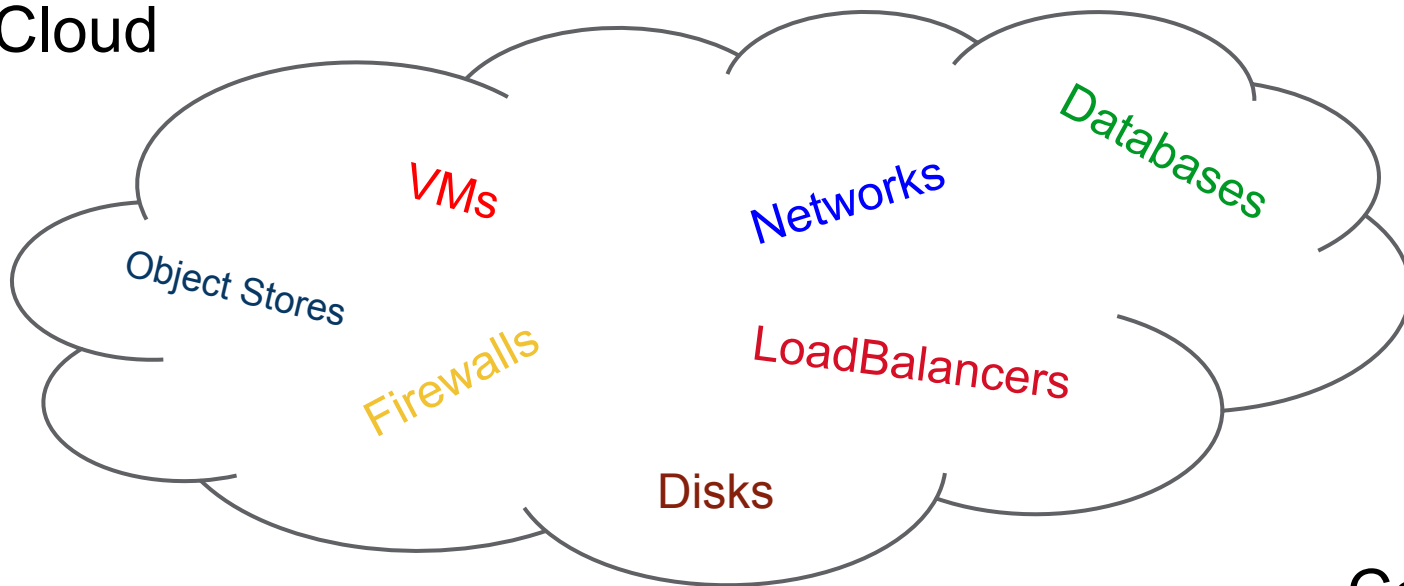




Deploying complex applications to Google Cloud

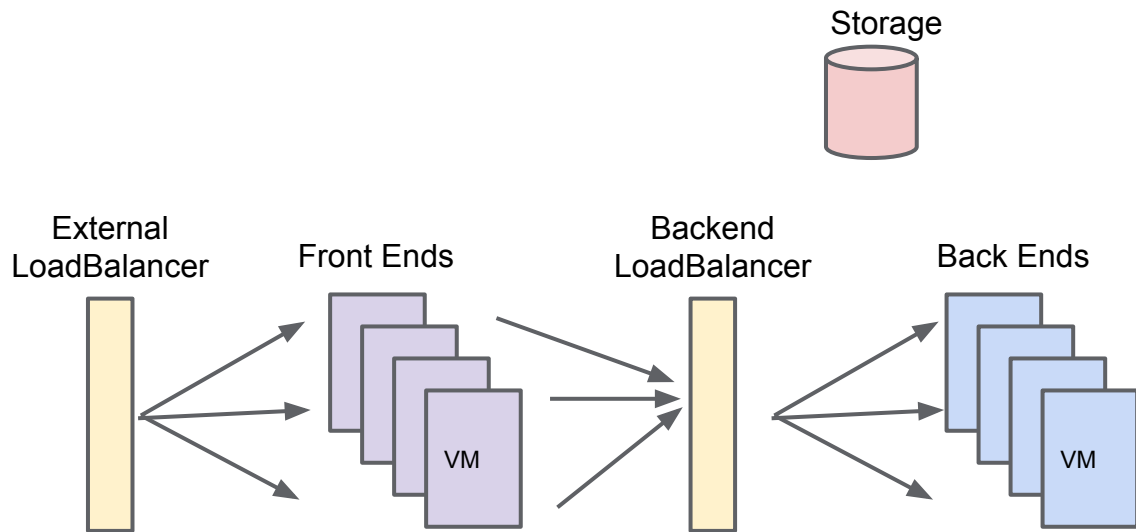
Olia Kerzhner
olia@google.com

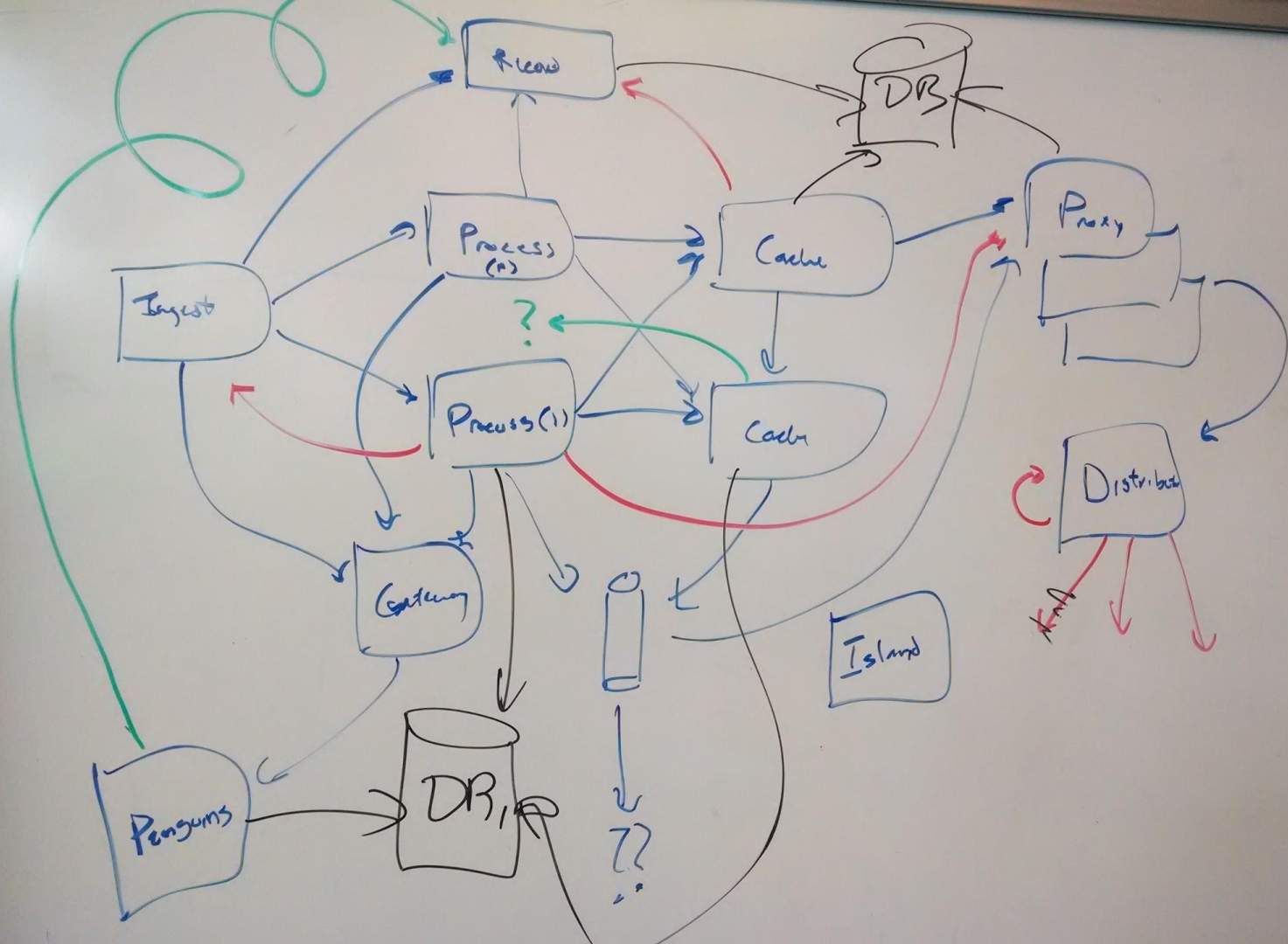
Cloud



Control ..?

Application stacks are complex





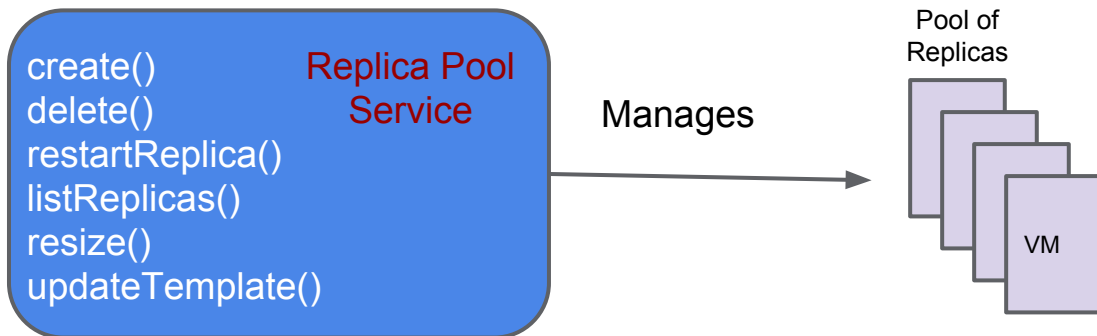
Building on top of Google Compute Engine APIs

- GCE API are great for manipulating individual resources
- any minimally complex application requires users to write scripts or other code to deploy
- using scripts or manually creating resources results in "snowflake" deployments
- many open-source or proprietary tools try to solve this problem

Introducing: Replica Pool API

Powerful new API to manage sets of homogenous VMs

- VMs are created based on a template
- declaratively specify what software to deploy
- VMs are monitored for health and are restarted on failed healthchecks
- the API supports the `resize()` operation that will add or subtract VMs to/from the set

