
OPEN
Compute Project
OCP U.S. SUMMIT 2017
Santa Clara, CA

Hardware Lifecycle at Scale

Brian Dodds, Craig Ross
Facebook

OPEN HARDWARE. OPEN SOFTWARE. OPEN FUTURE.

OPEN Compute Project

Agenda

1. Facebook’s Infrastructure Evolution
2. Hardware Lifecycle
3. Learnings
4. Wrap Up

Facebook's Infrastructure Evolution

Facebook’s Growth

2010	600M						
2012	1B	Intro		Acquisition			
2014	1.3B	200M		200M		Acquisition	
2016	1.65B	900M		500M		1B	

Facebook’s Scale Today

Each Day:

- Billions of photo and video uploads
- Trillions of user requests
- Tens of trillions of database queries
- 100s of trillions of cache queries

Huge demands on servers, storage, network, and power

Why Build Our Own Hardware?

Advantages

- Faster response to growth demands
- Optimize end-to-end (Application->Power->Thermal)
- Highest Operational Efficiency
- Commodity components

Be Open

The Facebook Datacenter

Infrastructure Evolution

2010

2011

Hardware
Compute

PRN

Open Compute
Project Launch

2012

FRC

2013

Hardware
Storage

LLA

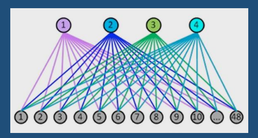
2014

ATN

2015

Hardware
Network

2016



Fabric

FTW, CLN

Hardware Evolution

2010

2011

Compute
Freedom

Rack & Power
Freedom triplet

2012

Compute
Windmill

