KubeCon

North America 2017

HA Services During Maintenance

Maisem Ali & Eric Tune (Google)

Maintenance

- Nodes require maintenance
- Disruptive
  - Removal of capacity
  - Shuffling of pods

// TODO: add diagram

Maintenance

- Planned
  - Upgrades
  - Repairs
  - Security patches
- Unplanned
  - Disk failure
  - Kernel panic

Graceful termination

- Termination grace period
  - SIGTERM
  - Wait for period
  - SIGKILL
- Propagation delay in service updates
  - In-flight requests should be fulfilled
  - Requests can still be delivered after receiving SIGTERM
  - Should be handled appropriately
    - e.g. GOAWAY

// TODO: add diagram of traffic routing

Pod Disruption Budgets

- Min Available
  - At least these many pods should be Ready
  - Example
    - at least 3 pods out of the 5 pods should be running to maintain quorum
- Max Unavailable
  - At max these many pods can be NotReady
  - Work only with controller
  - Example
    - at max 10% of the serving capacity can be unavailable

Eviction Subresource

- Verifies PDBs before deleting the pod
- Acts as a rate limiter
- Error codes
  - No PDBs violated or none defined
    - 200
  - Violates any associated PDB
    - 429

// TODO(maisem): add diagram

Stateful Sets

- Content
  - Content
    - Content

Draining a Node

- Mark it unschedulable
- Evict all the pods
  - Blocking for the PDBs
  - GKE timeouts after 1 hour
- Wait for all the pods to exit
  - Each pod should exit after the podTerminationGracePeriod and should be ignored after the threshold
  - GKE caps this at 1 hour
- Verify that there are no more pods

… or use `kubectl drain`

Draining a Node

//TODO: Add diagram

Demo

Q&A