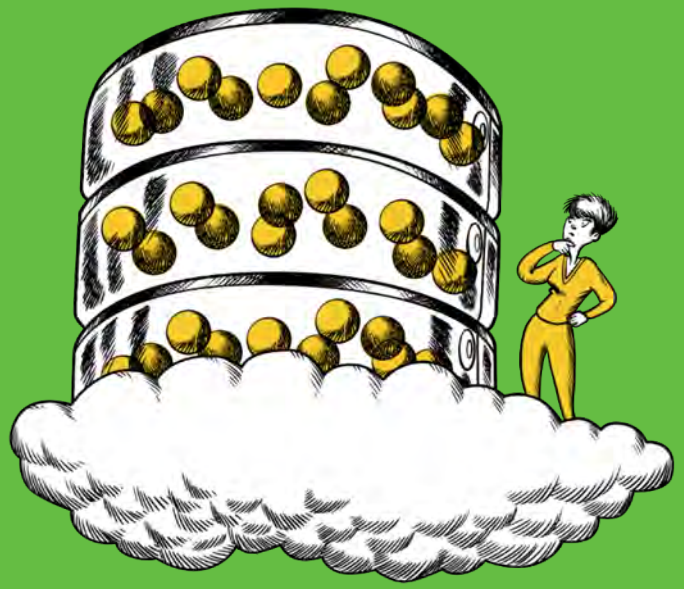
elastifile

Cross-Cloud Distributed File Systems: Bursting File System Dependent Workflows to the Public Cloud

Dr Allon Cohen, VP Products and Business Development, Elastifile

Everybody seems to be bursting their workloads to public clouds

- But, Why??



What the Cloud Offers

- **Elasticity *and* Scalability**
  - > Resources scale *and contract* to match bursty workload requirements
- **Ease of Use**
  - > Simple, intuitive management and universal application compatibility
- **Freedom of Choice**
  - > Data and apps are mobilized to run where you want them
- **Comparative Advantages for Analytics and AI Training**
  - > Cloud offers comparative advantage for certain workloads. Economies of scale, specialized HW, etc.

But, Cloud is a (Wonderfully) Strange Environment

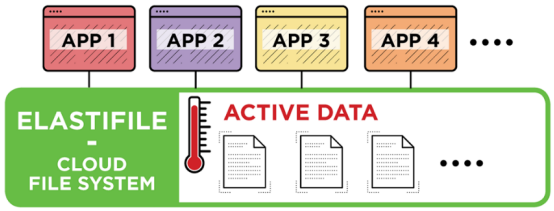
- **Everything *including Infrastructure* is dynamic**
  - > Clusters can change sizes in minutes (not months)
- **Scale seams infinite (from single enterprise perspective)**
  - > As many resources as you need available immediately
- **Underlying hardware is a mystery**
  - > Assume nothing
- **Different failure mechanisms**
  - > Zones, Live Migrations,
    Unpredictable resets by unhuman (AI) admins

So How do You Adapt a Distributed File System to Public Cloud?

You Don't!

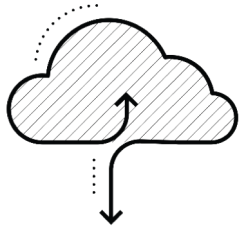
You Design and Build a New One

Optimization Point: Be the Best Data Platform for the Cloud Era Enterprise



Elastifile Cloud File System

Software-defined, Scalable File System



CloudConnect

Delivers Data Mobility Between File Systems and Object Storage

Architectural Base

- Scale Out
  - Support real-time & dynamic reconfiguration of the system
  - Consistent predictable performance (even under failures and noisy environments)
- Software only
  - Avoid unnecessary/uncommon hardware requirements like NVRAM, RDMA networks, etc
- File system core
  - Data level services & the best performance
  - Superset of block/object interfaces
  - Enables data sharing

