ABD310

AWS re:INVENT

Big Data, AWS, and Security
How FINRA Secures Its Big Data and Data Science Platform on AWS

Vincent Saulys
David Yacono

November 29, 2017

re:Invent

aws

Who is FINRA?

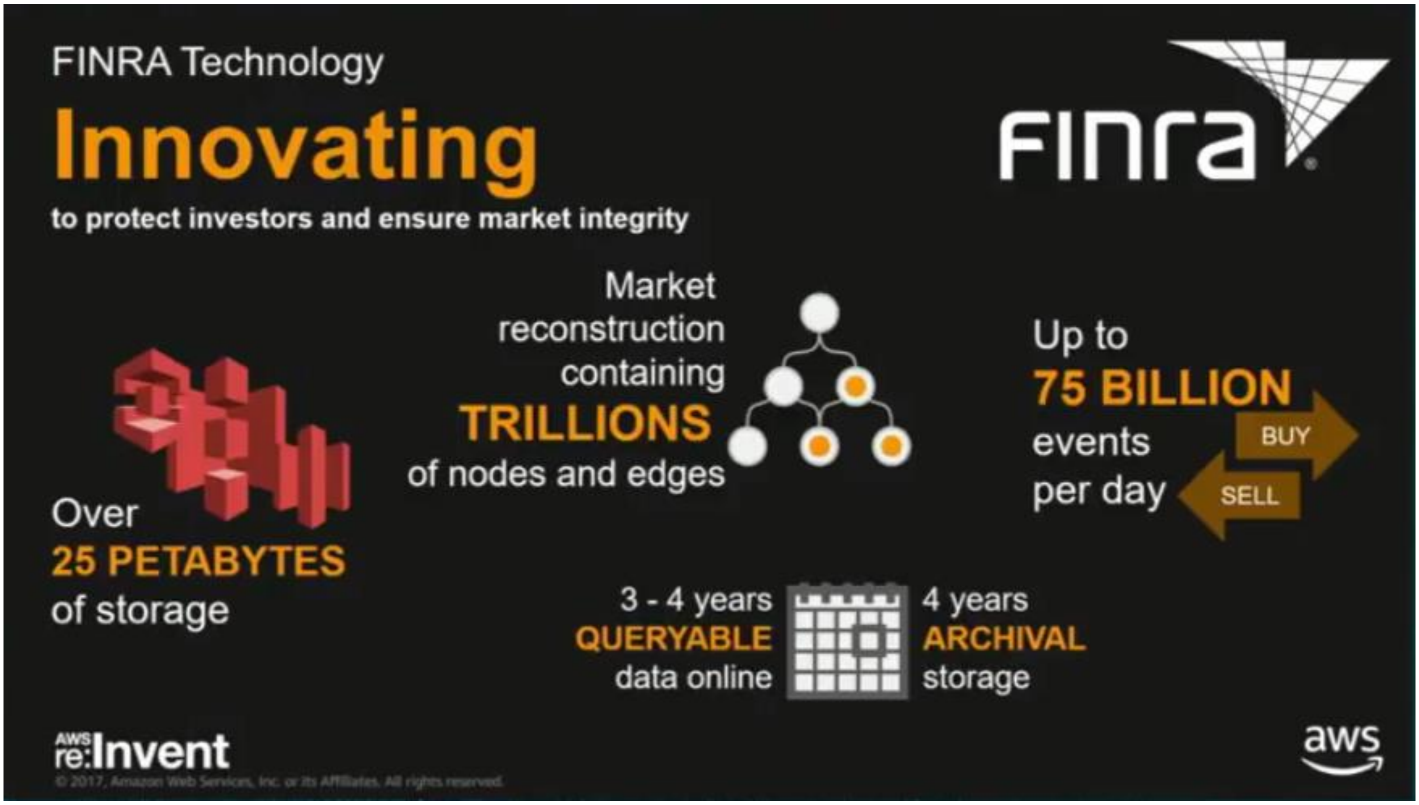
- Financial Industry Regulatory Authority.
- Our Mission: "Investor Protection—Market Integrity."
- We are a private sector not-for-profit organization authorized by Congress to protect America's investors.
- We do this by:
  - Writing and enforcing rules that govern the activities of 3,800 broker-dealers with 634,000 brokers.
  - Examining firms for compliance with those rules.
  - Fostering market transparency.
  - Educating investors.

And most significant to this discussion:

- FINRA uses big data and data science technologies to detect and analyze fraud, market manipulation, and insider trading across US capital markets.

re:Invent

aws



FINRA Technology

Innovating

to protect investors and ensure market integrity

FINRA

Over **25 PETABYTES** of storage

Market reconstruction containing **TRILLIONS** of nodes and edges

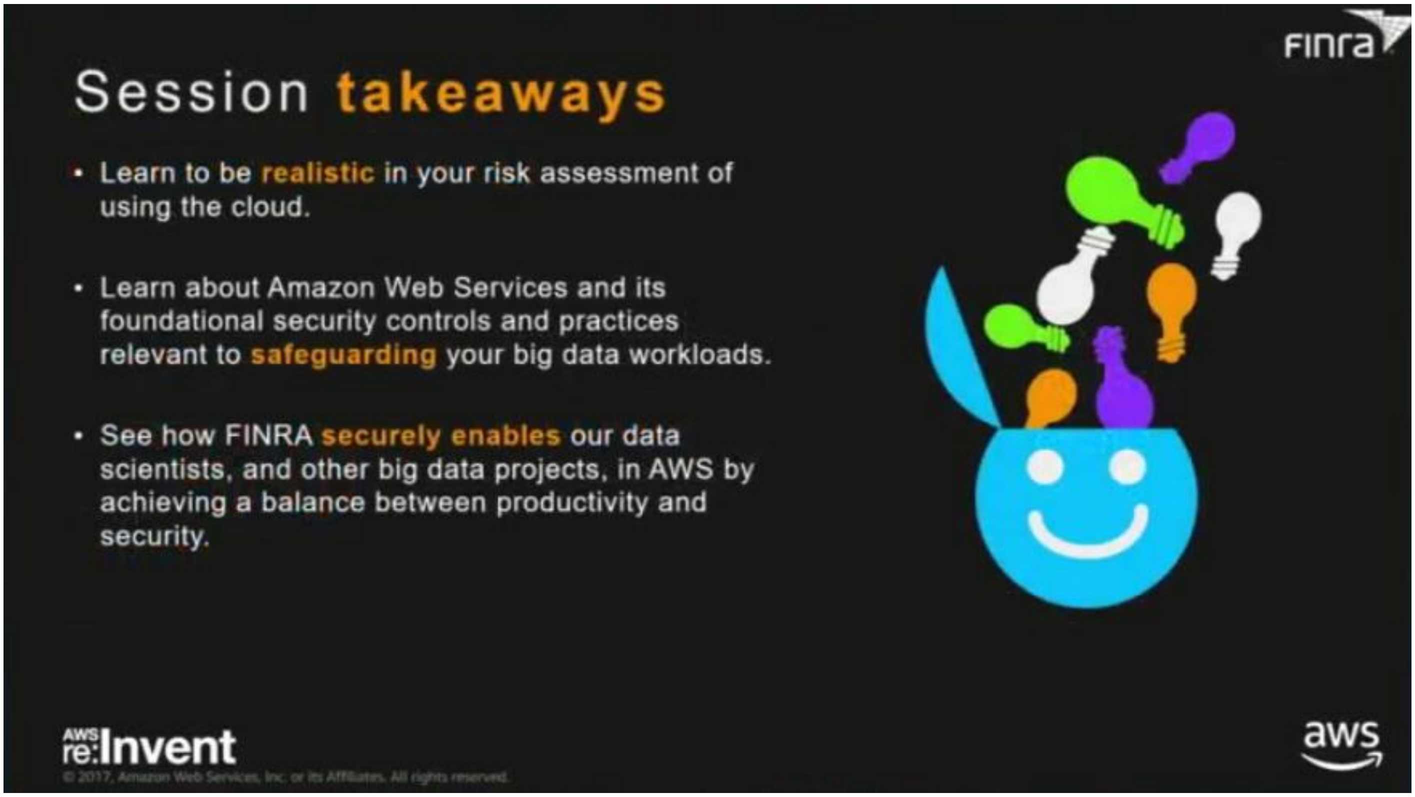
Up to **75 BILLION** events per day

BUY
SELL

3 - 4 years **QUERYABLE** data online

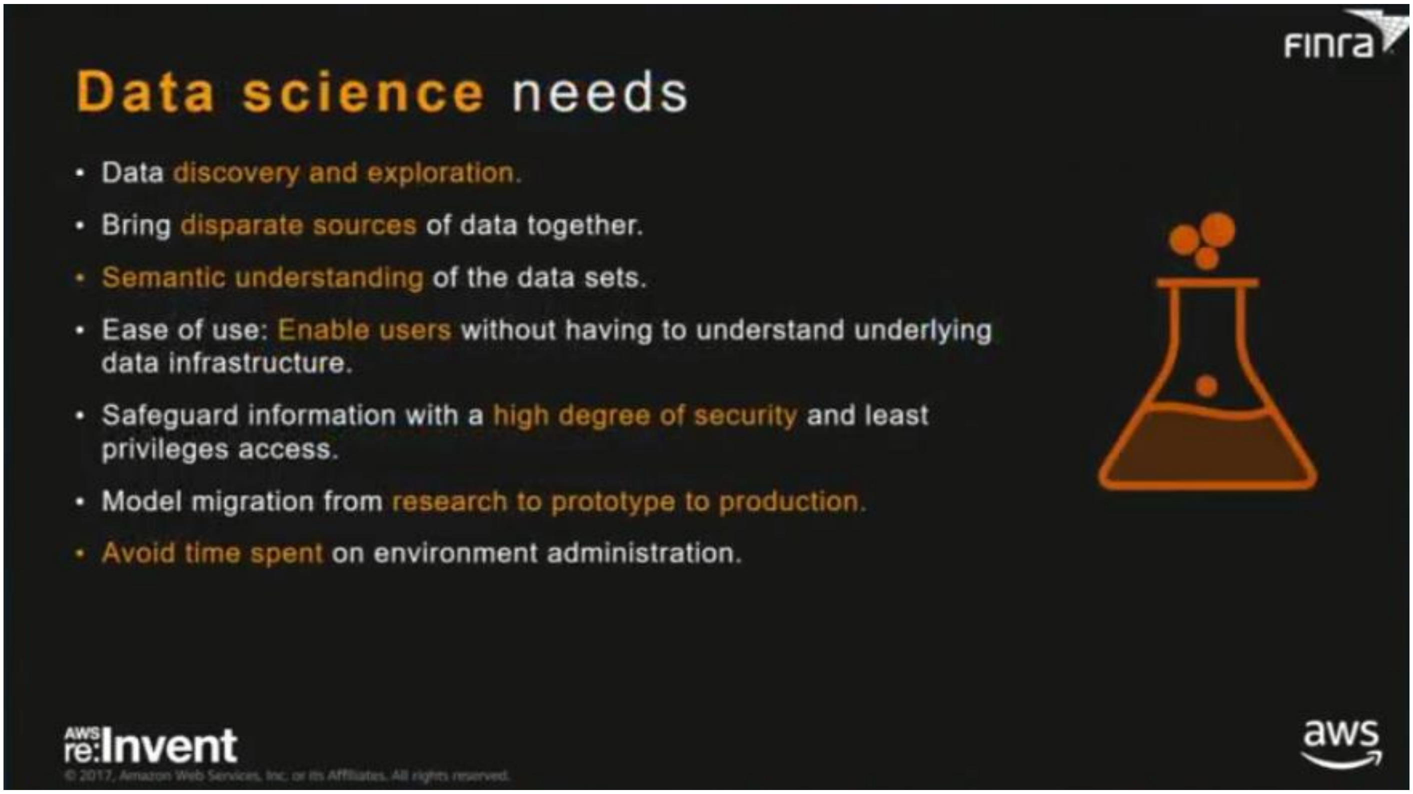
4 years **ARCHIVAL** storage

AWS re:Invent
aws



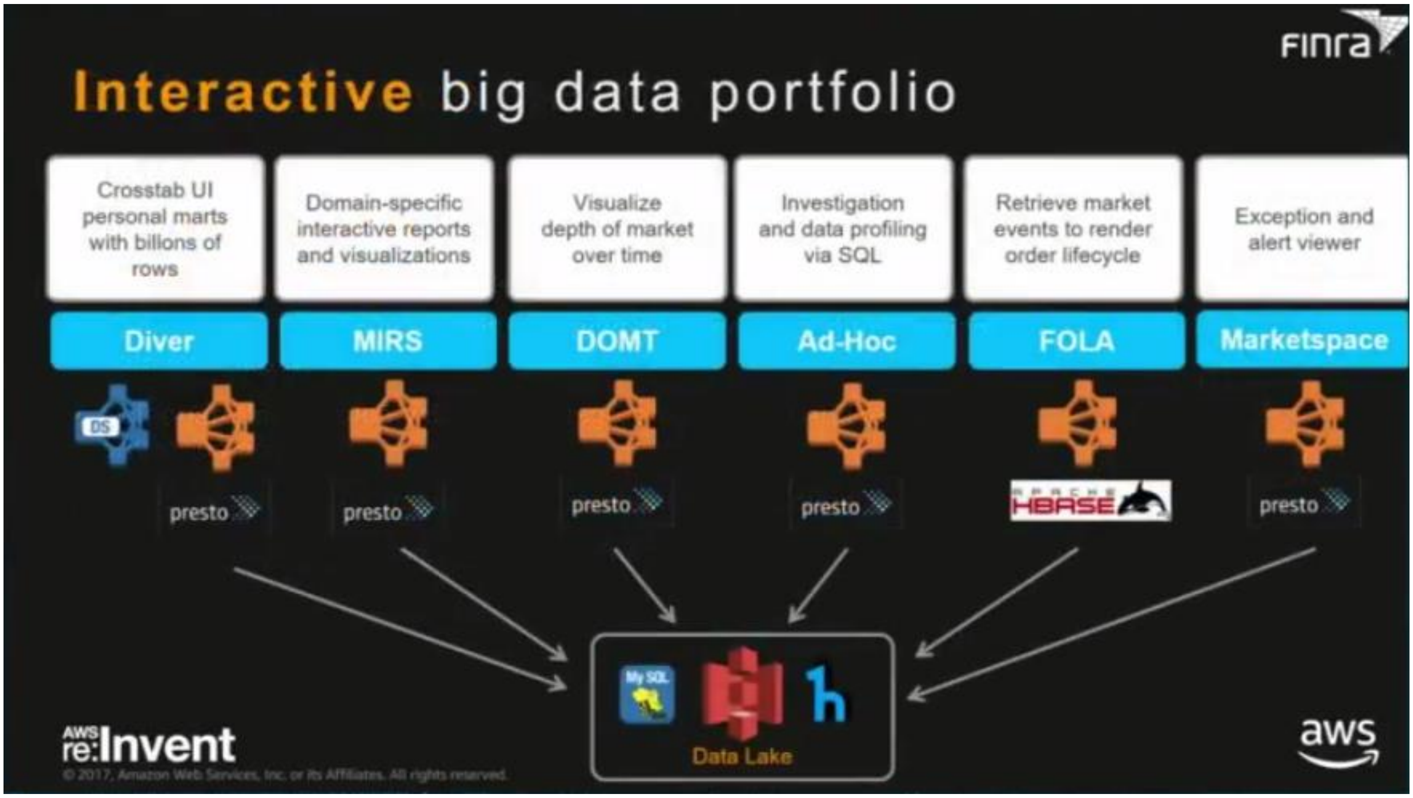
Session takeaways

- Learn to be **realistic** in your risk assessment of using the cloud.
- Learn about Amazon Web Services and its foundational security controls and practices relevant to **safeguarding** your big data workloads.
- See how FINRA **securely enables** our data scientists, and other big data projects, in AWS by achieving a balance between productivity and security.