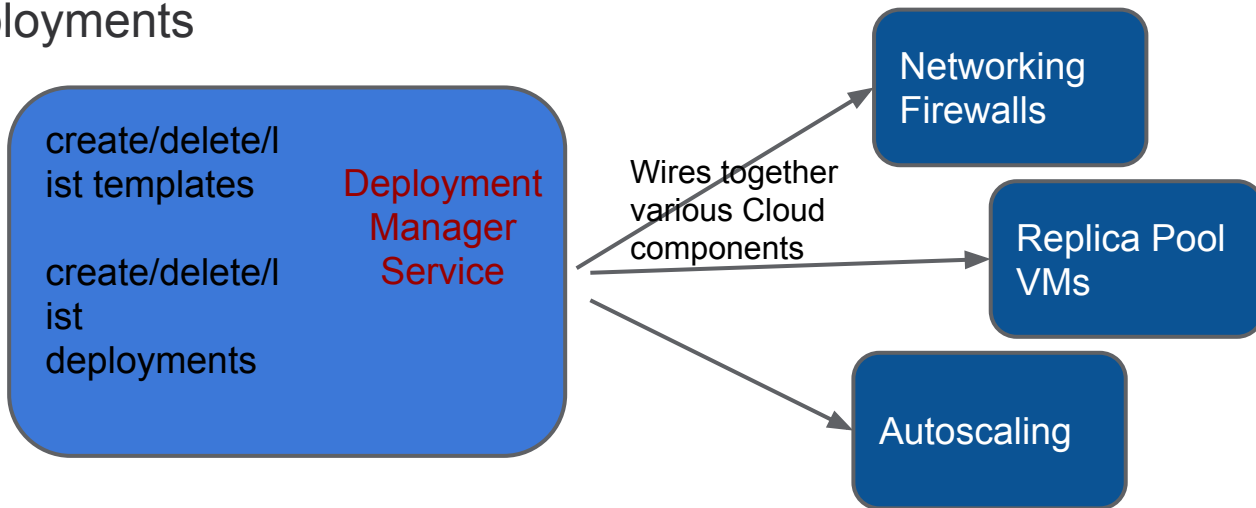


Managing more complexity

- need multiple sets of homogenous VMs: FrontEnds, BackEnds, Batch workers
- need load balancers, firewalls, networks, etc.
- need a way to tie the components together

Introducing: Deployment Manager

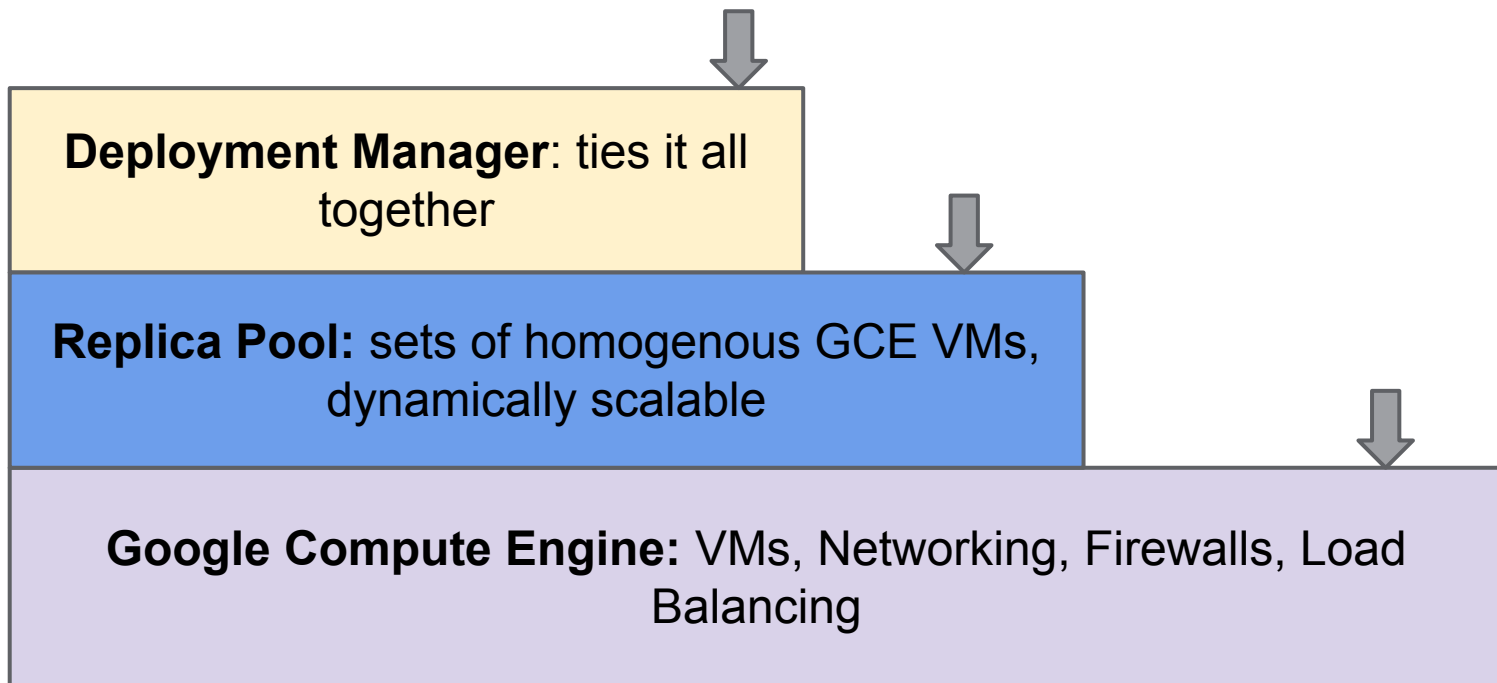
- a single declarative JSON template to tie together different Cloud components to define an application stack
- templates are reusable, overrides are supported for customization of deployments



Deployment Template

- a DECLARATIVE specification for how to deploy a set of Google Cloud resources together into an application stack
- defined in terms of "modules", each module corresponds to a Google Cloud API
- modules are connected together via module names and references

Google Cloud Layered APIs



Building Cloud-aware applications

What if you built your clustered application with the knowledge that it will run in the Cloud?