Examples of Unique Cloud Native Optimizations

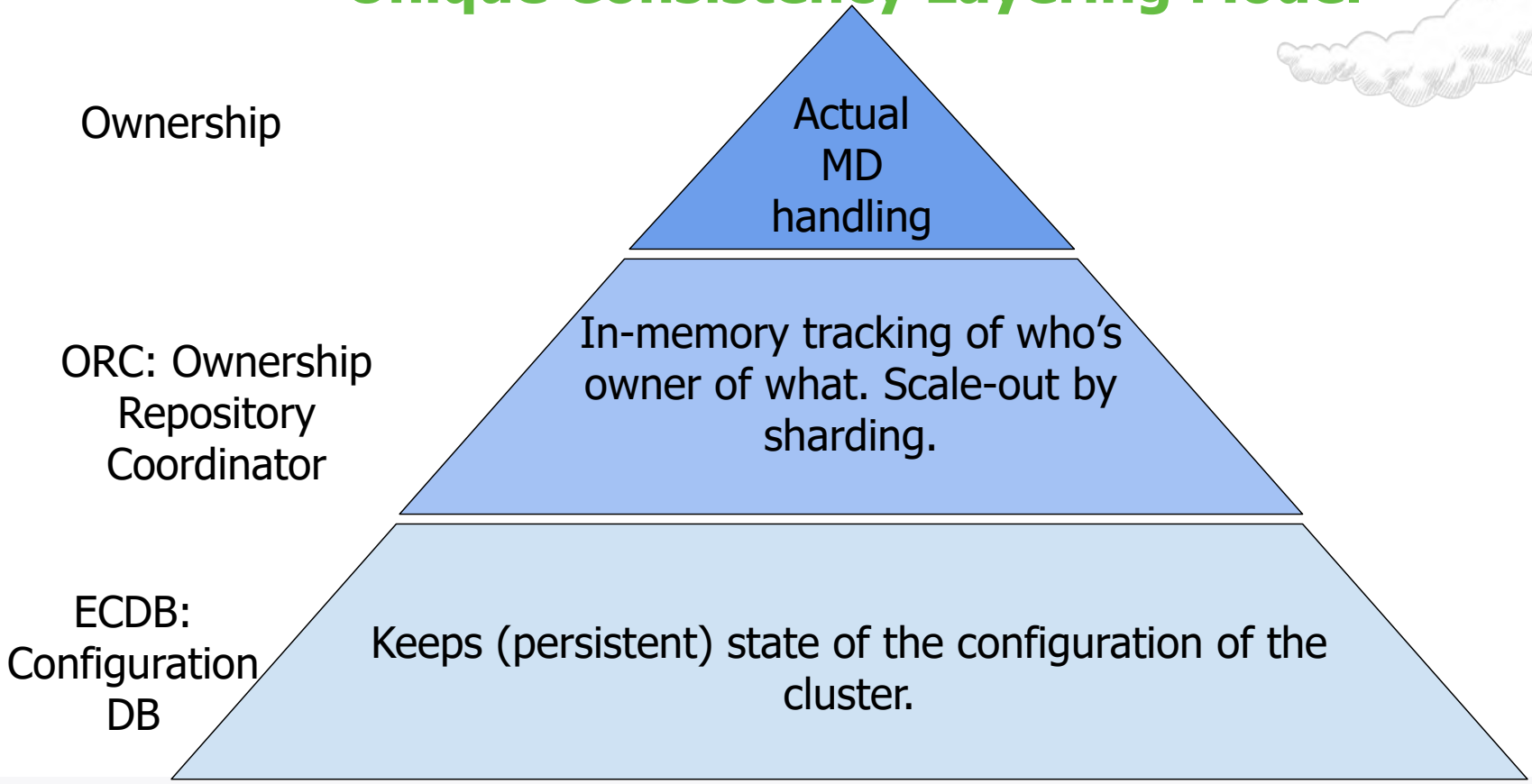
Noisy Environment Adaptations

- Write anywhere
- Hiccup support

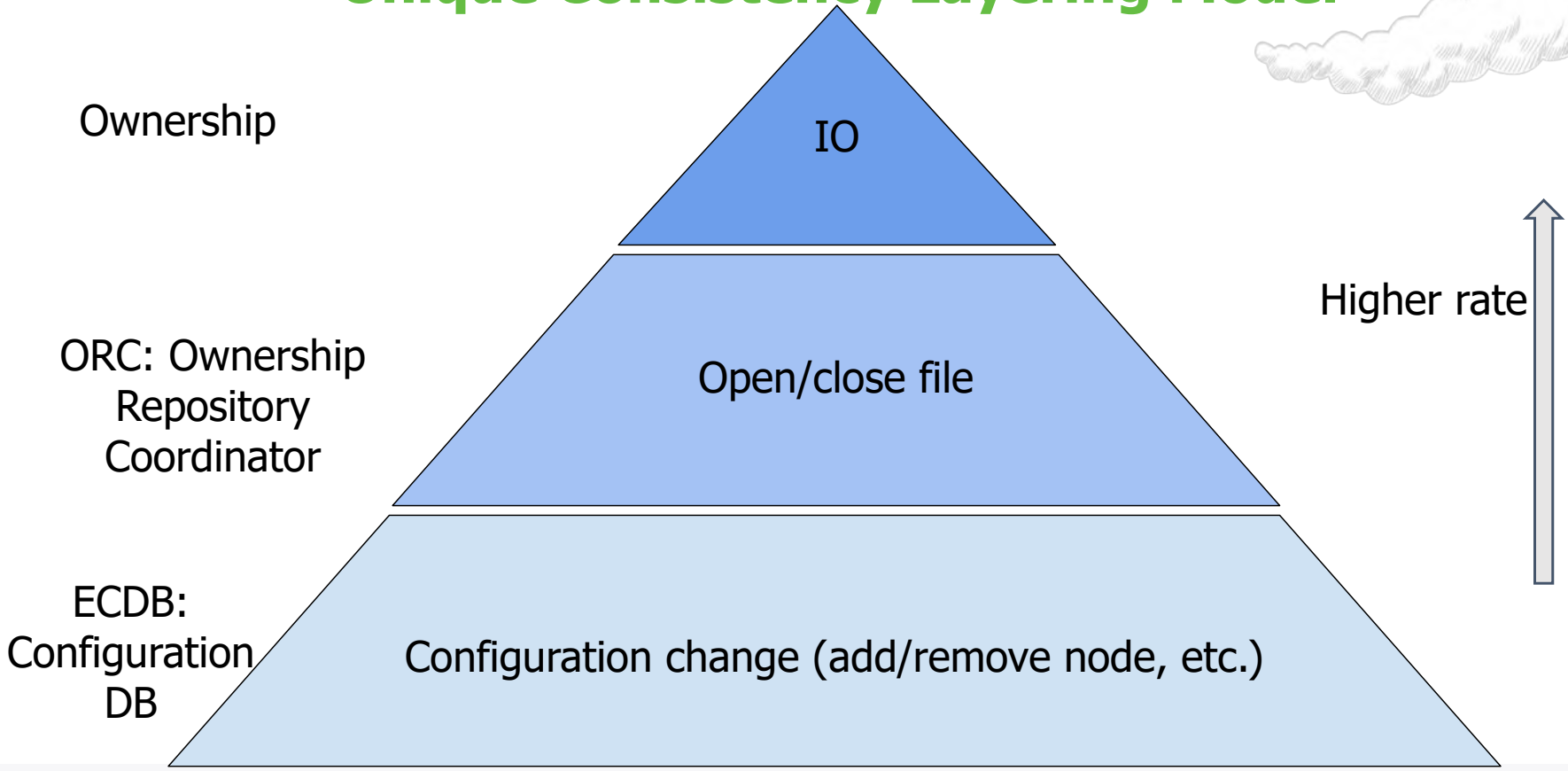
Scalable Metadata

- Scaling metadata is hard, due to the **consistency requirements** of file-systems
- Usually constancy comes at cost of performance
- We're using "consistency layering" to get a good combination of **consistency** and **performance at scale**