2013

Compute
Winterfell

Storage
Knox

Rack & Power
Open Rack V1

2014

Rack & Power
Open Rack V2

2015

Compute
Leopard

Storage
Honey Badger

Network Switch
Wedge

Storage
BluRay

2016

Compute
Yosemite

GPU
Big Sur

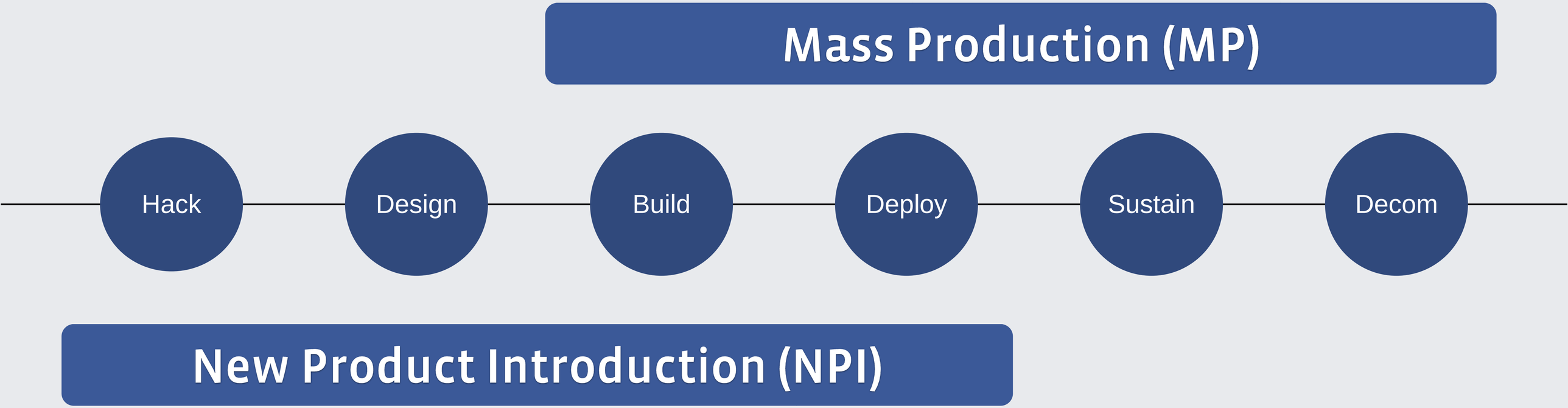
Network
Back Pack

Storage
Lightning

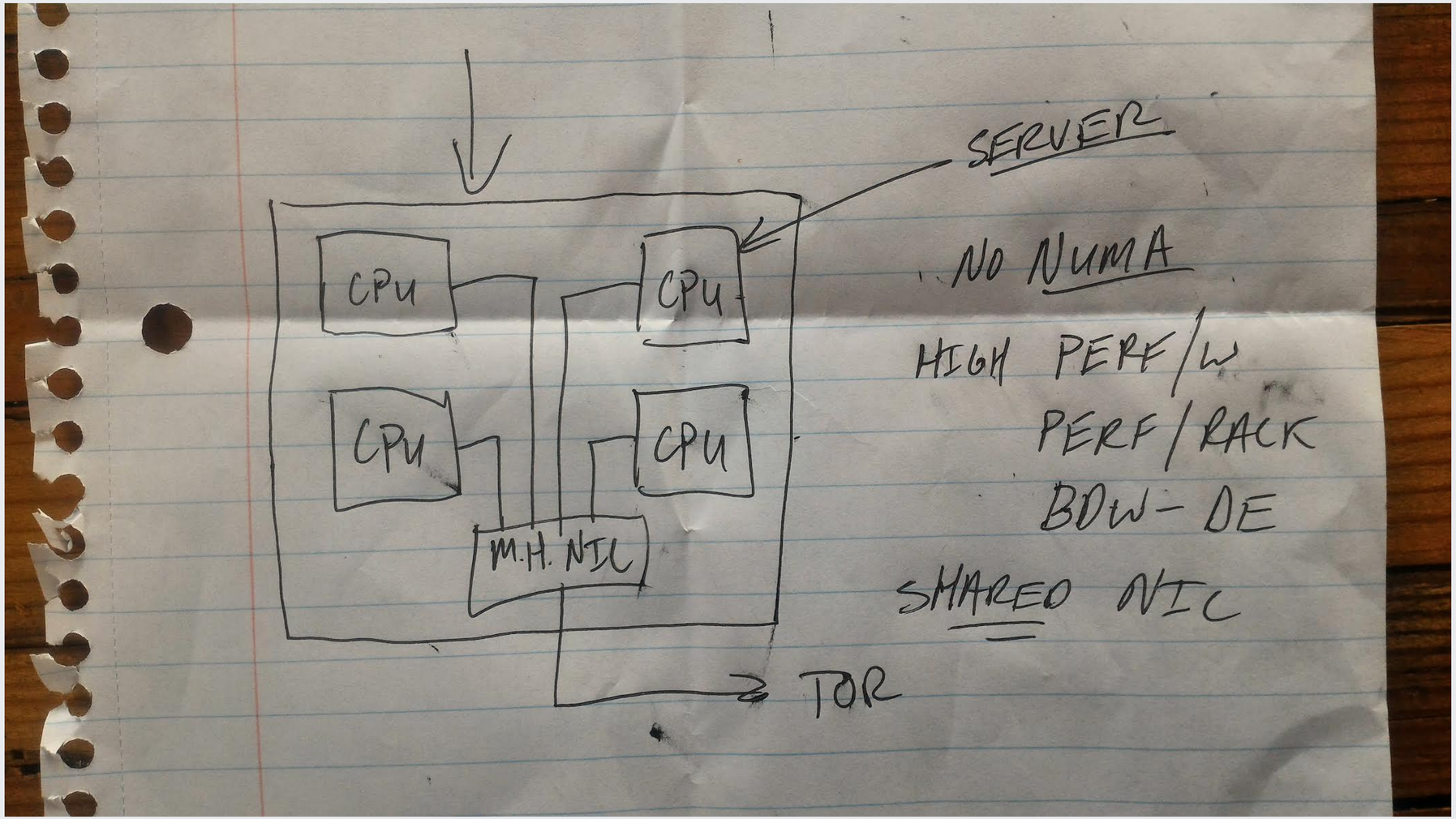
Facebook Datacenters

Hardware Lifecycle

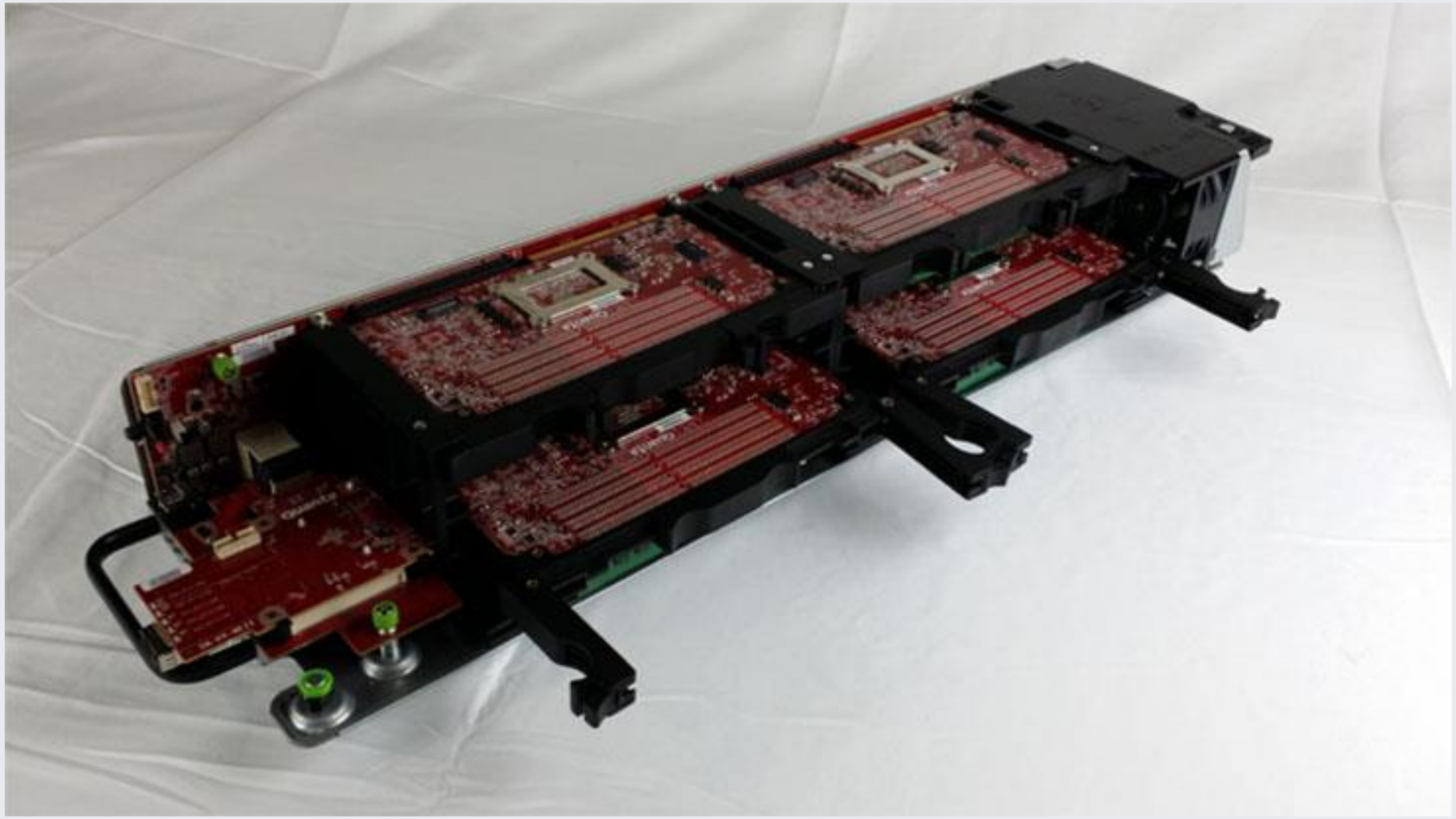
Infrastructure @ Scale


Mass Production (MP)
Hack
Design
Build
Deploy
Sustain
Decom
New Product Introduction (NPI)

Hack | Design | Build | Deploy | Sustain | Decom



Hack | **Design** | Build | Deploy | Sustain | Decom



Hack | **Design** | **Build** | Deploy | Sustain | Decom

EVT

Finalize Hardware Design
Build Systems!

DVT

Full Systems Integration
Build Racks!!

PVT

Deployment Ready
Build Cluster(s)!!!

Pilot

Small Scale Deployment
Deploy Faster!!!!

EVT

Finalize Hardware Design
Build Systems!

DVT

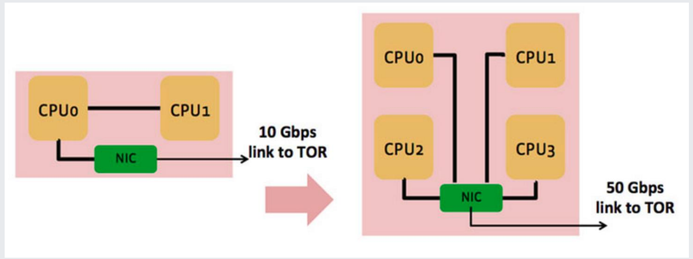
Full Systems Integration
Build Racks!!

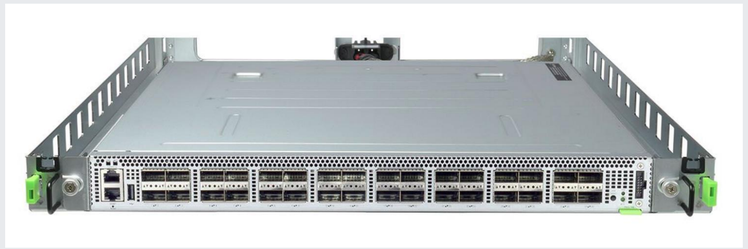
PVT

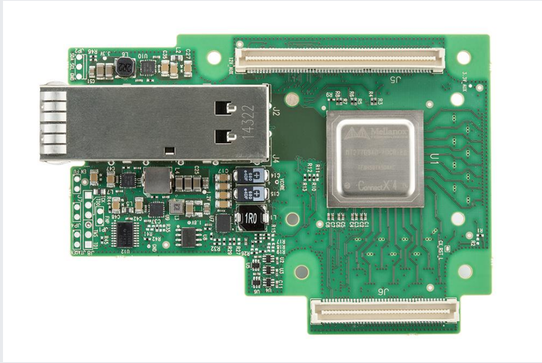
MFG & Deployment Ready
Build Cluster(s)!!!

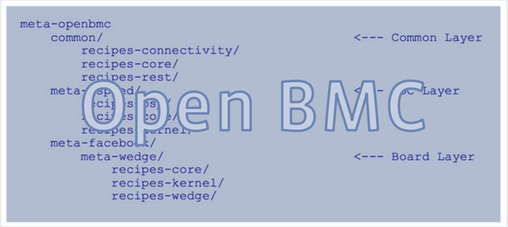
Pilot

Small Scale Deployment
Deploy Faster!!!!

