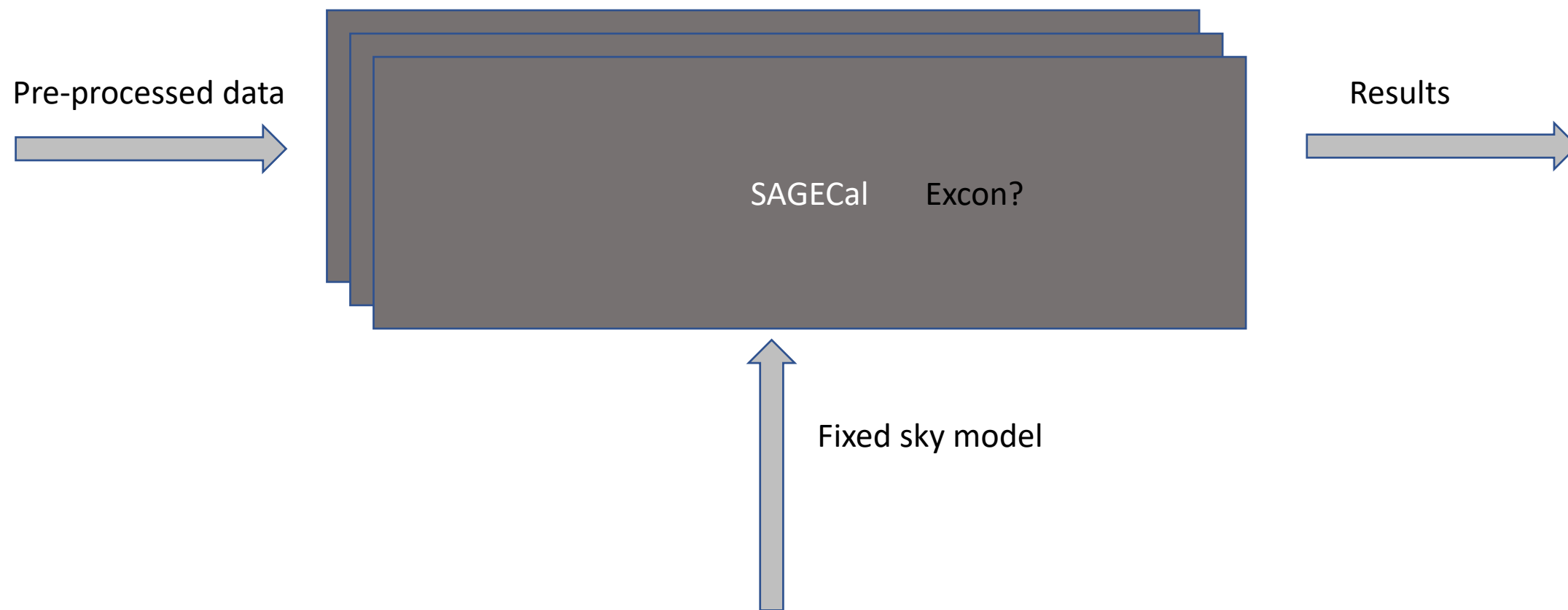
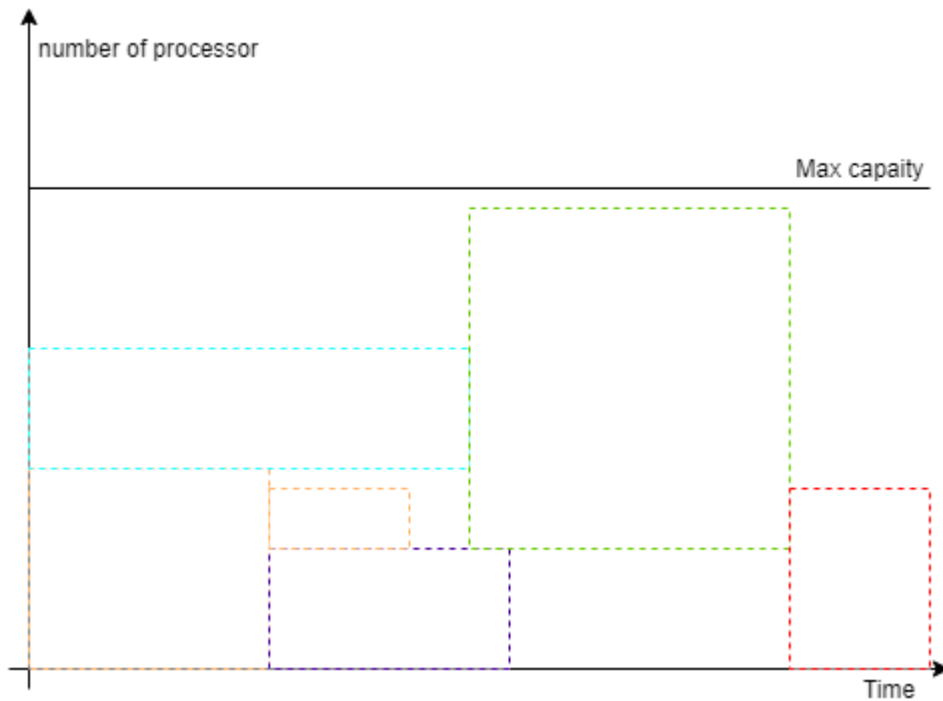


