# Trends from The Trenches

## 2016 BioIT World Conference & Expo

BIOTEAM
Enabling Science

Photo credit: Aaron Gardner

© Copyright BioTeam 2016. All rights reserved. bioteam.net/products

I’m Chris.
I’m an infrastructure geek
I work for the BioTeam.

Photo credit: Cindy Jessel

Honest. Objective. Vendor & Technology Agnostic.

15+ years “bridging the gap” between hardcore science, HPC & IT

Run by scientists forced to learn IT to “get science done”

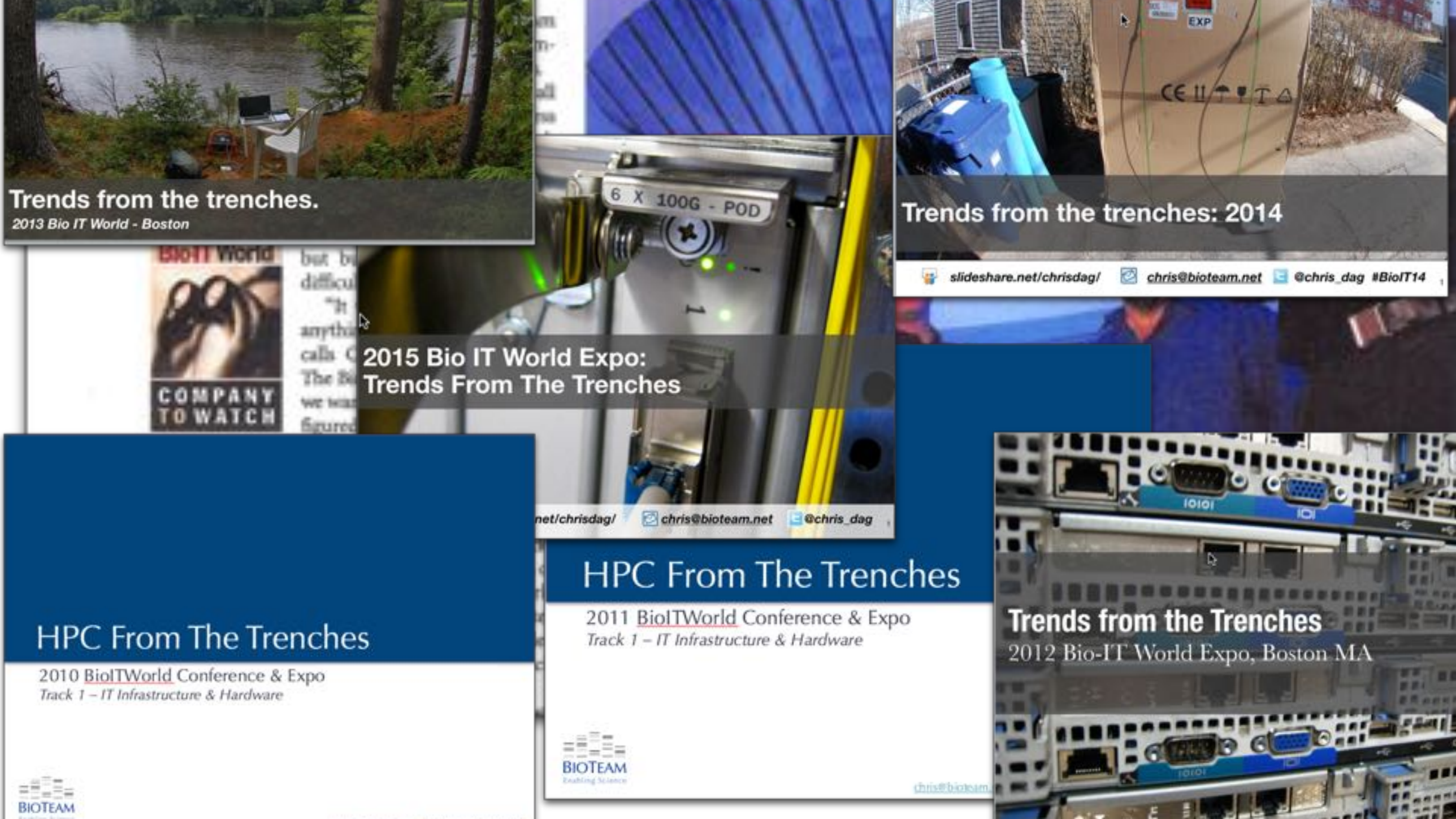
Virtual company with nationwide staff

Independent Consulting Shop

**www.BioTeam.net**

# BioTeam in '16

**missing: Cindy & Mark**


Trends from the trenches.
2013 Bio IT World - Boston
Trends from the trenches: 2014
slideshare.net/chrisdag/
chris@bioteam.net
@chris_dag #BioIT14
6 X 100G - POD
2015 Bio IT World Expo:
Trends From The Trenches
net/chrisdag/
chris@bioteam.net
@chris_dag
COMPANY TO WATCH
HPC From The Trenches
2010 BioITWorld Conference & Expo
Track 1 – IT Infrastructure & Hardware
BIOTEAM
HPC From The Trenches
2011 BioITWorld Conference & Expo
Track 1 – IT Infrastructure & Hardware
BIOTEAM
Trends from the Trenches
2012 Bio-IT World Expo, Boston MA

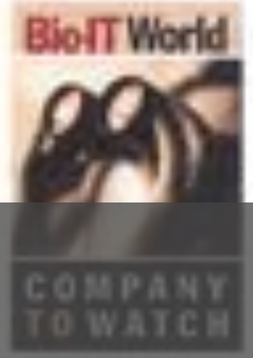
# The BioTeam: Riders of the Storm

The self-described SWAT team of bio-IT builds state-of-the-art infrastructure for life science organizations

■ BY JOHN RUSSELL

Early in 2002, the four founders of The BioTeam launched their fledgling bio-IT consulting company into the oncoming biotech industry storm.

Venture capital funding was shriveling for all but follow-on deals. The bioinformatics business model was foundering. ImClone Systems Inc.'s shockwaves rattled through the stock market. Service business models were trashed. And spending on IT was slowing.




It wasn't the Perfect Storm, but business conditions were difficult and worsening.

"It was a scary time to try anything entrepreneurial," recalls Chris Dagdigian, one of The BioTeam's founders. "[But] we wanted to [illegible] figured someone was going to land a permanent job very soon [illegible] one chance to see if we could do it. We worked together as a team before and were very successful."

All four — Dagdigian, Bill Van Etten, Michael Athanas, and Stan Gloss — were alumni of Blackstone Computing, where they comprised much of

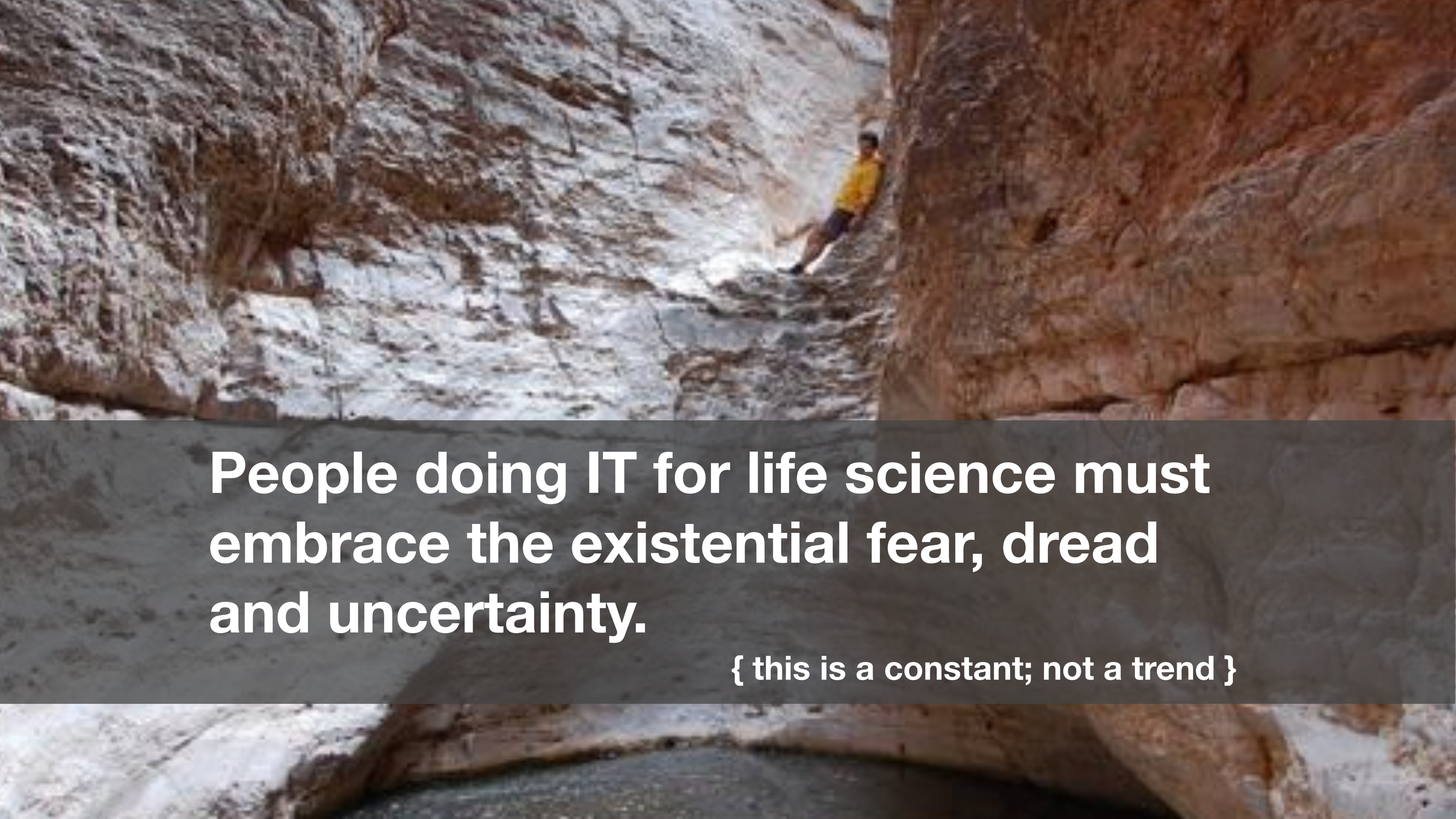
Been doing this stuff since 2002
Midlife crisis time …

Worried that this talk is becoming tired or too repetitive. Maybe we need to change the format and/or speaker.

6-question online survey is here:

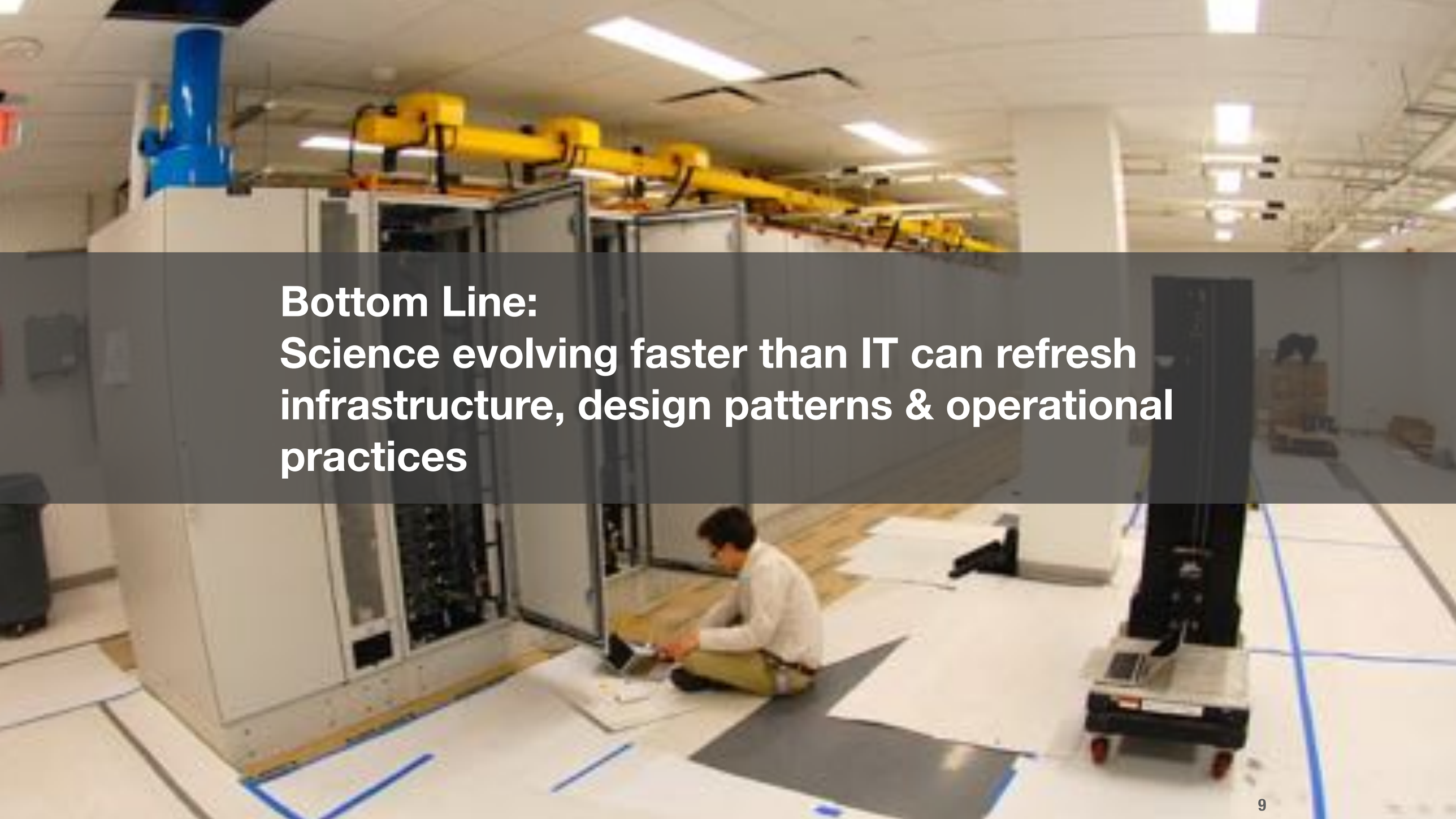
http://biote.am/survey16

*This scannable QR code links to the survey*



**People doing IT for life science must embrace the existential fear, dread and uncertainty.**

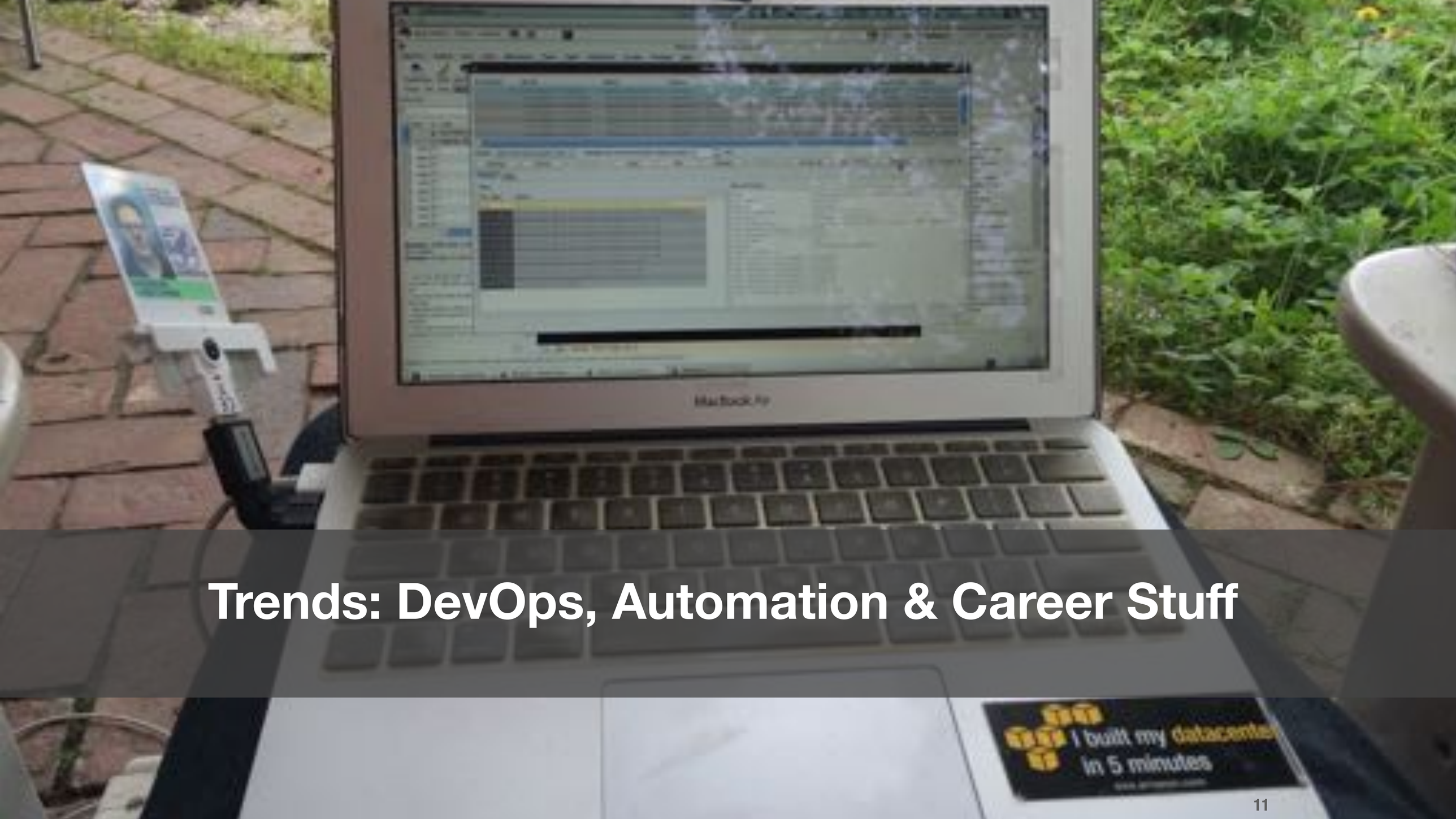
**{ this is a constant; not a trend }**



**Bottom Line:**
**Science evolving faster than IT can refresh infrastructure, design patterns & operational practices**

# We have to build stuff that runs for years

… to support researchers who can’t articulate their needs beyond a few months




# Trends: DevOps, Automation & Career Stuff

**This section 80% recycled from ’15 talk**

**… thought about skipping the old sides**

**but the topic is too important**

# SysAdmins: If you can’t script your upward career mobility is done. Period.

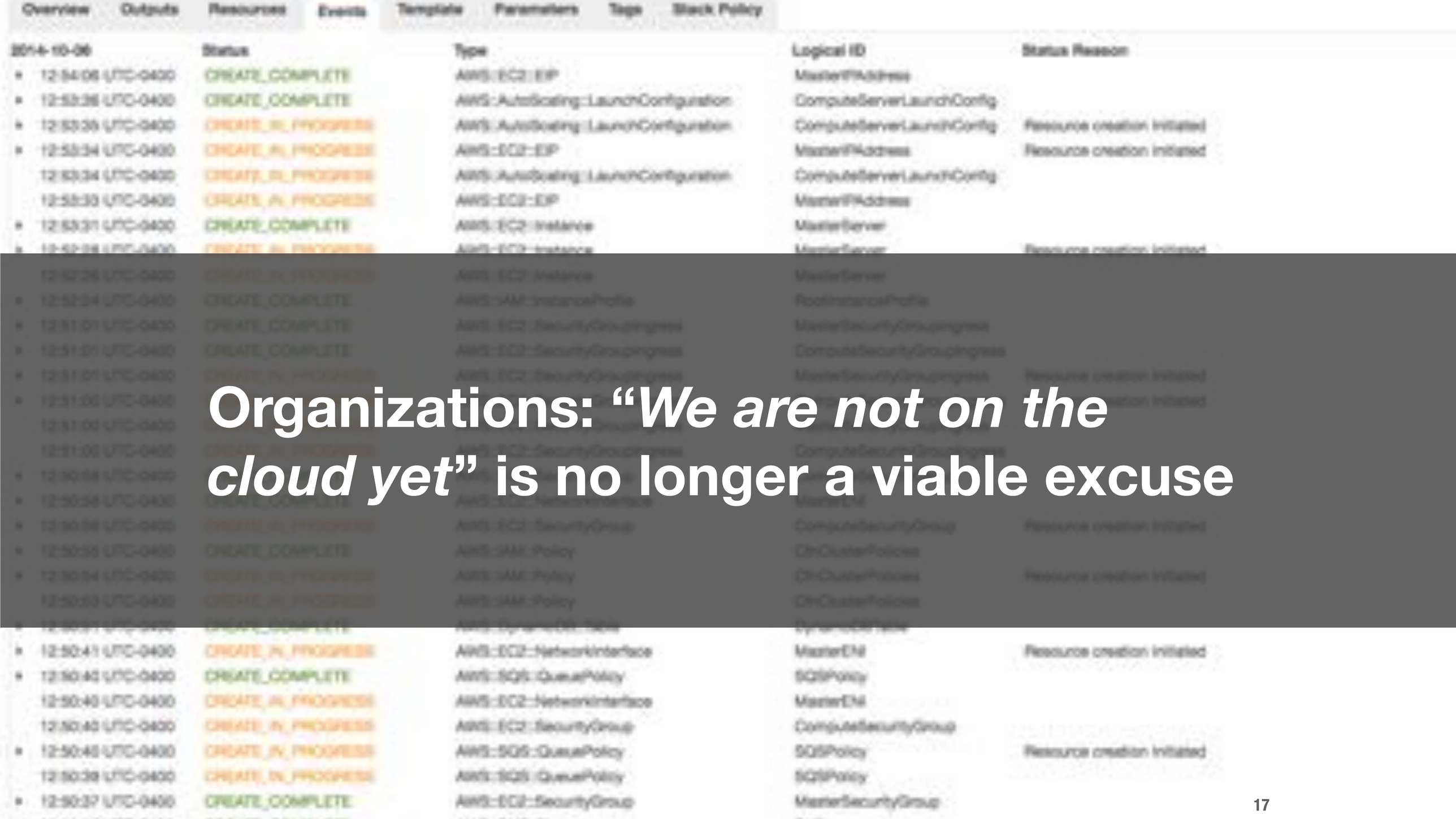
# It’s not “just” cloud & virtualization …
# Everything will have an API soon

# Orchestration, configuration management & automation stacks are already in your enterprise

# You will need to learn them.

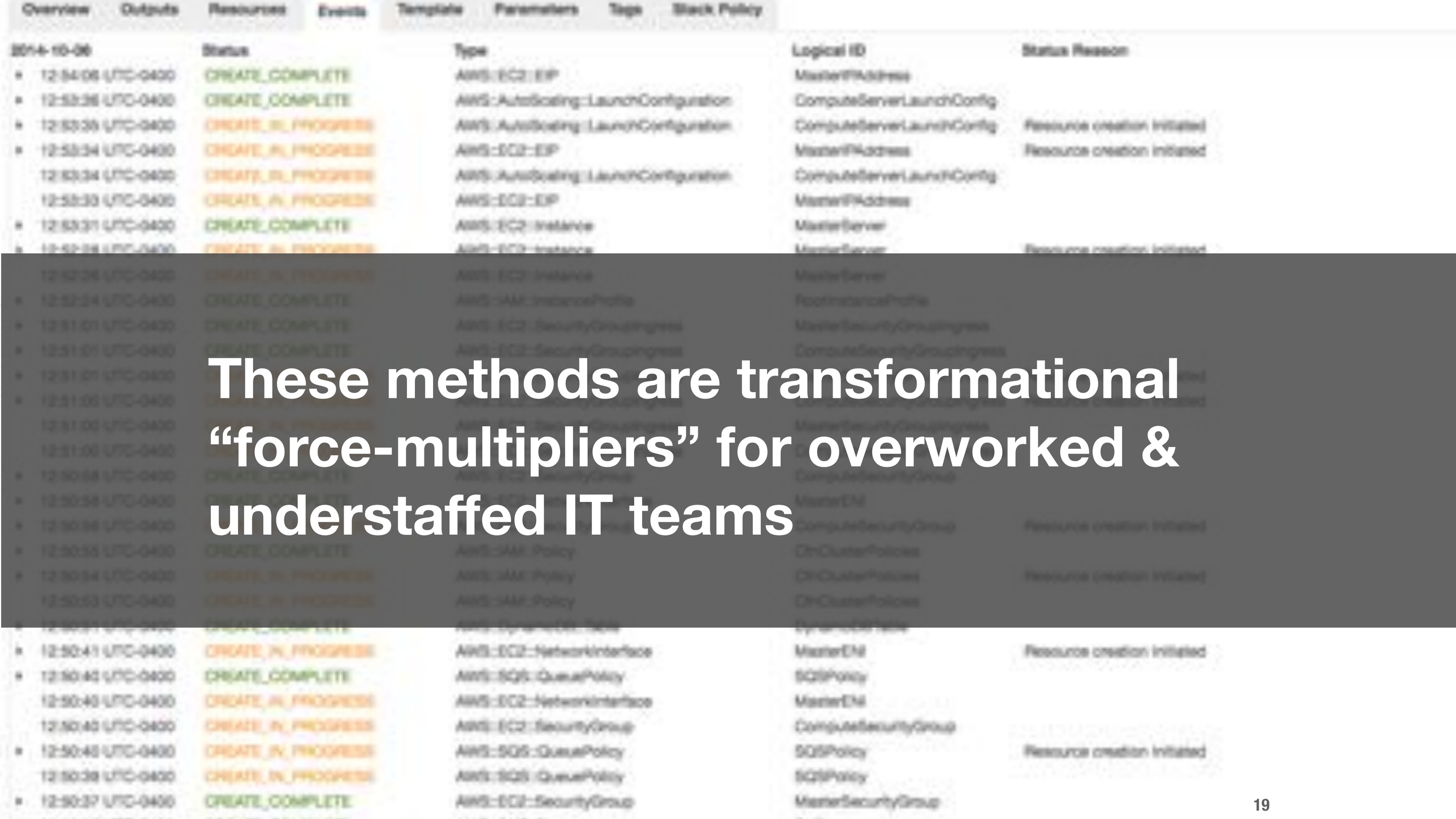
**Bonus: You will be 10x more productive**

**Bonus: Your LinkedIn profile views will go up 300%**

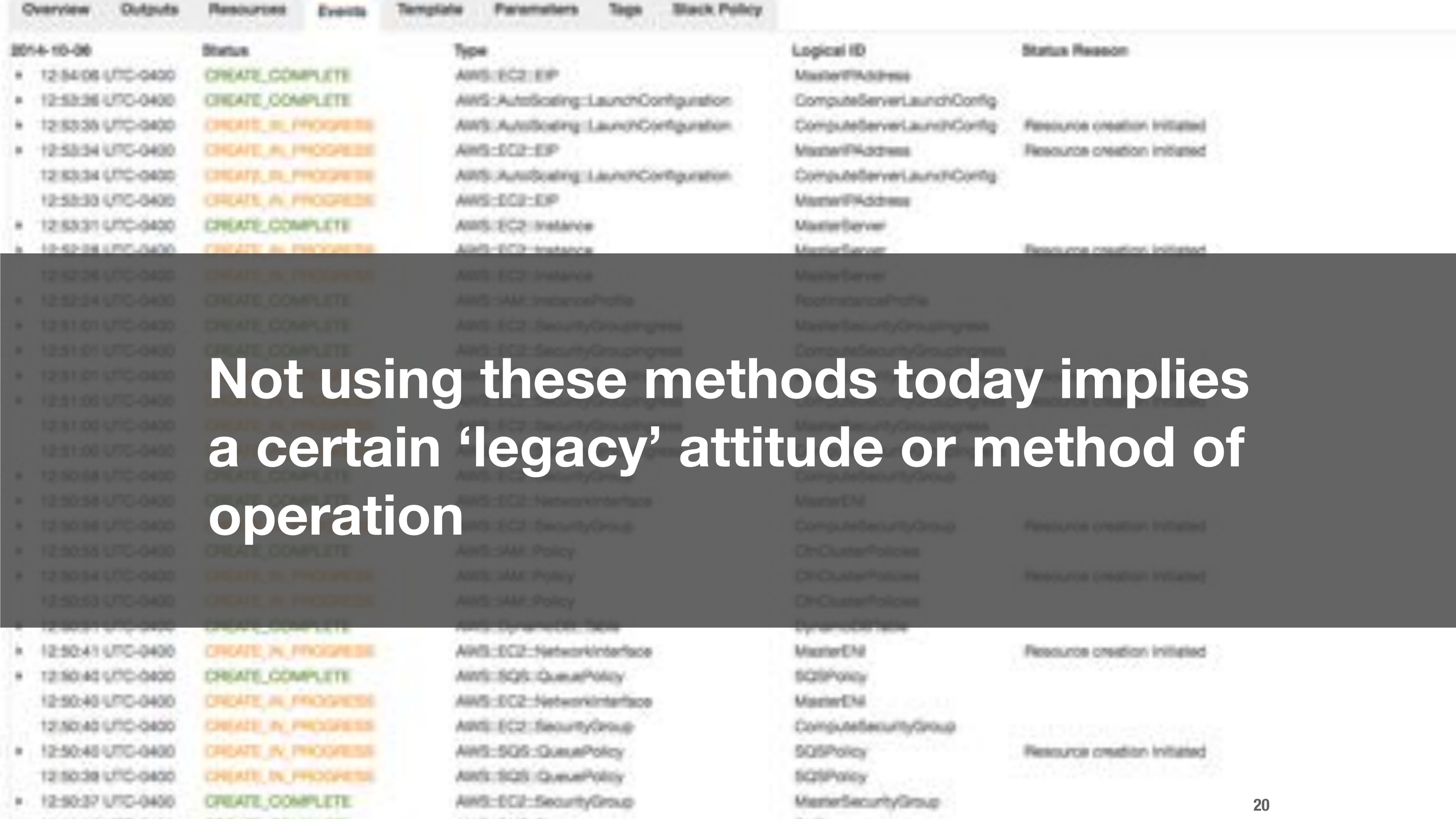



# Organizations: “*We are not on the cloud yet*” is no longer a viable excuse

# DevOps methods & infrastructure automation have been transforming on-premise IT for years at this point




# These methods are transformational “force-multipliers” for overworked & understaffed IT teams




# Not using these methods today implies a certain ‘legacy’ attitude or method of operation

Overview | Outputs | Resources | Events | Template | Parameters | Tags | Stack Policy