Data science needs

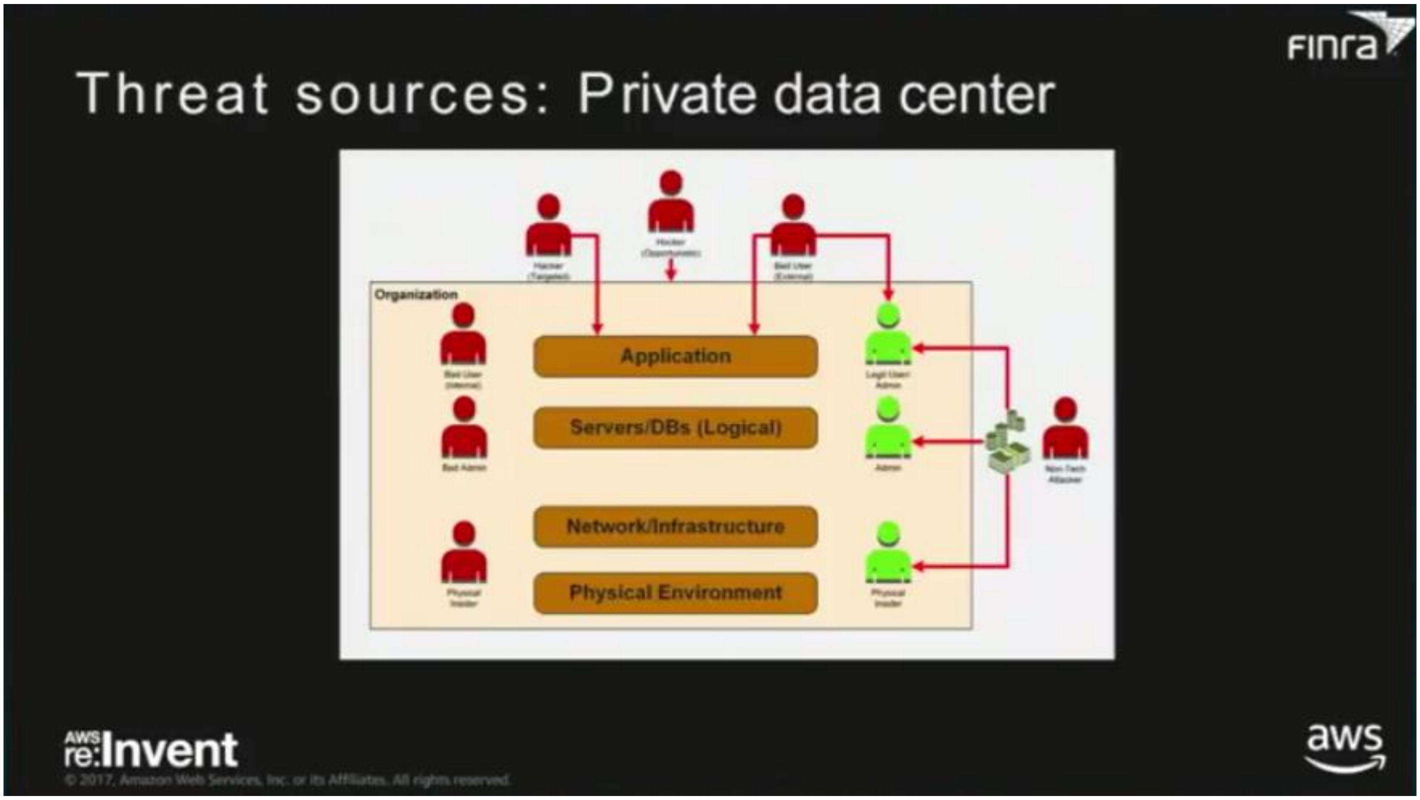
- Data discovery and exploration.
- Bring disparate sources of data together.
- Semantic understanding of the data sets.
- Ease of use: Enable users without having to understand underlying data infrastructure.
- Safeguard information with a high degree of security and least privileges access.
- Model migration from research to prototype to production.
- Avoid time spent on environment administration.



Interactive big data portfolio

Crosstab UI personal marts with billons of rows
Diver
Domain-specific interactive reports and visualizations
MIRS
Visualize depth of market over time
DOMT
Investigation and data profiling via SQL
Ad-Hoc
Retrieve market events to render order lifecycle
FOLA
Exception and alert viewer
Marketspace
DS
presto
presto
presto
presto
APACHE HBASE
presto
My SQL
h
Data Lake

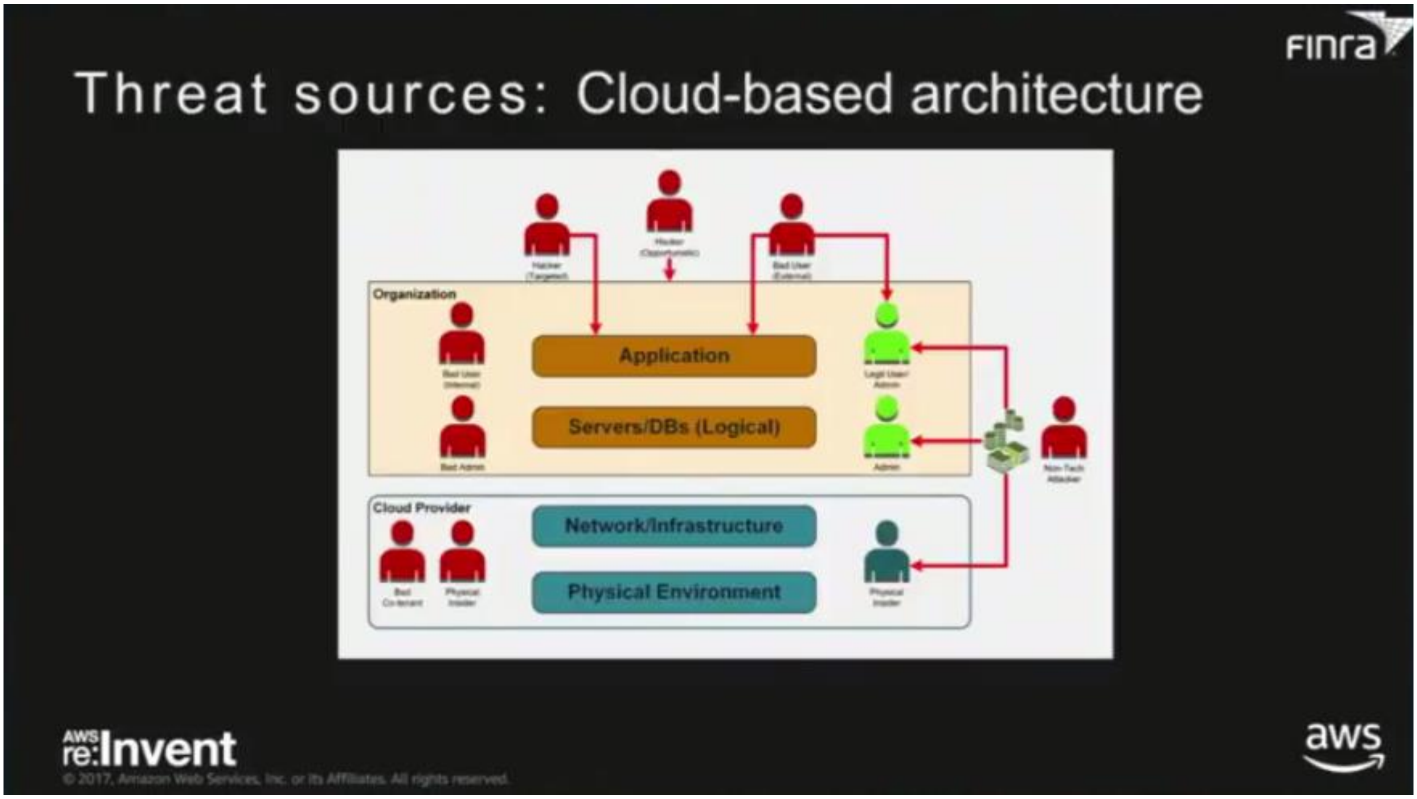
These are all the query engines that we built and the applications that we have built on top the engines. We built our own data lake (~25 Petabytes) using S3, we also have our OSS data catalog called Herd that helps store the metadata about our data including all previous data history. We generally use Presto (and HBase) to query the data lake from the many apps we build.



Threat sources: Private data center

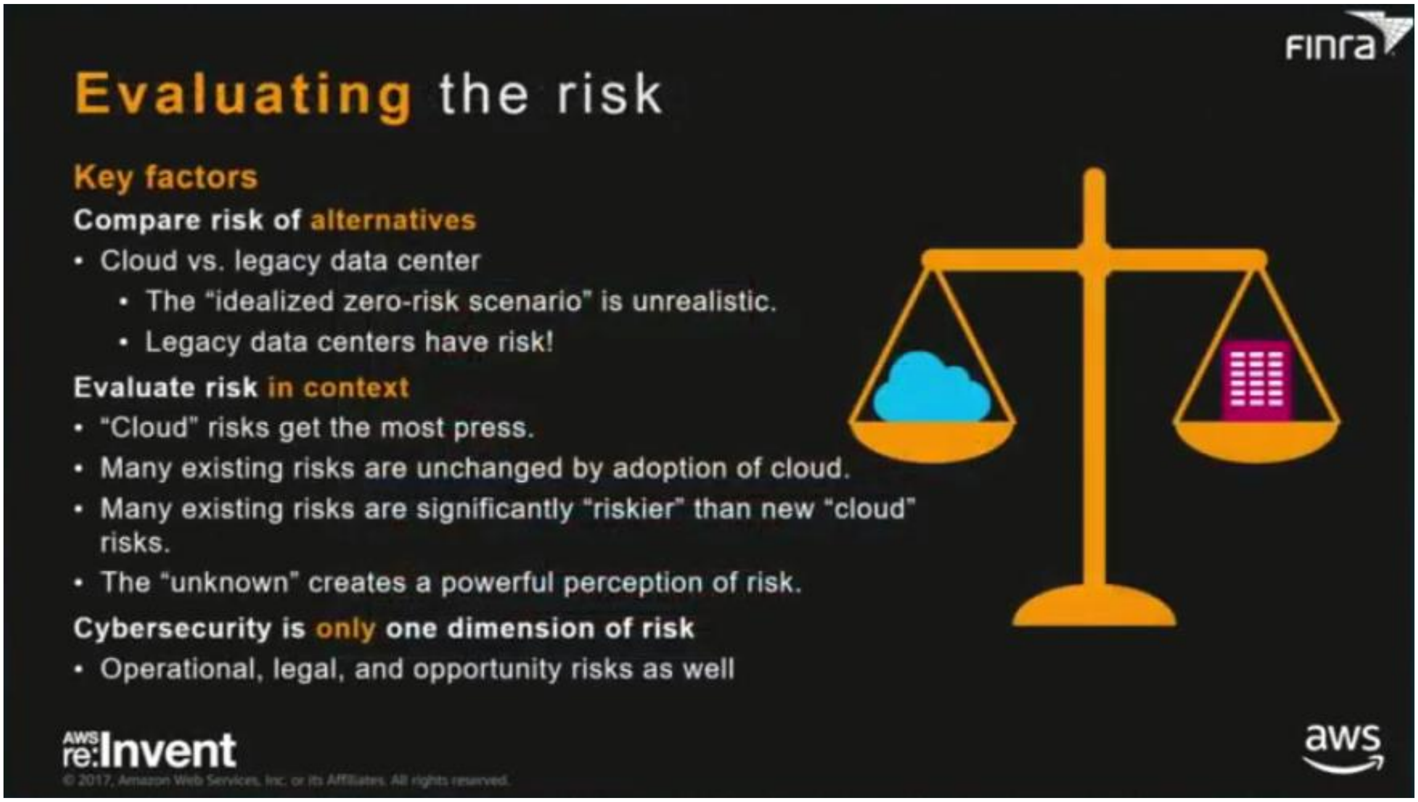
re:Invent

There are security needs for big data.



Threat sources: Cloud-based architecture

re:Invent



Evaluating the risk

Key factors

Compare risk of alternatives

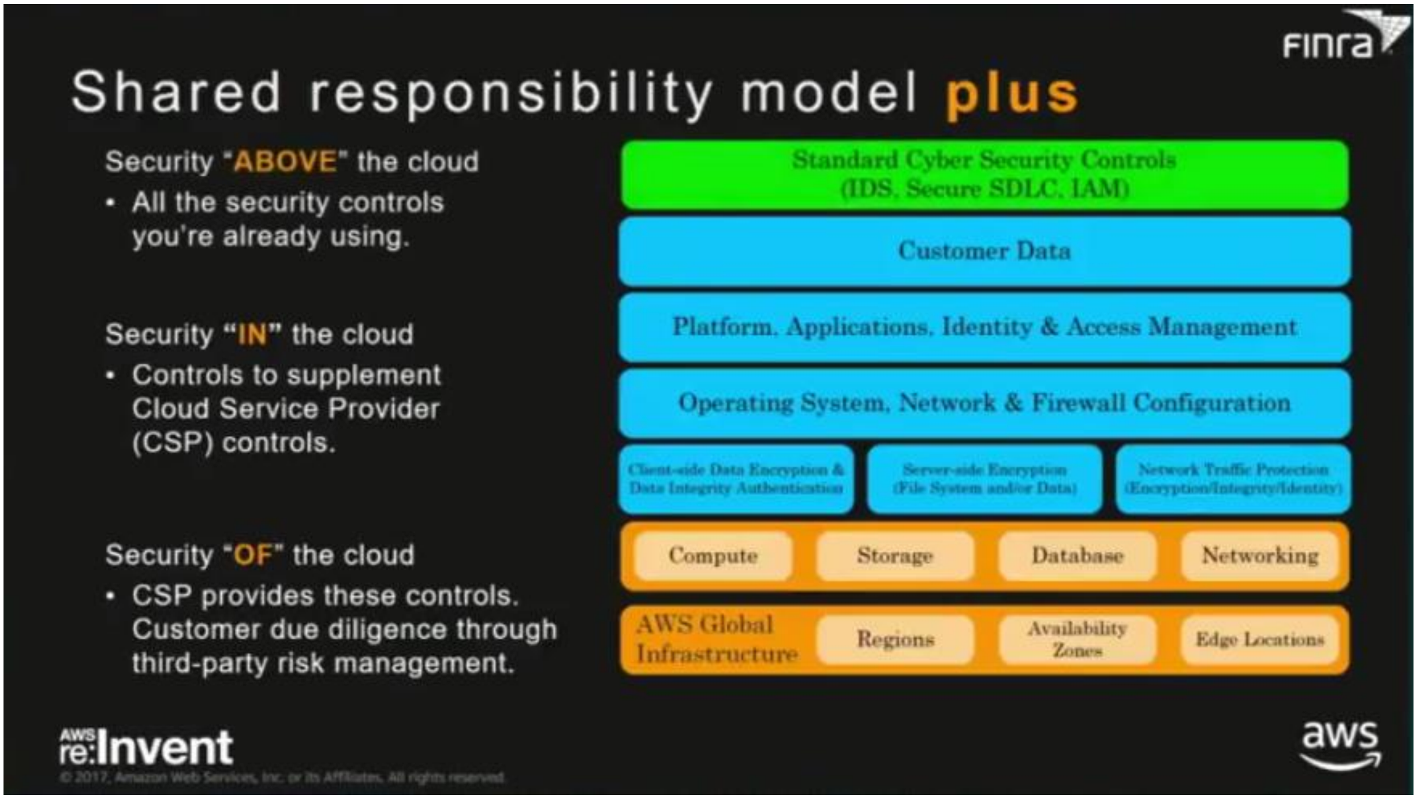
- Cloud vs. legacy data center
  - The "idealized zero-risk scenario" is unrealistic.
  - Legacy data centers have risk!

Evaluate risk in context

- "Cloud" risks get the most press.
- Many existing risks are unchanged by adoption of cloud.
- Many existing risks are significantly "riskier" than new "cloud" risks.
- The "unknown" creates a powerful perception of risk.

Cybersecurity is only one dimension of risk

- Operational, legal, and opportunity risks as well



Shared responsibility model plus

Security "ABOVE" the cloud

- All the security controls you're already using.

Security "IN" the cloud

- Controls to supplement Cloud Service Provider (CSP) controls.

Security "OF" the cloud

- CSP provides these controls. Customer due diligence through third-party risk management.

Standard Cyber Security Controls
(IDS, Secure SDLC, IAM)

Customer Data

Platform, Applications, Identity & Access Management

Operating System, Network & Firewall Configuration

Client-side Data Encryption & Data Integrity Authentication

Server-side Encryption (File System and/or Data)

Network Traffic Protection (Encryption/Integrity/Identity)

Compute

Storage

Database

Networking

AWS Global Infrastructure

Regions

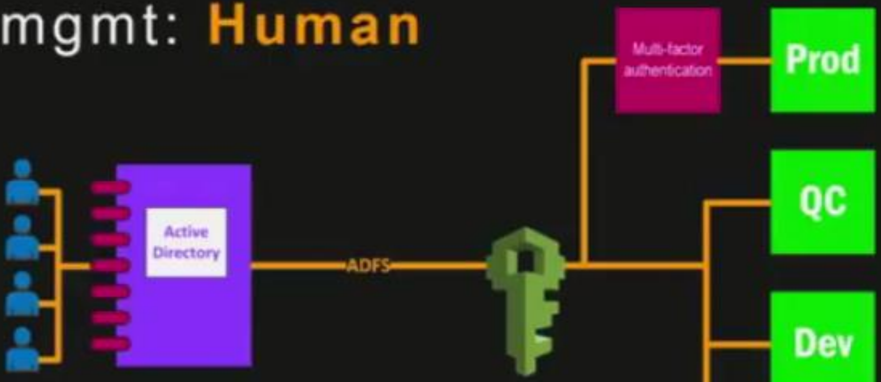
Availability Zones

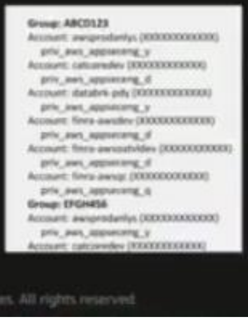
Edge Locations

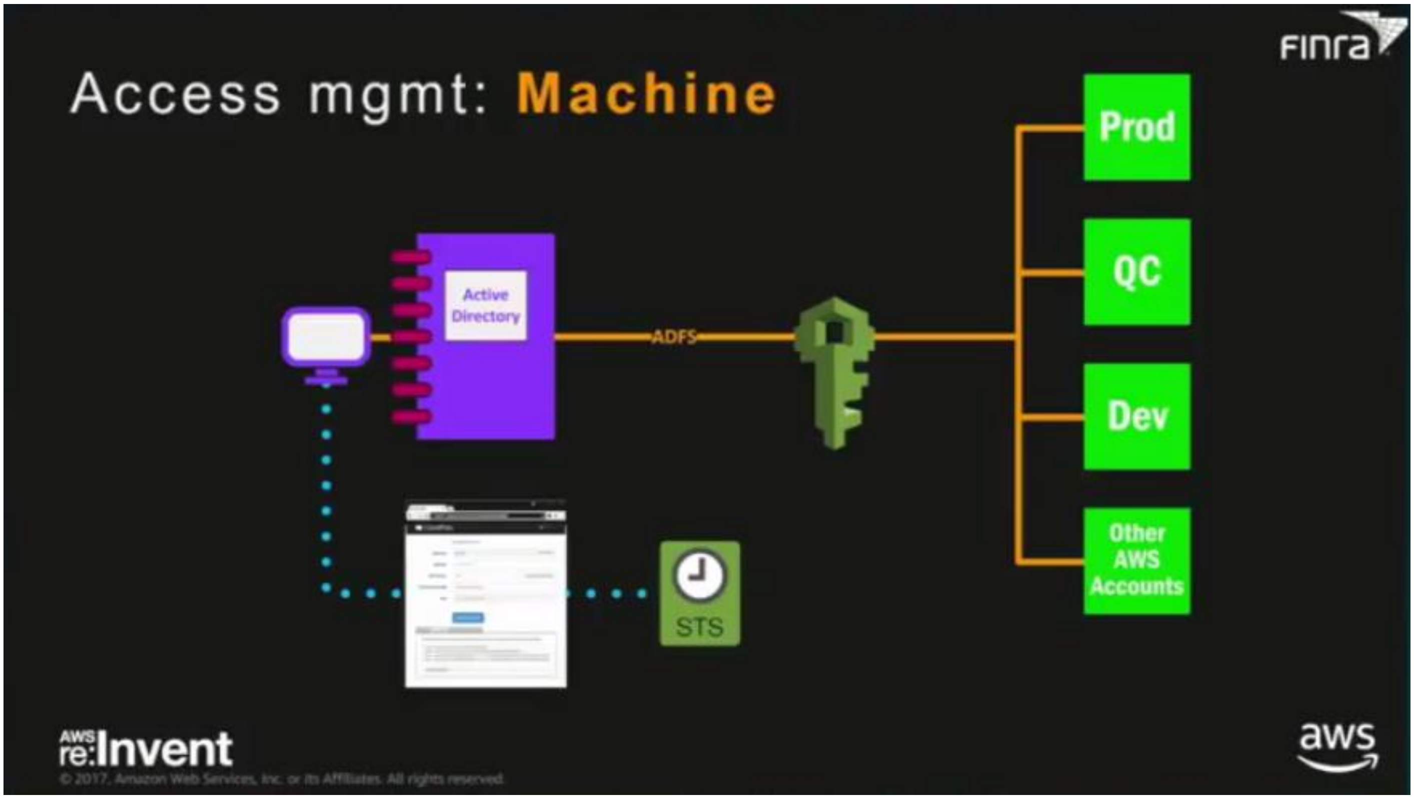
Foundational security controls

- AWS security best practices
- Access management
  - Authentication
  - Authorization/Separation of Duties (SoD)
  - Policy enforcement and oversight
- Logging/monitoring/alerting/UEBA
  - Controls for Economic DoS
- Network architecture
- Encryption and Key Management (KMS)
- Governance