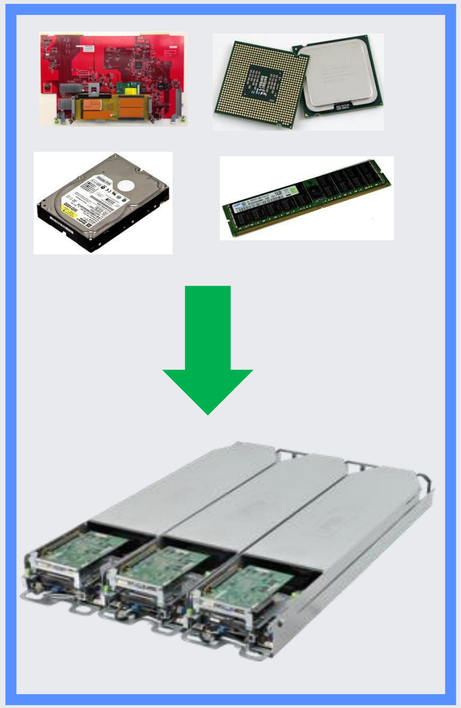




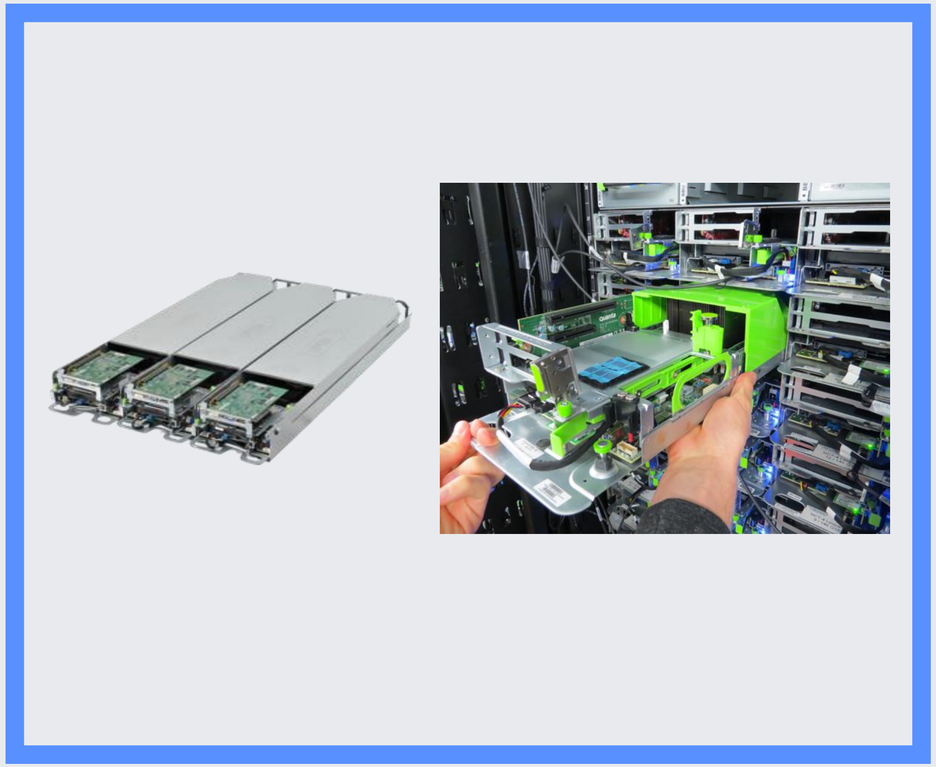




Hack | Design | **Build** | Deploy | Sustain | Decom



Chassis Level Assembly

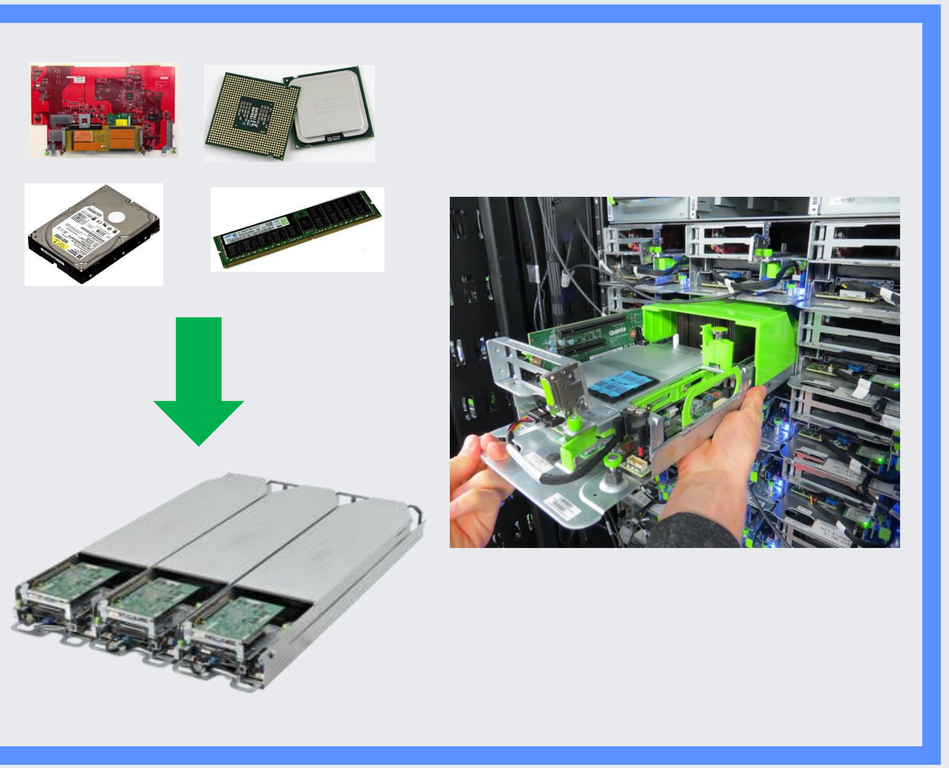


Rack Assembly (in Region)

Data Centers

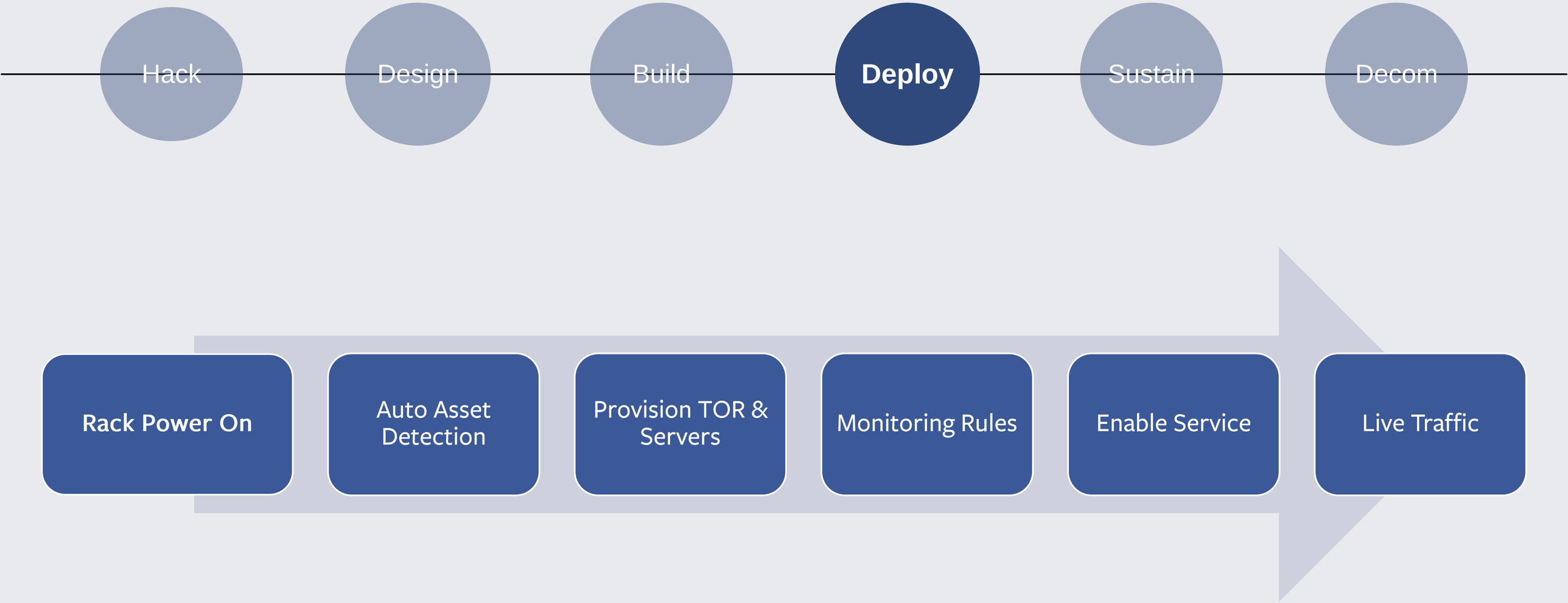
Hack | Design | **Build** | Deploy | Sustain | Decom

Component Level Manufacturing



Chassis + Rack Level Assembly (in Region)

Data Centers


Hack
Design
Build
Deploy
Sustain
Decom
Rack Power On
Auto Asset Detection
Provision TOR & Servers
Monitoring Rules
Enable Service
Live Traffic

Yosemite Deployment

SW Load Balancer

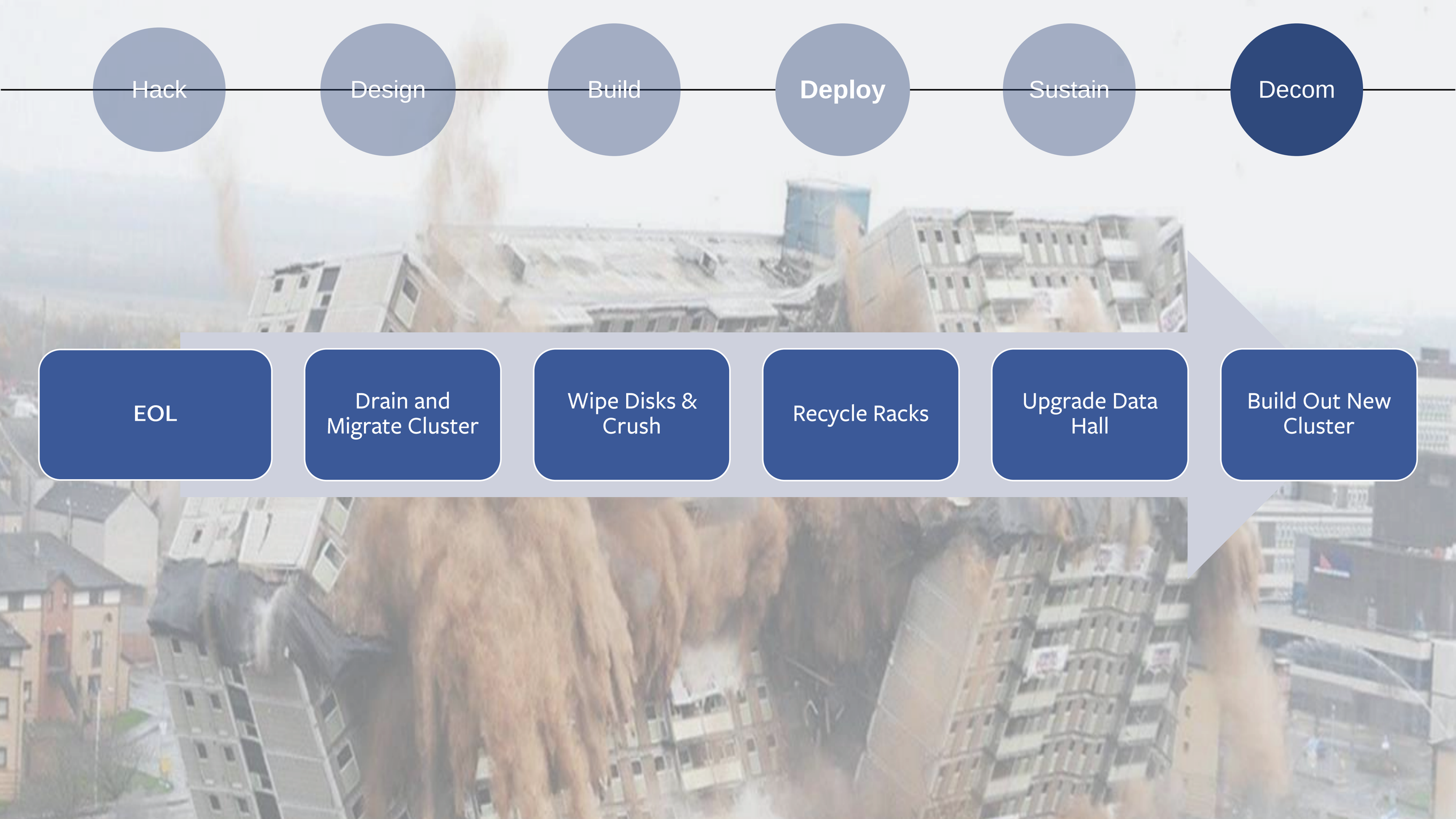
1S vs 2S Tuning

Perf Variations

Hack | Design | Build | **Deploy** | Sustain | Decom

Ensure smooth operation after deployment

Yosemite:
OpenBMC – OOM!


