Multiple region support

Goals:

- Be able to balance workload between two regions
- Workloads from same tenant can communicate if needed
- Region internal logic (network, compute, storage) remains as it, i.e. each region is still a functional unit

Assumptions:

- Each region maintains its own control plane(network, compute, storage)
- Each region had different caps, supported services/features
- Networking physically disconnect, but can be connected via gateway

Potential Models:

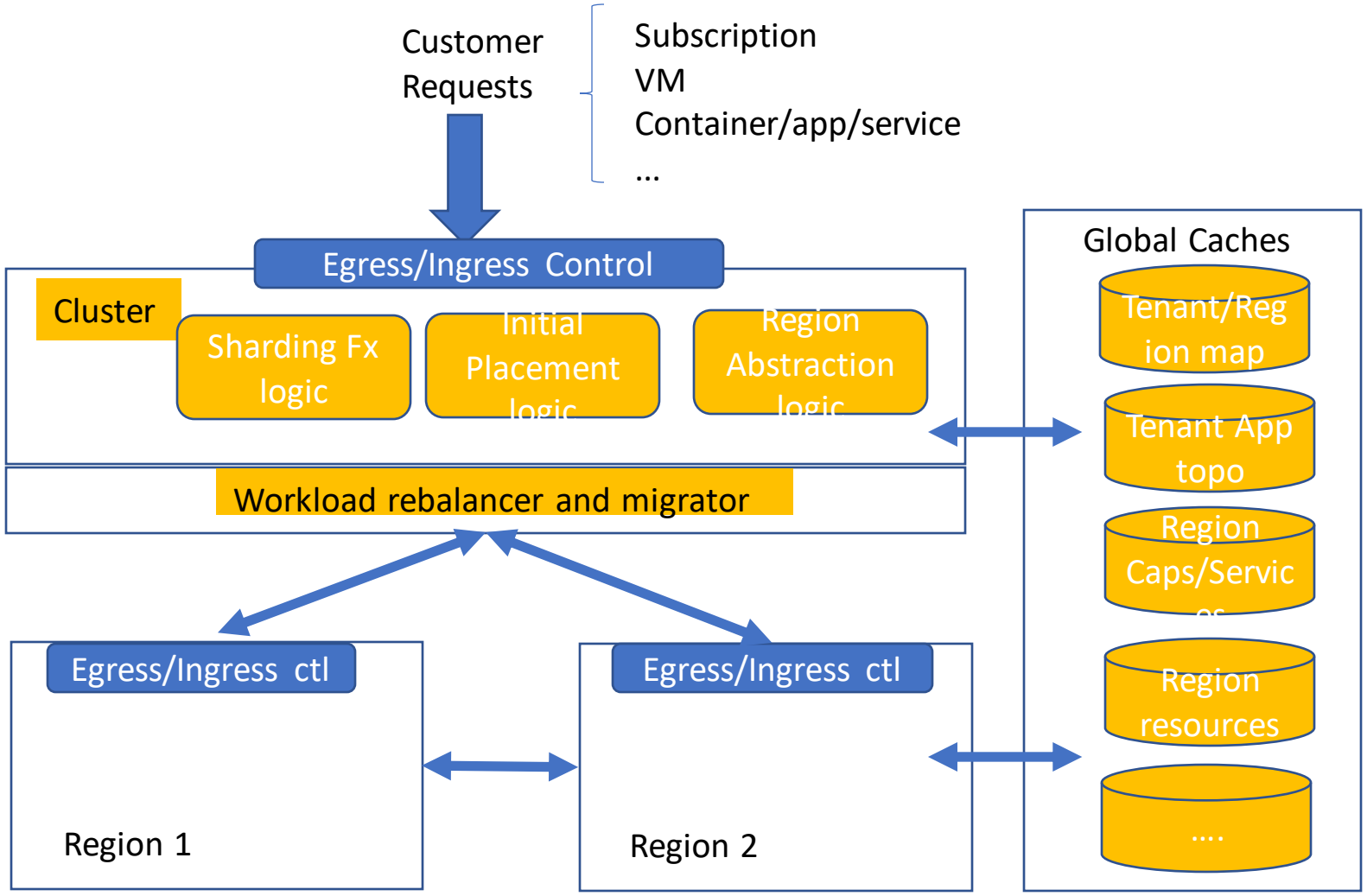
- Tenants were assigned to a particular region and all tenant's workload placed to this region
- Tenants can cross regions. Tenant's workloads are placed to all regions, affinity/anti-affinity enforced
- Number 2 plus, dynamically adjust workload per policies: e.g. region's capacity, tenant workloads SLA, migrate workloads if needed

Abstraction & Aggression

- Aggression of region resources, compute, network(public IPs), storage
- Abstraction of workload, requests

FUTUREWEI Technologies

CloudPaaS: A High Level View



FUTUREWEI Technologies

Challenges

- Maintain multiple region level tenant/app topology map for ideal placements of tenant workloads
- Ensure data locality with sharding of user applications
- Abstraction of regional level requests/workloads
- Support for both in-provider and inter-provider regions

Pending issues

- Define a controllable and manageable scope, e.g. initial step with single cloud provider to avoid the compatibility, billing, policy, type abstraction etc.
- Estimate, assess the cost of the new layers, validate the design approach
- Networking between regions, can regions merge to one under one set of control plane ?
- Two level schedule or single level (scheduling logic are all in the region level)