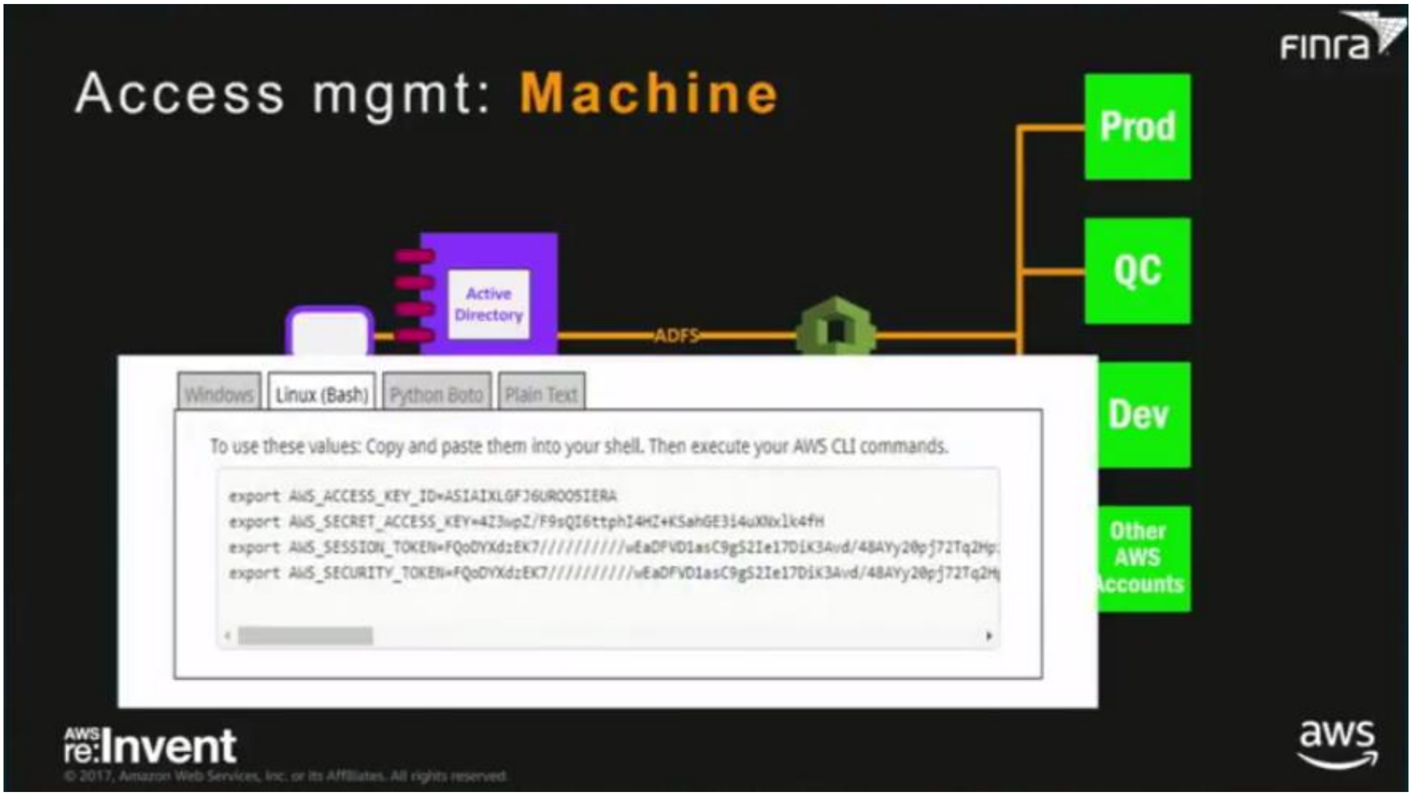
Access mgmt: Human




Group: ABCD123
Account: awsprodanlys (XXXXXXXXXXXX)
priv_aws_appseceng_y
Account: catcoredev (XXXXXXXXXXXX)
priv_aws_appseceng_d
Account: databrk-pdy (XXXXXXXXXXXX)
priv_aws_appseceng_y
Account: finra-awsdev (XXXXXXXXXXXX)
priv_aws_appseceng_d
Account: finra-awsoatvldev (XXXXXXXXXXXX)
priv_aws_appseceng_d
Account: finra-awsqc (XXXXXXXXXXXX)
priv_aws_appsecеng_q
Group: EFGH456
Account: awsprodanlys (XXXXXXXXXXXX)
priv_aws_appsecеng_y
Account: catcoredev (XXXXXXXXXXXX)



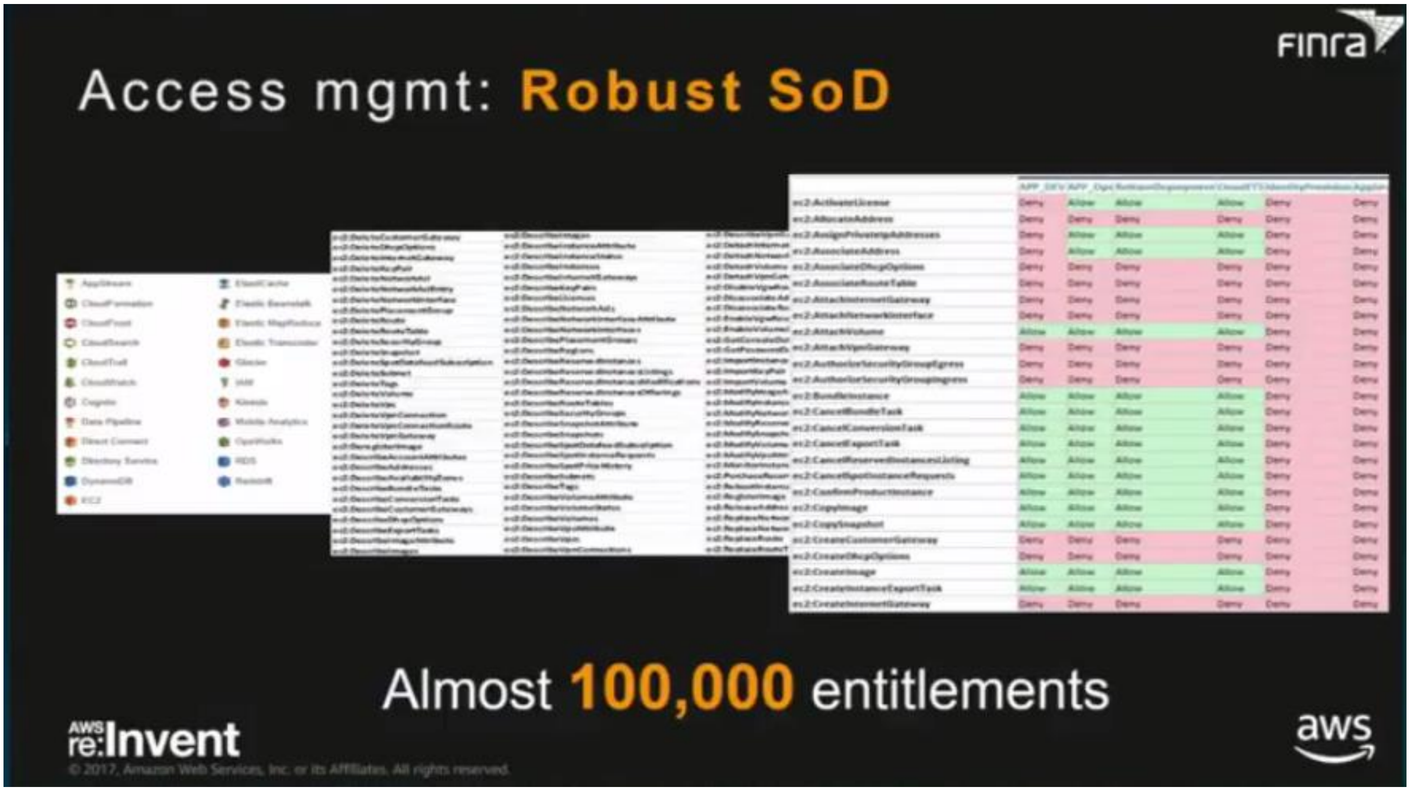
Access mgmt: Machine



Access mgmt: Machine

To use these values: Copy and paste them into your shell. Then execute your AWS CLI commands.

```
export AWS_ACCESS_KEY_ID=ASIAIXLGFJ6UROO5IERA
export AWS_SECRET_ACCESS_KEY=4Z3wpZ/F9sQI6ttphI4HZ+KSahGE3i4uXNxlk4fH
export AWS_SESSION_TOKEN=FQoDYXdzEK7//////////wEaDFVD1asC9gS2Ie17DiK3Avd/48AYy20pj72Tq2Hp:
export AWS_SECURITY_TOKEN=FQoDYXdzEK7//////////wEaDFVD1asC9gS2Ie17DiK3Avd/48AYy20pj72Tq2H
```



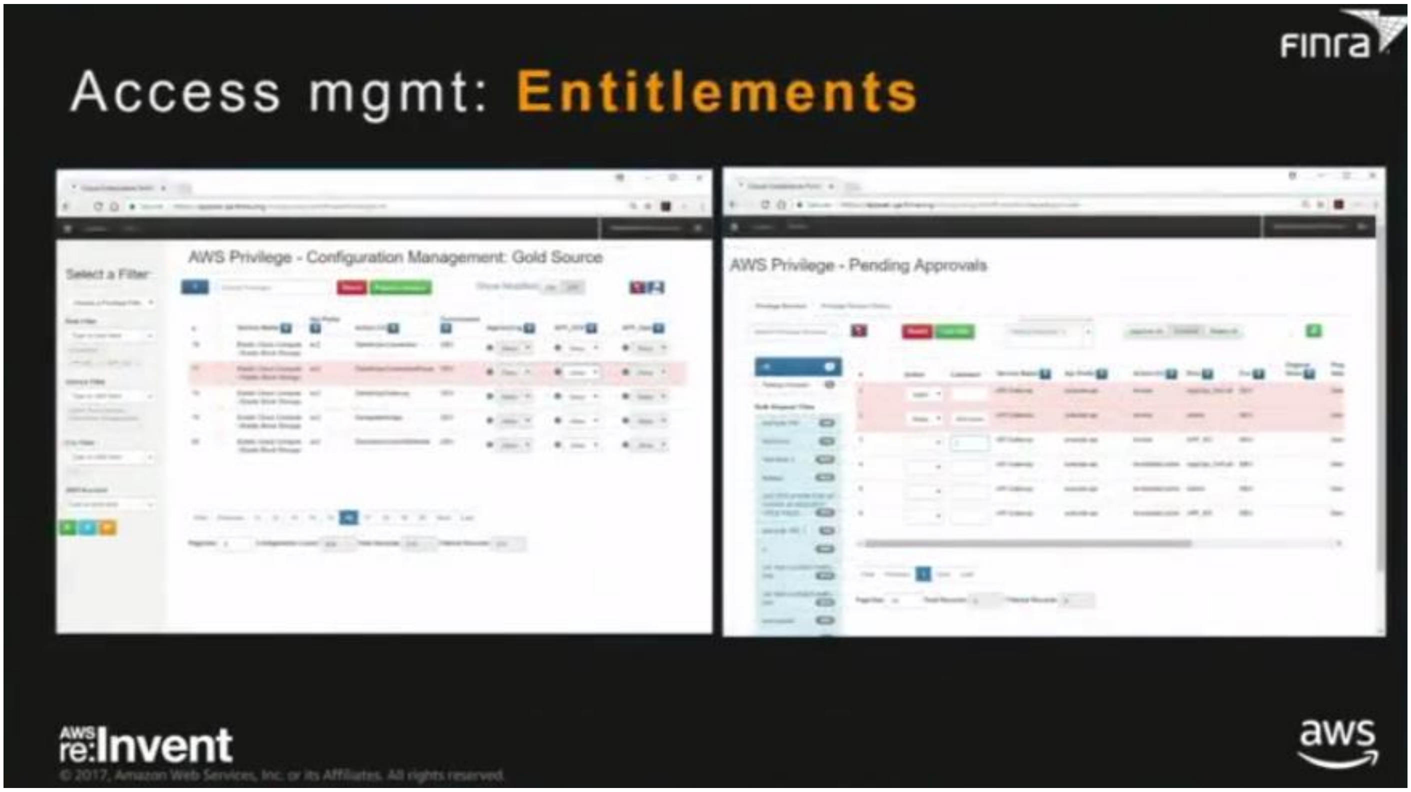
Access mgmt: Robust SoD

- AppStream
- CloudFormation
- CloudFront
- CloudSearch
- CloudTrail
- CloudWatch
- Cognito
- Data Pipeline
- Direct Connect
- Directory Service
- DynamoDB
- EC2
- ElastiCache
- Elastic Beanstalk
- Elastic MapReduce
- Elastic Transcoder
- Glacier
- IAM
- Kinesis
- Mobile Analytics
- OpsWorks
- RDS
- Redshift

	APP_DEV	APP_Ops	[illegible]	[illegible]	[illegible]	[illegible]
ec2:ActivateLicense	Deny	Allow	Allow	Allow	Deny	Deny
ec2:AllocateAddress	Deny	Deny	Deny	Deny	Deny	Deny
ec2:AssignPrivateIpAddresses	Deny	Allow	Allow	Allow	Deny	Deny
ec2:AssociateAddress	Deny	Allow	Allow	Allow	Deny	Deny
ec2:AssociateDhcpOptions	Deny	Deny	Deny	Deny	Deny	Deny
ec2:AssociateRouteTable	Deny	Deny	Deny	Deny	Deny	Deny
ec2:AttachInternetGateway	Deny	Deny	Deny	Deny	Deny	Deny
ec2:AttachNetworkInterface	Deny	Deny	Deny	Deny	Deny	Deny
ec2:AttachVolume	Allow	Allow	Allow	Allow	Deny	Deny
ec2:AttachVpnGateway	Deny	Deny	Deny	Deny	Deny	Deny
ec2:AuthorizeSecurityGroupEgress	Deny	Deny	Deny	Deny	Deny	Deny
ec2:AuthorizeSecurityGroupIngress	Deny	Deny	Deny	Deny	Deny	Deny
ec2:BundleInstance	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CancelBundleTask	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CancelConversionTask	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CancelExportTask	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CancelReservedInstancesListing	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CancelSpotInstanceRequests	Allow	Allow	Allow	Allow	Deny	Deny
ec2:ConfirmProductInstance	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CopyImage	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CopySnapshot	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CreateCustomerGateway	Deny	Deny	Deny	Deny	Deny	Deny
ec2:CreateDhcpOptions	Deny	Deny	Deny	Deny	Deny	Deny
ec2:CreateImage	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CreateInstanceExportTask	Allow	Allow	Allow	Allow	Deny	Deny
ec2:CreateInternetGateway	Deny	Deny	Deny	Deny	Deny	Deny