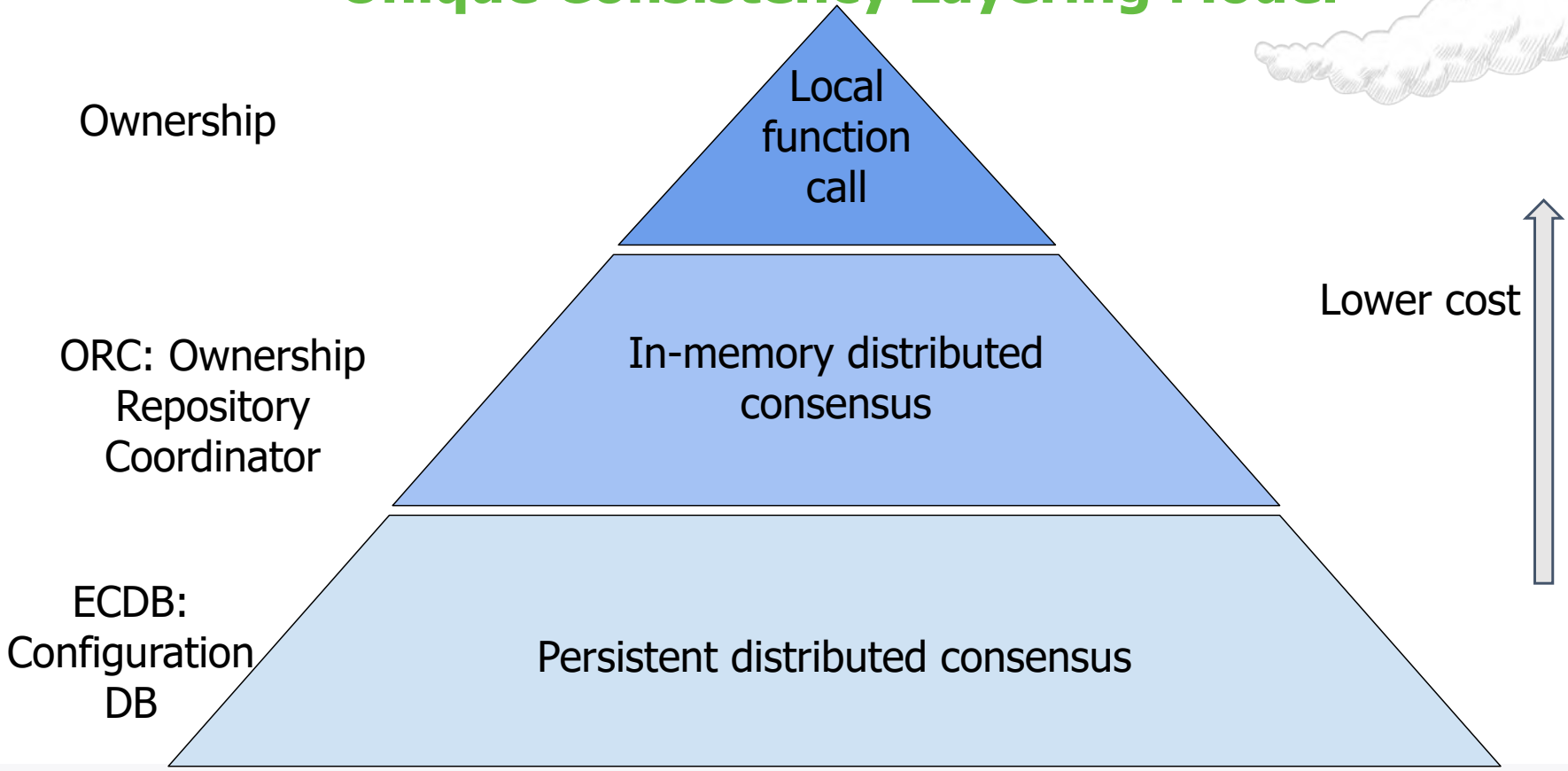
Unique Consistency Layering Model



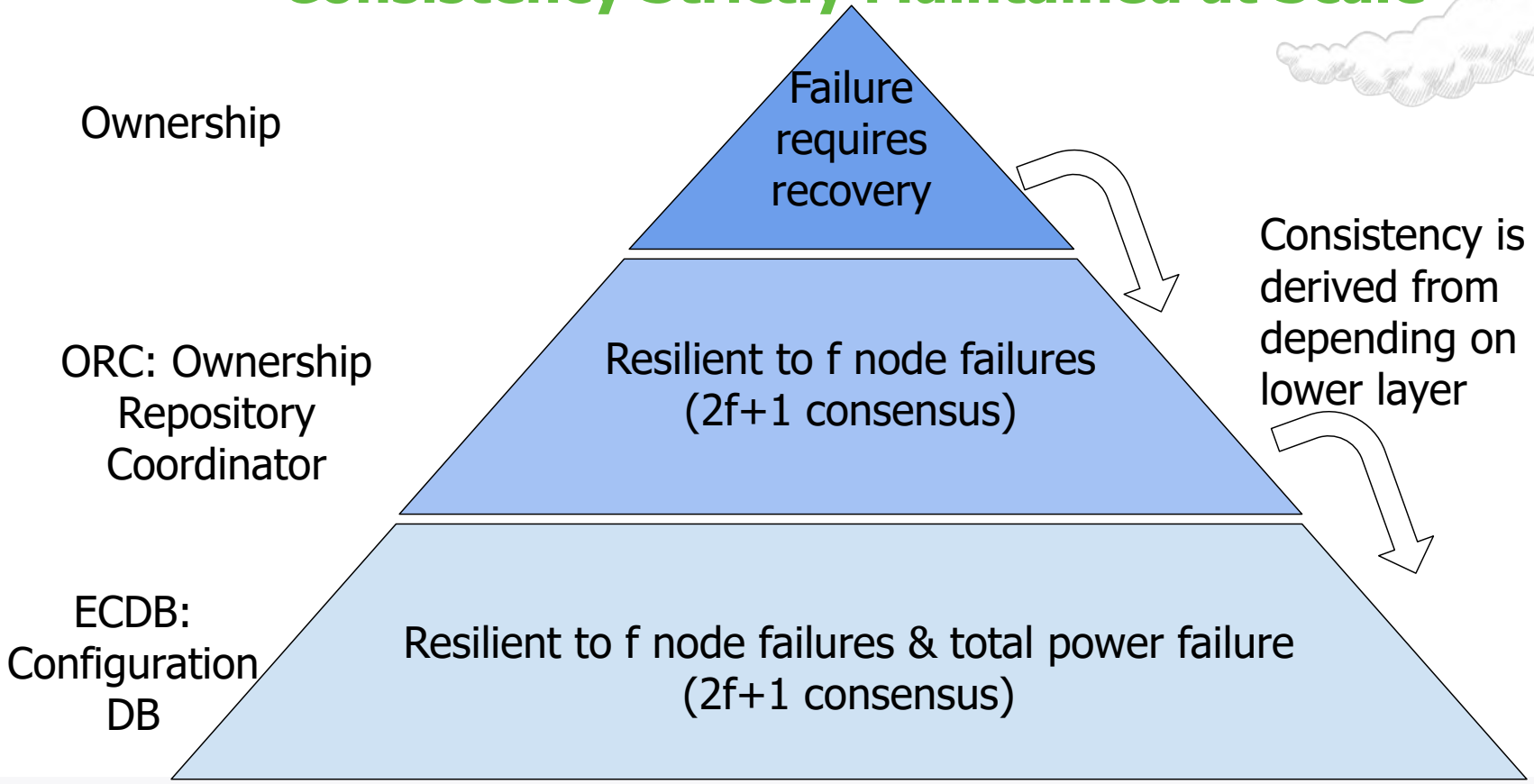
Unique Consistency Layering Model



Unique Consistency Layering Model

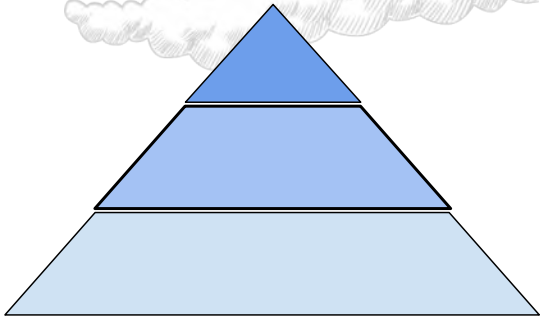


Consistency Strictly Maintained at Scale



Bizur: A Key-value Consensus Algorithm

Specialized consensus algorithm (instead of Paxos / Raft)
Optimized for high concurrency and low latency, especially during failures