Hack
Design
Build
Deploy
Sustain
Decom
EOL
Drain and Migrate Cluster
Wipe Disks & Crush
Recycle Racks
Upgrade Data Hall
Build Out New Cluster

Learnings

Hardware Evolution

2010

2011

Compute
Freedom

Rack & Power
Freedom triplet

2012

Compute
Windmill

2013

Compute
Winterfell

Storage
Knox

Rack & Power
Open Rack V1

2014

Rack & Power
Open Rack V2

2015

Compute
Leopard

Storage
Honey Badger

Network Switch
Wedge

Storage
BluRay

2016

Compute
Yosemite

GPU
Big Sur

Network
Six Pack

Storage
Lightning

Learnings - Sensors

2010

2011

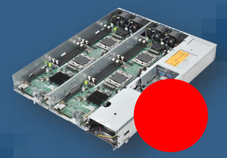
Compute
Freedom

Rack & Power
Freedom triplet

Issues: BMC and PSU monitoring woes

Learnings: Improve monitoring of critical sensors.

2012



Compute
Windmill

2013

Compute
Winterfell

Storage
Knox

Rack & Power
Open Rack V1

2014

Rack & Power
Open Rack V2

2015

Compute
Leopard

Storage
Honey Badger

Network Switch
Wedge

Storage
BluRay

2016

Compute
Yosemite

GPU
Big Sur

Network
Six Pack

Storage
Lightning

Learnings – Supply Chain/Application

2010

2011

Compute
Freedom

Rack & Power
Freedom triplet

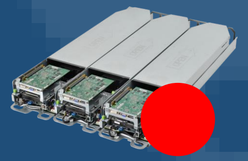
Issues: Single-sourced epidemic failure. App performance issues. Row Hammer.

Learnings: Multi-source components, robust app testing @ scale, improve component monitoring.