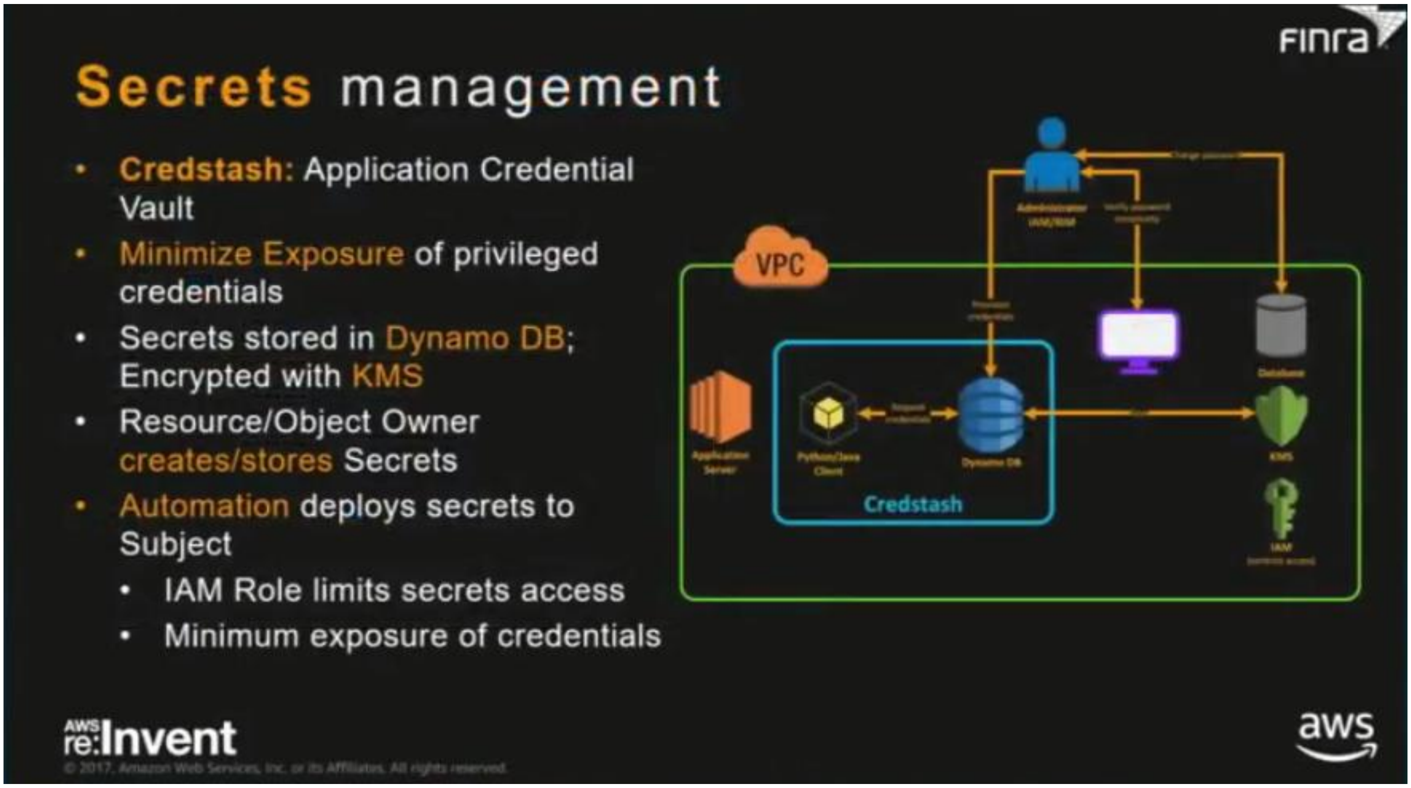
Almost **100,000** entitlements

We have access granted by roles per environment. No IAM role has Delete role on Production assets like S3 buckets or EC2 instances. We follow the principle of least entitlement where an individual is given the minimum set of role entitlements that they need to get the job done.



Access mgmt: Entitlements

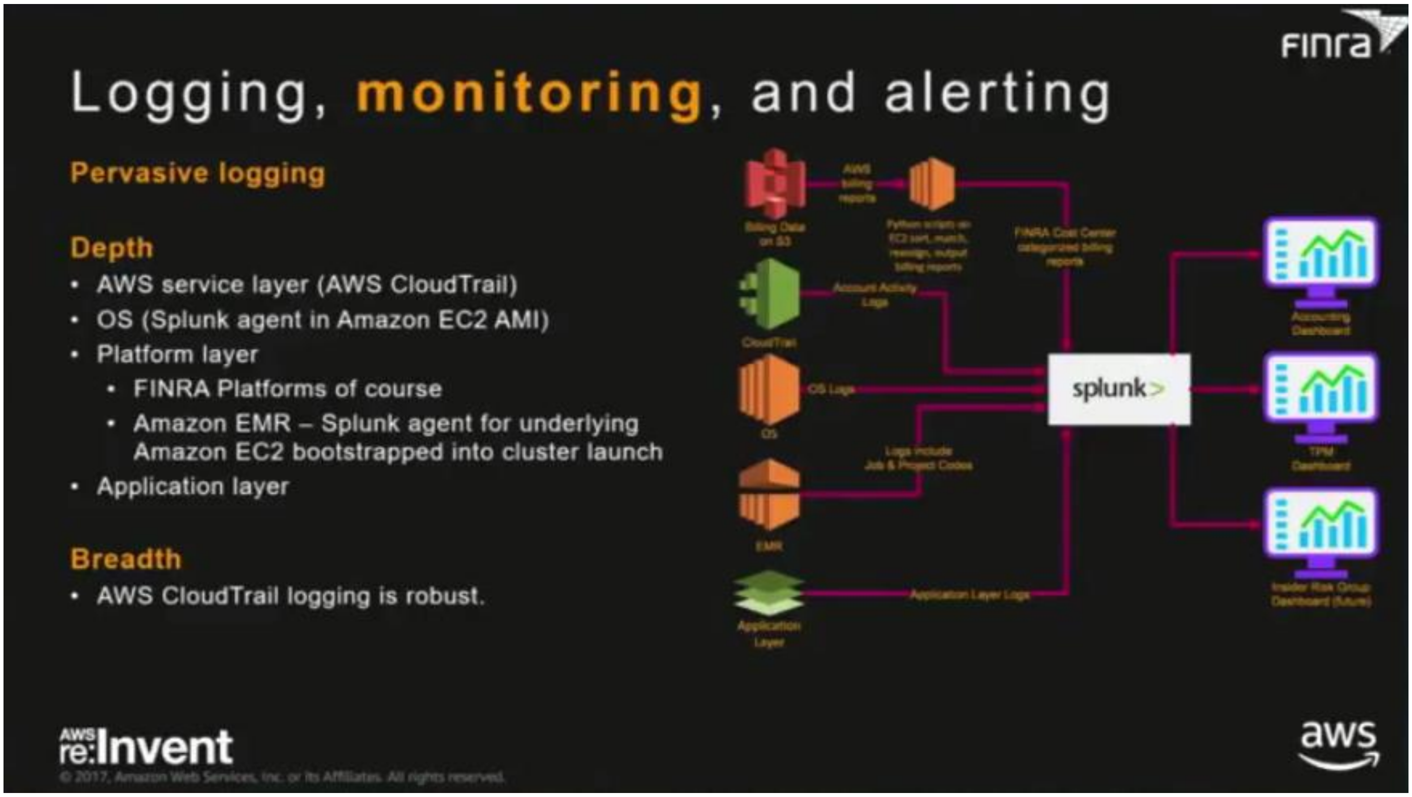
Select a Filter
AWS Privilege - Configuration Management: Gold Source
AWS Privilege - Pending Approvals



Secrets management

- Credstash: Application Credential Vault
- Minimize Exposure of privileged credentials
- Secrets stored in Dynamo DB; Encrypted with KMS
- Resource/Object Owner creates/stores Secrets
- Automation deploys secrets to Subject
  - IAM Role limits secrets access
  - Minimum exposure of credentials

We use an OSS application credential vault called Credstash that implements KMS encryption of the secrets and stores them in DynamoDB. The resource (DB admin) creates the secrets and passwords, they store it in Credstash, automation on the user's system (Python or Java client) pulls those secrets at deploy time without human intervention, the target system is IAM role restricted for seeing the credentials.



Logging, monitoring, and alerting

Pervasive logging

Depth

- AWS service layer (AWS CloudTrail)
- OS (Splunk agent in Amazon EC2 AMI)
- Platform layer
  - FINRA Platforms of course
  - Amazon EMR – Splunk agent for underlying Amazon EC2 bootstrapped into cluster launch
- Application layer

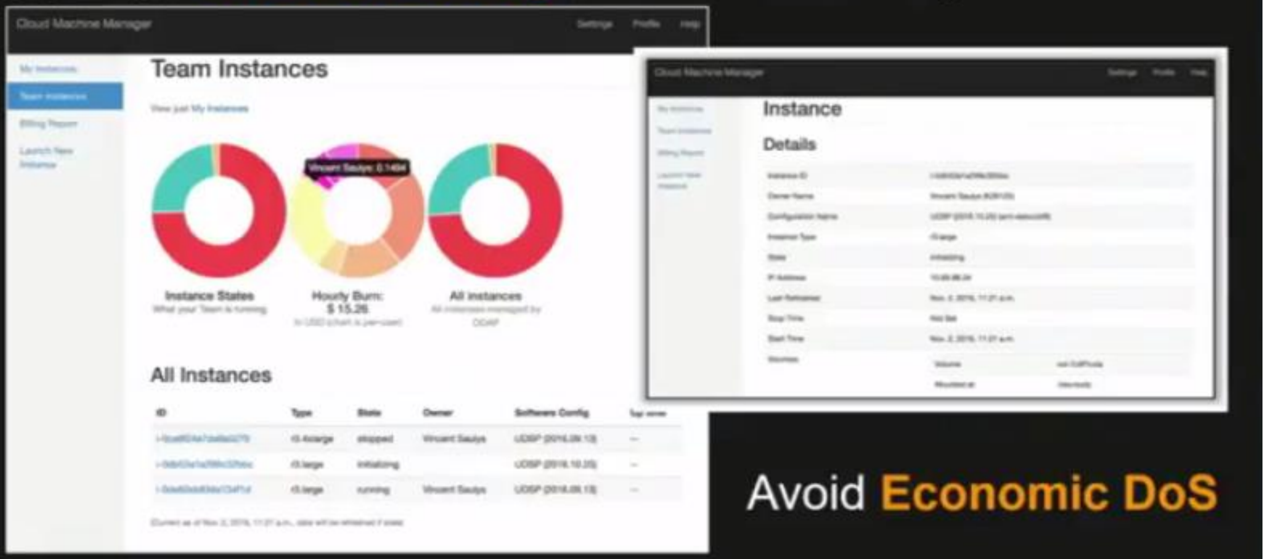
Breadth

- AWS CloudTrail logging is robust.

Billing Data on S3
AWS billing reports
Python scripts on EC2 sort, match, reassign, output billing reports
FINRA Cost Center categorized billing reports
CloudTrail
Account Activity Logs
OS
OS Logs
EMR
Logs include Job & Project Codes
Application Layer
Application Layer Logs
splunk>
Accounting Dashboard
TPM Dashboard
Insider Risk Group Dashboard (future)

FINRA

Logging, monitoring, and alerting



Cloud Machine Manager

Settings Profile Help

My Instances
Team Instances
Billing Report
Launch New Instance

Team Instances

View just My Instances

Instance States
What your Team is running

Hourly Burn:
$ 15.26
In USD (chart is per-user)

All instances
All instances managed by CCAP

All Instances

ID	Type	State	Owner	Software Config	[illegible]
[illegible]	r3.4xlarge	stopped	Vincent Saulys	UDSP (2016.09.13)	—
[illegible]	r3.large	initializing		UDSP (2016.10.25)	—
[illegible]	r3.large	running	Vincent Saulys	UDSP (2016.09.13)	—

Current as of Nov 2, 2016, 11:21 a.m.

Instance

Details

Avoid **Economic DoS**

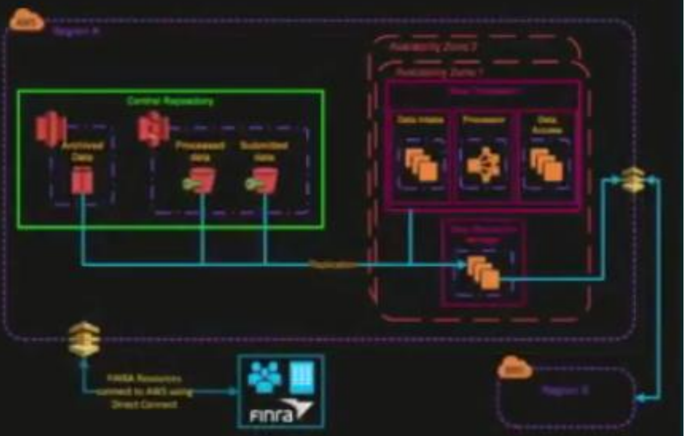
re:Invent

aws

FINRA

Network architecture

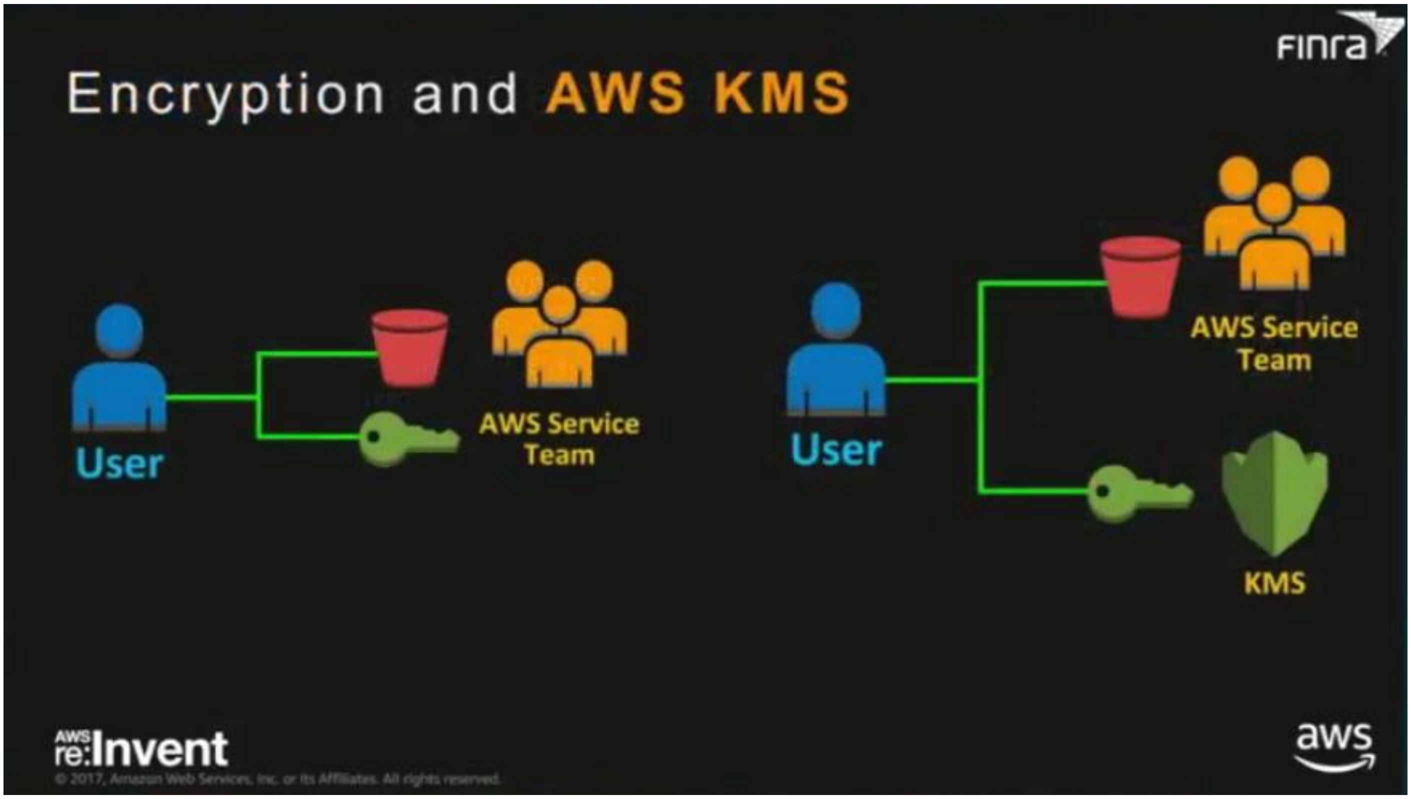
- Architecture distributed across multiple **Availability Zones**
- **Cross-region** replication used where:
  - Geographic dispersion is needed
  - Single region data durability is not considered sufficient



re:Invent

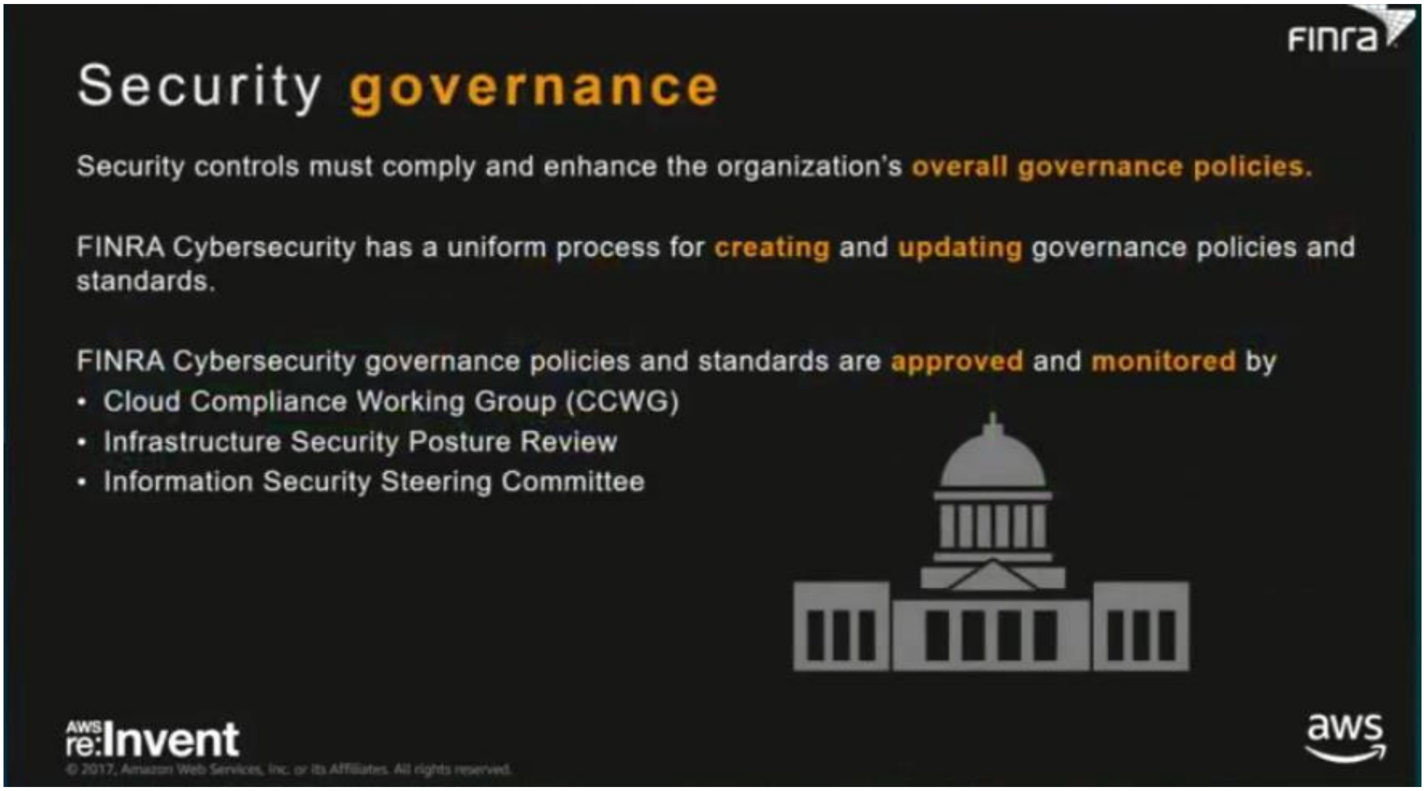
aws

This is a sample deployment architecture we use at FINRA, we use up to 4 AZs.