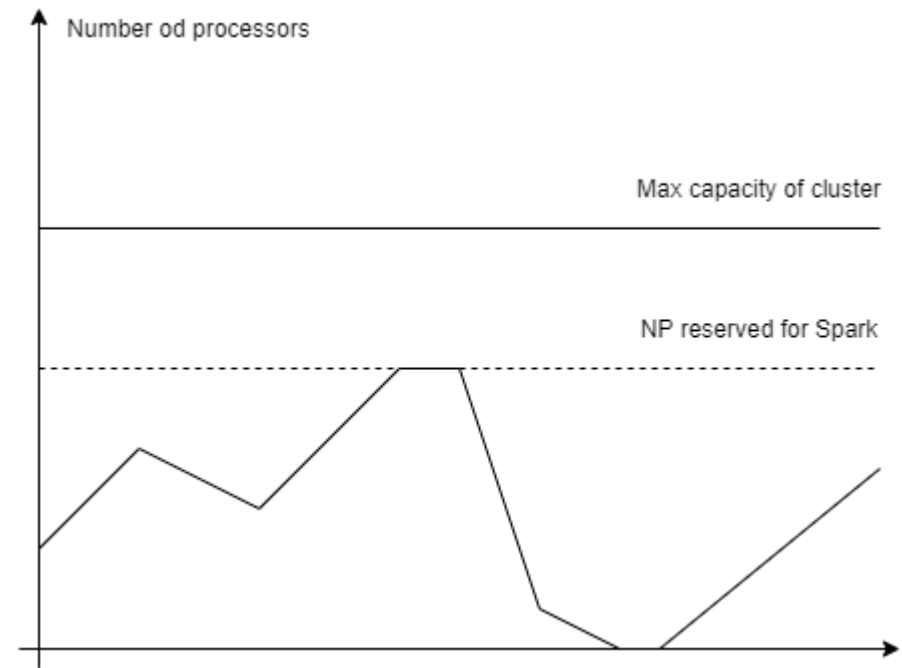
# SAGECaI



# Issues of MPI and Spark



The resources can not be utilized completely



# Possible solution?

- MPI:
  - Smart job queue managed by cluster
  - Finer or uniformed granularity
- Spark
  - Containerization (current solution)
  - Buy Databricks auto-scaling solution
- Other solution?
  - Kubernetes
  - VMs