| 2014-10-06 | Status | Type | Logical ID | Status Reason |
|---|---|---|---|---|
| 12:54:06 UTC-0400 | CREATE_COMPLETE | AWS::EC2::EIP | MasterIPAddress | |
| 12:53:38 UTC-0400 | CREATE_COMPLETE | AWS::AutoScaling::LaunchConfiguration | ComputeServerLaunchConfig | |
| 12:53:35 UTC-0400 | CREATE_IN_PROGRESS | AWS::AutoScaling::LaunchConfiguration | ComputeServerLaunchConfig | Resource creation Initiated |
| 12:53:34 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::EIP | MasterIPAddress | Resource creation Initiated |
| 12:53:34 UTC-0400 | CREATE_IN_PROGRESS | AWS::AutoScaling::LaunchConfiguration | ComputeServerLaunchConfig | |
| 12:53:33 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::EIP | MasterIPAddress | |
| 12:53:31 UTC-0400 | CREATE_COMPLETE | AWS::EC2::Instance | MasterServer | |
| 12:52:38 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::Instance | MasterServer | Resource creation Initiated |
| 12:52:36 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::Instance | MasterServer | |
| 12:52:34 UTC-0400 | CREATE_COMPLETE | AWS::IAM::InstanceProfile | RootInstanceProfile | |
| 12:51:01 UTC-0400 | CREATE_COMPLETE | AWS::EC2::SecurityGroupIngress | MasterSecurityGroupIngress | |
| 12:51:01 UTC-0400 | CREATE_COMPLETE | AWS::EC2::SecurityGroupIngress | ComputeSecurityGroupIngress | |
| 12:51:01 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::SecurityGroupIngress | MasterSecurityGroupIngress | Resource creation Initiated |
| 12:51:00 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::SecurityGroupIngress | ComputeSecurityGroupIngress | Resource creation Initiated |
| 12:51:00 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::SecurityGroupIngress | MasterSecurityGroupIngress | |
| 12:51:00 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::SecurityGroupIngress | ComputeSecurityGroupIngress | |
| 12:50:58 UTC-0400 | CREATE_COMPLETE | AWS::EC2::SecurityGroup | ComputeSecurityGroup | |
| 12:50:58 UTC-0400 | CREATE_COMPLETE | AWS::EC2::NetworkInterface | MasterENI | |
| 12:50:56 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::SecurityGroup | ComputeSecurityGroup | Resource creation Initiated |
| 12:50:55 UTC-0400 | CREATE_COMPLETE | AWS::IAM::Policy | OhClusterPolicies | |
| 12:50:54 UTC-0400 | CREATE_IN_PROGRESS | AWS::IAM::Policy | OhClusterPolicies | Resource creation Initiated |
| 12:50:53 UTC-0400 | CREATE_IN_PROGRESS | AWS::IAM::Policy | OhClusterPolicies | |
| 12:50:51 UTC-0400 | CREATE_COMPLETE | AWS::DynamoDB::Table | DynamoDBTable | |
| 12:50:41 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::NetworkInterface | MasterENI | Resource creation Initiated |
| 12:50:40 UTC-0400 | CREATE_COMPLETE | AWS::SQS::QueuePolicy | SQSPolicy | |
| 12:50:40 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::NetworkInterface | MasterENI | |
| 12:50:40 UTC-0400 | CREATE_IN_PROGRESS | AWS::EC2::SecurityGroup | ComputeSecurityGroup | |
| 12:50:40 UTC-0400 | CREATE_IN_PROGRESS | AWS::SQS::QueuePolicy | SQSPolicy | Resource creation Initiated |
| 12:50:39 UTC-0400 | CREATE_IN_PROGRESS | AWS::SQS::QueuePolicy | SQSPolicy | |
| 12:50:37 UTC-0400 | CREATE_COMPLETE | AWS::EC2::SecurityGroup | MasterSecurityGroup | |

# .. that your competition may not have

# Chef, Puppet, Ansible, SaltStack …

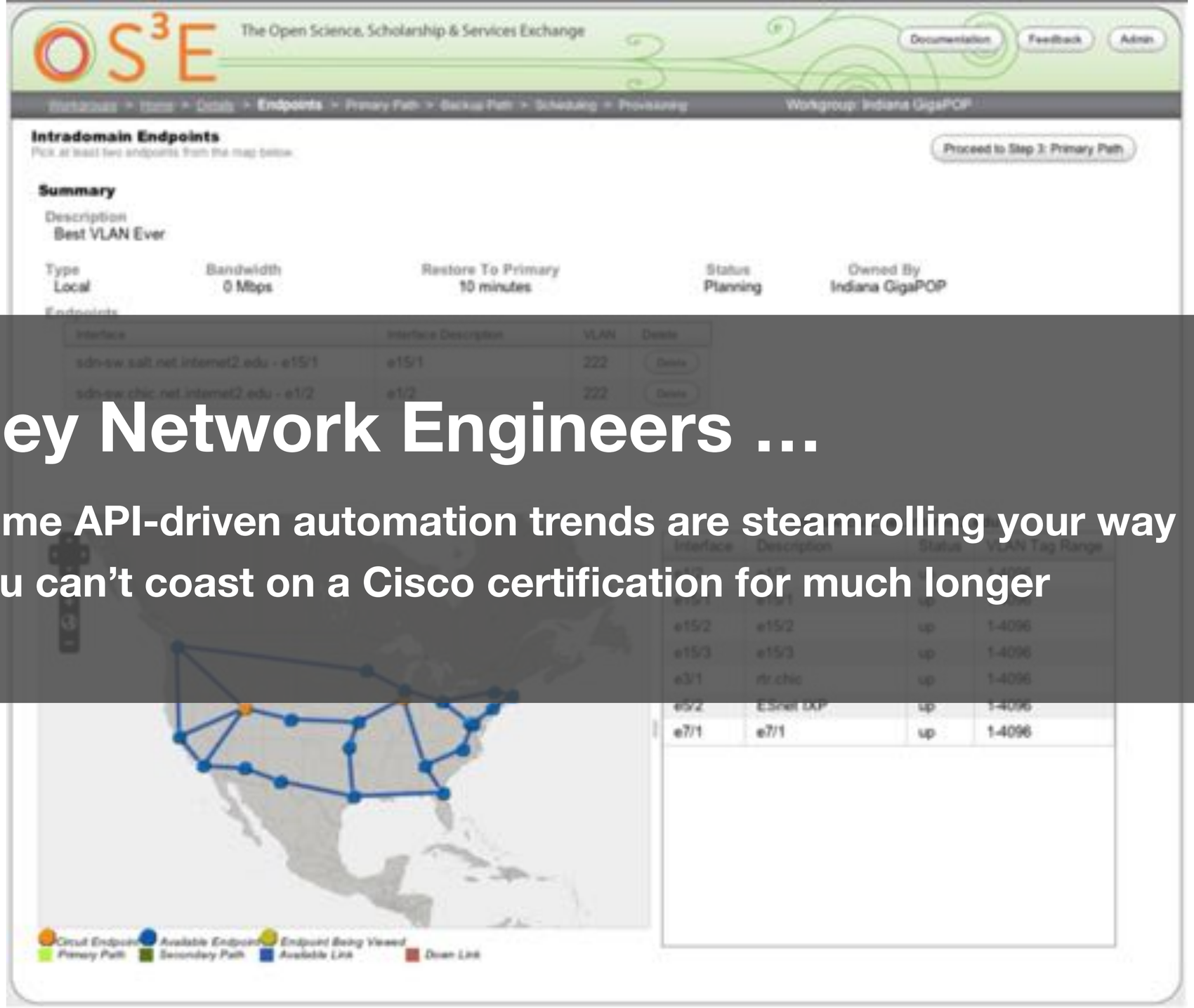
Pick what works for you (and that all can agree on)

… and commit to learning, using and evangelizing it

… ideally across the enterprise (not just Research)

# Happy to chat/demo this stuff

… and if you stop by our booth ask to see the awesome way we are doing automated and simple appliance deployments and updates via USB wireless dongle, iphone/tablet & automation magic …




# Hey Network Engineers …

**Same API-driven automation trends are steamrolling your way**

**You can’t coast on a Cisco certification for much longer**

# New for 2016: Bench Scientists …

‘big data’ is becoming mainstream. Spreadsheets and Excel macros are not going to be viable for some of you very quickly.

May have to learn new data analytic techniques to get the best info out of your most interesting data

# New for 2016: Data Scientist Roles

Sexy new occupation. High demand/salaries across many industries

Warning: We may be seeing signs of rot in the talent pool similar to ‘bioinformatician’ craze in the ‘90s …

Blunt truth: Analytic mastery requires BOTH skill AND domain expertise. Watch out for certificate/diploma mills. You don’t need the mythical “unicorn” but you need people who understand the domain and not just the toolset.

# New for 2016: Data Engineer Role

[Emerging Trend ]
Not so sexy and not near as much hype as ‘data scientist’
… but becoming increasingly essential in organizations with high-scale, high-velocity data. Especially with diverse data types.

Blunt truth:  Specialist skills are needed to manage data ingest, provenance, transformation, movement and manipulation. It’s a waste of resources to have a scientist do this and Enterprise IT does not have the skillset.

# Trends: Compute

## Just the new/emerging stuff …

# Emerging compute pain / capability gap

**Bench scientists losing out on workforce mobility initiatives because they are‘tethered’ to labside analysis workstations**

# The issue:

- More and more lab side instruments require large workstations to process results and look at data
  - These users do not need HPC and may have had little interaction with the ‘big compute’ people
  - … but the hardware requirements for the vendor software far exceed what their personal workstation is capable of running
  - Corporate VM / VDI teams have no idea

# The result:

- Highly skilled, highly recruited and highly-paid PhD level scientists are ‘tethered’ to wet lab workstations for many hours a day
  - … they CAN’T work from their office/desk
  - … they CAN’T work remotely via VPN or VDI
- Productivity and morale issue today. Potential to become an employee retention problem in the future

# Example: Flow Cytometry gets hardcore.

# Compute Trend:

## HPC Across the Enterprise + HPC as a Service

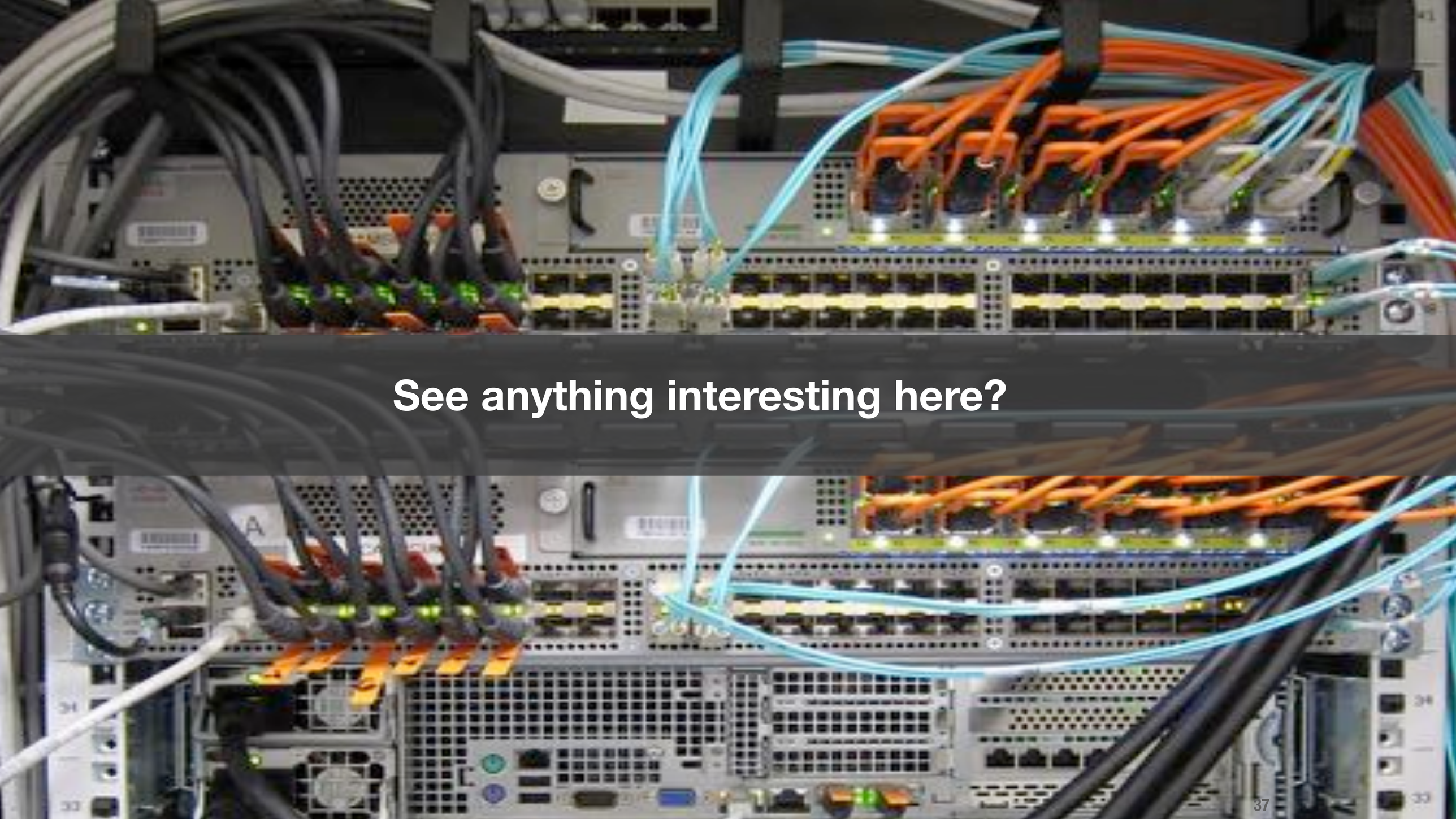
# HPC for Enterprise; Not Just Research

- Recurring trend in recent HPC assessment projects
  - Biotech, Pharma and Government
- HPC now rebuilt, rebranded, operated and advertised as a service consumable by anyone in the organization
  - Available via traditional methods
  - … but also via exposed APIs that enable HPC to be integrated in various upstream workflows
- Stretching our Infrastructure & Ops in interesting ways