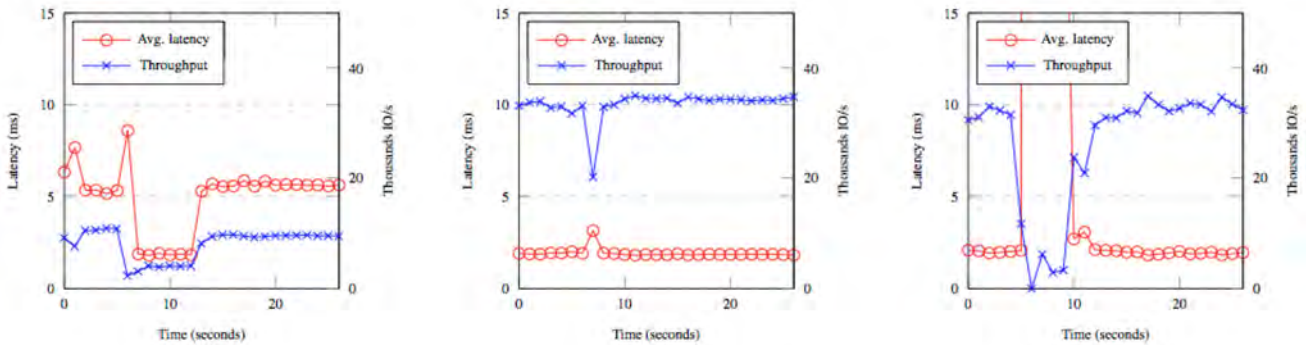
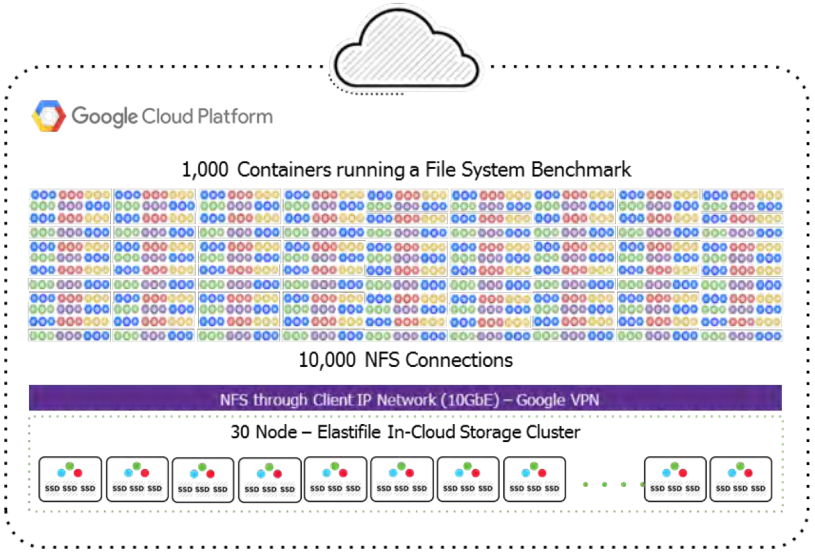