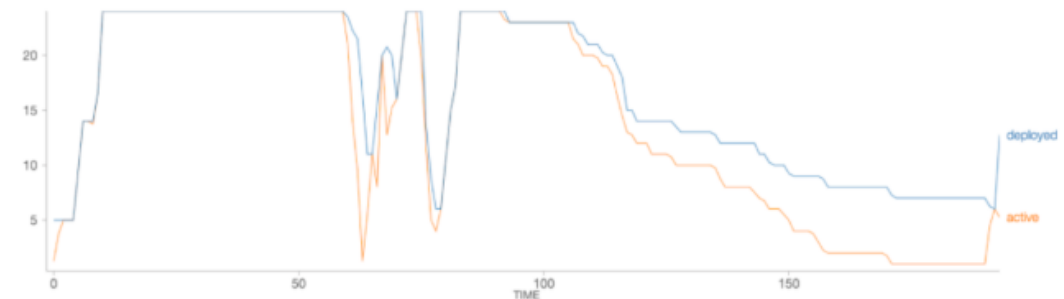
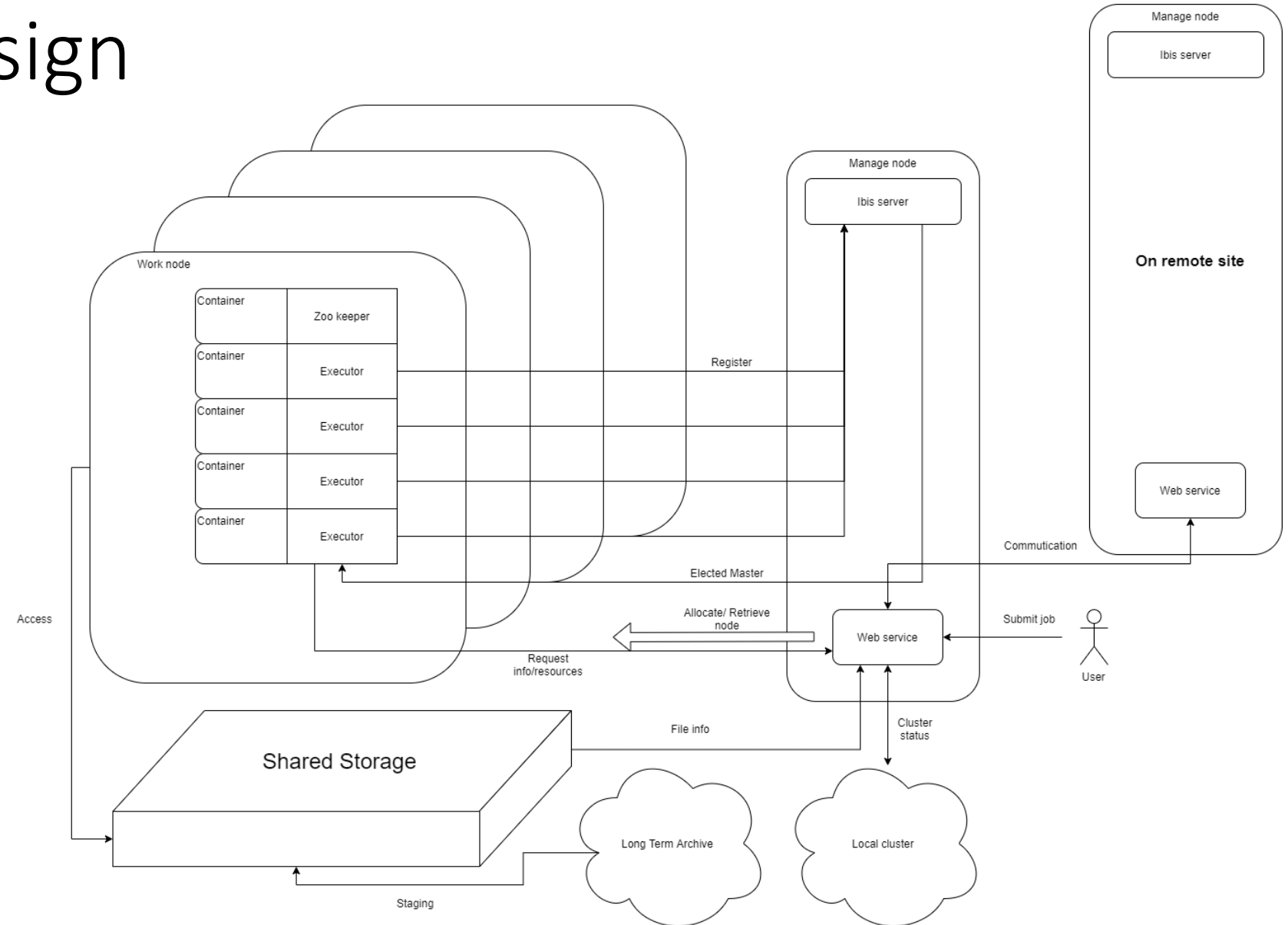