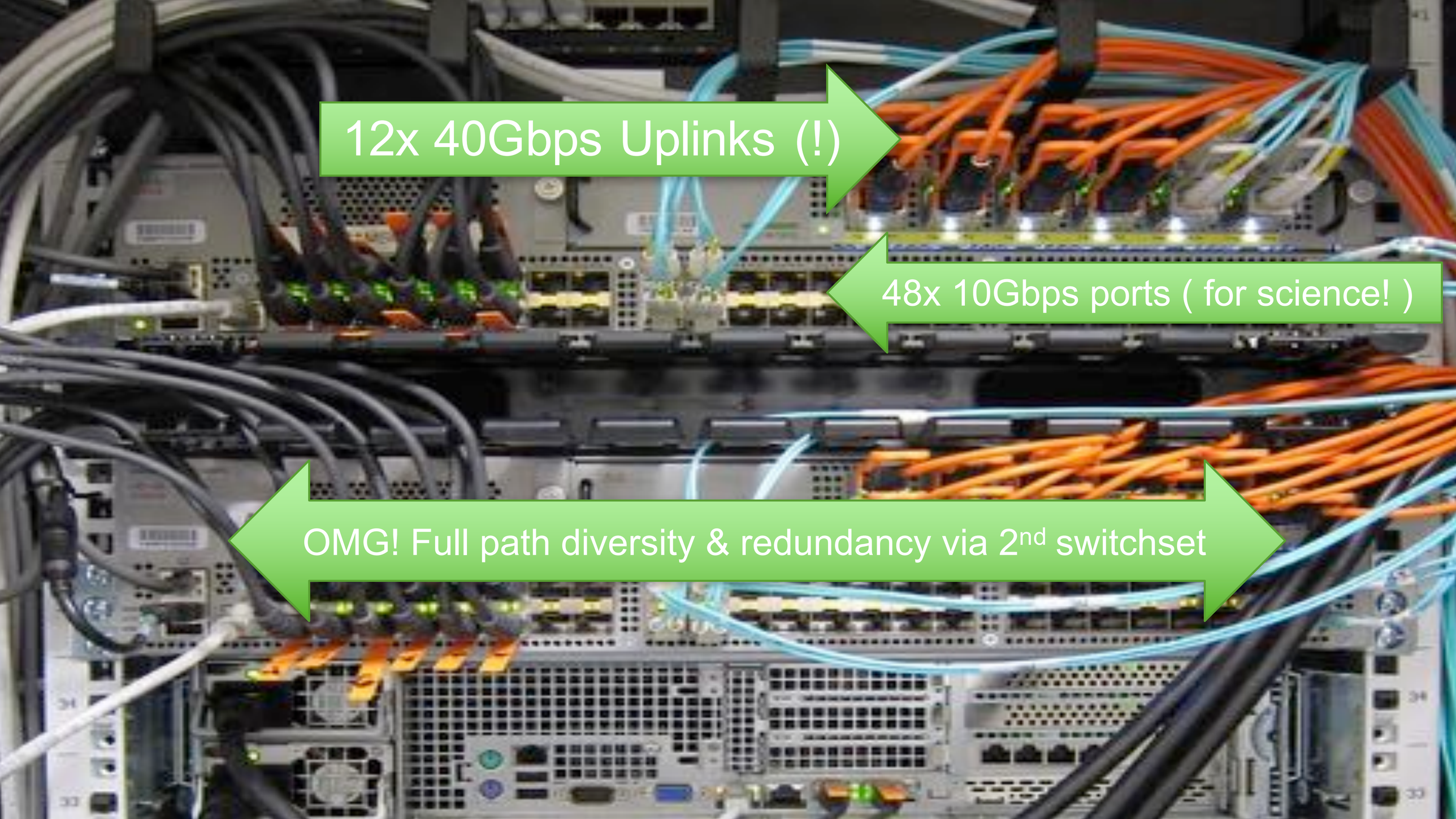
# HPC for Enterprise; Not Just Research

- Requires significant investment in user outreach, onboarding, training, knowledge transfer & mentoring
- Increased focus on portals and common workflow templates (cli via SSH is not for everyone …)
- For APIs we are seeing:
  - Homegrown, Univa and http://agaveapi.co
- This also affects hardware infrastructure and operations model a little bit …

- History Time:
  - Discovery oriented research shops often intentionally sacrifice high availability to save vast piles of money
  - It may be annoying but we can tolerate downtime in exchange for bigger/faster/better or more capable …
  - This approach does not translate well into the Enterprise world with SLAs and 24x7 uptime needs
  - Lets look at one real world example …



See anything interesting here?


12x 40Gbps Uplinks (!)
48x 10Gbps ports ( for science! )
OMG! Full path diversity & redundancy via 2nd switchset

# Compute Trend:

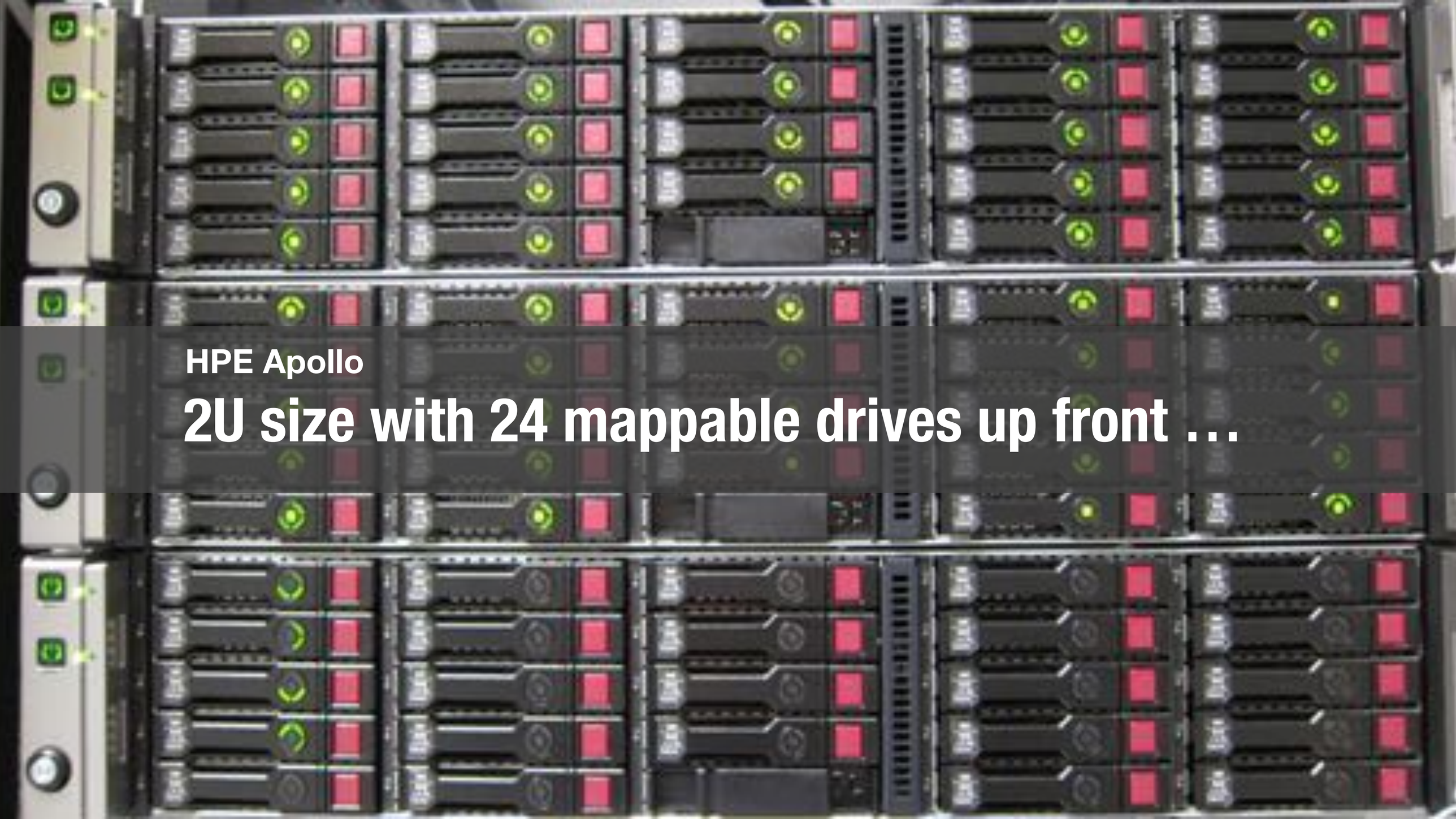
## Common/Reconfigurable Hardware for HPC and Hadoop

# Compute 2:  Common Kit for HPC/Hadoop

[ Remember the ‘meta’ problem] — IT has to design infrastructures that will live for years in face of unknowable scientific requirement

“Big Data” may still be 70% vapid hype and vendor lies but the 30% that is actually useful is showing excellent promise within the groups that are deploying it.

HDFS fundamentally wants lots of local disk spindles as close as possible to the CPU. This is starting to significantly affect the types of servers that forward-looking BioIT people are procuring.



HPE Apollo

# 2U size with 24 mappable drives up front …

# And 4x very dense servers packed into the rear

**This is an interesting packaging mix, particularly for those seeking to support 'traditional' HPC and Hadoop/BigData on common hardware.**

# Compute Trend & Emerging Pain Point:

## Data Lakes & Multi-tennant secure Hadoop

# Compute 3: Hadoop 2.0 & Data Lakes

- Significant Trend in 2016
- Feels like we've passed the 50% threshold of clients currently running real workloads on "big data" platforms
- And just like HPC, these platforms are (often) being marketed beyond just the research organization
- Interesting: Roughly 70:30 split between cloud and onsite stacks
- Significant potential for wasted resources
  - Hype may drive adoption in absence of realistic need
  - Approach with caution (and with the right skills!)

# Compute 3:  Hadoop 2.0 & Data Lakes

- [Longstanding Issue] Just like “grid computing” in the 90s we are still in the “empty hype”, “vapid press release” and “endless lies from marketing” phase of Hadoop/BigData/DataLake
- [Blunt Advice] Despite vendor claims, this stuff is currently neither easy, simple or approachable by mere mortals.
  - Not magic, you just have to work for it. Unicorns/ninjas not required.
  - Getting useful, actionable results out of these systems REQUIRES a user that is knowledgeable of the domain, data types, data structure, common data transformations, SQL, statistics, visualization, etc. etc. etc.
    - Pre-sales people are often not truthful about this (!)

# Apache Hadoop Ecosystem

**YARN, Spark, Hive, Tez, Pig,  Zookeeper, HDFS, Ambari, etc. etc.**

- https://projects.apache.org/projects.html?category
- Fantastic open source building blocks with multiple commercial & free options
- Two quick anecdotes / war stories

# Apache Hadoop Ecosystem

## Adventures with Zeppelin

- Apache Hadoop Ecosystem consists of MANY independent open source projects
- There are a fantastic number of moving parts that need consideration when you are trying to build a functional and integrated stack
- There is a lot of coordination but each of these projects tends to have varying levels of developer support, release cycles and maturity
- One example: Apache Zeppelin
  - Web based notebook for interactive data analytics
  - https://zeppelin.incubator.apache.org/
  - Incredibly approachable very useful - end users LOVE it

http://localhost:19901/#/

localhost:19901/#/notebook/2BAFV9HXS

Search

Zeppelin Notebook Interpreter Connected

cdagdig-new-lo

+ Create new note

cdagdig-new-location-W3K4VUTJ4

Zeppelin Tutorial

default

```
sc.version
res0: String = 1.4.1
Took 11 seconds
```

FINISHED

FINISHED

```
import org.apache.commons.io.IOUtils
import java.net.URL
import java.nio.charset.Charset

// Zeppelin creates and injects sc (SparkContext) and sqlContext (HiveContext or SqlContext)
// So you don't need create them manually

// Load bank data
val bankText = sc.parallelize(
    IOUtils.toString(
        new URL("https://s3.amazonaws.com/apache-zeppelin/tutorial/bank/bank.csv"),
        Charset.forName("UTF8"),).split("\n"))

case class Bank(age: Integer, job: String, marital: String, education: String, balance: Integer)

val bank = bankText.map(s => s.split(";")).filter(s => s(0) != "\"age\"").map(
    s => Bank(s(0).toInt,
            s(1).replaceAll("\"", ""),
            s(2).replaceAll("\"", ""),
            s(3).replaceAll("\"", ""),
            s(5).replaceAll("\"", "").toInt
        )
).toDF()
bank.registerTempTable("bank")
```

# Apache Hadoop Ecosystem

## Adventures with Zeppelin

- Almost every data analyst who sees Zeppelin loves it
- May become a standard for web-based ‘data science notebooks’
- However … it’s still in “incubator” stage within Apache Foundation
- And this presents some challenges with deployment
  - User interest outpaces development roadmap