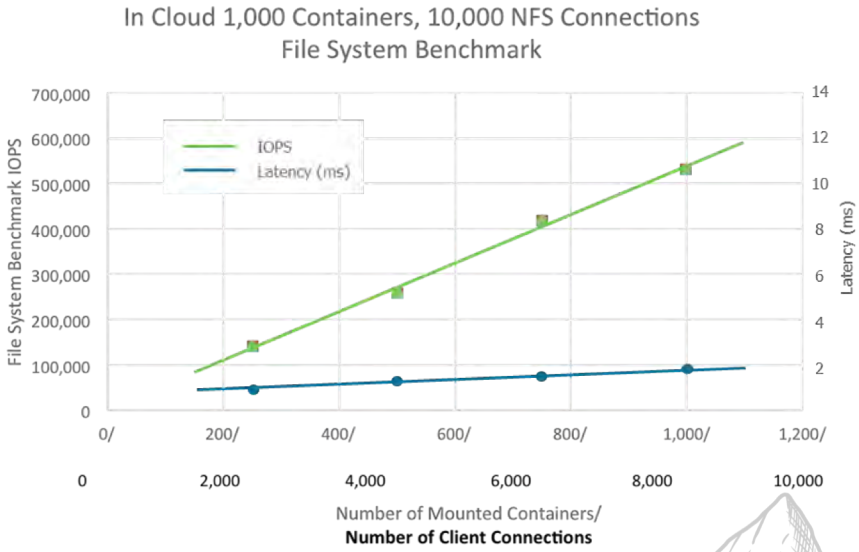
Figure 4: Effect of Leader Failure

Delivering Scalable, High-Performance In-Cloud File Services

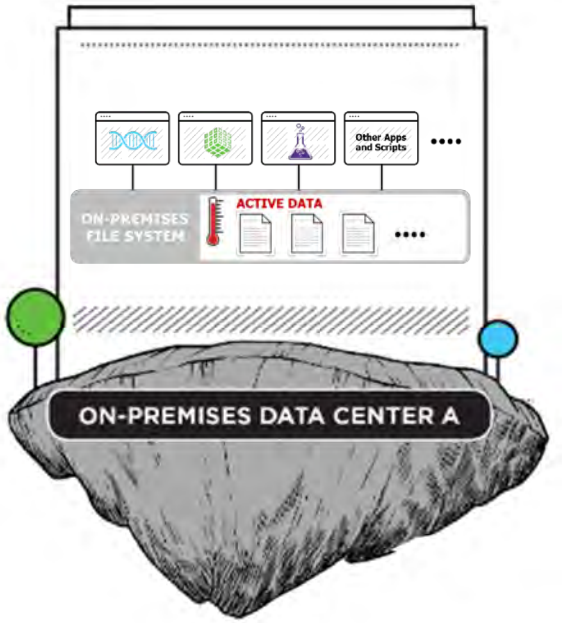
Environment Setup

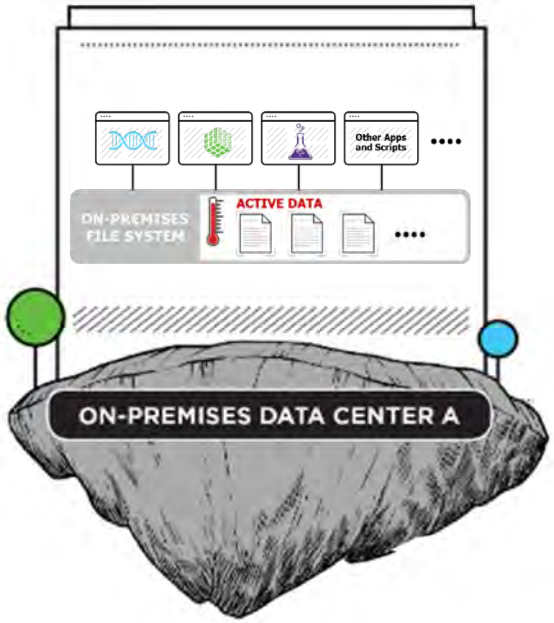


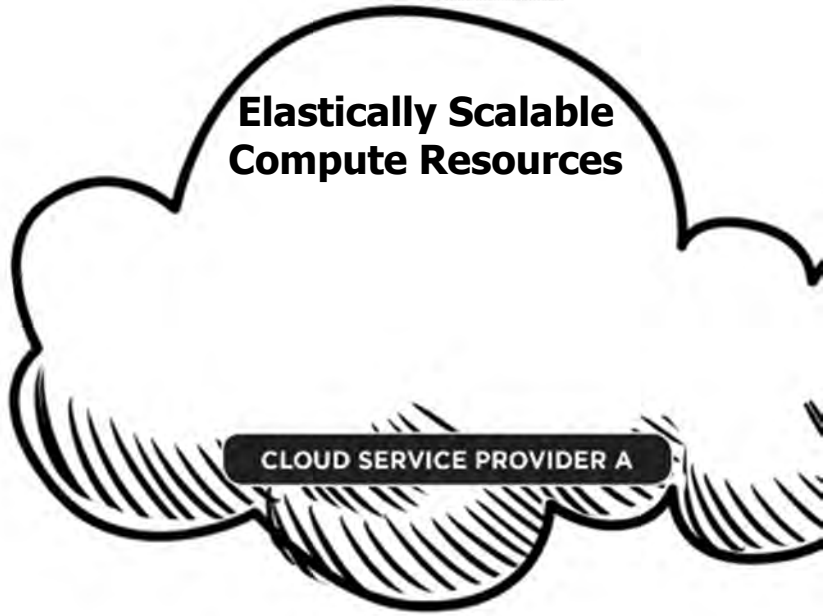
Results