Figure 2. Databricks' optimized autoscaling: Active executors vs total executors

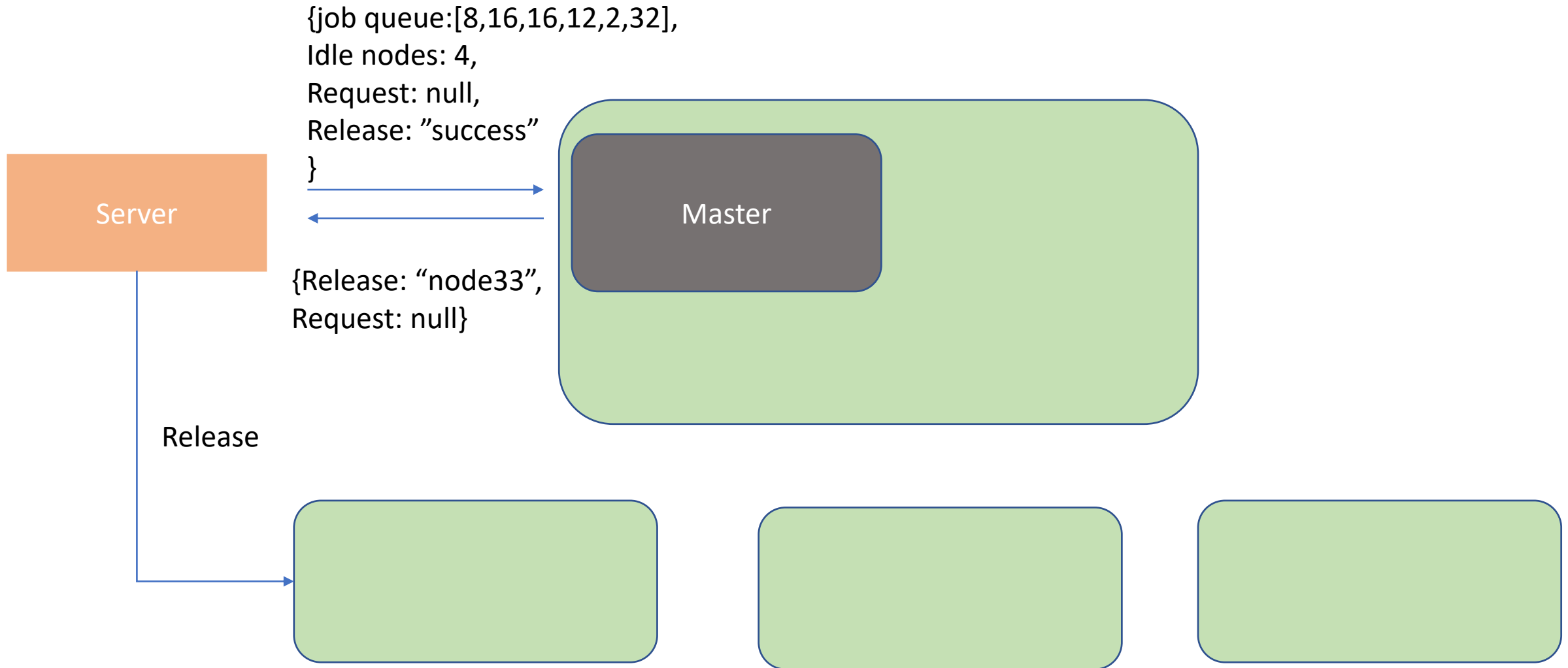
# Overview design

Tackle on THREE dimensions

- Self-adjusted provisioning
- Cross region
- Data locality

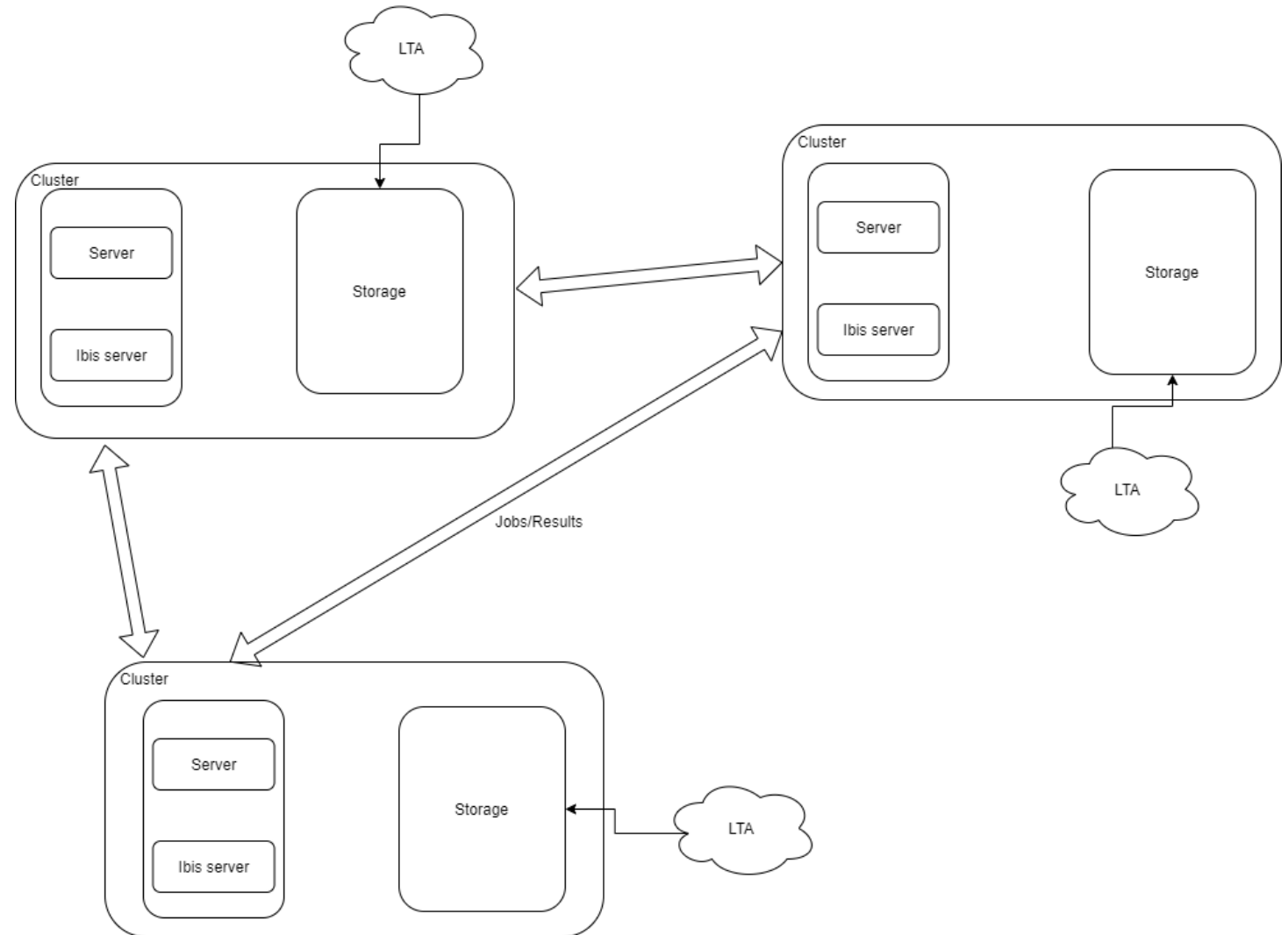


# Self-adjusted provisioning



# Cross region & Data locality

- Utilizing remote resources
- Backup to each other
- Submit anywhere, run at close place
- Staging and processing at same time



# Stages plan

Phase 1: Ibis +Sagecal with docker container on one node

Phase 2: resource management at one cluster(multiple nodes)

Phase 3: cross-region

Extra 1: parallelize staging from tape and processing