Encryption and AWS KMS

We use SSL in-transit encryption with all the AWS service endpoints we have, we also use encryption at rest and S3 and KMS encryption.



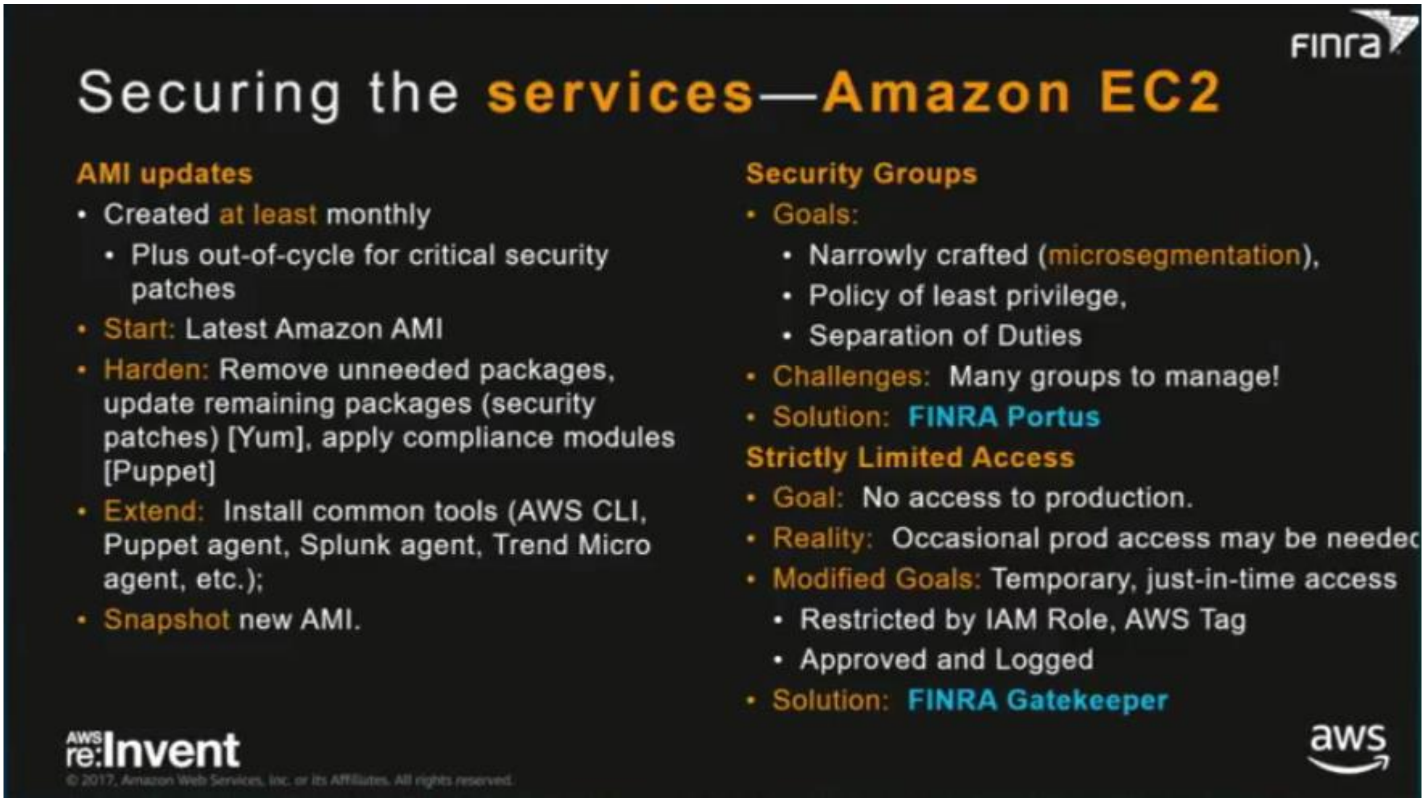
Security governance

Security controls must comply and enhance the organization's **overall governance policies.**

FINRA Cybersecurity has a uniform process for **creating** and **updating** governance policies and standards.

FINRA Cybersecurity governance policies and standards are **approved** and **monitored** by

- Cloud Compliance Working Group (CCWG)
- Infrastructure Security Posture Review
- Information Security Steering Committee



Securing the services—Amazon EC2

AMI updates

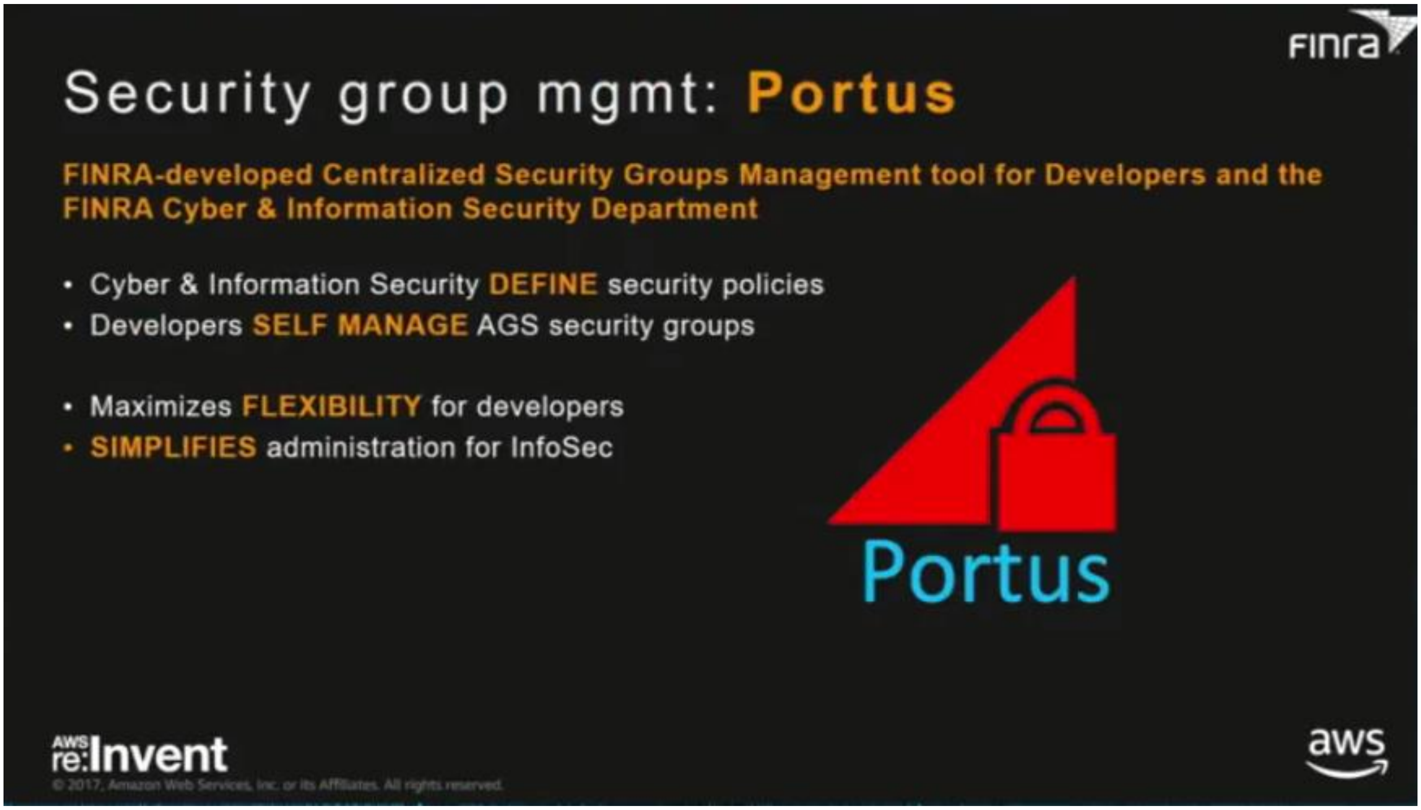
- Created at least monthly
  - Plus out-of-cycle for critical security patches
- Start: Latest Amazon AMI
- Harden: Remove unneeded packages, update remaining packages (security patches) [Yum], apply compliance modules [Puppet]
- Extend: Install common tools (AWS CLI, Puppet agent, Splunk agent, Trend Micro agent, etc.);
- Snapshot new AMI.

Security Groups

- Goals:
  - Narrowly crafted (microsegmentation),
  - Policy of least privilege,
  - Separation of Duties
- Challenges: Many groups to manage!
- Solution: FINRA Portus

Strictly Limited Access

- Goal: No access to production.
- Reality: Occasional prod access may be neede
- Modified Goals: Temporary, just-in-time access
  - Restricted by IAM Role, AWS Tag
  - Approved and Logged
- Solution: FINRA Gatekeeper



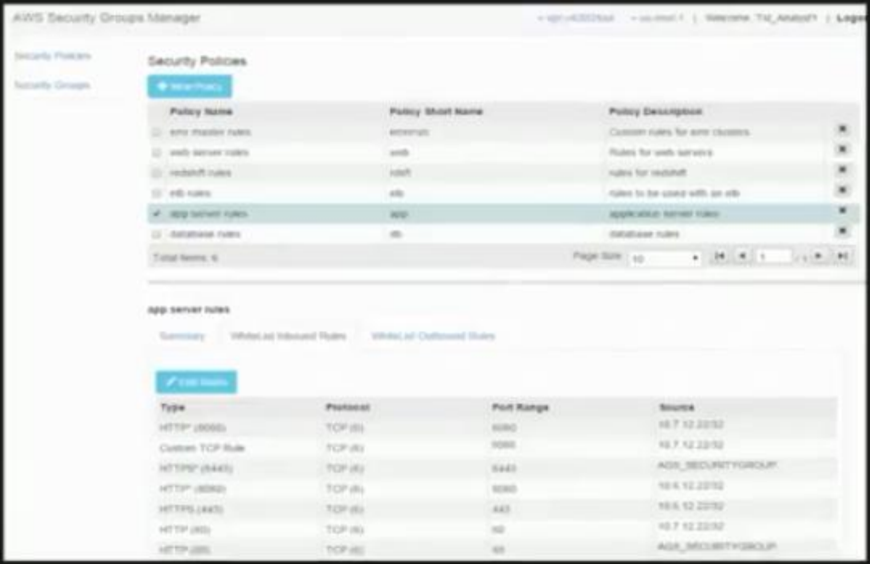
Security group mgmt: Portus

FINRA-developed Centralized Security Groups Management tool for Developers and the FINRA Cyber & Information Security Department

- Cyber & Information Security DEFINE security policies
- Developers SELF MANAGE AGS security groups

- Maximizes FLEXIBILITY for developers
- SIMPLIFIES administration for InfoSec

Portus dashboard



AWS Security Groups Manager — Welcome: TIZ_Analyst1 | Logout

Security Policies

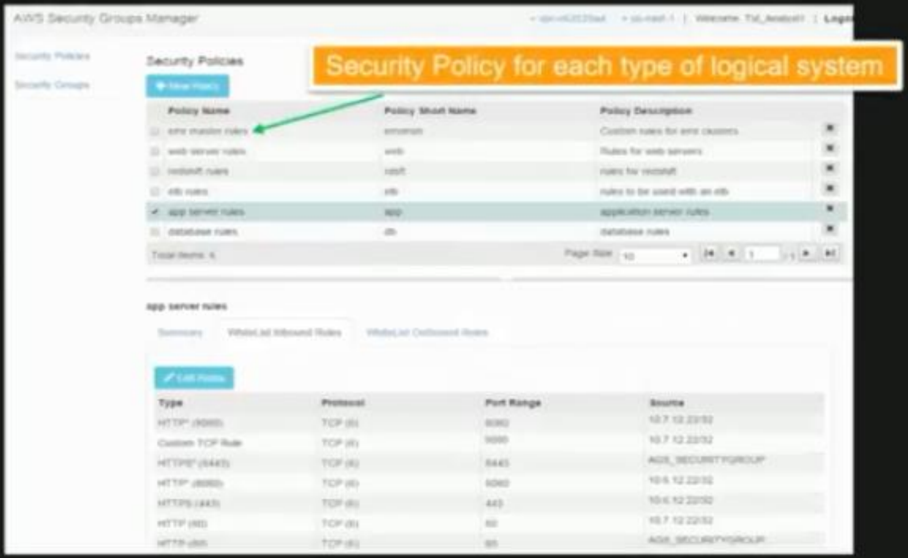
	Policy Name	Policy Short Name	Policy Description	
□	emr master rules	emrmstr	Custom rules for emr clusters	✖
□	web server rules	web	Rules for web servers	✖
□	redshift rules	rdsft	rules for redshift	✖
□	elb rules	elb	rules to be used with an elb	✖
✓	app server rules	app	application server rules	✖
□	database rules	db	database rules	✖

Total Items: 6 — Page Size: 10 — 1 / 1

app server rules

Type	Protocol	Port Range	Source
HTTP* (8080)	TCP (6)	8080	10.7.12.22/32
Custom TCP Rule	TCP (6)	9080	10.7.12.22/32
HTTPS* (8443)	TCP (6)	8443	AGS_SECURITYGROUP
HTTP* (8080)	TCP (6)	8080	10.6.12.22/32
HTTPS (443)	TCP (6)	443	10.6.12.22/32
HTTP (80)	TCP (6)	80	10.7.12.22/32
HTTP (85)	TCP (6)	85	AGS_SECURITYGROUP

Portus dashboard



Security Policy for each type of logical system

AWS Security Groups Manager — Welcome: TIZ_Analyst1 | Logout

Security Policies

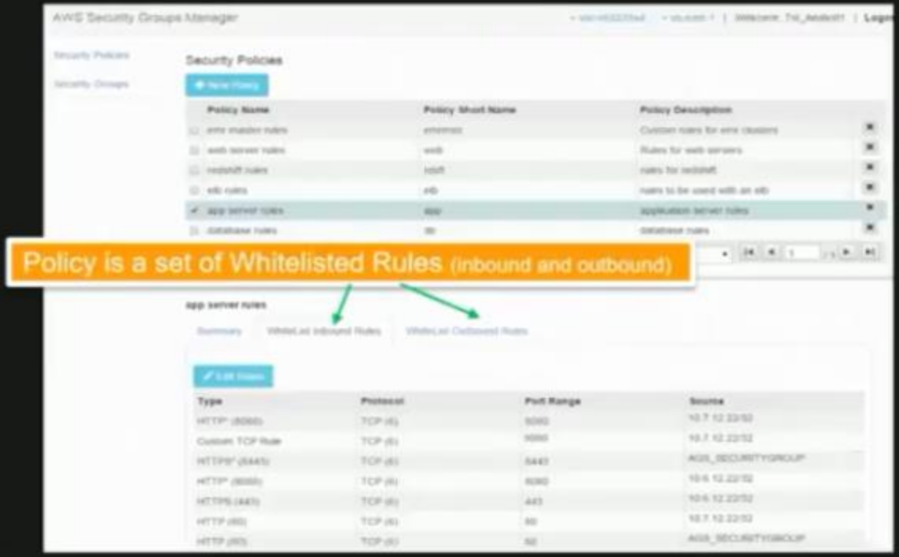
	Policy Name	Policy Short Name	Policy Description	
□	emr master rules	emrmstr	Custom rules for emr clusters	✖
□	web server rules	web	Rules for web servers	✖
□	redshift rules	rdsft	rules for redshift	✖
□	elb rules	elb	rules to be used with an elb	✖
✓	app server rules	app	application server rules	✖
□	database rules	db	database rules	✖