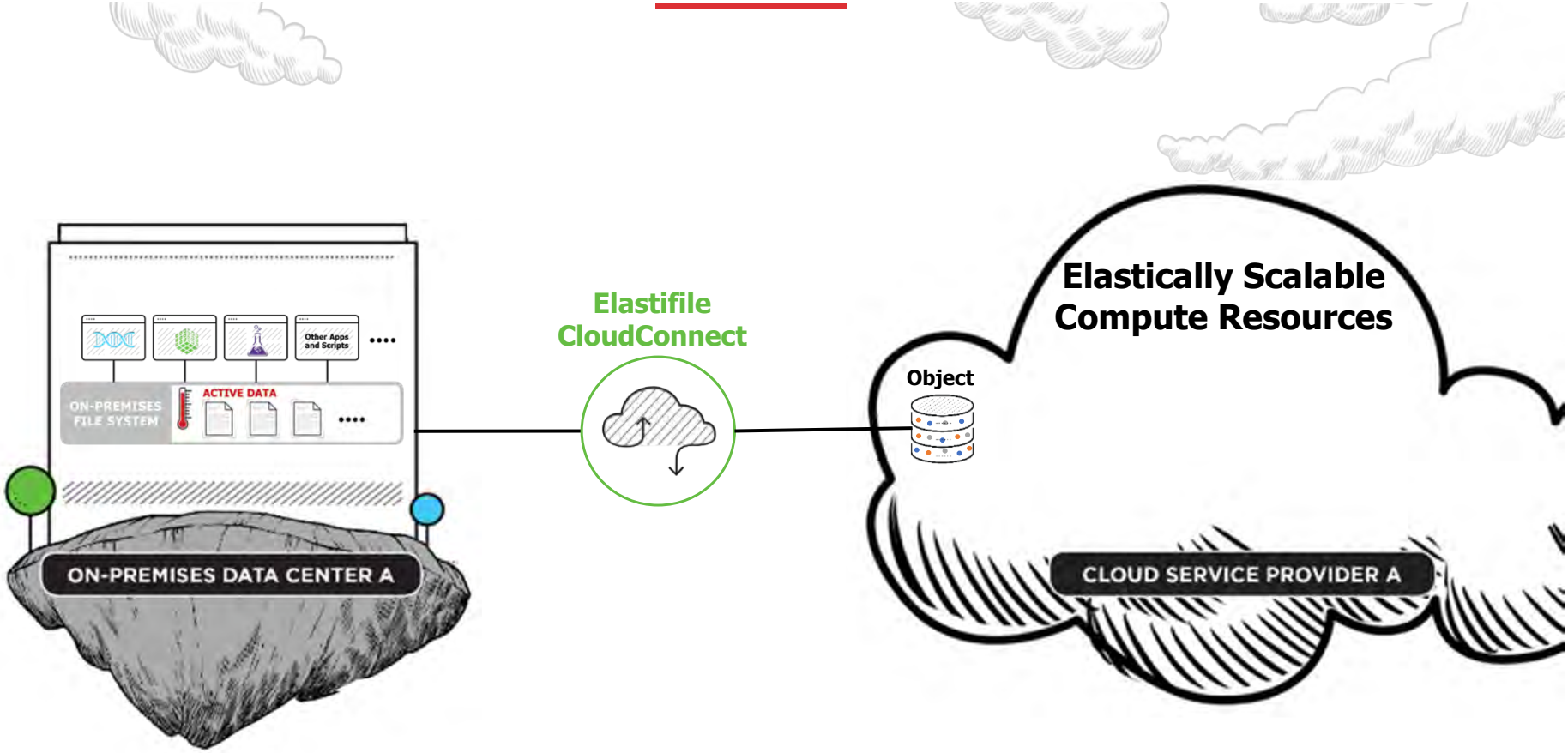


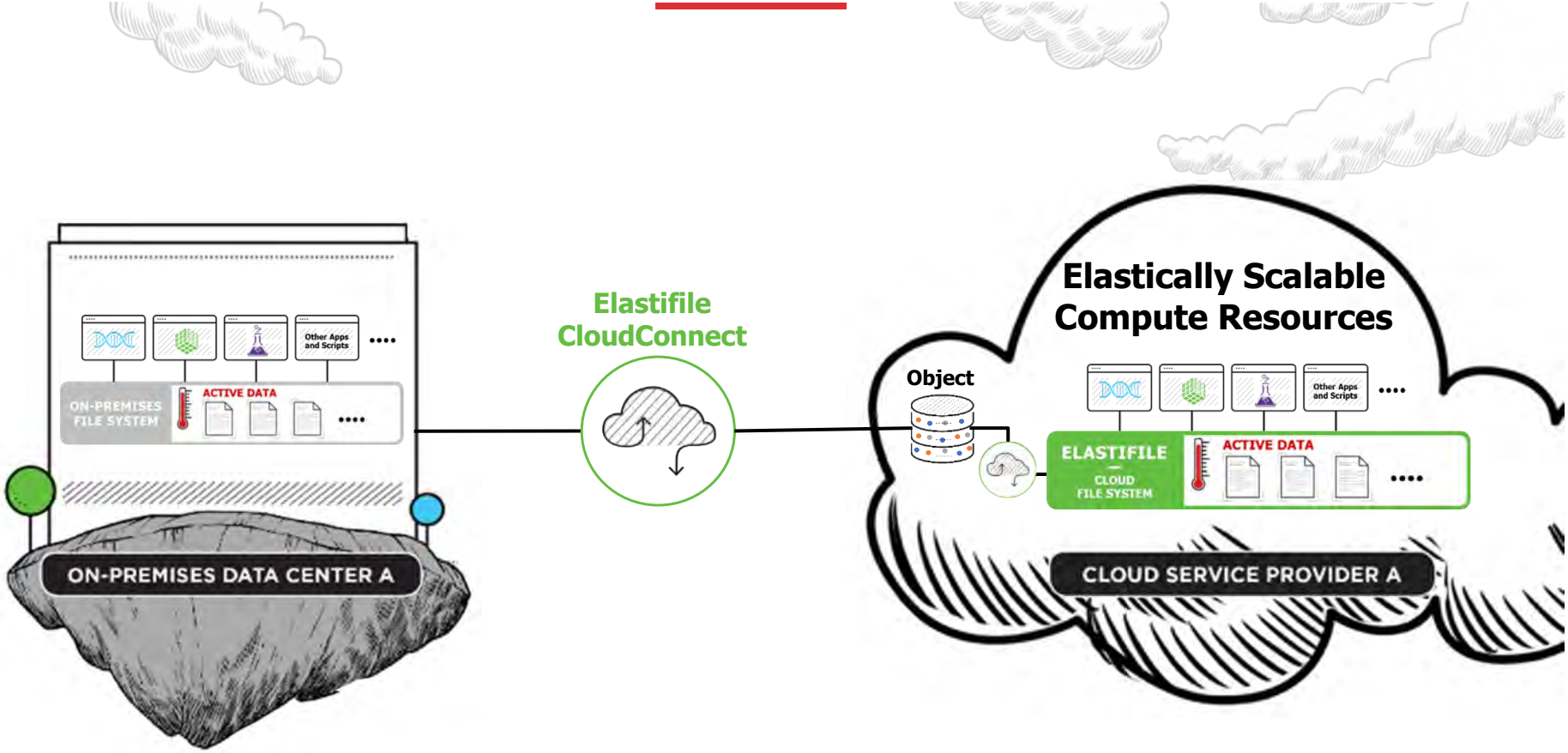
Burst Use Case Examples

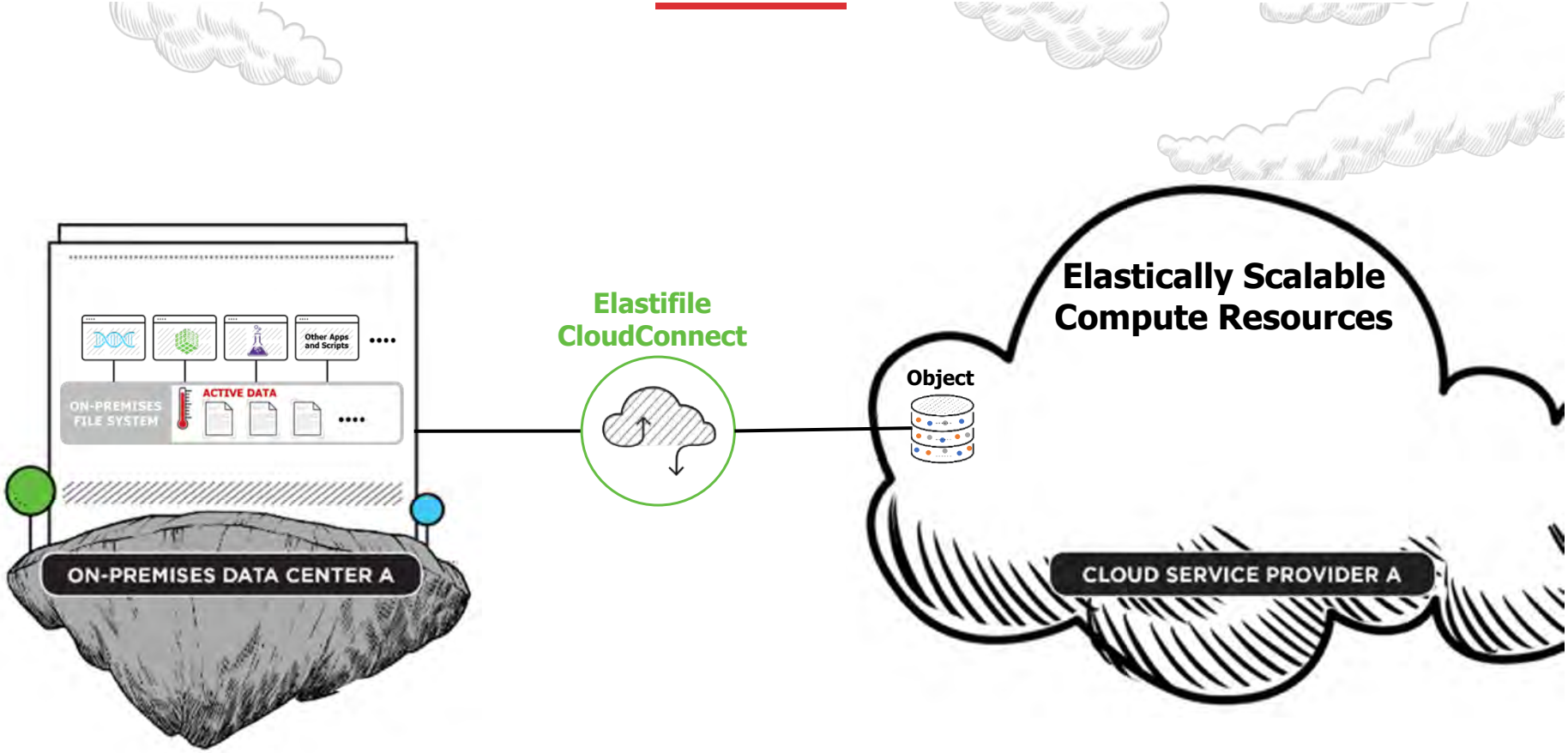

Other Apps and Scripts
ON-PREMISES FILE SYSTEM
ACTIVE DATA
ON-PREMISES DATA CENTER A


Other Apps and Scripts
ON-PREMISES FILE SYSTEM
ACTIVE DATA
ON-PREMISES DATA CENTER A