2012

Compute
Windmill

2013



Compute
Winterfell

Storage
Knox

Rack & Power
Open Rack V1

2014

Rack & Power
Open Rack V2

2015

Compute
Leopard

Storage
Honey Badger

Network Switch
Wedge

Storage
BluRay

2016

Compute
Yosemite

GPU
Big Sur

Network
Six Pack

Storage
Lightning

Learnings – DC Tooling

2010

2011

Compute
Freedom

Rack & Power
Freedom triplet

2012

Compute
Windmill

2013

Compute
Winterfell

Storage
Knox

Rack & Power
Open Rack V1

2014

Rack & Power
Open Rack V2

2015

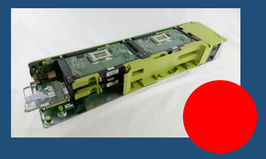
Compute
Leopard

Storage
Honey Badger

Network Switch
Wedge

Storage
BluRay

2016



Compute
Yosemite

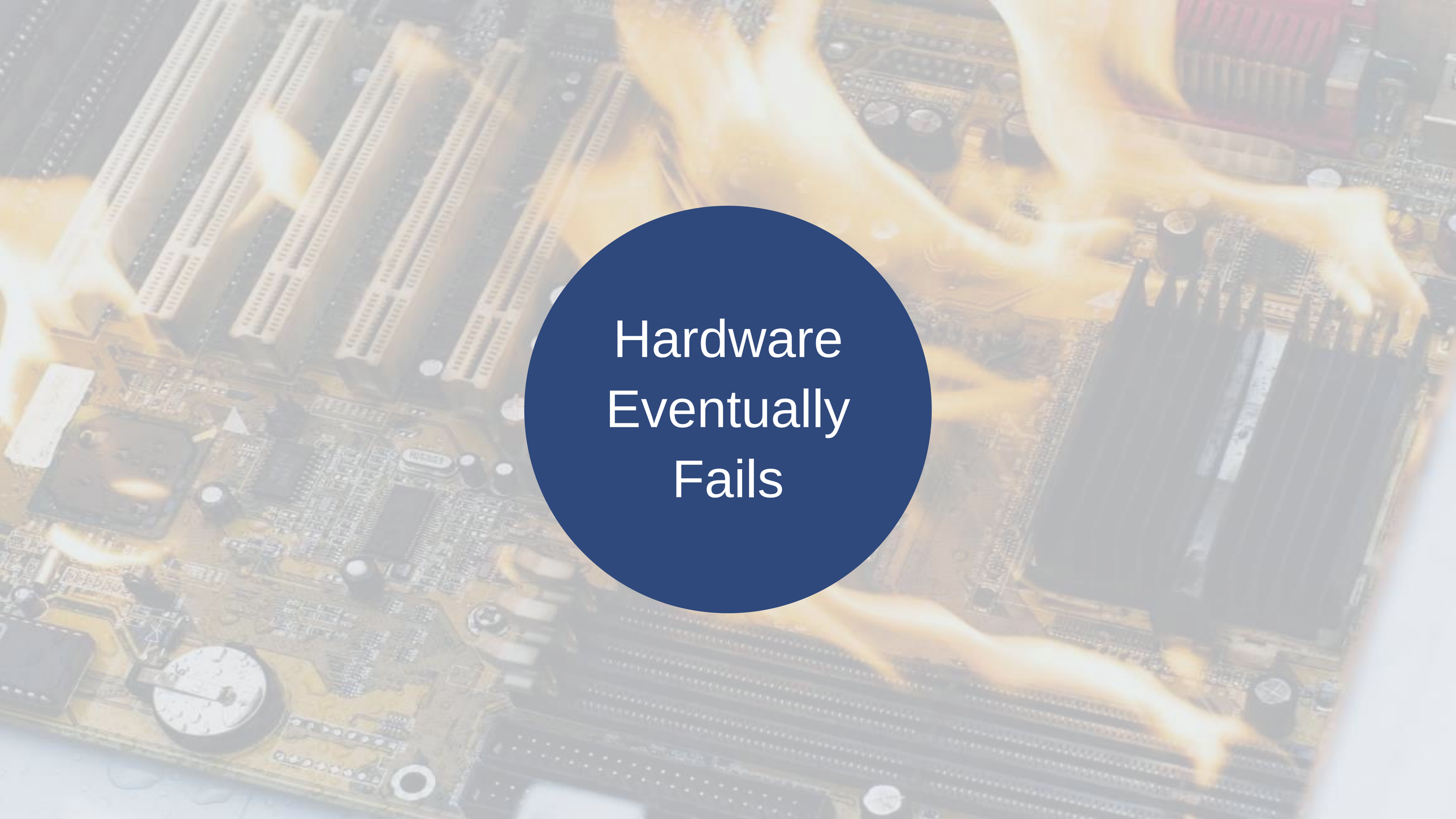
GPU
Big Sur

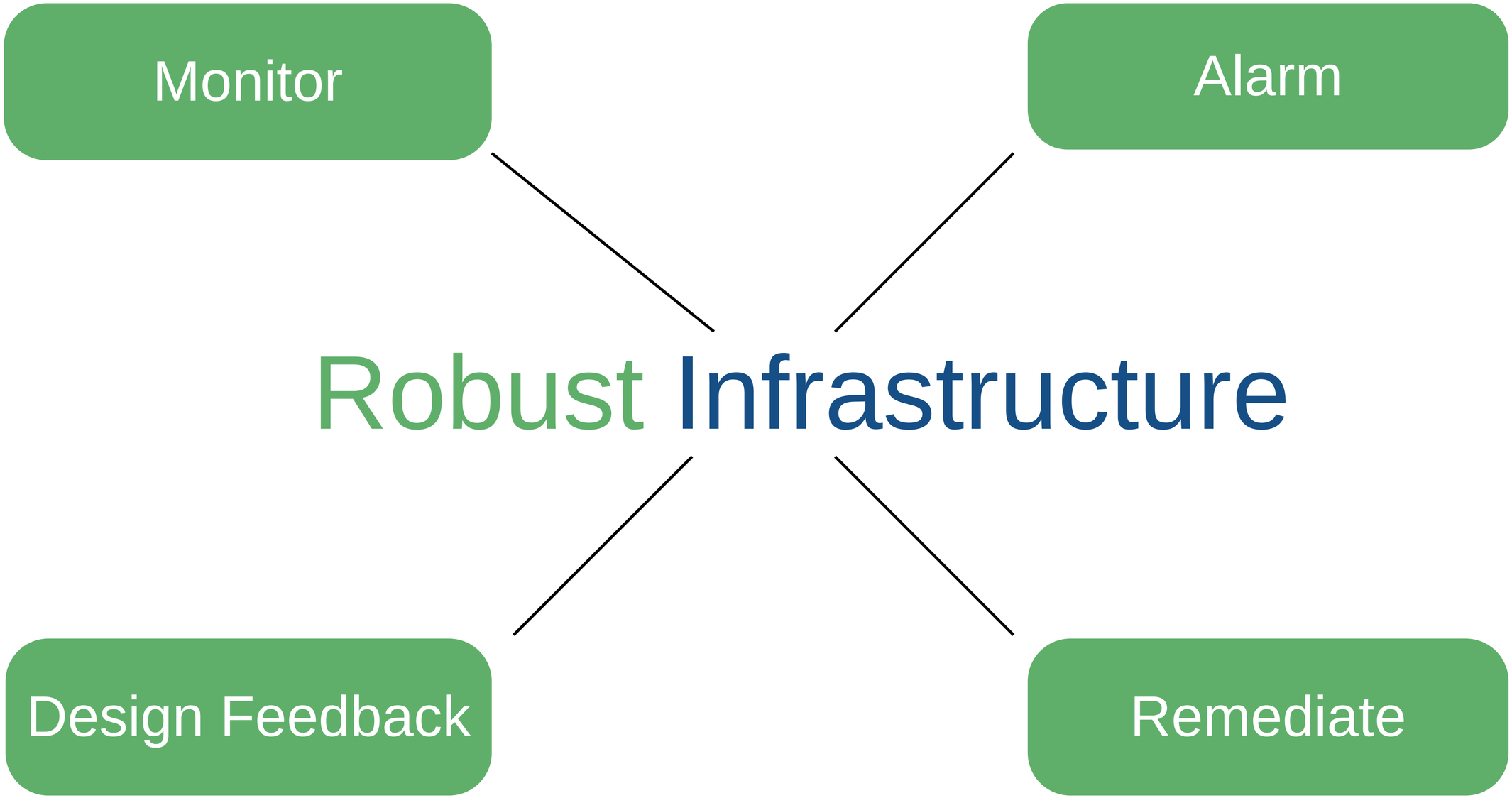
Network
Six Pack

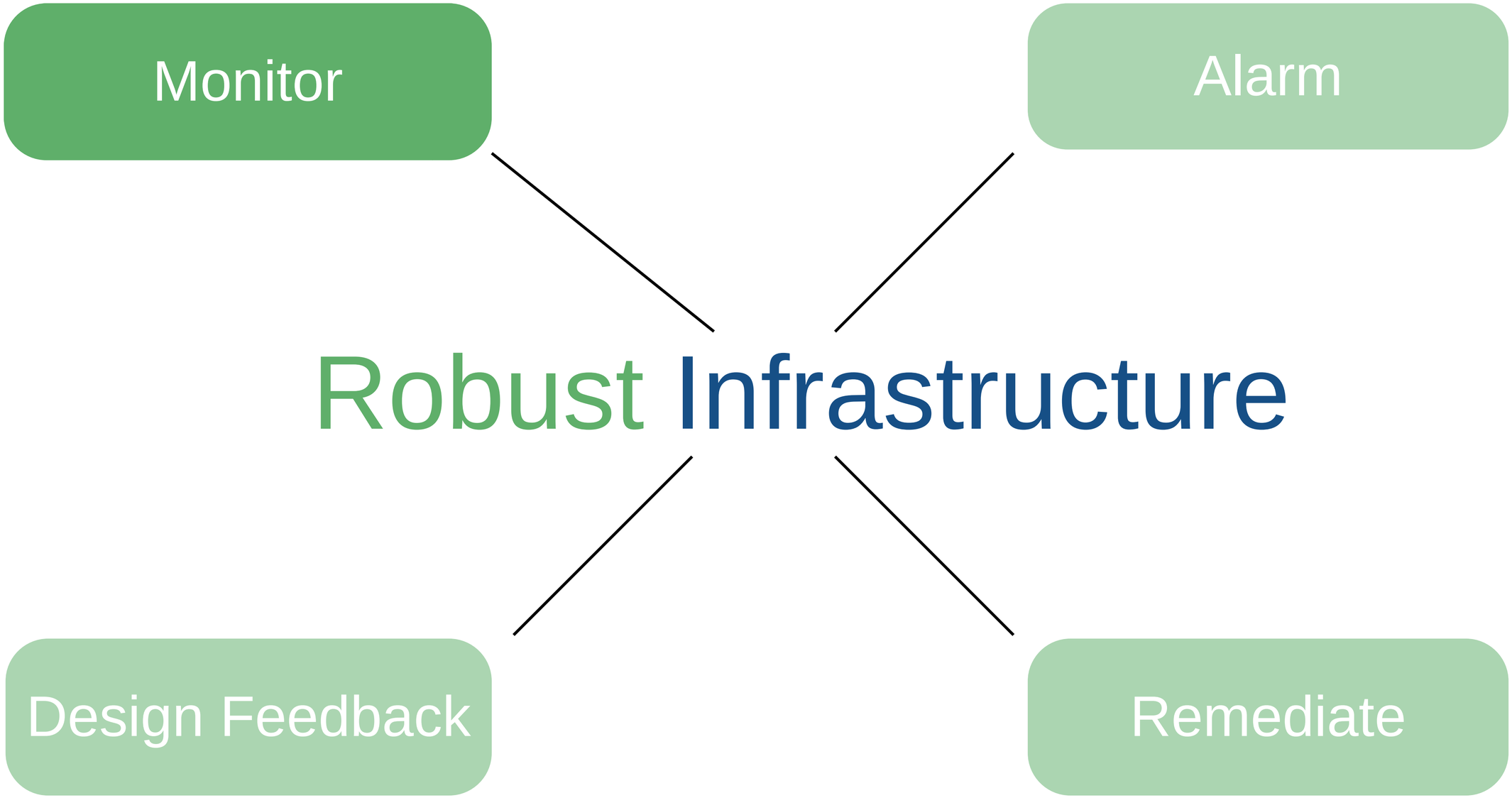
Storage
Lightning

Issues: Shipped hardware before all tooling was finished – Idle HW.

Learnings: Make tooling a first-class citizen for phase exit.