Total Items: 6 — Page Size: 10 — 1 / 1

app server rules

Type	Protocol	Port Range	Source
HTTP* (8080)	TCP (6)	8080	10.7.12.22/32
Custom TCP Rule	TCP (6)	9080	10.7.12.22/32
HTTPS* (8443)	TCP (6)	8443	AGS_SECURITYGROUP
HTTP* (8080)	TCP (6)	8080	10.6.12.22/32
HTTPS (443)	TCP (6)	443	10.6.12.22/32
HTTP (80)	TCP (6)	80	10.7.12.22/32
HTTP (85)	TCP (6)	85	AGS_SECURITYGROUP

Portus dashboard



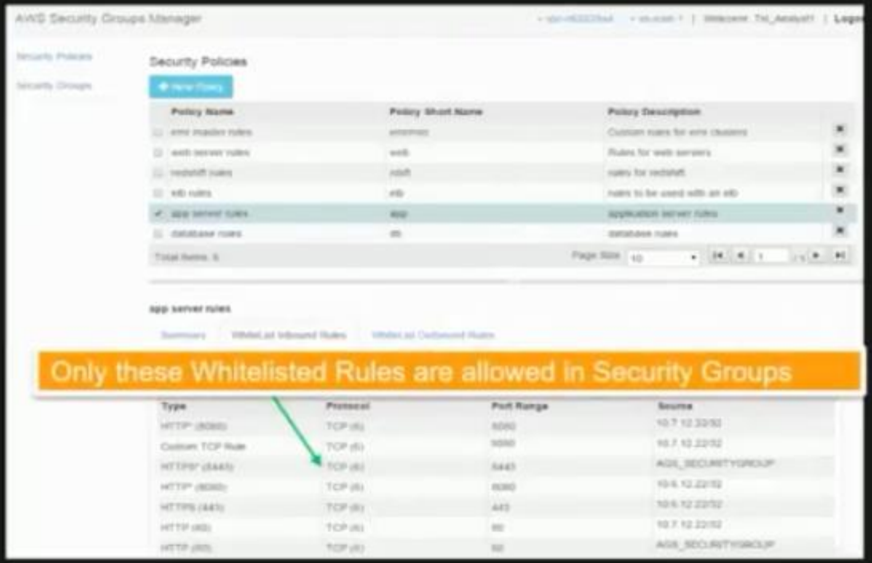
Security Policies

	Policy Name	Policy Short Name	Policy Description
□	emr master rules	emrmstr	Custom rules for emr clusters
□	web server rules	web	Rules for web servers
□	redshift rules	rdsft	rules for redshift
□	elb rules	elb	rules to be used with an elb
✓	app server rules	app	application server rules
□	database rules	db	database rules

Policy is a set of Whitelisted Rules (inbound and outbound)

Type	Protocol	Port Range	Source
HTTP* (8080)	TCP (6)	8080	10.7.12.22/32
Custom TCP Rule	TCP (6)	9090	10.7.12.22/32
HTTPS* (8443)	TCP (6)	8443	AGS_SECURITYGROUP
HTTP* (8080)	TCP (6)	8080	10.6.12.22/32
HTTPS (443)	TCP (6)	443	10.6.12.22/32
HTTP (80)	TCP (6)	80	10.7.12.22/32
HTTP (80)	TCP (6)	80	AGS_SECURITYGROUP

Portus dashboard



Security Policies

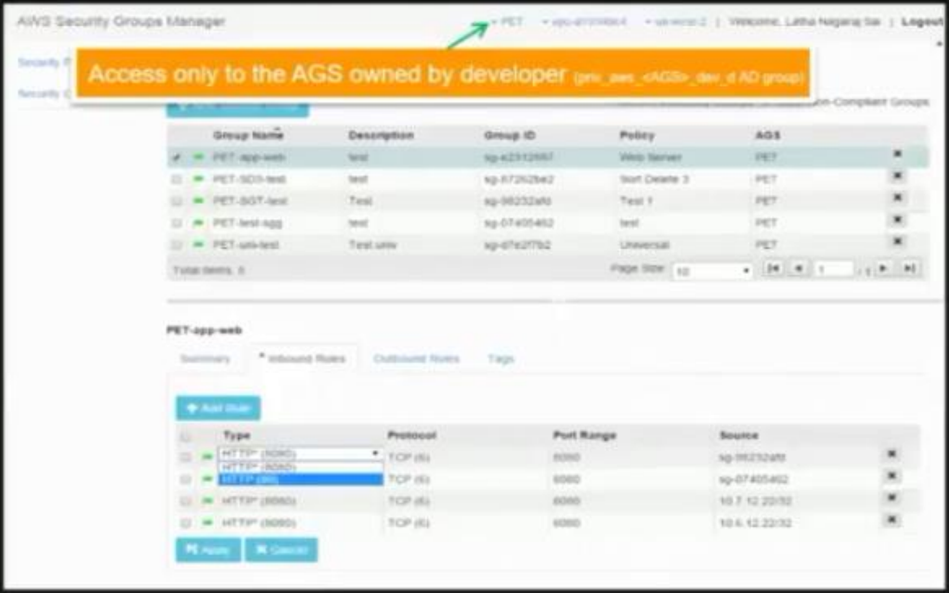
	Policy Name	Policy Short Name	Policy Description
□	emr master rules	emrmstr	Custom rules for emr clusters
□	web server rules	web	Rules for web servers
□	redshift rules	rdsft	rules for redshift
□	elb rules	elb	rules to be used with an elb
✓	app server rules	app	application server rules
□	database rules	db	database rules

Total Items: 6

Only these Whitelisted Rules are allowed in Security Groups

Type	Protocol	Port Range	Source
HTTP* (8080)	TCP (6)	8080	10.7.12.22/32
Custom TCP Rule	TCP (6)	9090	10.7.12.22/32
HTTPS* (8443)	TCP (6)	8443	AGS_SECURITYGROUP
HTTP* (8080)	TCP (6)	8080	10.6.12.22/32
HTTPS (443)	TCP (6)	443	10.6.12.22/32
HTTP (80)	TCP (6)	80	10.7.12.22/32
HTTP (80)	TCP (6)	80	AGS_SECURITYGROUP

Portus **developer** dashboard



AWS Security Groups Manager ▾ PET ▾ vpc-a1f3f4bc4 ▾ us-west-2 | Welcome, Latha Nagaraj Sai | Logout

Access only to the AGS owned by developer (pric_aws_<AGS>_dev_d AD group)

		Group Name	Description	Group ID	Policy	AGS	
✓	▬	PET-app-web	test	sg-e2312667	Web Server	PET	✖
□	▬	PET-SD3-test	test	sg-872626e2	Sort Delete 3	PET	✖
□	▬	PET-SGT-test	Test	sg-98232afd	Test 1	PET	✖
□	▬	PET-test-sgg	test	sg-07405462	test	PET	✖
□	▬	PET-uni-test	Test univ	sg-d7e2f7b2	Universal	PET	✖

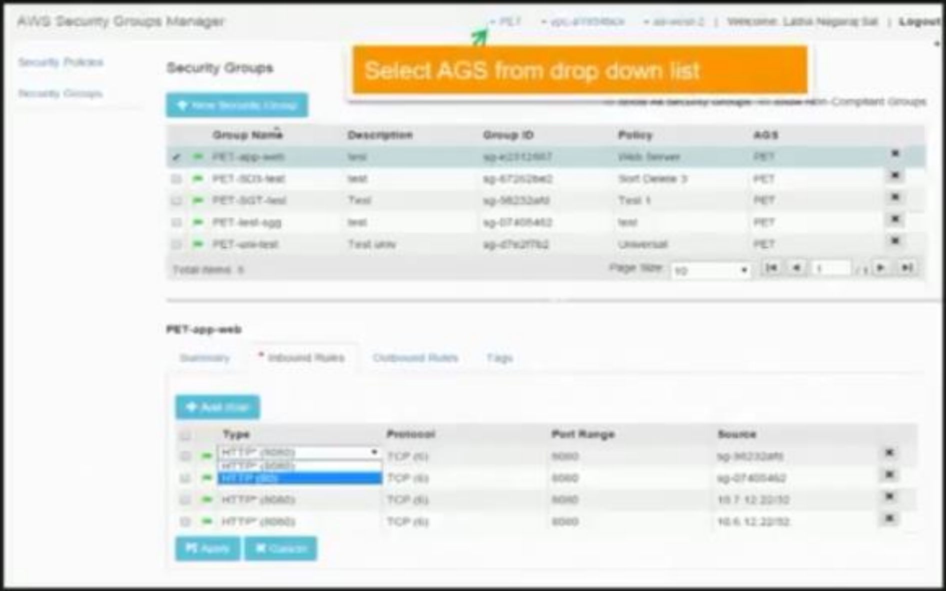
Total items: 5 Page Size: 10 ▾ 1 / 1

PET-app-web

Summary | * Inbound Rules | Outbound Rules | Tags

□		Type	Protocol	Port Range	Source	
□	▬	HTTP* (8080) ▾ HTTP* (8080) HTTP (80)	TCP (6)	8080	sg-98232afd	✖
□	▬		TCP (6)	8080	sg-07405462	✖
□	▬	HTTP* (8080)	TCP (6)	8080	10.7.12.22/32	✖
□	▬	HTTP* (8080)	TCP (6)	8080	10.6.12.22/32	✖

Portus **developer** dashboard



AWS Security Groups Manager ▾ PET ▾ vpc-a1f3f4bc4 ▾ us-west-2 | Welcome, Latha Nagaraj Sai | Logout

Select AGS from drop down list

Security Groups

		Group Name	Description	Group ID	Policy	AGS	
✓	▬	PET-app-web	test	sg-e2312667	Web Server	PET	✖
□	▬	PET-SD3-test	test	sg-872626e2	Sort Delete 3	PET	✖
□	▬	PET-SGT-test	Test	sg-98232afd	Test 1	PET	✖
□	▬	PET-test-sgg	test	sg-07405462	test	PET	✖
□	▬	PET-uni-test	Test univ	sg-d7e2f7b2	Universal	PET	✖

Total items: 5 Page Size: 10 ▾ 1 / 1

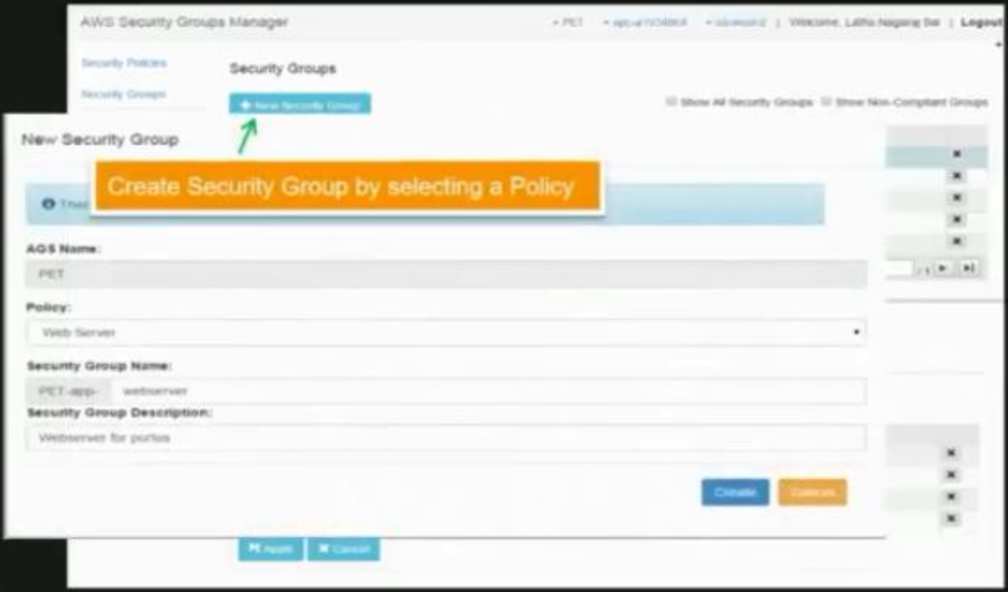
PET-app-web

Summary | * Inbound Rules | Outbound Rules | Tags

□		Type	Protocol	Port Range	Source	
□	▬	HTTP* (8080) ▾ HTTP* (8080) HTTP (80)	TCP (6)	8080	sg-98232afd	✖
□	▬		TCP (6)	8080	sg-07405462	✖
□	▬	HTTP* (8080)	TCP (6)	8080	10.7.12.22/32	✖
□	▬	HTTP* (8080)	TCP (6)	8080	10.6.12.22/32	✖

FINRA

Portus **developer** dashboard


AWS Security Groups Manager
• PET
Welcome, Latha Nagaraj Sai
Logout
Security Policies
Security Groups
+ New Security Group
☐ Show All Security Groups ☐ Show Non-Compliant Groups
Create
Cancel
Apply
Cancel

Security Groups

New Security Group

AGS Name: PET

Policy: Web Server

Security Group Name: PET-app- webserver

Security Group Description: Webserver for portus

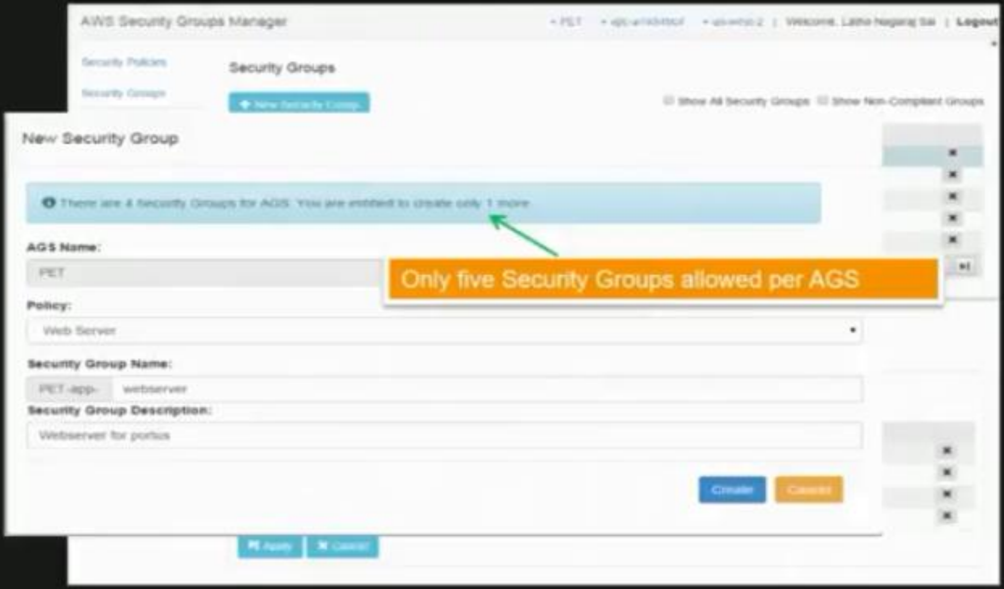
Create Security Group by selecting a Policy

AWS re:Invent

aws

FINRA

Portus **developer** dashboard


AWS Security Groups Manager
• PET
Welcome, Latha Nagaraj Sai
Logout
Security Policies
Security Groups
+ New Security Group
☐ Show All Security Groups ☐ Show Non-Compliant Groups
Create
Cancel
Apply
Cancel

Security Groups

New Security Group

There are 4 Security Groups for AGS. You are entitled to create only 1 more

AGS Name: PET

Policy: Web Server

Security Group Name: PET-app- webserver

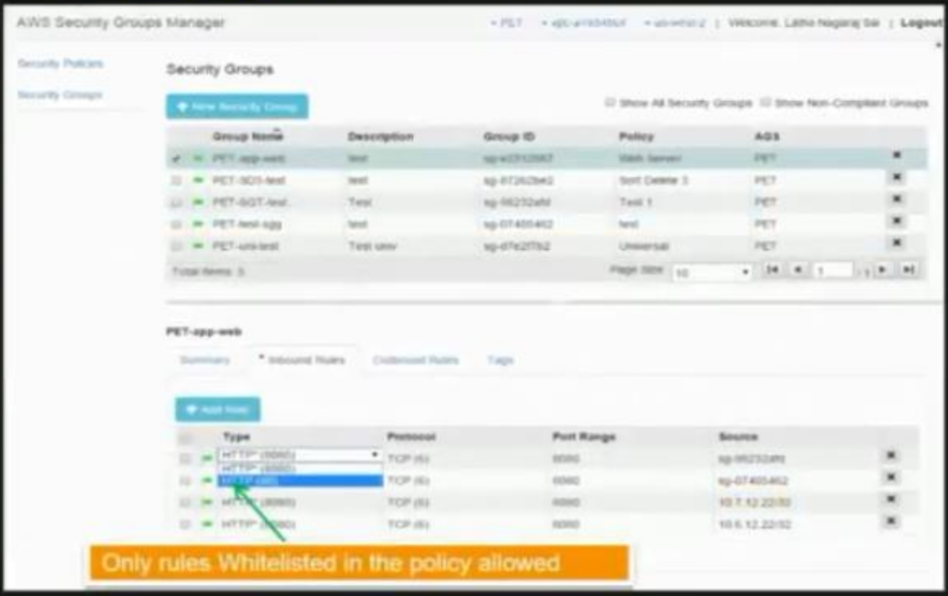
Security Group Description: Webserver for portus