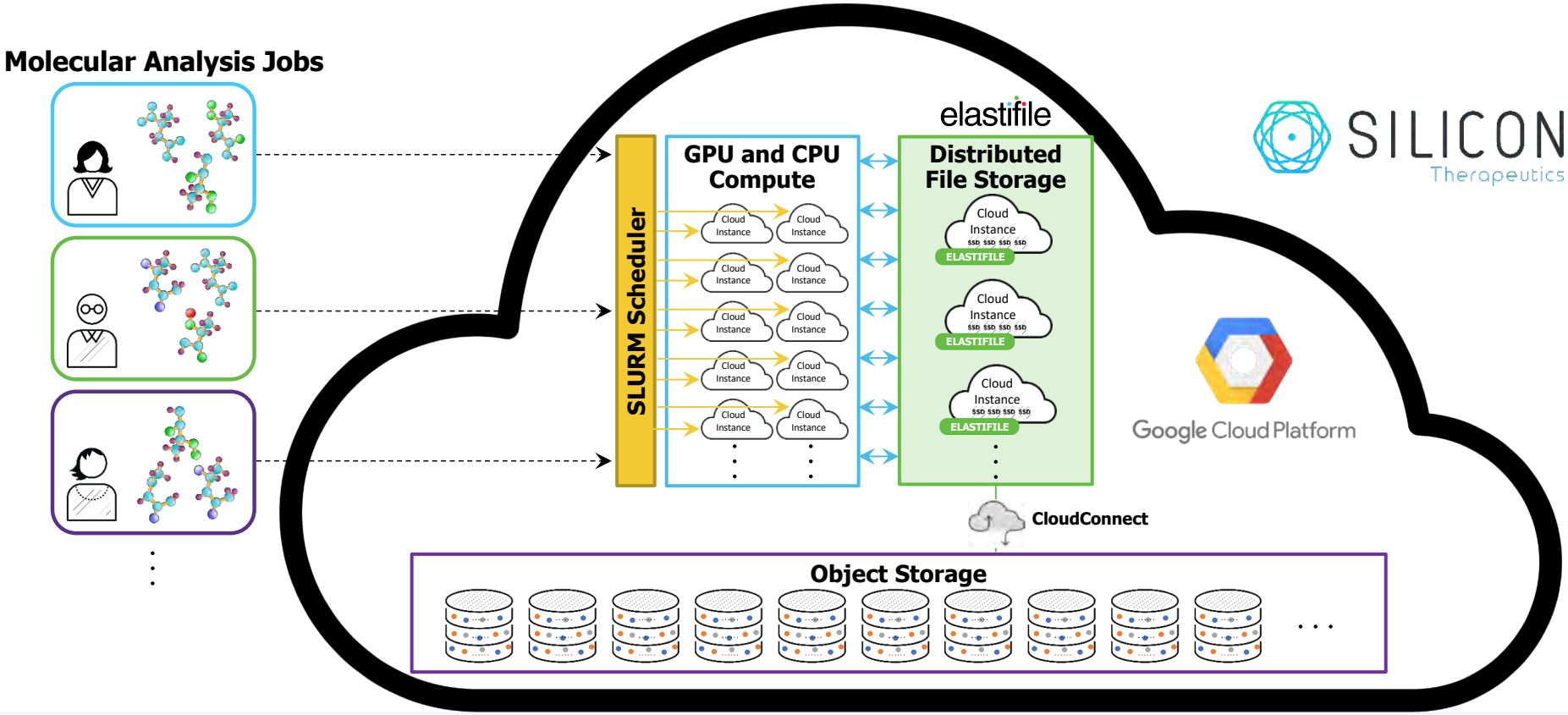

Elastically Scalable Compute Resources
CLOUD SERVICE PROVIDER A

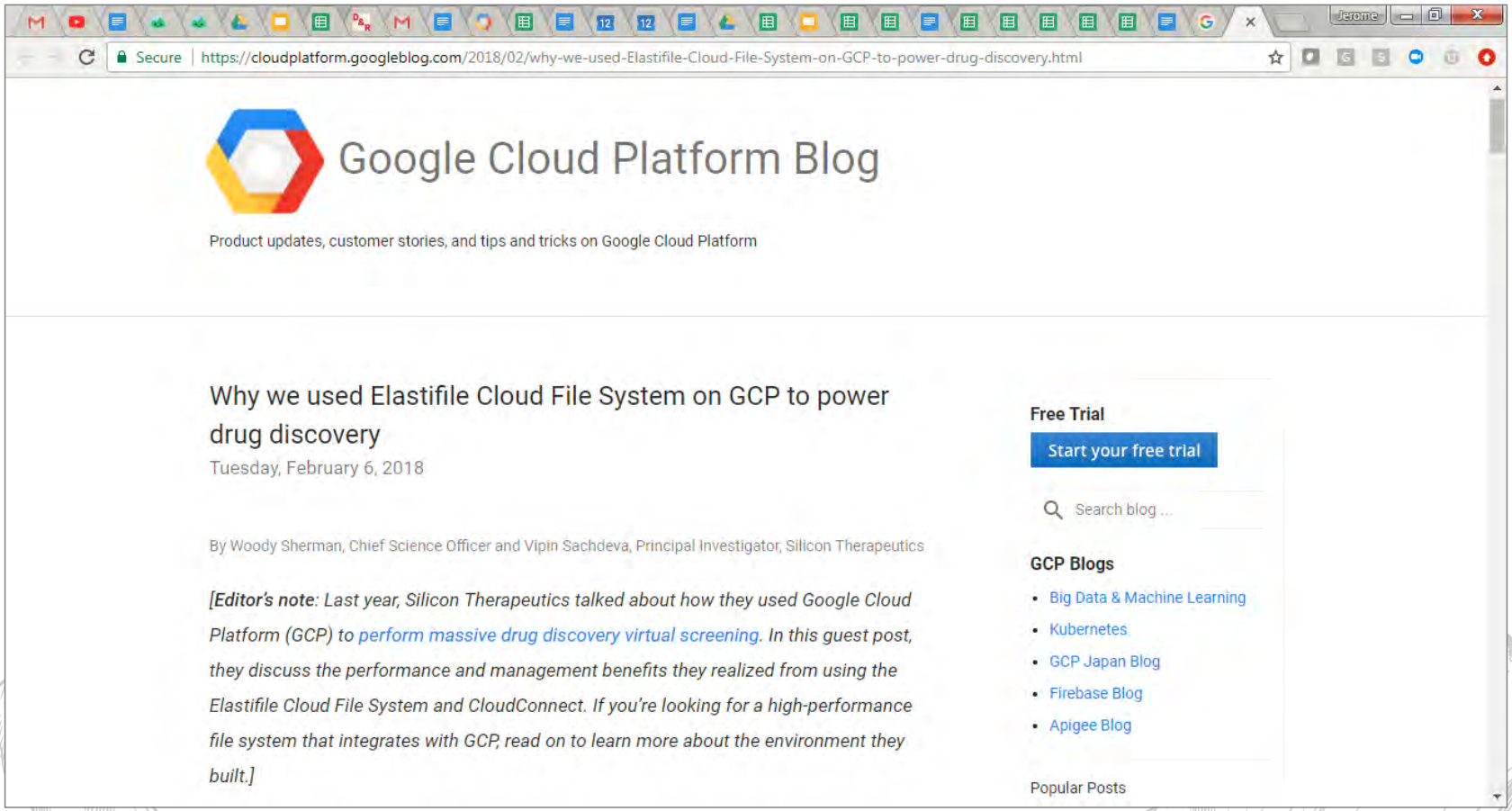

Other Apps and Scripts
ON-PREMISES FILE SYSTEM
ACTIVE DATA
ON-PREMISES DATA CENTER A
Elastifile CloudConnect
Elastically Scalable Compute Resources
Object
CLOUD SERVICE PROVIDER A


Elastifile
CloudConnect
Elastically Scalable
Compute Resources
Object
Other Apps and Scripts
ON-PREMISES FILE SYSTEM
ACTIVE DATA
ON-PREMISES DATA CENTER A
ELASTIFILE
CLOUD FILE SYSTEM
ACTIVE DATA
CLOUD SERVICE PROVIDER A