# Apache Hadoop Ecosystem

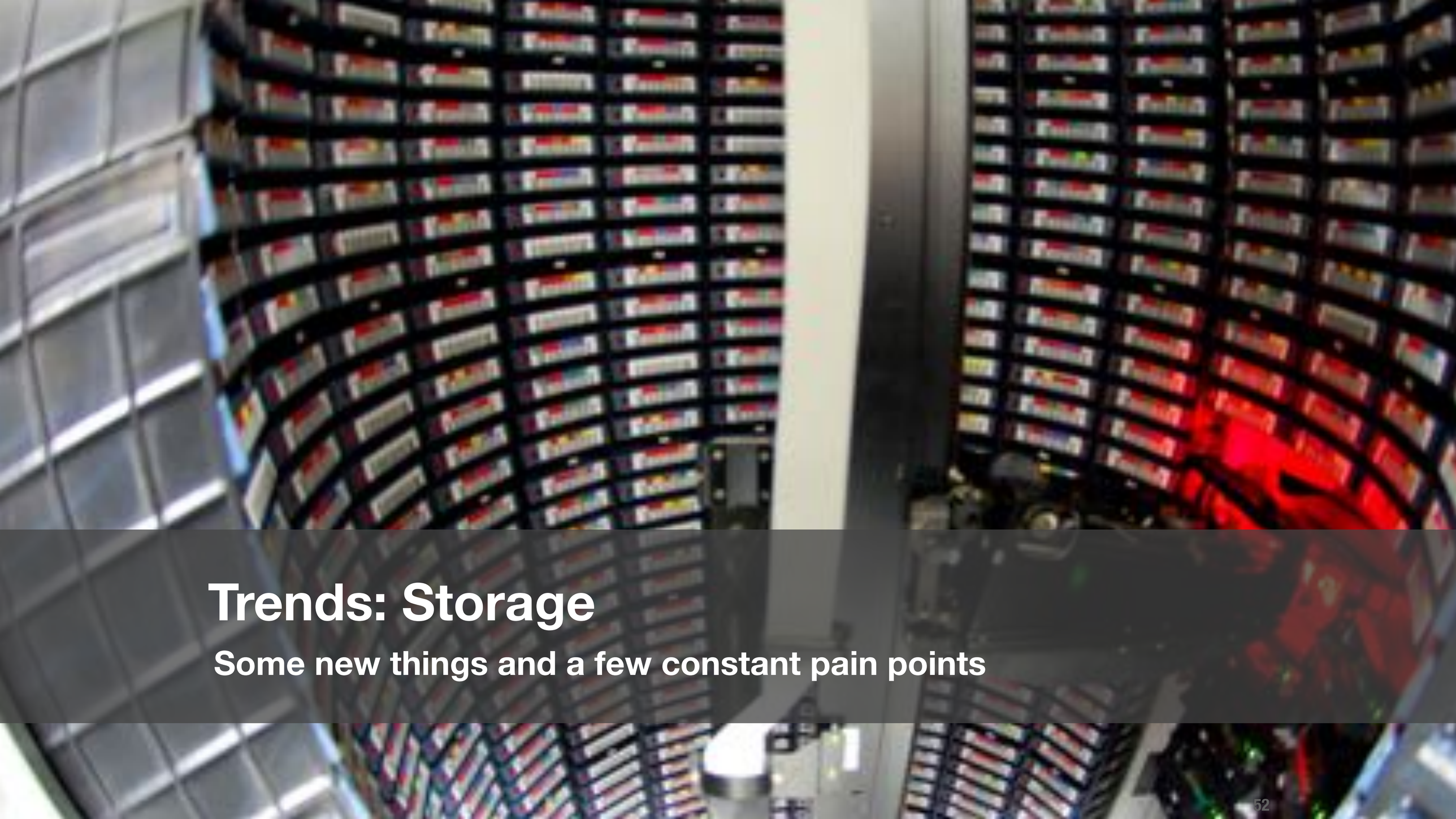
## Adventures with Zeppelin

- When we 1st tried to deploy zeppelin for users …
  - It had no integrated support for authentication or more than 1 user
  - All notebooks, all data and all results were accessible and editable by anyone with access to the (non secured, non-authenticated) web interface
- So we had to install multiple zeppelin servers
  - Each user gets a private, customized install
  - And it was our responsibility to create and enforce access controls
- Honestly not a hard issue to resolve
  - But a good example of “interest outpacing development” in this space

# Apache Hadoop Ecosystem

## Securing for Enterprise Use

- Deployment is smooth with no sensitive data or a single-user workflow
  - Cloud environments are great as you can spin up bespoke clusters
- Bit harder when Enterprise SecOps compliance checkboxes are involved:
  - Multi-user / Active Directory & Kerberos integration
  - Encryption everywhere (at-rest, in-flight) for data and communication
  - Role based access control for services and HDFS filesystem contents
  - Audit trails & compliance/governance testing
  - Authenticated API calls and secure API gateway
- Not all ecosystem components support hardened operation
  - May be forced to remove elements (Hive, etc.) that can't play in the hardened environment



# Trends: Storage

Some new things and a few constant pain points

# Still have ALL the file types

This is what makes life science an interesting market

- Massive text files
- Massive binary files
- Flatfile ‘databases’
- Spreadsheets everywhere
- Directories w/ 6 million files
- Large files: 600GB+
- Small files: 30kb or smaller

- This is problematic because many storage products are engineered to to excel at only 1-2 of these things …

# Data Curation Still Important

- Organizations with 1PB+ or more data under management that DO NOT have a full-time employee managing/curating/monitoring things
- … are wasting more in storage CapEx than the fully loaded cost of the human resource

# Data Ingest Still Risky & Hard

Shipping disk drives fraught with risk

- **Highly skilled, highly compensated PhD's still wasting hours or days moving data around**
- **Easy to talk about; Hard to implement sensibly**
  - Operational timesink for human staff
  - SOPs for manual ingest often non-existent
  - Does IT even know when this occurs?
  - Are scientists doing MD5 checksums each time a file moves from location A to B?
- **Network based data movement still safer/superior**

# Cloud Storage Strategy Still Required

- **Cloud storage cannot be ignored even by organizations committed to a 100% on-premise footprint**
  - Even at institutions where space/power/cooling costs are hidden or “free” to research users
- **Amazon S3 and similar have become important:**
  - For data ingest and data exchange
  - Long term data dissemination via “downloader pays” model

# Storage: Few trends becoming more real in ‘16

# Reality #1 - Ease of Generation/Access

- The friction required to ACQUIRE or GENERATE vast amounts of scientific data is LOWER than it has ever been
- Far easier to acquire data than to store/manage it
- Petabytes of open-access data available via Internet and legit reasons why a researcher may want to bring significant portions in-house for analysis
- We need governance in place now; Research users must get used to providing biz/science justification for storage consumption

# Reality #2 - Data Diffusion

- We have lost the battle to centralize large-scale data providers and data consumers in managed locations
- Data-intensive instruments will continue to diffuse throughout EVERY part of the organization

# Reality #3 - Avere Systems

- **POCs are turning into procurement**
- **Single product tackles multiple pain points**

1. Make cheap storage fast so I don't have to buy the spendy stuff
2. Multi-vendor NAS behind a single consistent namespace
3. Multi-vendor NAS-to-NAS data migration
4. POSIX interface for local object storage
5. POSIX interface for multi-cloud object storage
6. Multi-cloud replication/sync for HPC "cloud bursting"

# Storage: New/Interesting in 2016

# Storage Related Current Event:

Great post-mortem csc.fi writeup on a massive Lustre storage crash where 1.7 petabytes / 850 million files were lost before being heroically (*mostly*) recovered

Short URL: http://biote.am/bq

Image source: https://csc.fi/image/blogs/entry?img_id=252729

# New: Massive NAS Refresh/Replace

2016 is going to be a VERY INTERESTING year

- **Huge install base of 3-5 year old Scale-out NAS**
- **… all approaching refresh/replace cycle time**
- **BioTeam alone has 4+ large-scale NAS re-assessment projects going on right now - all options open**
- **Drivers coming from all sorts of directions**
  - Object, cloud, commodity SDS, disruptive kit, etc.
- **2016 could see a massive shift in the on-premise storage landscape as organizations reevaluate storage plaforms, plans and architectures**

# New: Storage landscape getting interesting

Very interesting implications for 2016+ storage refresh projects

- **Formally novel/disruptive stuff is gaining acceptance**
  - New crop of scale-out NAS players (Qumulo, etc.)
  - Software-based options (Scality, Rozo, etc.)
  - Next-gen object store options (Igneous, etc.)
  - Flash/SSD pricing becoming viable in our space
- **And the DIY folks keep doing cool things**
  - Example: Pineda lab doing 2PB+ Lustre-on-ZFS green storage at prices lower than the big cloud providers
  - Link: http://biote.am/be

## New: “Data about our Data” is now essential

Critical topic when storage refresh projects being considered

- **STORAGE = CHEAP**
- **METADATA = PRECIOUS**

# Quick Detour
# A few war stories from the “what the heck are we keeping on these systems” battlefront …

# Metrics #1: Robinhood w/ EMC Isilon

About

- **Core storage: Multi-PB EMC Isilon**
- **Method: Robinhood 3.0-alpha1 on a dedicated server with MariaDB on XFS as the database. Full filesystem walk via NFSv3 and auto mount**
- **Produced:**
  - Took **7.8 days** to walk the filesystem and index 283 million entries. Final database size was 175GB
  - Re-indexing takes LONGER due to database interactions

# Metrics #1: Robinhood w/ EMC Isilon

Basic findings

- **273 million files**
  - 1.29 PB total space consumed
  - 5.07 MB average file size
- **3.64 million directories**
- **321,000 symbolic links**

# Metrics #1: Robinhood w/ EMC Isilon

Basic findings, continued

- **Top User owned 327 TB ( ~28% of all data!)**
- **Top 10 Users responsible for ~93 % of all data**
- **Only about 30TB of 1.2PB is really active**
- **Majority of data written is NEVER read**
  - Raw instrument data
  - Copies of public and partner data

# Metrics #1: Robinhood w/ EMC Isilon

## Hard lessons learned

- Gathering this data was hard work but it may drive a fundamental change in the enterprise scientific storage platform
- Concrete evidence that scientists are using expensive Tier-1 storage as a dumping ground for files that are never looked at again
- A ~30TB active data size has interesting implications for future architectures and potential use of SSD/Flash

# Method #2: pwalk + redis vs Isilon/Avere

An interesting noSQL approach to the same problem

- Can’t share results as this work is ongoing right now
- This is a group w/ ~5PB of Isilon and an Avere based client access layer
- This group had previously tried filesystem scanning and SQL data stores but hit the wall at 500 million records
- So they are doing something new …

# Method #2: pwalk + redis vs Isilon/Avere

Basic Method

- **Scan phase:**
  - Filesystem heiarchy broken up into walkable segments
  - Individual pwalk tasks submitted as normal HPC jobs
  - Manual tuning required as directories vary by size and depth
  - Each pwalk job writes to a unique log file

# Method #2: pwalk + redis vs Isilon/Avere

Basic Method

- **Load phase:**
  - pwalk jobs are read in parallel and sent to redis noSQL store
  - Work is chunked to keep redis pegged across SMP cores
  - Primary key = full path to filesystem object/directory/pointer
  - Record = file stats + hash map of parent/child relationships

# Method #2: pwalk + redis vs Isilon/Avere

## Basic Method

- **Analysis phase:**
  - Massive secondary processing within redis to generate vast new sets of keypair:record entries
  - Effectively pre-computing the answer to all desired storage metric questions

# Method #2: pwalk + redis vs Isilon/Avere

Basic Method

- **Future Optimization:**
  - Avere exposes some pretty useful APIs
  - Current idea is to use Avere API to report on which part of the big filesystem is experiencing heavy activity
  - … then spawn pwalk jobs that JUST traverse the recently active files and folders
  - … avoiding or reducing the need for full filesystem crawls

# Getting better “data about our data” - Future

- **Crawling filesystems is painful or not viable above a certain scale or data churn level**
- **Expect to see more of:**
  - Storage platforms that offer native insight/analytics
  - Storage platforms continuing to expose useful APIs
- **Expect commercial offerings as well**
  - BioTeam will be looking at ClarityNow! from dataframeworks.com in 2016

I’m still saying this:

Object storage is the future of scientific data at rest.

Period.

# Storage: Object Is the Future

This what my metadata looks like on a POSIX filesystem:

```
Owner
Group membership
Read/write/execute permissions based on Owner or Group
File size
Creation/Modification/Last-access Timestamps
```

# Storage: Object Is the Future

## This is what I WOULD LIKE TO TRACK on a PER-FILE basis:

What instrument produced this data?
What funding source paid to produce this data?
What revision was the instrument/flowcell at?
Who is the primary PI or owner of this data? Secondary?
What protocol was used to prepare the sample?
Where did the sample come from?
Where is the consent information?
Can this data be used to identify an individual?
What is the data retention classification for this file?
What is the security classification for this file?
Can this file be moved offsite?
etc. etc. etc.
…

# Storage: Object Predictions

- It will still be many years (and a long transition) before we are native object for scientific data at rest (you've got time!)
- Storage refresh projects will drive testing and POCs in 2016+
  - Metadata, cost and erasure coding will be the primary drivers
- Expect to see
  - A lot of POSIX/Object gateways and transition products
  - Data or image-heavy commercial software platforms will begin to support native object stores

# Trends: Cloud

A few new things …

# Cloud: Recap

- Core drivers behind cloud adoption remain unchanged
  - Remember, it's about CAPABILITY not COST for many
  - Cloud footprints being designed for security, data exchange and multi-party collaboration
- In 2015-16 BioTeam has seen a large increase in
  - Hybrid cloud architectures
  - Multi-cloud design patterns and architectures
  - Clients procuring direct cloud connectivity (1Gbps, 10Gbps)

# Cloud: New for 2016

- BioTeam is predicting a new era of CLOUD SOBRIETY
- We expect to see a few major cloud pullbacks in 2016-2017
- Root cause:
  - C-level executives ordering “cloud first” without doing the math
  - Long-time cloud users now have actionable cost & spend data
  - Commodity price trends for on-premise kit looking more interesting when cloud storage bill exceeds $100,000/month

# Cloud: New for 2016

- The cloud pullback/clawback projects will be interesting
- Some of them may go elsewhere rather than back to on-premise
- Interesting movements within the traditional large supercomputing centers. Some offering services to Industry in order to obtain new revenue/funding/cost-recovery streams
- 2016-2017 may see workloads pulled from IaaS cloud platforms and redeployed at places like NCSA iForge rather than on-premise or in a standard colo suite

# Cloud: New for 2016

- Starting to see interesting new criteria for Cloud vendor selection
- Now you have to ask these questions:
  - *“What are the chances that my cloud provider is going to start a pharmaceutical or healthcare company that will directly compete with me?”*
  - *“Are product teams monitoring my API usage? Who are they communicating with? Do they use this information to make my life easier or are they collecting competitive intelligence that may be used to compete against me in the future?”*
- My cynical (personal) opinion: Cloud providers must be screened for competitive risk in 2016 and beyond. I’m getting concerned & suspicious.