Only five Security Groups allowed per AGS

AWS re:Invent

aws

Portus developer dashboard



AWS Security Groups Manager

Security Groups

	Group Name	Description	Group ID	Policy	AGS	
✓	PET-app-web	test	sg-e2312667	Web Server	PET	✖
□	PET-SDS-test	test	sg-87262be2	Sort Delete 3	PET	✖
□	PET-SGT-test..	Test	sg-06232afd	Test 1	PET	✖
□	PET-test-sgg	test	sg-07405462	test	PET	✖
□	PET-uni-test	Test univ	sg-d7e2f7b2	Universal	PET	✖

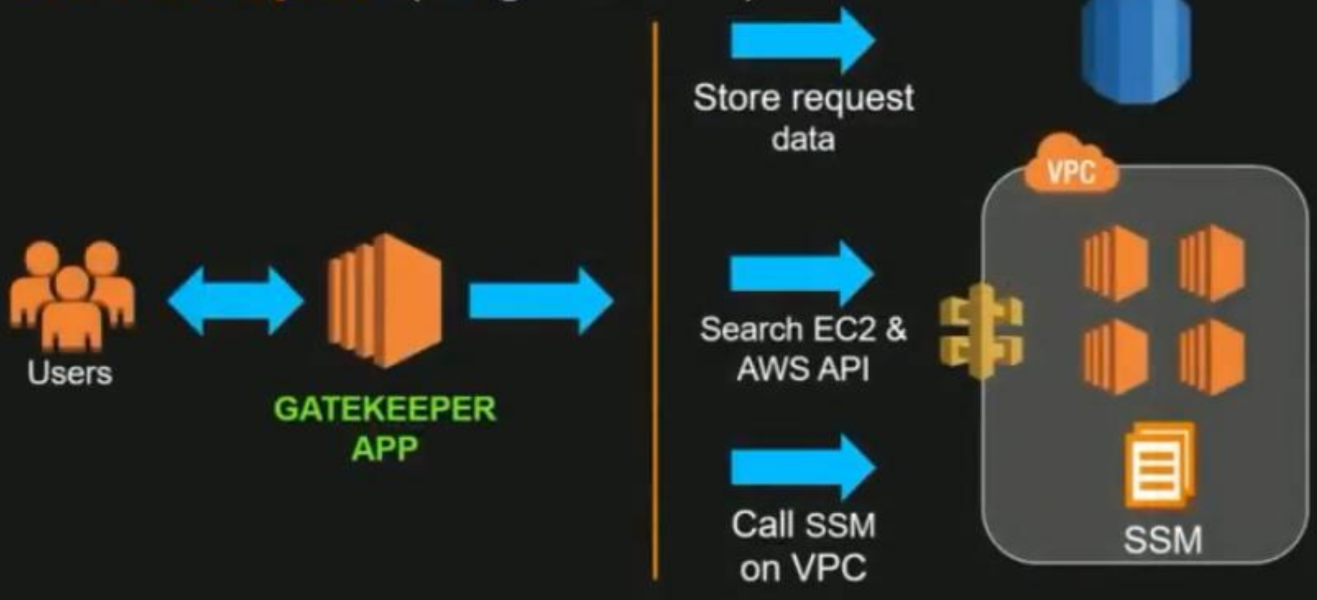
Total items: 5 Page Size: 10

PET-app-web

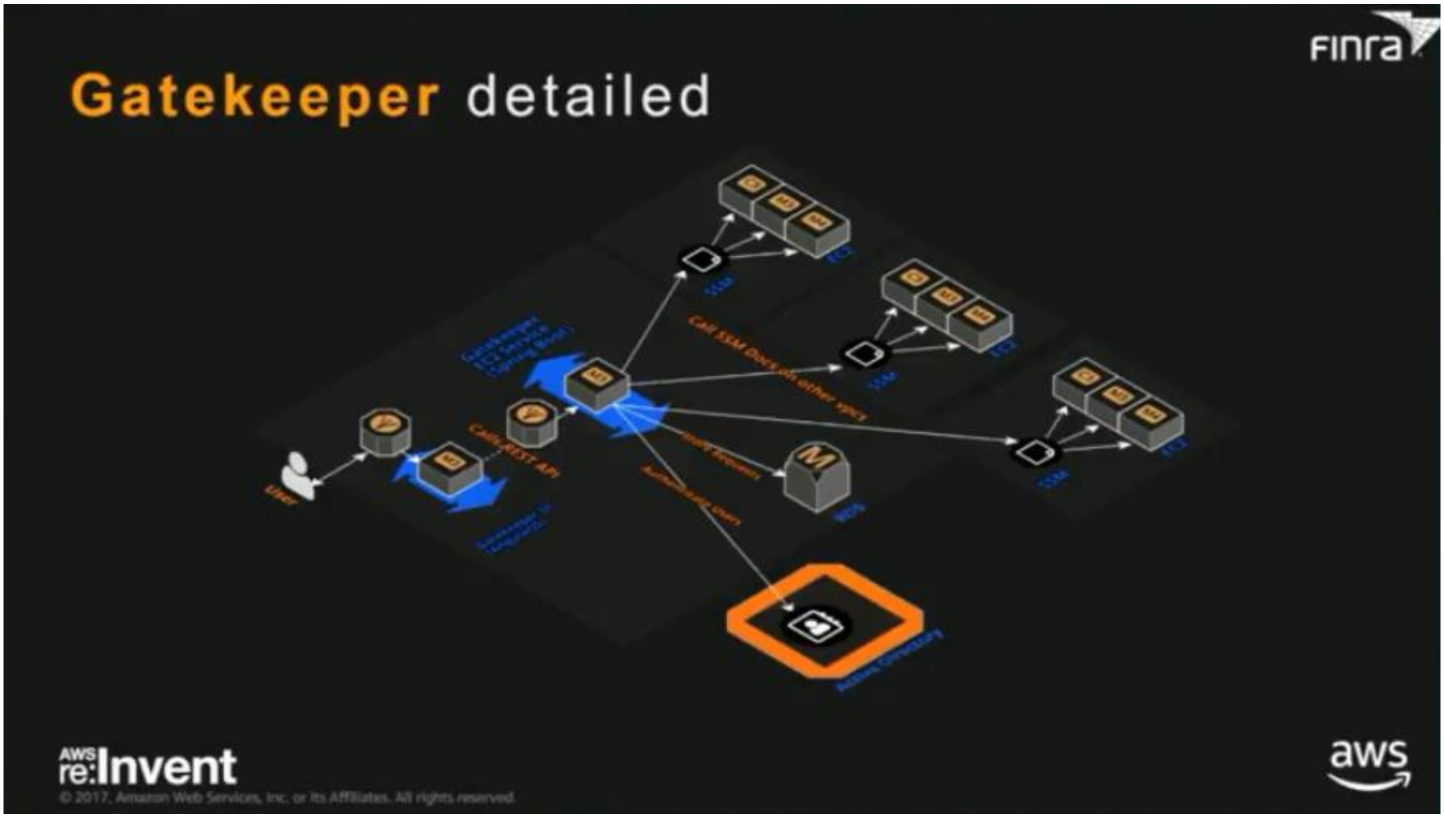
	Type	Protocol	Port Range	Source	
□	HTTP* (8080)	TCP (6)	8080	sg-06232afd	✖
□	HTTP (80)	TCP (6)	8080	sg-07405462	✖
□	HTTP* (8080)	TCP (6)	8080	10.7.12.22/32	✖
□	HTTP* (8080)	TCP (6)	8080	10.6.12.22/32	✖

Only rules Whitelisted in the policy allowed

Gatekeeper (High level)

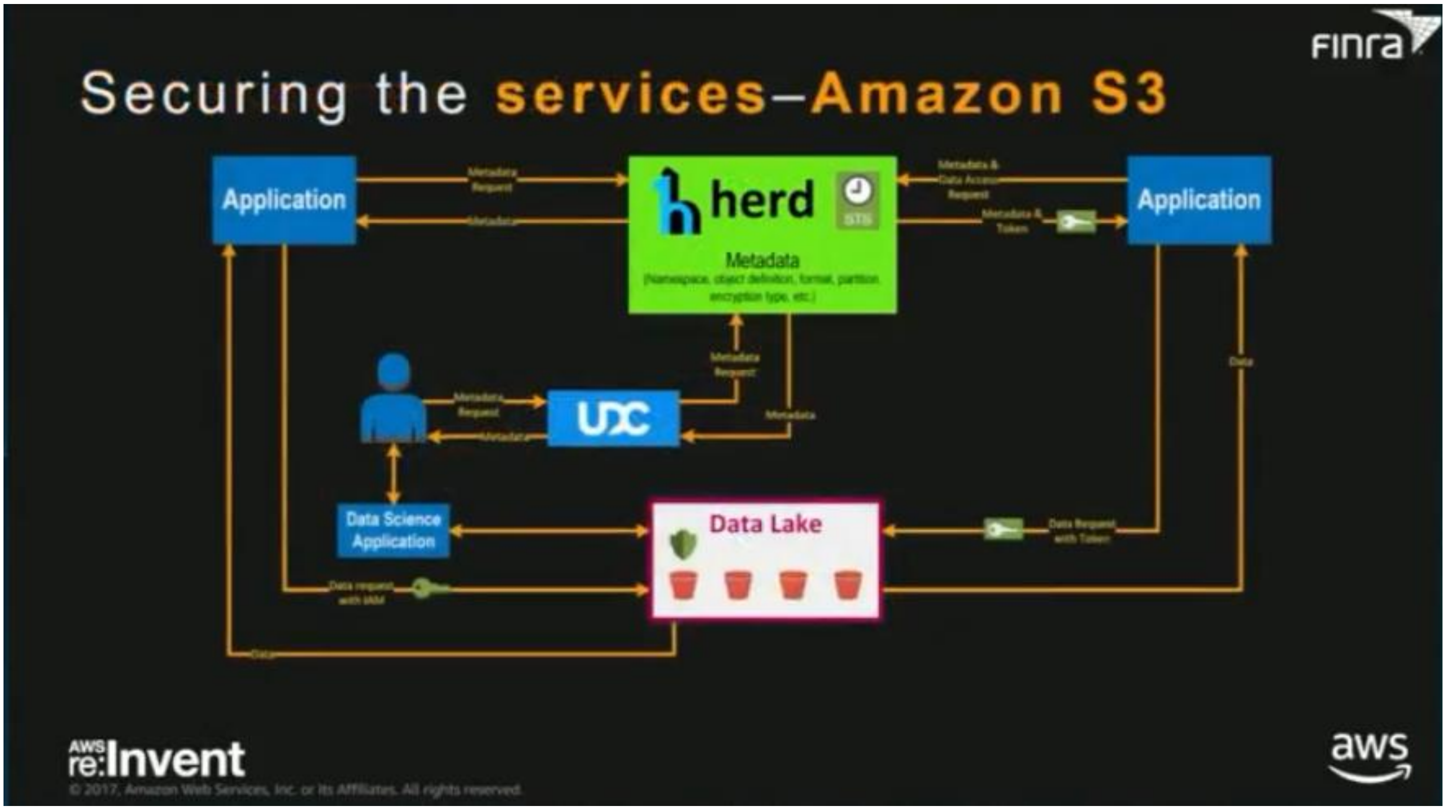


Gatekeeper is another system that we built that allows that Just in time, limited production environment access to an EC2 instance.



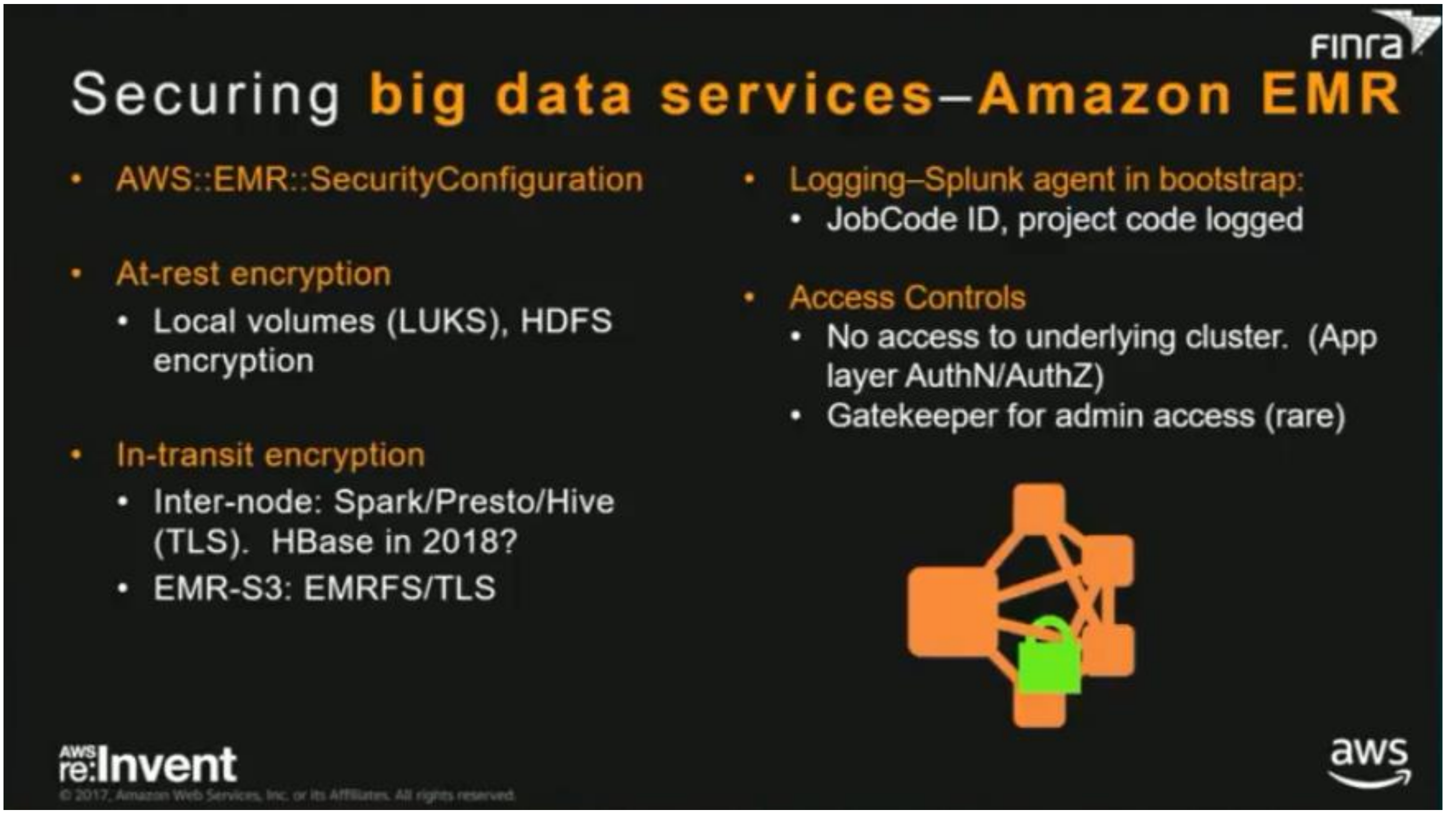
Gatekeeper detailed

This further minimizes the opportunity for credentials to be exposed.



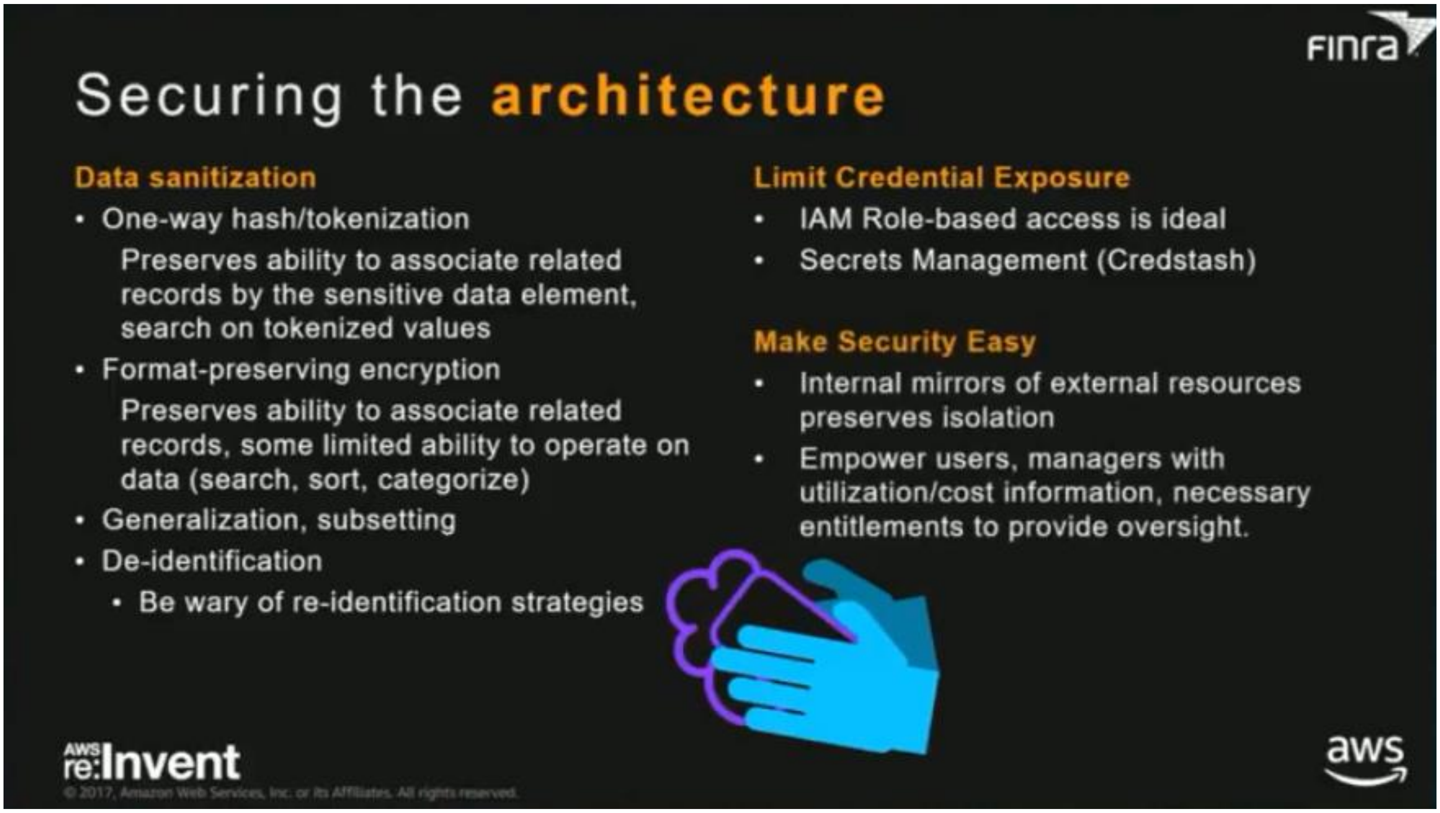
Securing the services–Amazon S3

The Herd system helps us to manage access to the information to the data lake in S3, it uses metadata about the data like namespace, object definition, format, partition, encryption type, etc.