Other Apps and Scripts
ON-PREMISES FILE SYSTEM
ACTIVE DATA
ON-PREMISES DATA CENTER A
Elastifile CloudConnect
Elastically Scalable Compute Resources
Object
CLOUD SERVICE PROVIDER A

A Real-World, Cloud-Integrated Drug Discovery Solution





elastifile

https://cloudplatform.googleblog.com/2018/02/why-we-used-Elastifile-Cloud-File-System-on-GCP-to-power-drug-discovery.html

Genomics-as-a-Service

A Non-Profit Research Institute Needed to leverage hybrid IT infrastructure

Challenge

Merging the Best Aspects of On-Premises and Public Cloud Infrastructure

Data needs to be durable and accessible across environments

Applications should be capable of running the location that makes sense

Need to align costs with workload requirements

Solution

Hybrid Architecture to Deliver Data Mobility Across Environments

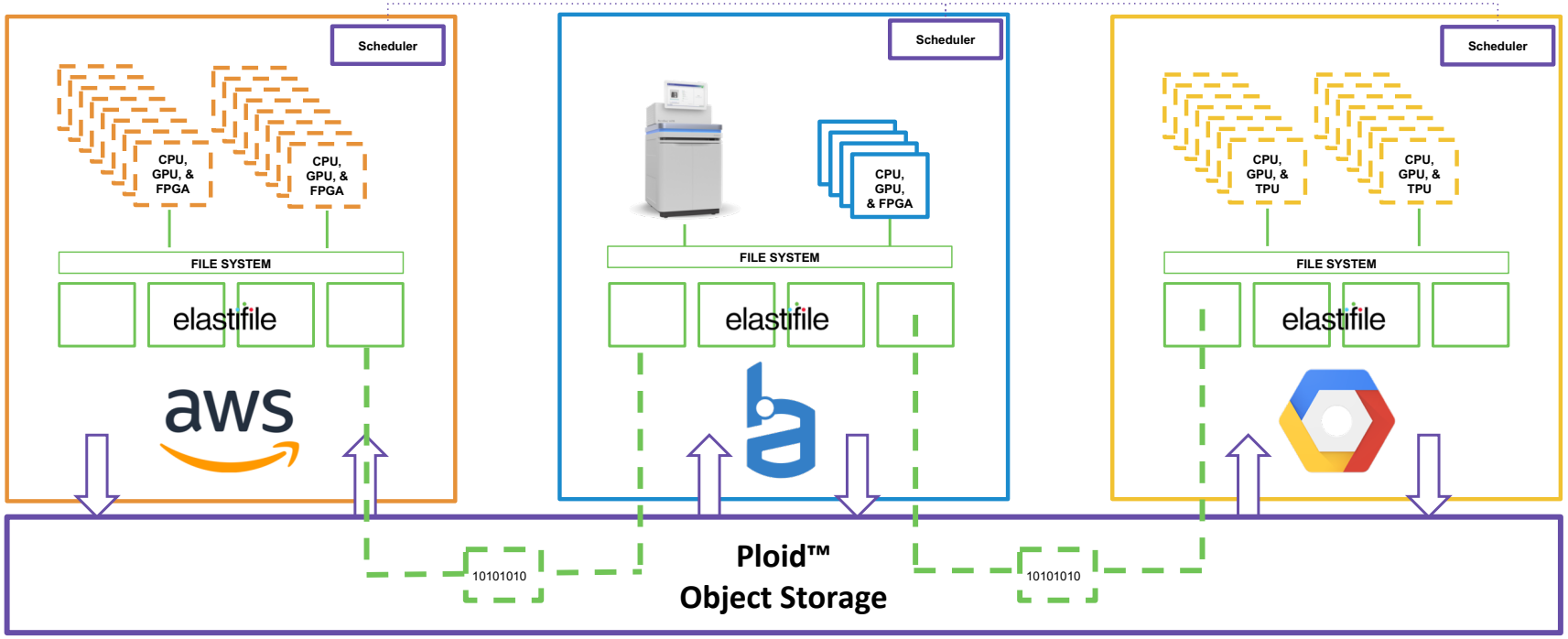
Elastifile's POSIX-compliant file system for a consistent interface on-premises and in-cloud

Elastifile CloudConnect for data transport between on-premises and cloud

Elastifile CloudConnect for data tiering between file storage and low-cost object

GxaaS Architecture Strategy

Mitigate the bad in each cloud … maximize the good … make it look a really big cloud … (seems easy right?)



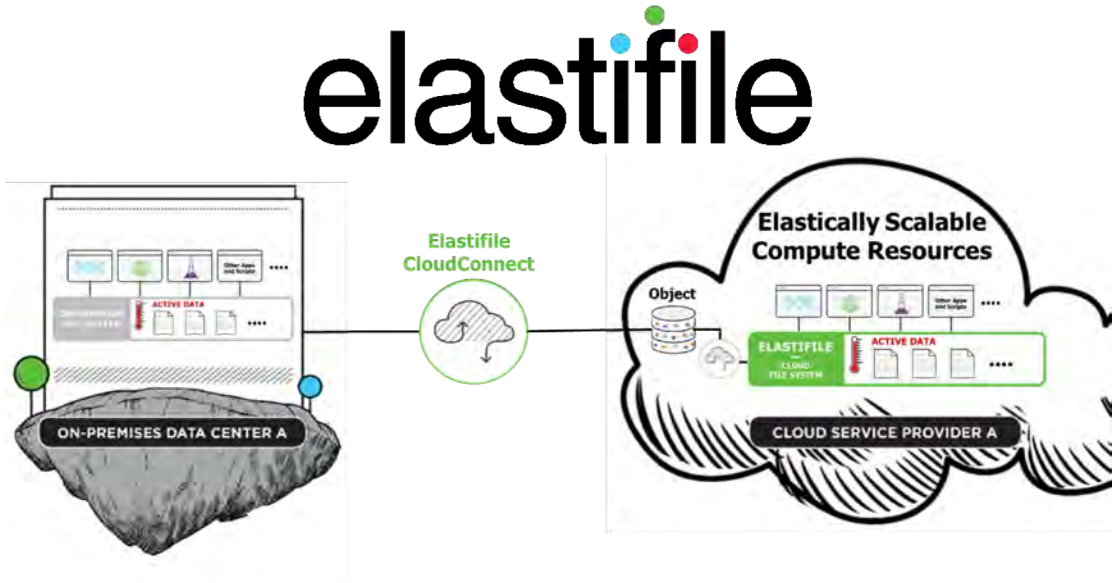
Elastifile: Enabling Cloud-Integrated File Workloads

Burst to cloud to expand resources or offload on-premises infrastructure

Delivering cloud application compatibility

Delivering in-cloud data management

Delivering hybrid cloud data mobility



Free Your Data,
Free Your Business

THANK YOU!

elastifile

Cross-Cloud Data Fabric