# Transition & Wrap-up Time.

Lots of recycled slides from last year

Keeping them because it’s the most important section of the talk.

This really is going to be our hardest challenge over the next few years

# A confluence of unfortunate events …

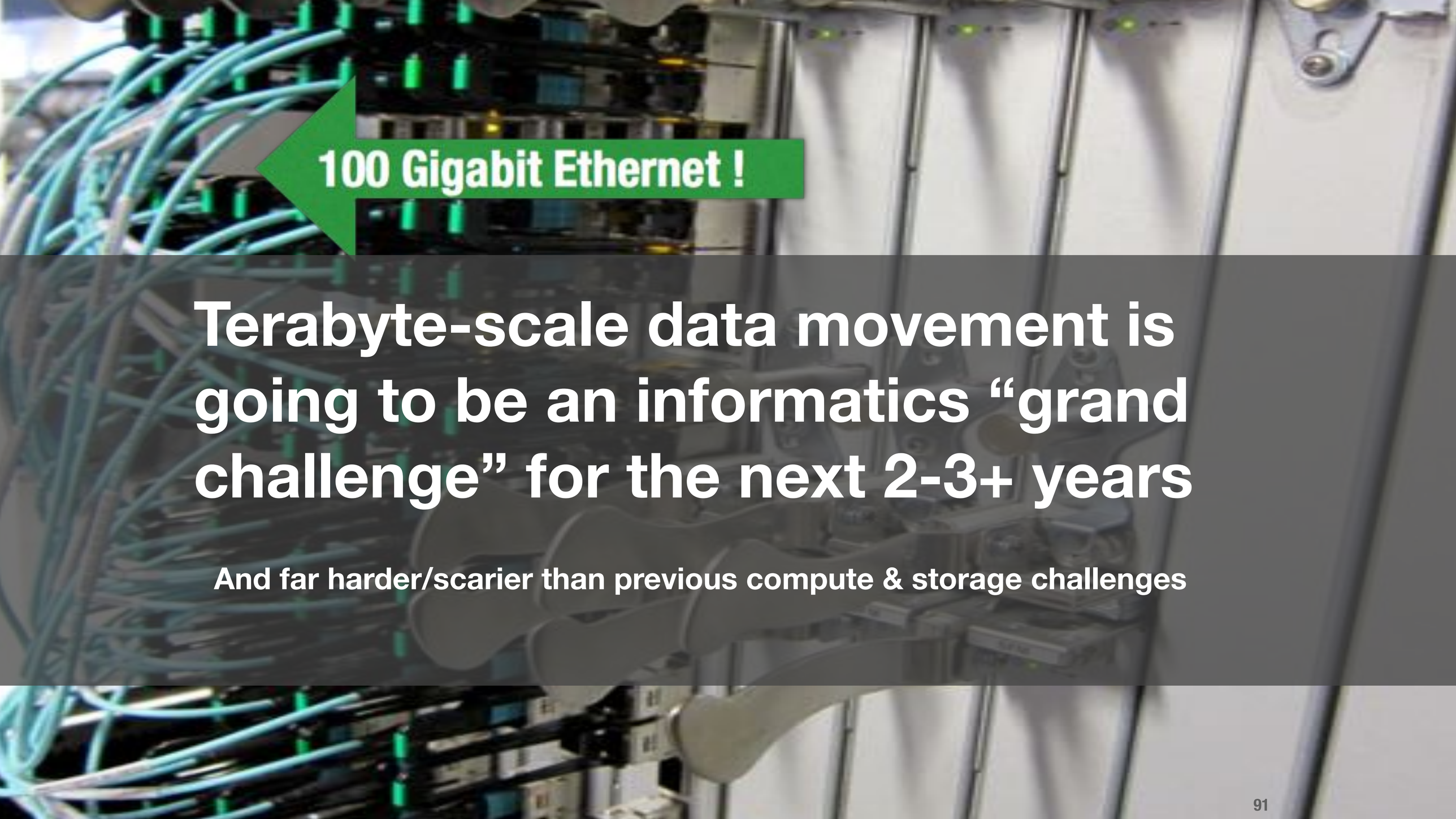
- Terabyte-scale wet lab instruments are everywhere
- Centralizing high-data-rate instruments = umpossible :)
- Petabytes of interesting open-access data on Internet
- Petabytes of interesting data w/ collaborators

- Tera|Petabyte scale “data spread” seems unavoidable
- Known: data will span sites, buildings and clouds

# Data Intensive Science Will Require

- Ludicrous bandwidth @ the network core
- Switches & routers that don't drop packets
- Massive bandwidth delivered to Top of Rack
- Very fast (40g or 100g) to campus buildings
- Very fast (10g or 40g) to lab spaces
- 1Gbps or 10Gbps Internet/Internet2 access
- Security controls that can handle single-stream 10Gbps + data flows without melting down

# It all boils down to …




# Terabyte-scale data movement is going to be an informatics “grand challenge” for the next 2-3+ years

And far harder/scarier than previous compute & storage challenges

# Issue #1

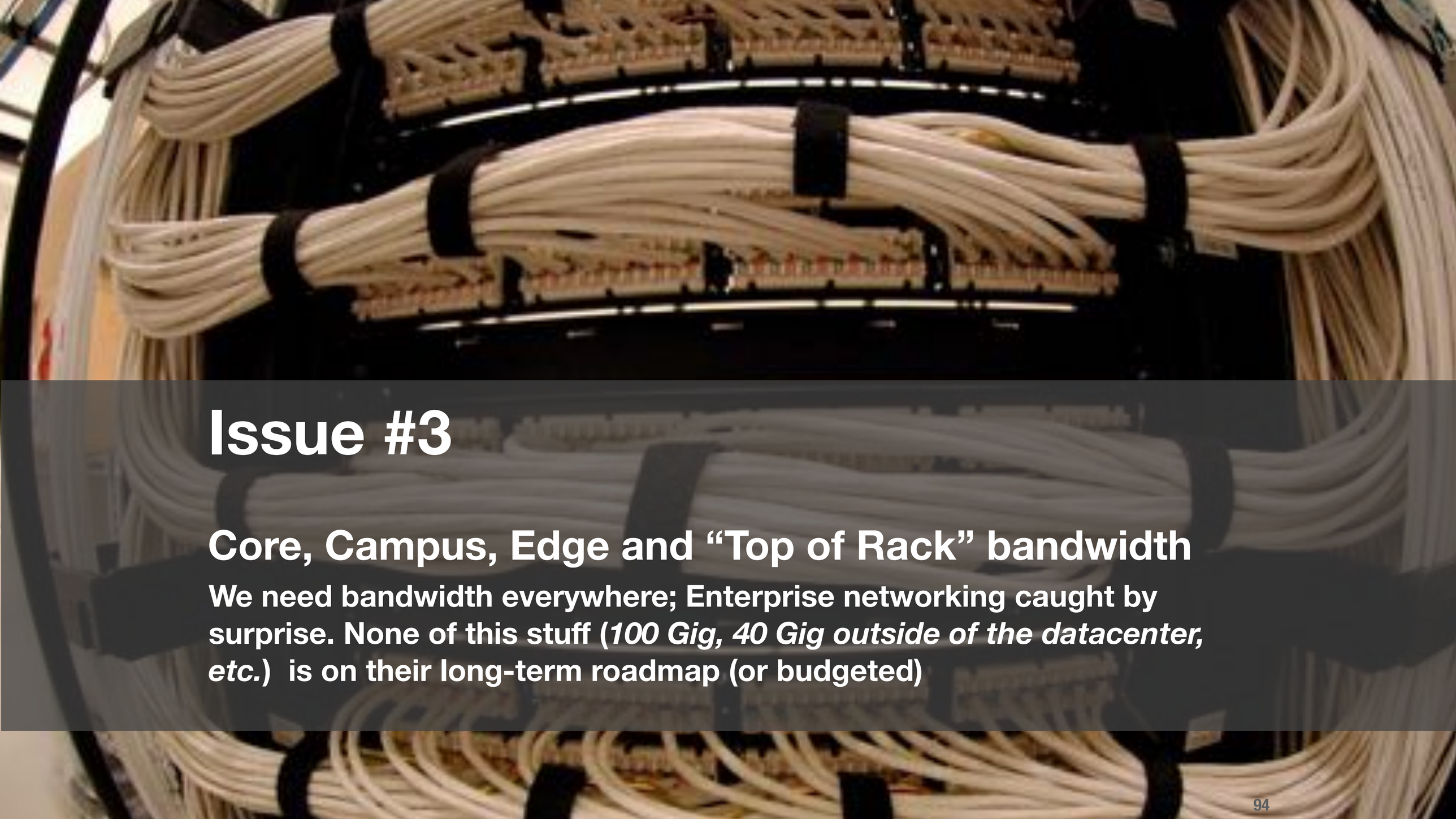
## Current LAN/WAN stacks bad for emerging use case

**Existing technology we've used for decades has been architected to support many small network flows; not a single 'elephant' data flow**

# Issue #2

## Ratio of LAN:WAN bandwidth is out of whack

We will need faster links to “outside” than most organizations have anticipated or accounted for in long-term technology planning



# Issue #3

## Core, Campus, Edge and “Top of Rack” bandwidth

We need bandwidth everywhere; Enterprise networking caught by surprise. None of this stuff (*100 Gig, 40 Gig outside of the datacenter, etc.*) is on their long-term roadmap (or budgeted)

# Issue #4

## Bigger blast radius when stuff goes wrong

Compute & storage can be logically or physically contained to minimize disruption/risk when Research does stupid things.

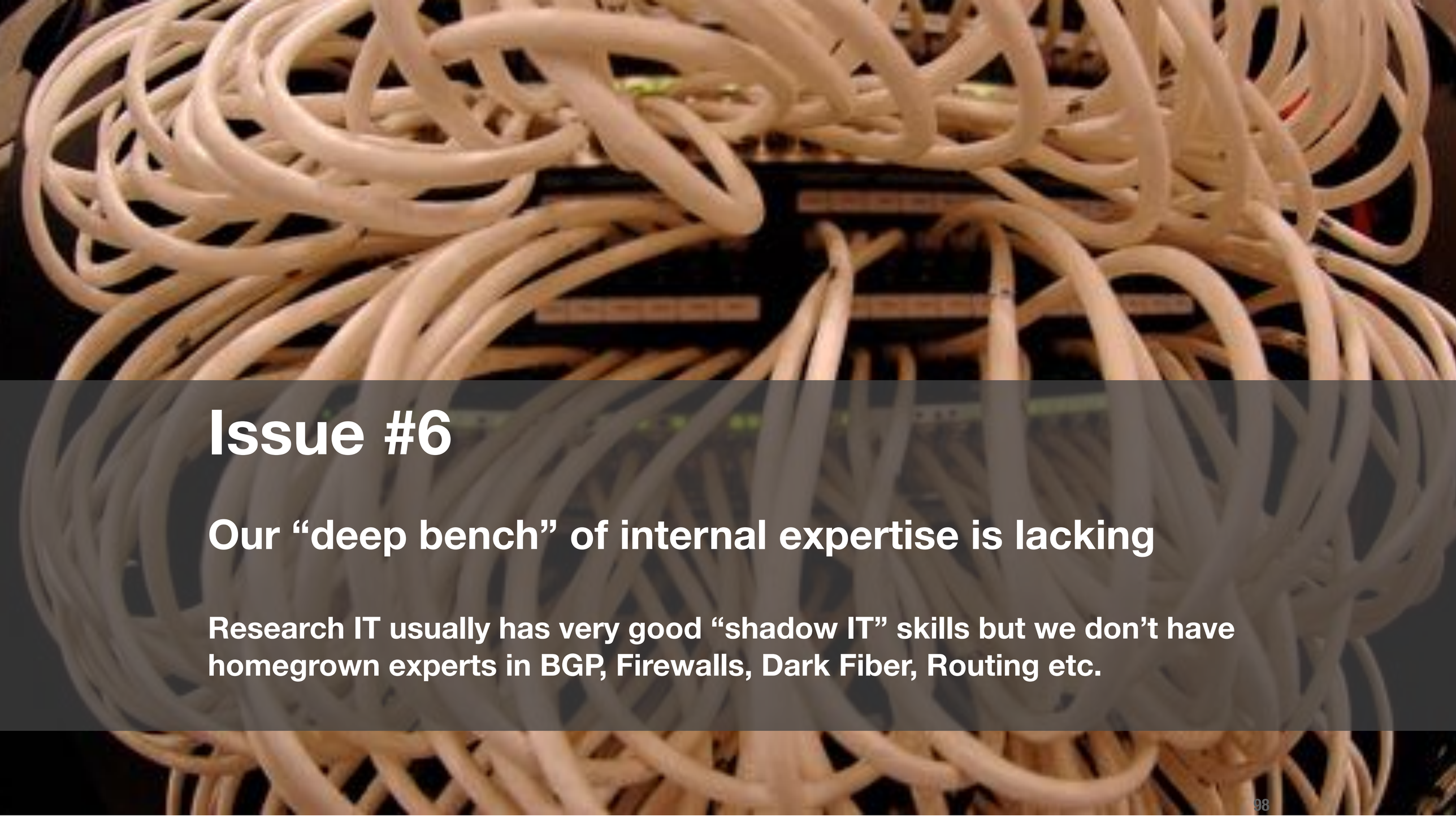
Networks, however, touch EVERYTHING EVERYWHERE. Major risk.