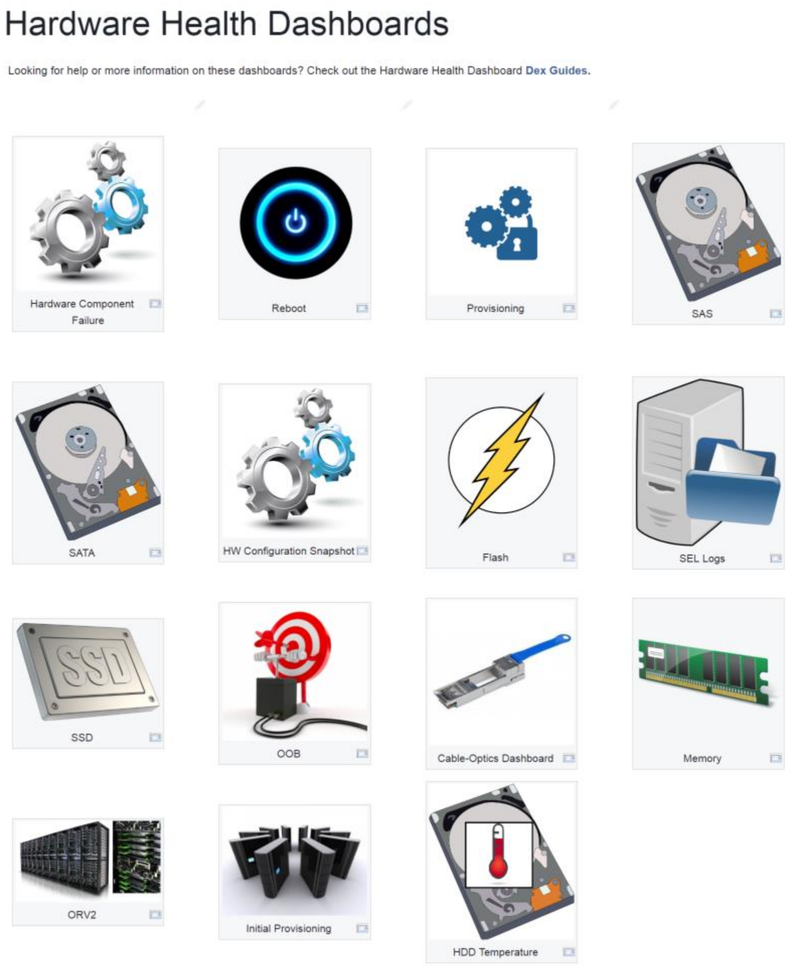

Hardware
Eventually
Fails


Monitor
Alarm
Robust Infrastructure
Design Feedback
Remediate


Monitor
Alarm
Robust Infrastructure
Design Feedback
Remediate

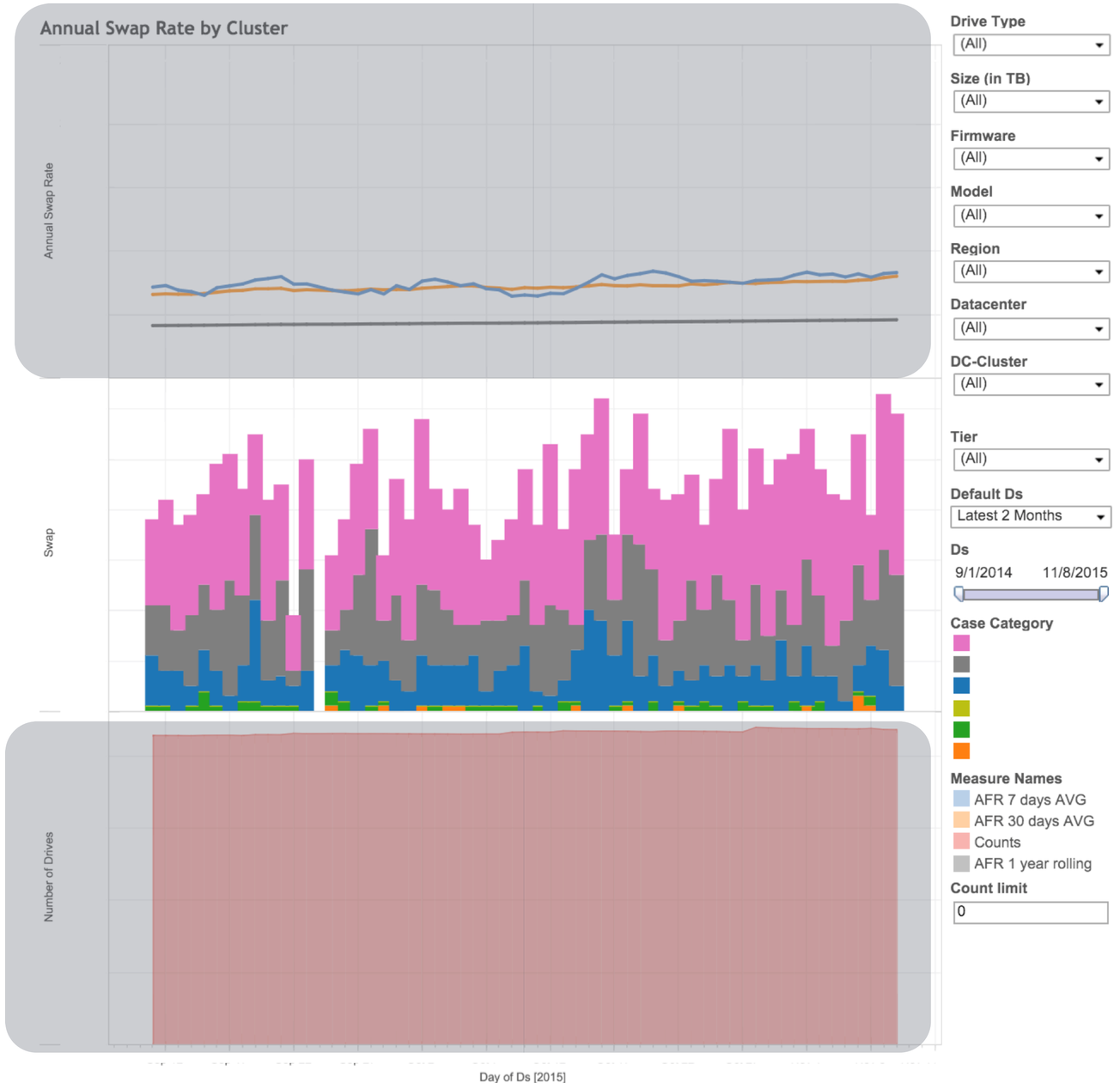
Monitoring

Many servers, components, services, and regions



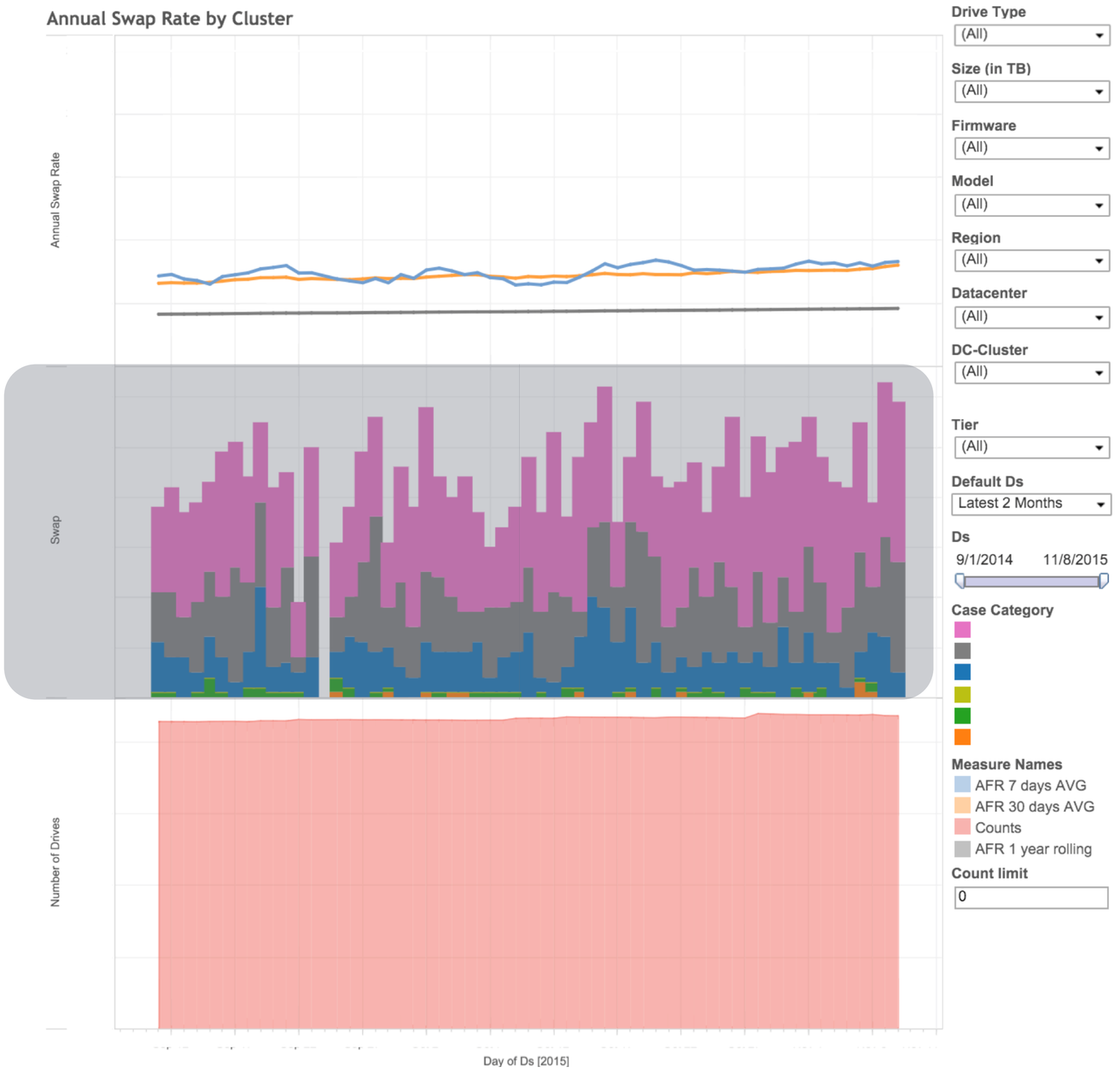
Monitoring

Failure Rate



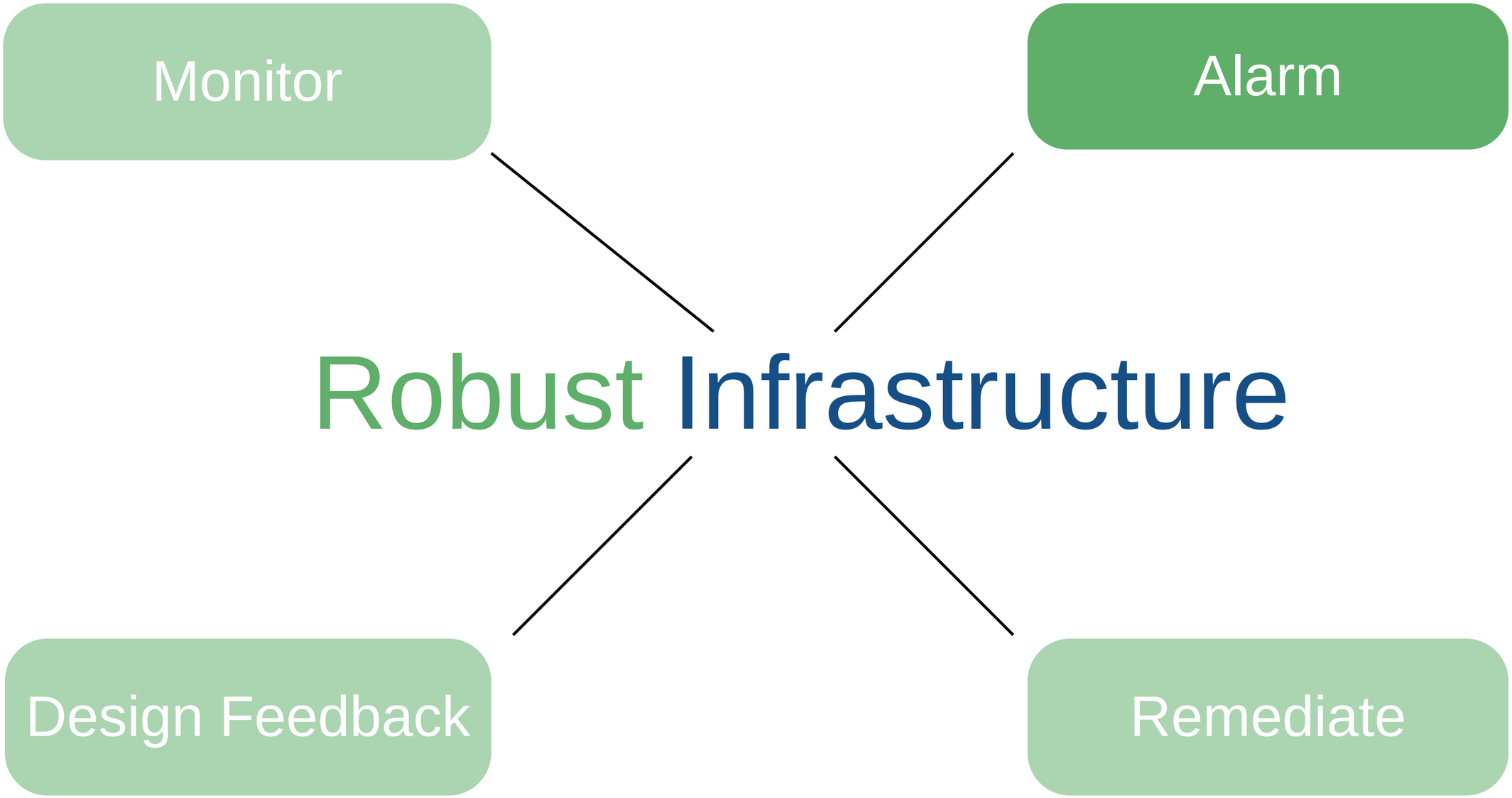
Monitoring

Error Types



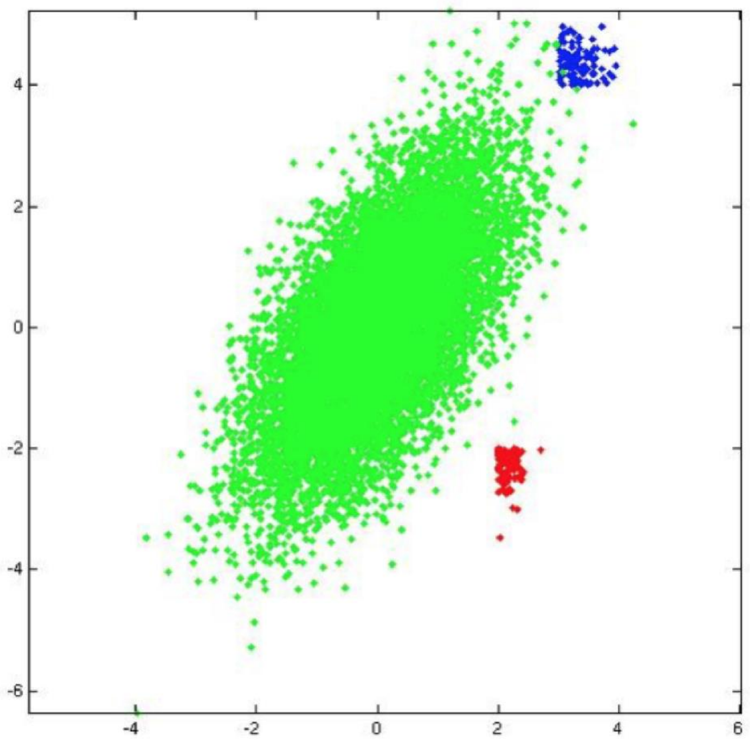
Monitoring

Filters