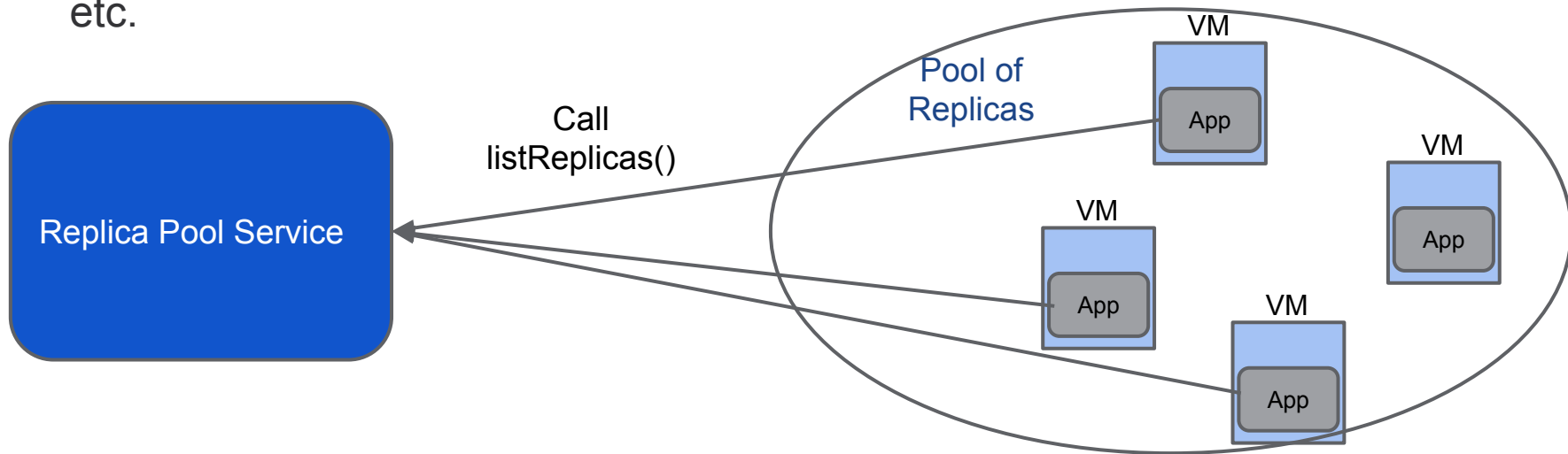
Enabling Cloud-aware cluster setup

Traditional Cluster setup approaches:

- a hard-coded list of nodes supplied in the cluster on initialization
- connect peer-to-peer for discovery (complex configuration for peer connections)
- designating a master for set management
- brittle, complex, not scalable

Replica Pool provides Set membership discovery

- there is now a REST endpoint that lists members (replicas) in a set
- adjusted dynamically as the set grows or shrinks
- Cloud-aware applications can use it to set up clusters, find peers, shard, etc.



Summary

- **Replica Pool** is used to deploy scalable sets of homogenous VMs
- **Deployment Manager** is used to deploy and tie together multiple Replica Pools and other Cloud resources
- Setting up a cluster using Replica Pool allows for easy and robust cluster setup

Give it a try!

- released to Limited Preview at the Google Cloud event in March
- documentation is public, usage requires whitelisting
- <https://developers.google.com/deployment-manager>
- <https://developers.google.com/compute/docs/replica-pool>

Try the sample templates, try writing your own. Send us your feedback!