# Why this will be difficult to achieve

# Issue #5

## Social, trust & cultural issues

We lack the multi-year relationship and track record we've built with facility, compute & storage teams. We are "strangers" to many WAN and SecurityOps types



# Issue #6

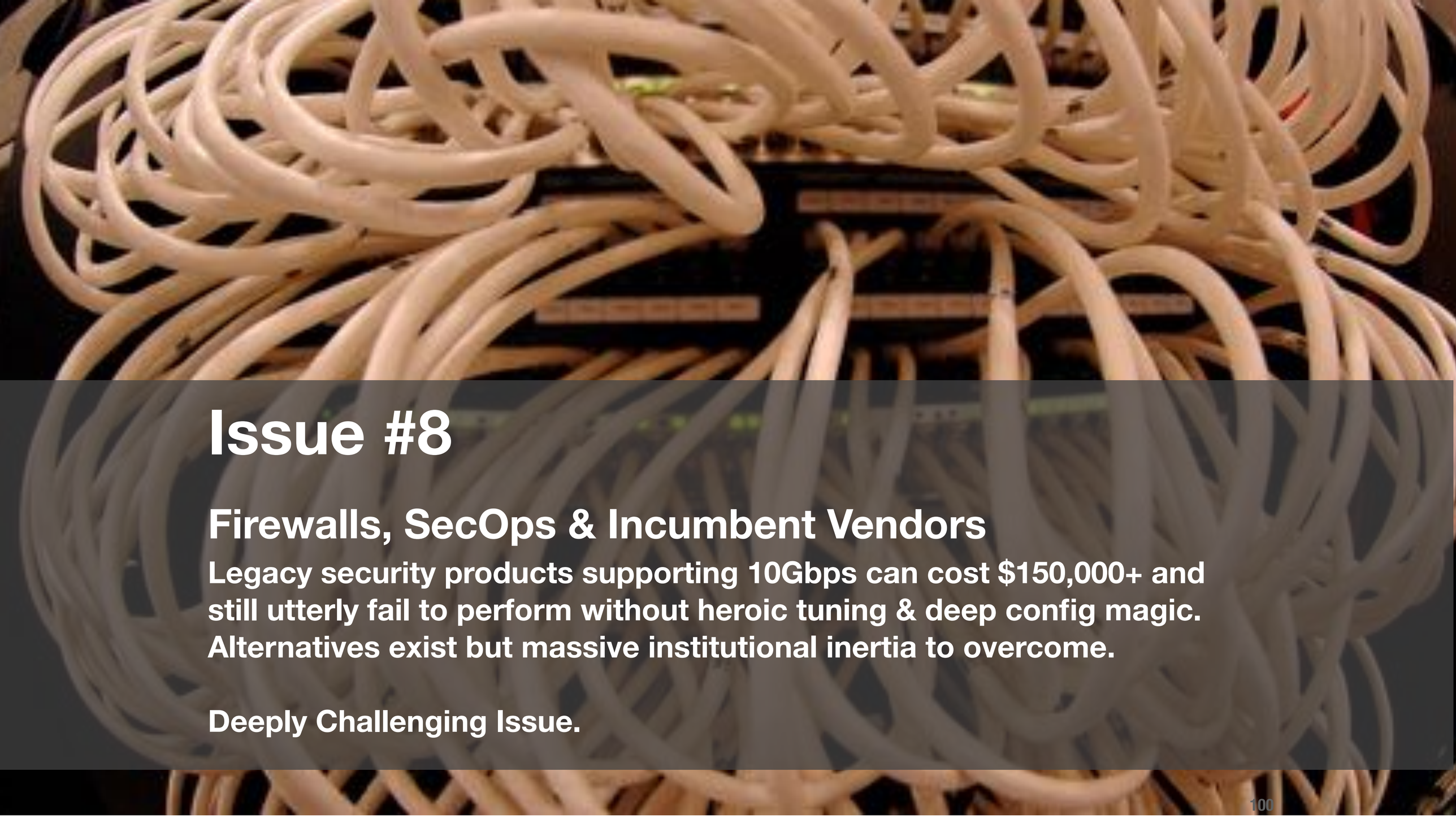
## Our “deep bench” of internal expertise is lacking

Research IT usually has very good “shadow IT” skills but we don’t have homegrown experts in BGP, Firewalls, Dark Fiber, Routing etc.

# Issue #7

## Cost. Cost. Cost.

Have you seen what Cisco charges for a 100Gbps line card?



# Issue #8

## Firewalls, SecOps & Incumbent Vendors

**Legacy security products supporting 10Gbps can cost $150,000+ and still utterly fail to perform without heroic tuning & deep config magic. Alternatives exist but massive institutional inertia to overcome.**

**Deeply Challenging Issue.**

# Gulp.
# So what do we do?

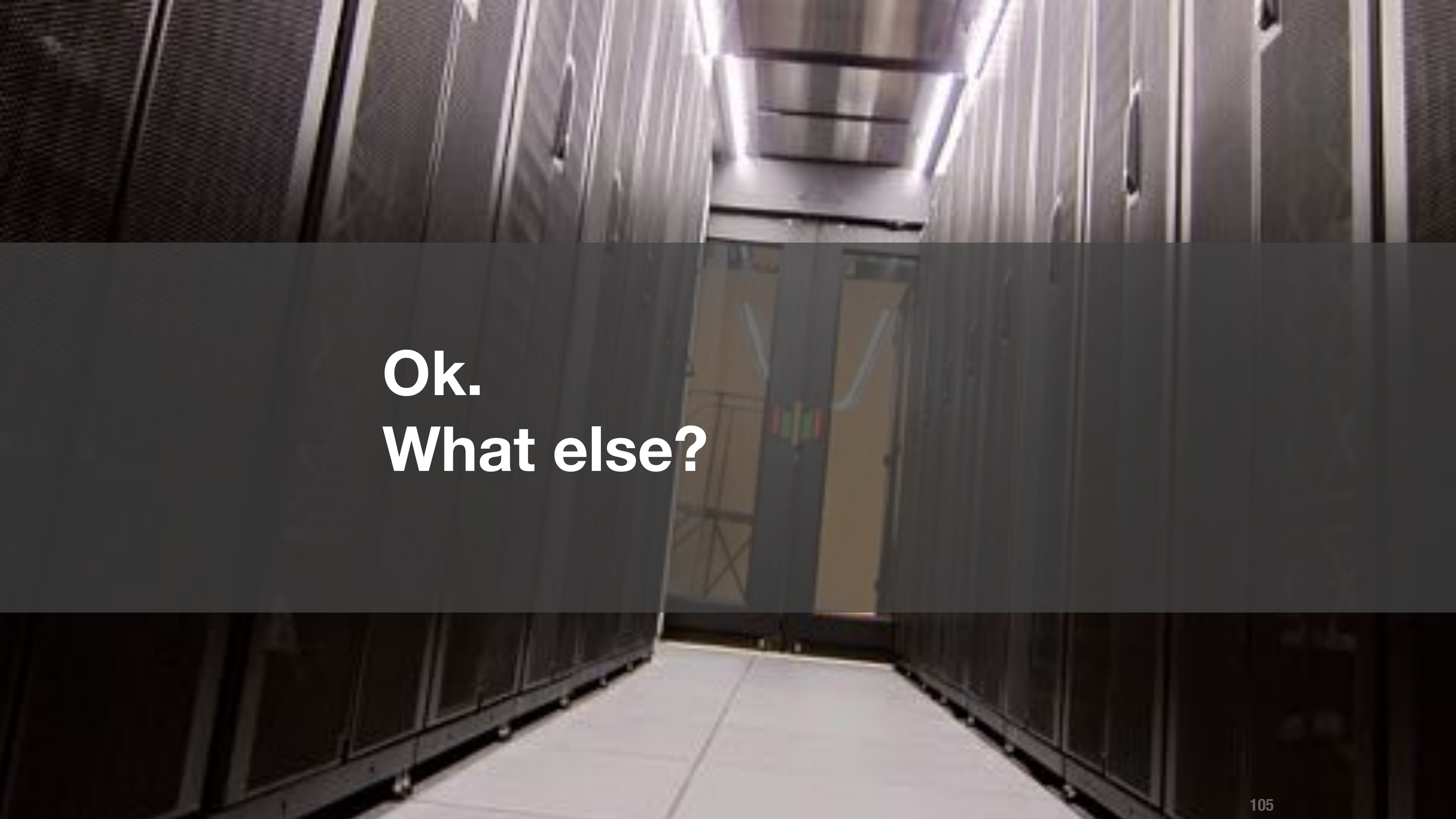
- Acknowledge that we are entering a new chapter in an era of “data intensive science”
- Understand that we now have high scale data producers and consumers diffusing everywhere
  - *… Including clouds, partners and collaborators*
- Recognize that new methods are needed and that this may include new vendors/platforms

# Advice #1

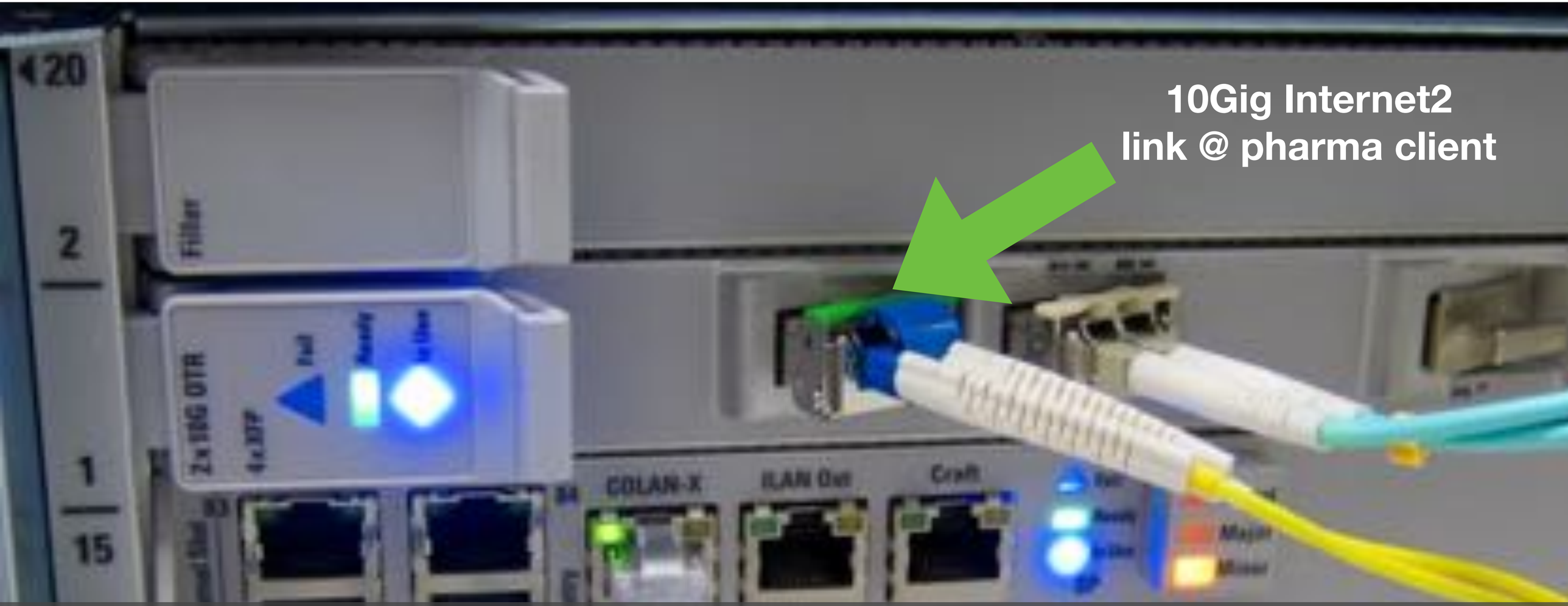
- Engage ASAP with enterprise networking, WAN and security teams - they now need to be part of the scientific computing family
- Survey networking roadmap, budget and capabilities
  - *Historically network engineering groups are under-resourced relative to Research*
- [Research Leadership]  If you want to make new friends in networking, become the outside champion arguing that they need more resources

# Advice #2

- Survey your own labs, floors and buildings to document the large-scale data producers and consumers
- Talk to PIs about their instrument procurement plans
- Talk to business leadership to understand collaboration efforts and use of external partners or data providers
- Actively seek out the folks moving data on physical disks



Ok.
What else?




# Internet2 Is Open for Business

# Internet2 and High Speed R&E Networks

- Did you see Dan's talk yesterday?
- Since 2015 we have seen significant increase in # of BioTeam clients choosing to make use of Internet2 for high speed access to collaborators and clouds
  - At least ~6 now spanning commercial and US.gov
- Pro: 100 Gig network purpose built for data intensive science
- Con:  I2 is not an ISP so connecting is non-trivial, especially if you want Layer-3 access and don't have a /24 public IPv4 block to spare

And this brings us to …

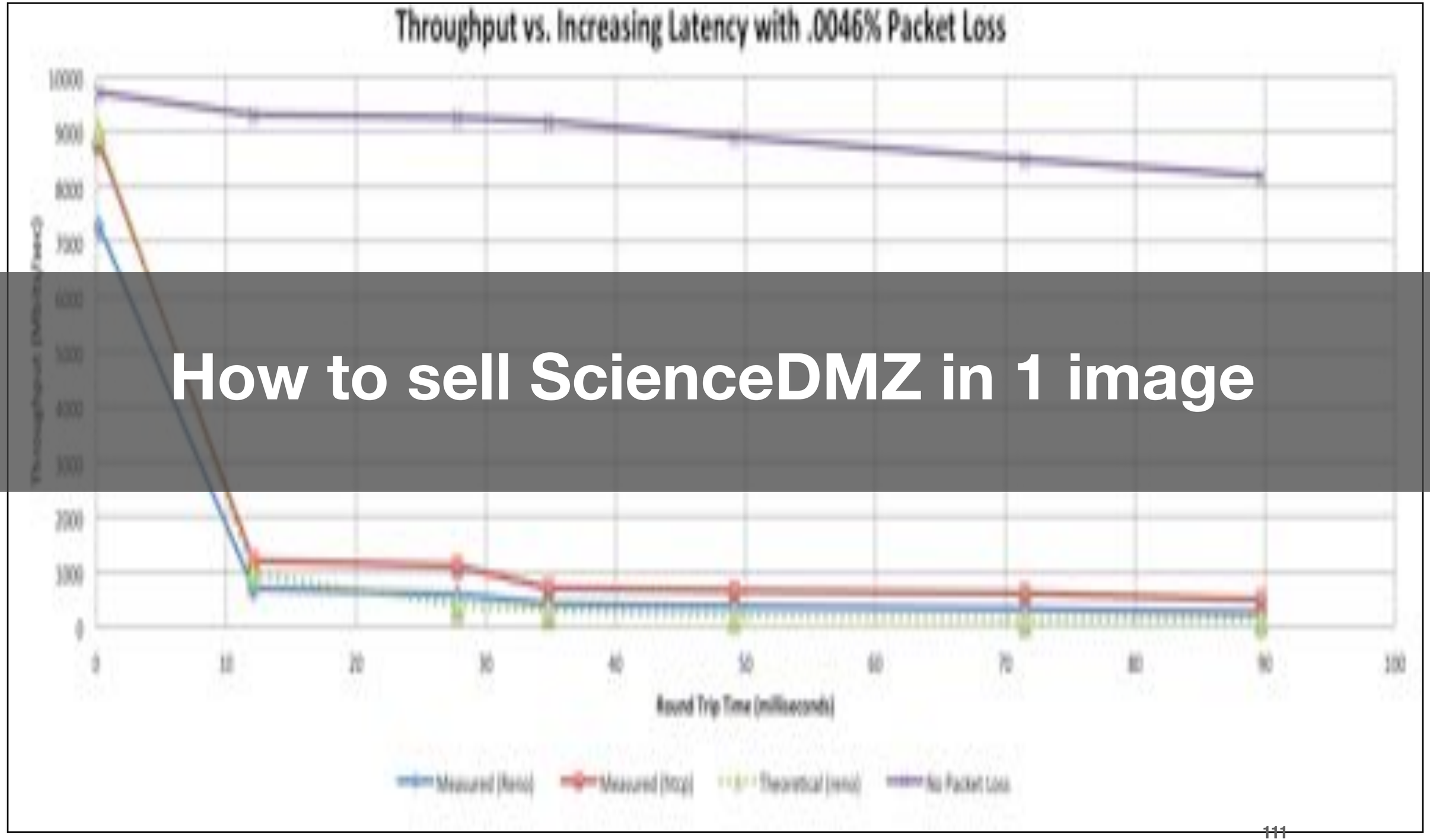
# ScienceDMZ

# ScienceDMZ #1

- The “data intensive science” era may finally trigger a process by which organizations logically or physically separate the “business network” from the “science/data” network
- Huge deal. This is a fundamental shift in practices that have largely been static for decades