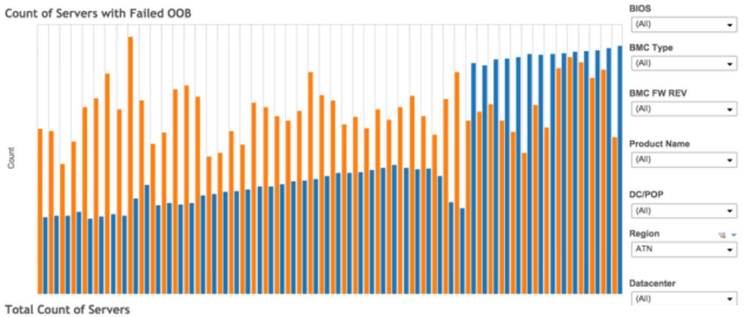

Monitor
Alarm
Robust Infrastructure
Design Feedback
Remediate

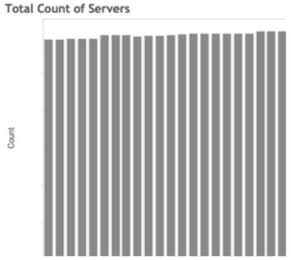
Alarms

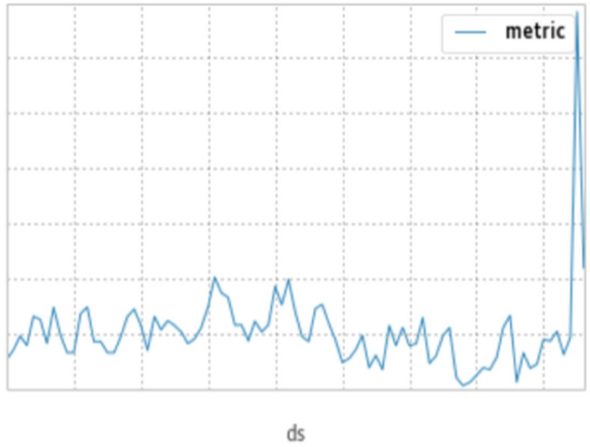
Anomaly Detection



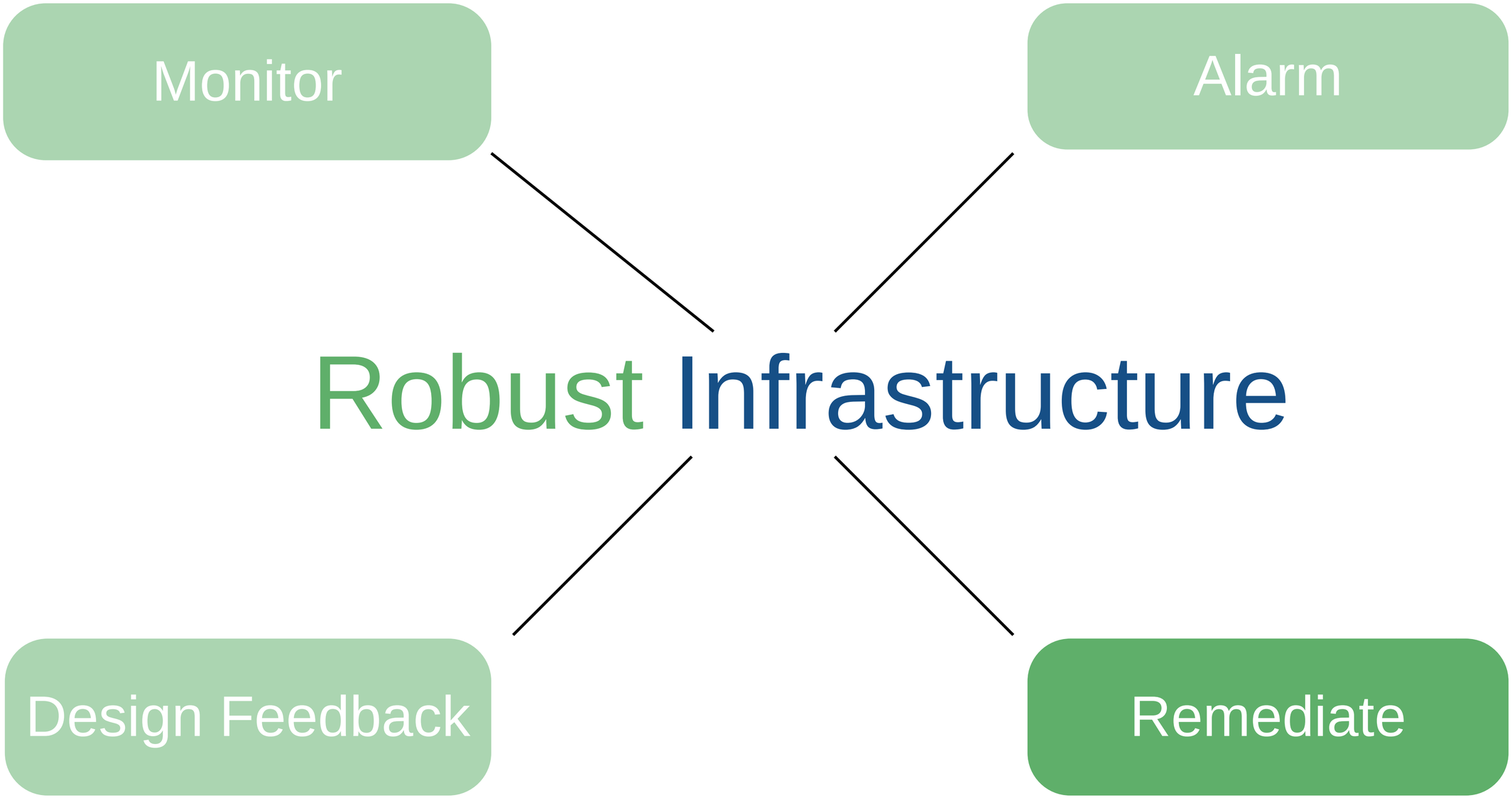
Anomaly Within Cohorts







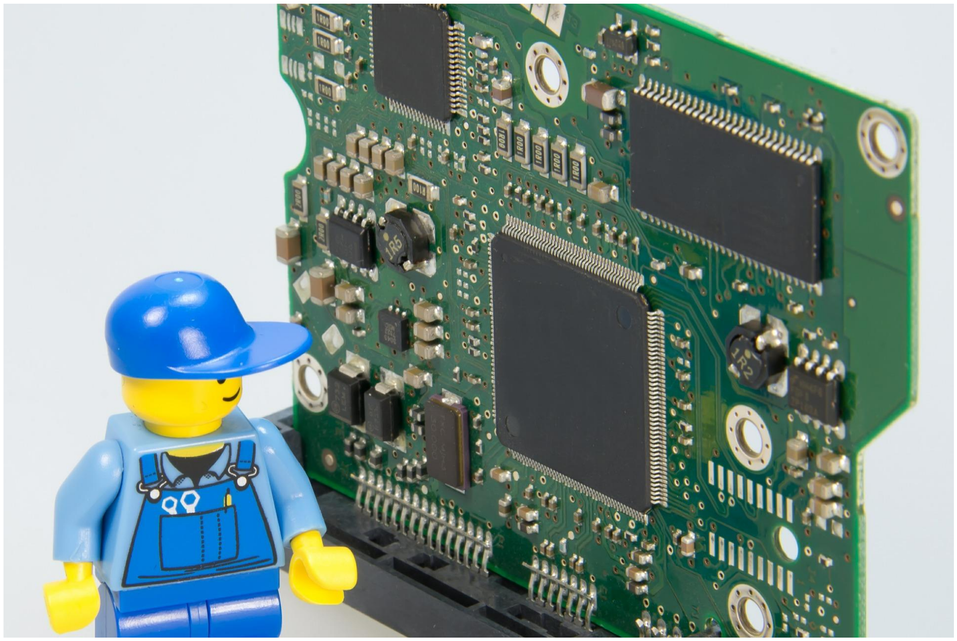
Gradual Increases
And
Sudden Spikes

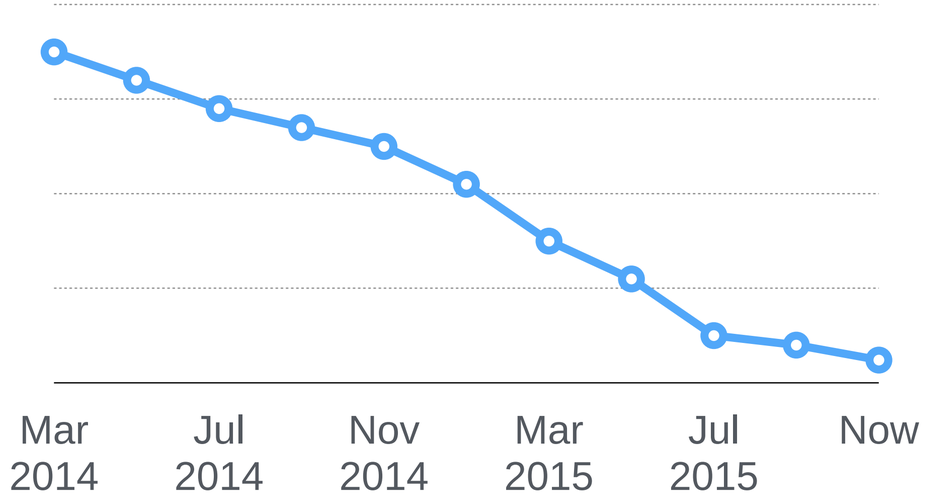

Monitor
Alarm
Robust Infrastructure
Design Feedback
Remediate

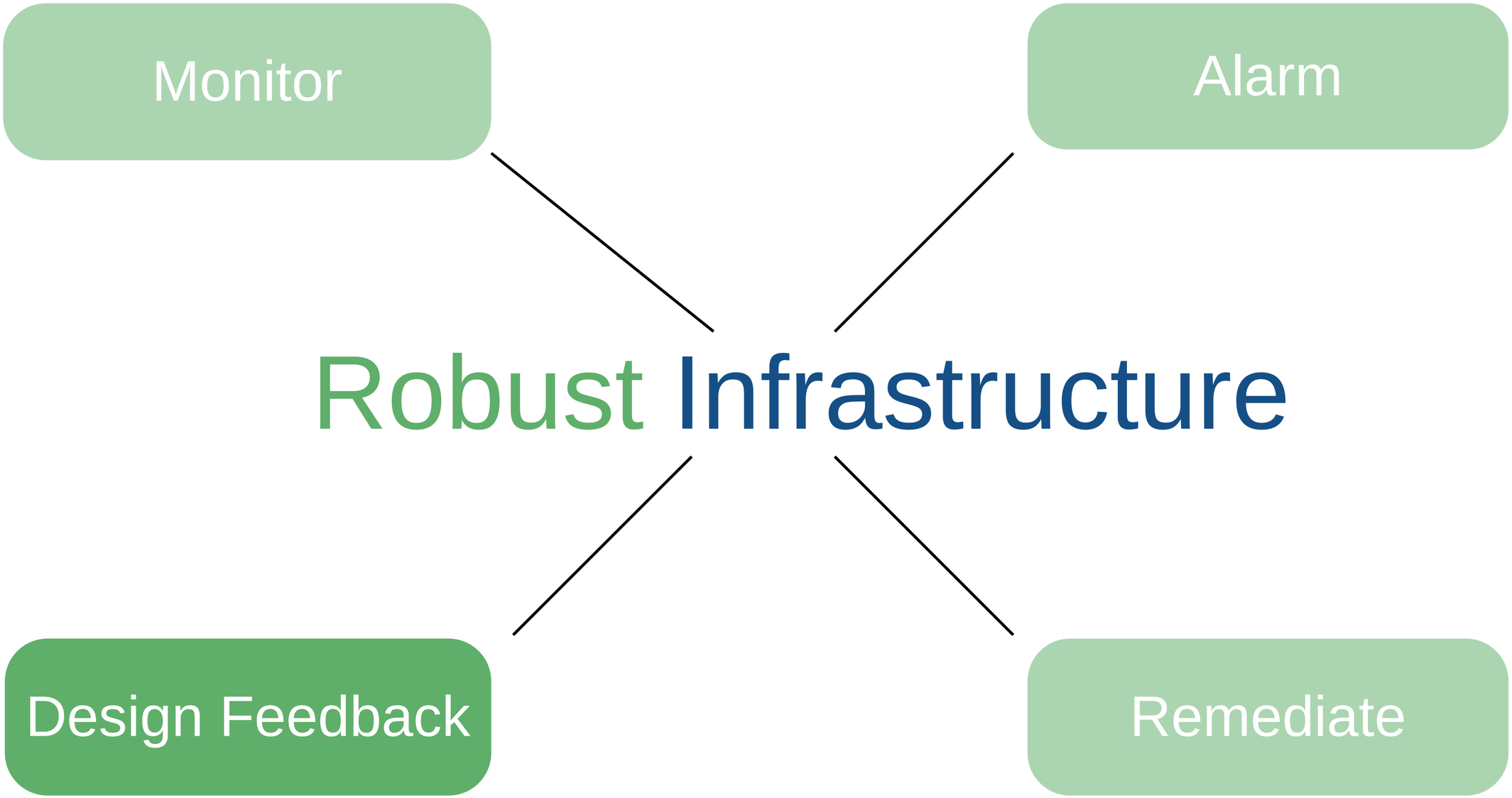
Remediation

The Journey is 1% Finished

- Phase 1: Root Cause Analysis
- Phase 2: Review Remediation Plan
- Phase 3: Implement Remediation

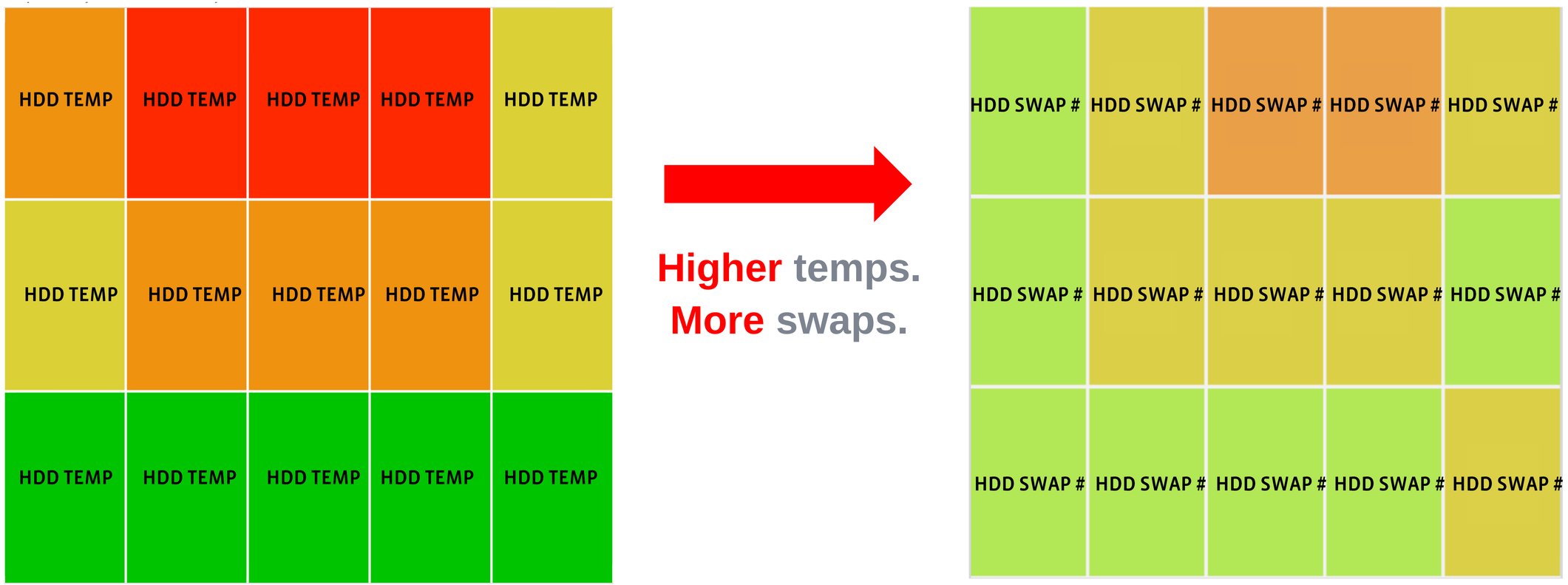





Monitor
Alarm
Robust Infrastructure
Design Feedback
Remediate

Design Improvements

HDD Slot Temperature vs. Swap Rate



Wrap Up

Key takeaways

- FB scale is growing. Infrastructure needs to innovate
- Move fast and adapt with robust HW lifecycle
- Everything fails – minimize impact with tooling


OPEN
Compute Project