# ScienceDMZ #2

- Other fields (High Energy Physics, etc.) have faced this problem before and solved it
- The philosophy, concepts and methods they use have been broadly grouped under the descriptive term “ScienceDMZ”
  - Evangelized by nice folks at http://fasterdata.es.net
- In use today in large scale production environments. No BS. It works.




# How to sell ScienceDMZ in 1 image

# Practical ScienceDMZ Elements

- Keep large flows off of business networks
- Build TCP networks that do not drop packets
- Instrument at multiple locations via perfSONAR
- Use grown-up xfer tools, not FTP, RSYNC or SCP
- Completely rethink how network security and intrusion detection will be performed

# What BioTeam is doing

- Our second dedicated Converged IT summit will be in San Diego this year - October 24-26, 2016
- We have a 100 Gig test lab hosted by TACC in Austin, TX and supported by many of our vendor friends - including a 100 gig Internet2 link
- We intend to build, test, benchmark and talk about various types of ScienceDMZ network designs
- We intend to share this info freely and (ideally) open our lab up for others to use/test/research

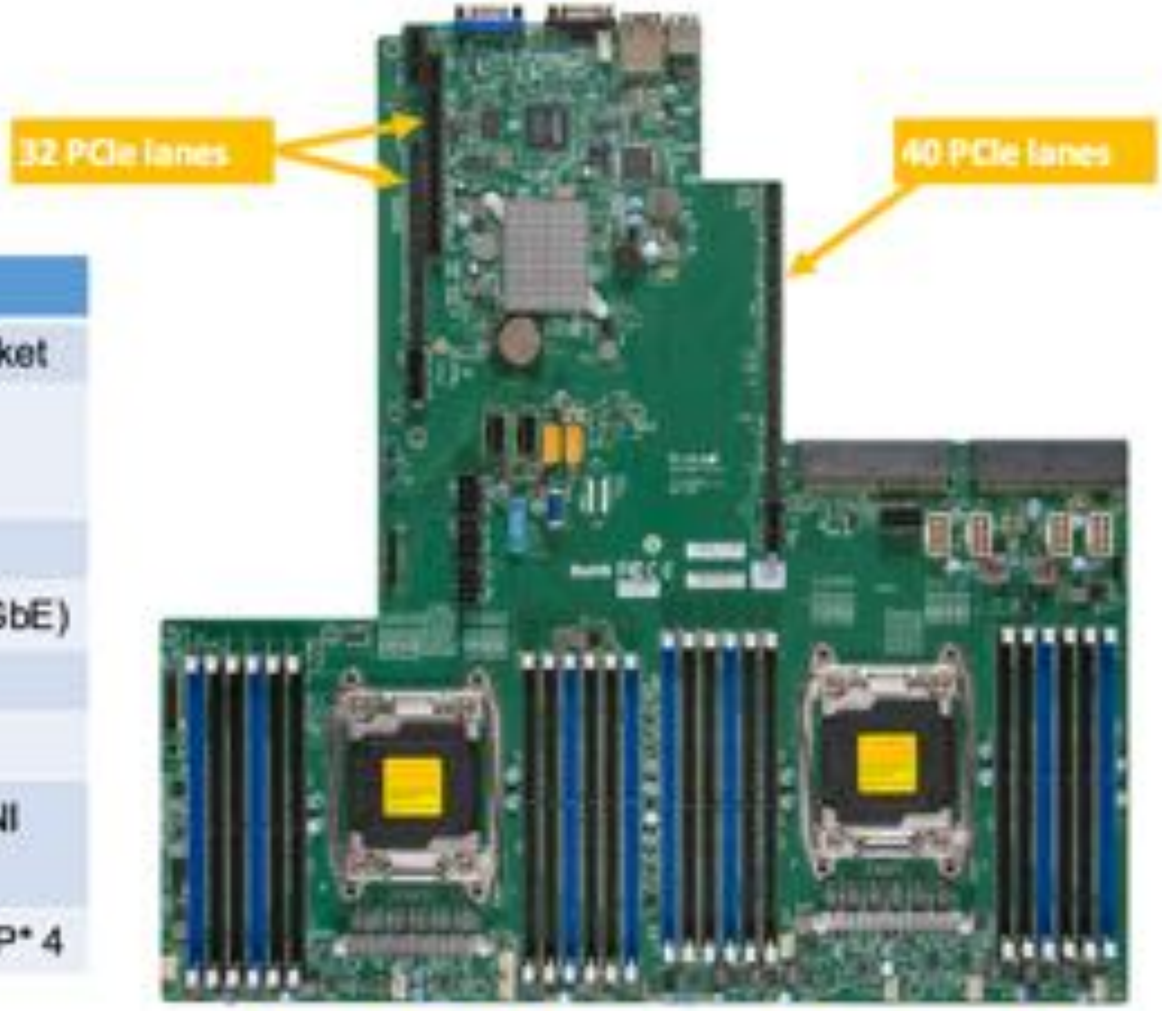
# Happening right this instant …

## BioTeam + Intel + Aspera + Internet2 100 Gbps data transfer trials

- BioTeam 100 Gig testbed @ TACC in Austin, TX
- 2x Intel servers at each end of the Internet2 VLAN, each with 3x 40GB NICs
- 100 Gig circuit from TACC to NSCA via Internet2
- Aspera data-xfer software + Intel DPDK tech
- Goal: See how fast we can move real data server-to-server across a 100 Gig link

# Data Transfer Node (DTN) Hardware Overview

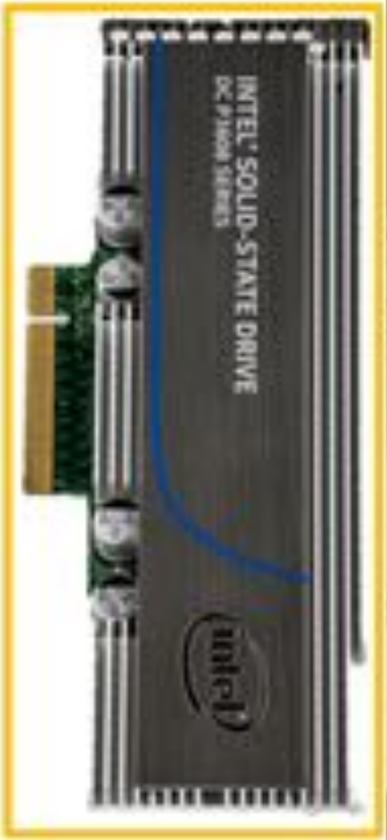
| SYSTEM CONFIGURATION | |
|---|---|
| System: | SuperMicro X10DRU-i+ motherboard, 2-socket |
| CPU: | Intel® Xeon® processor E5-2687W v3, 10 cores, 3.10GHz, 25M Cache |
| RAM: | 256GB DDR4 memory (2133 MHz) |
| NIC: | 3 x 1-port Intel® XL710 Enet Controller (40GbE) |
| Storage: | 4 x Intel® P3608 4TB SSDs |
| Switch: | Juniper 100Gbps QFX-5100 switch. |
| Software: | Intel® DPDK 2.2, Intel® DDIO, Intel® AES-NI GCM CentOS 7.1 |
| | Development version based on Aspera FASP* 4 |

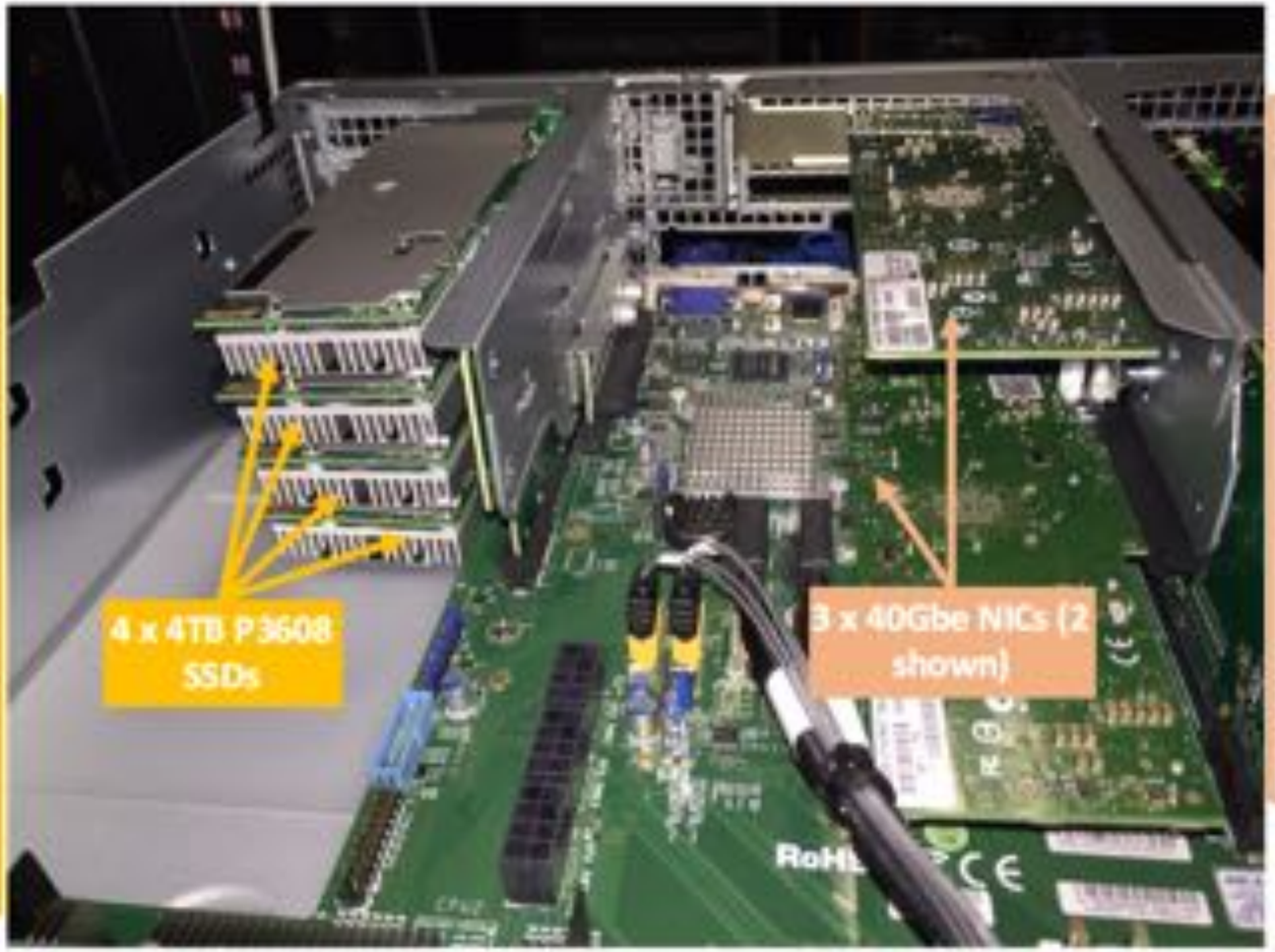



SuperMicro X10DRU-i+ motherboard

1

# Inside Look at the Data Transfer Node (DTN)



2

# History of DTN Attempts

| ATTEMPT | NODE | SSD | CPU | RAM | NIC | GOAL | RESULT |
|---|---|---|---|---|---|---|---|
| **BioIT 2013** | Intel motherboard | 910 (2) 800GB | E5-2687W 3.1GHz | 32GB | X540 (1) 10GbE | ≤10Gbps | 9.50Gbps |
| **SC2013** | Intel motherboard | S3700 (12) 800GB | E5-2650v2 2.6GHz | 128GB | X520 (2) 10GbE | ≤40Gbps | 37.75Gbps |
| **SC2014** | Intel motherboard | P3700 (5) 800GB | E5-2697v3 2.3GHz | 128GB | XL710 (2) 40GbE | ≤80Gbps | 65.41Gbps |
| **BioIT 2016** | Supermicro motherboard | P3608 (4) 4TB | E5-2687Wv3 3.1GHz | 256GB | XL710 (3) 40GbE | ≤96Gbps | **??? (Visit our booth!)** |

- Current status
  - Still sorting out hardware issues and tuning the OS, NIC and driver stack for best possible throughput
  - Currently seeing 73Gbit/sec Memory-To-Memory
  - Will do disk-to-disk at BioIT’16 conference this week
  - Very interesting process and some unexpected results
    - Thermal management is apparently a big deal
    - Testing teased out non-trivial hardware issues: disk, PCI risers, NICs and multiple motherboards

6-question online survey is here:

http://biote.am/survey16

*This scannable QR code links to the survey*

# end; Thanks!