Securing big data services–Amazon EMR

- AWS::EMR::SecurityConfiguration
- At-rest encryption
  - Local volumes (LUKS), HDFS encryption
- In-transit encryption
  - Inter-node: Spark/Presto/Hive (TLS). HBase in 2018?
  - EMR-S3: EMRFS/TLS
- Logging–Splunk agent in bootstrap:
  - JobCode ID, project code logged
- Access Controls
  - No access to underlying cluster. (App layer AuthN/AuthZ)
  - Gatekeeper for admin access (rare)



Securing the architecture

Data sanitization

- One-way hash/tokenization

  Preserves ability to associate related records by the sensitive data element, search on tokenized values
- Format-preserving encryption

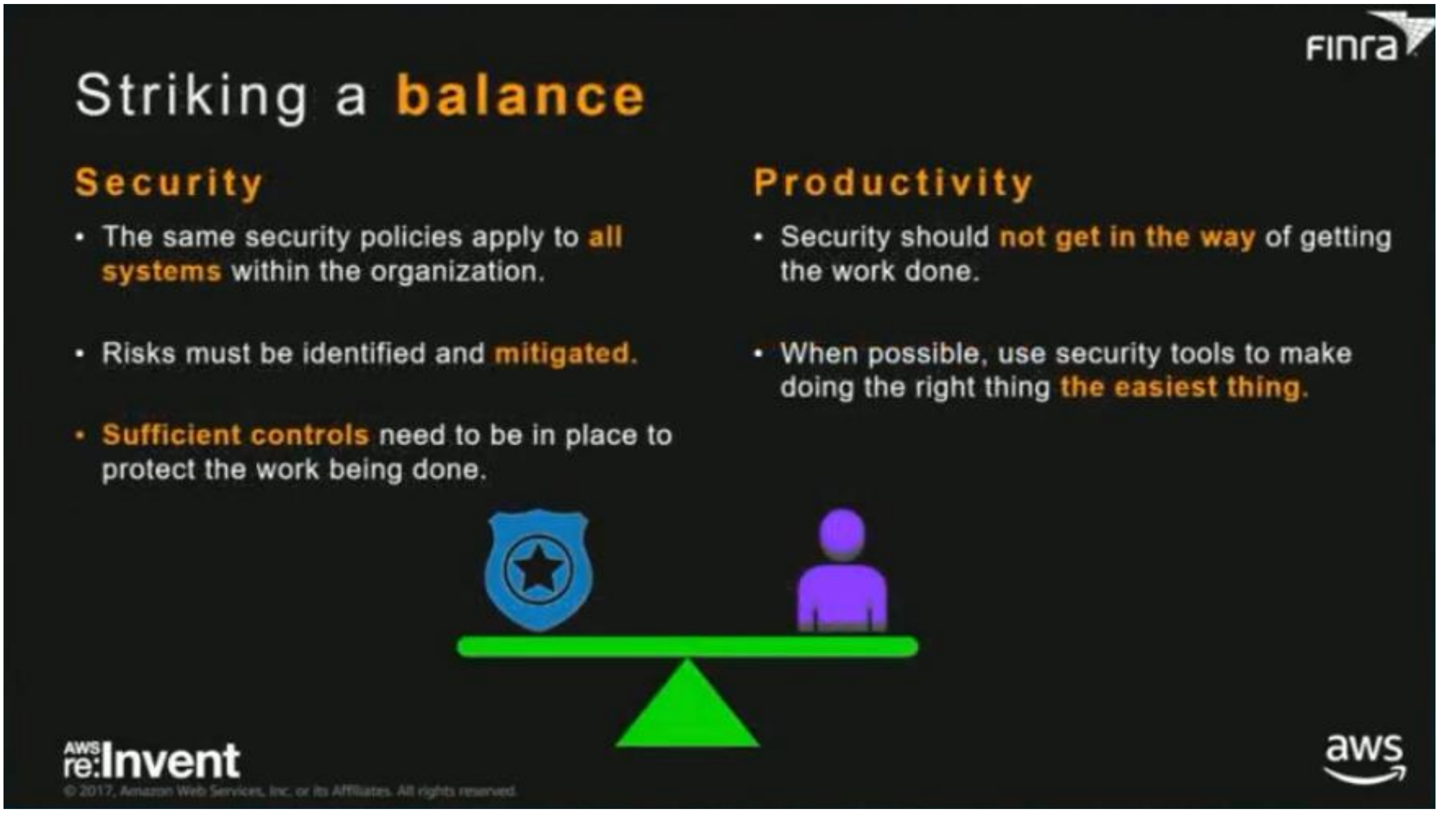
  Preserves ability to associate related records, some limited ability to operate on data (search, sort, categorize)
- Generalization, subsetting
- De-identification
  - Be wary of re-identification strategies

Limit Credential Exposure

- IAM Role-based access is ideal
- Secrets Management (Credstash)

Make Security Easy

- Internal mirrors of external resources preserves isolation
- Empower users, managers with utilization/cost information, necessary entitlements to provide oversight.



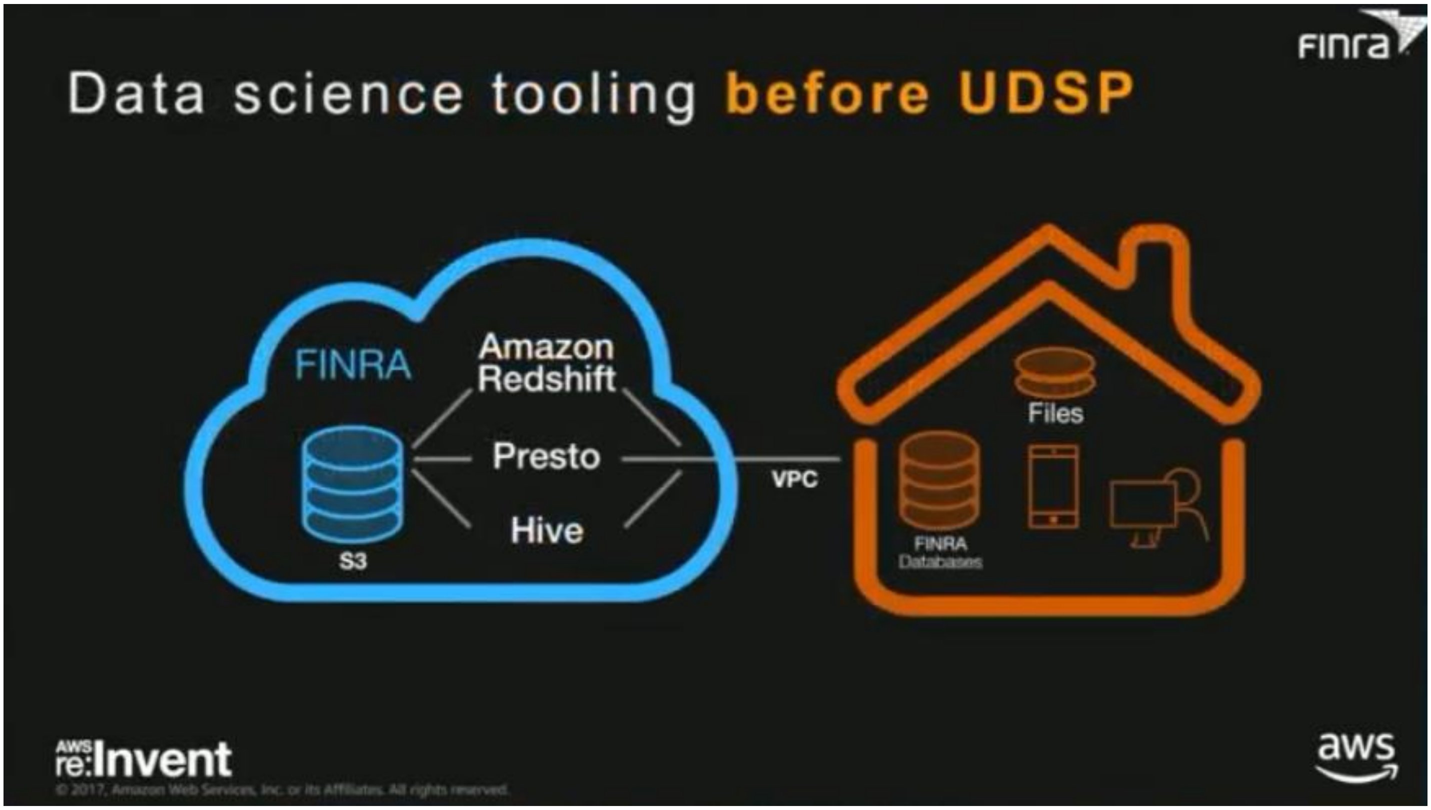
Striking a balance

Security

- The same security policies apply to **all systems** within the organization.
- Risks must be identified and **mitigated.**
- **Sufficient controls** need to be in place to protect the work being done.

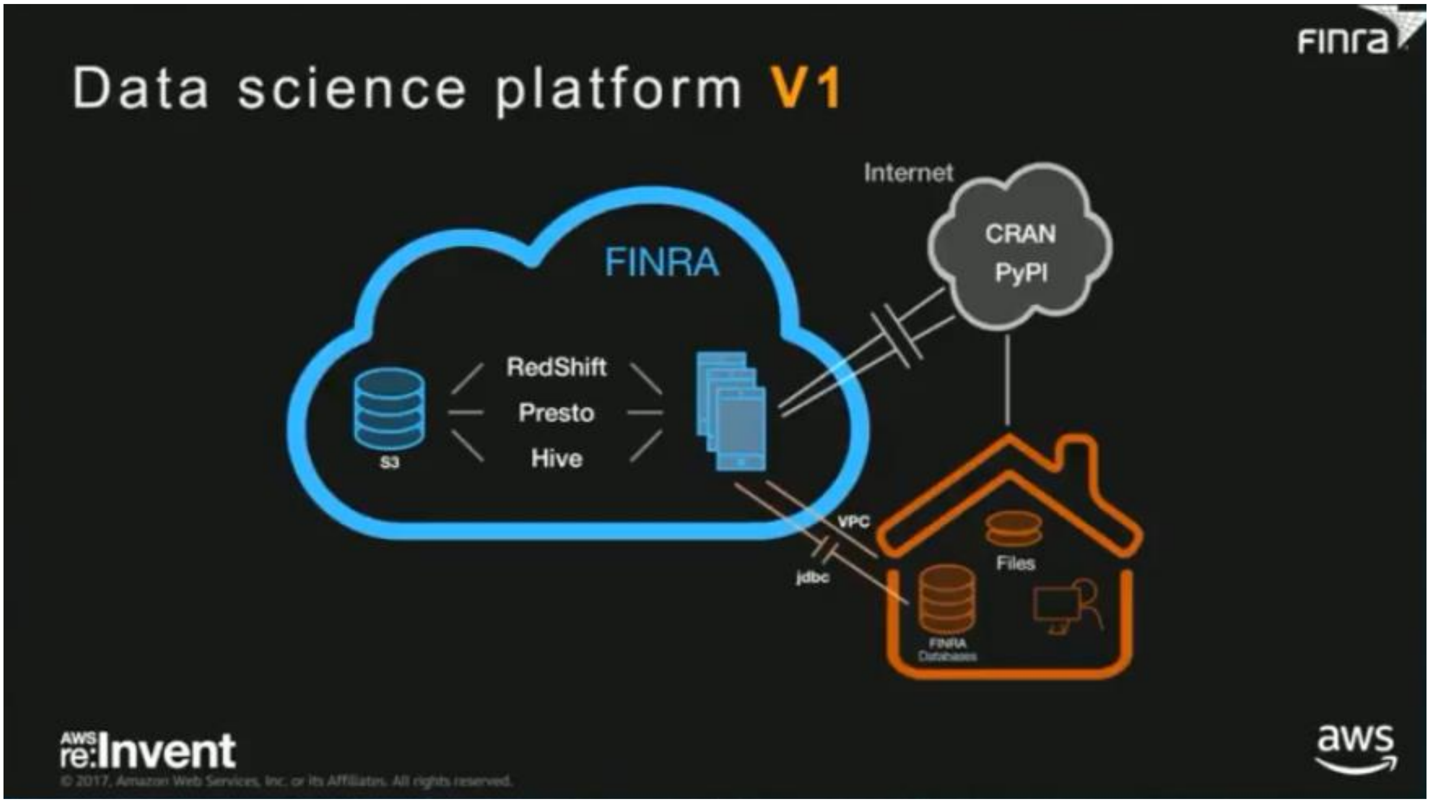
Productivity

- Security should **not get in the way** of getting the work done.
- When possible, use security tools to make doing the right thing **the easiest thing.**



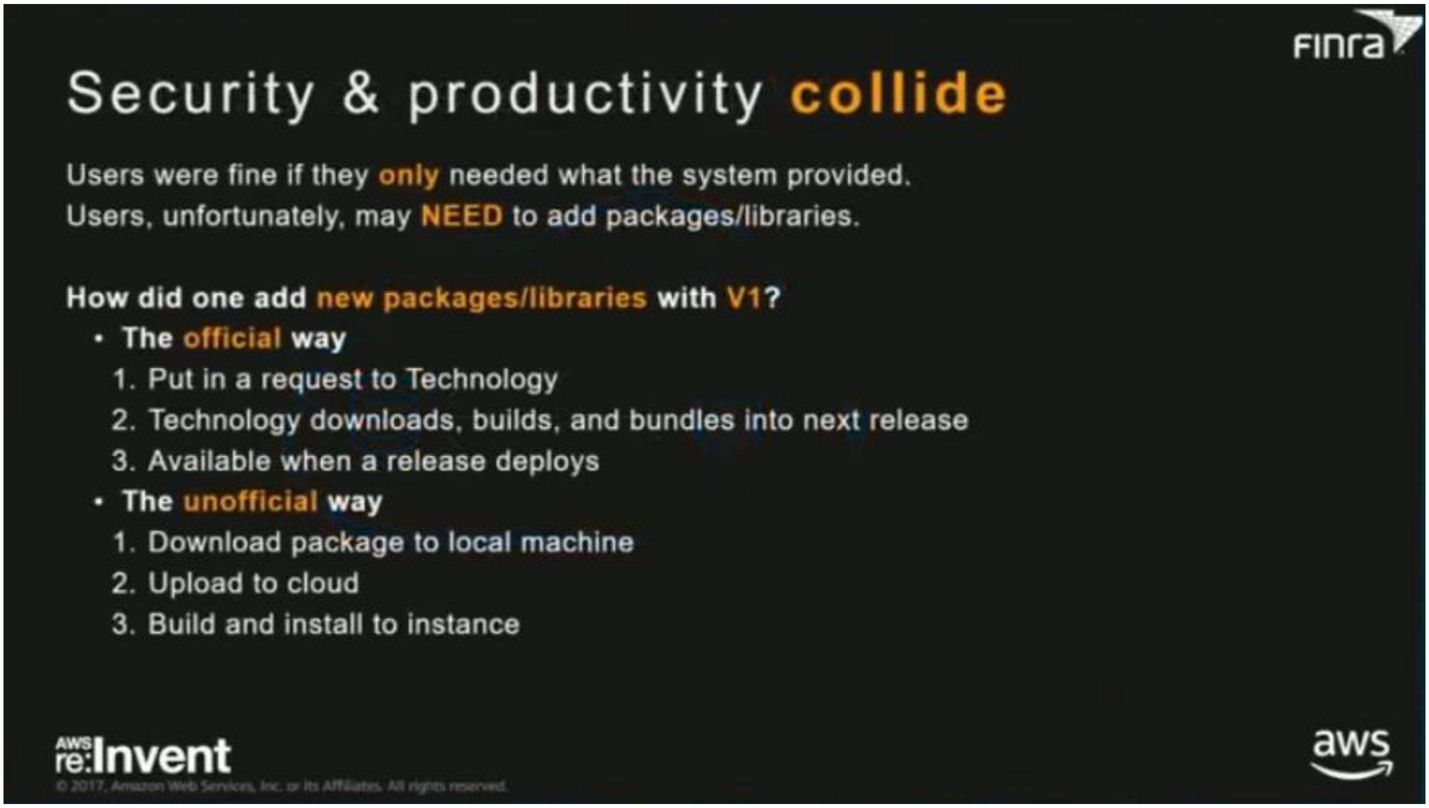
Data science tooling before UDSP

We moved all our data into S3 and provide access to it through the various SQL engines like Redshift, Presto, Hive, and added access to on-premise data in databases and files via the VPC.



Data science platform V1

We now move the compute into the cloud by providing EC2 instances with the SQL engines for querying and working with the S3 data using some Python libraries.

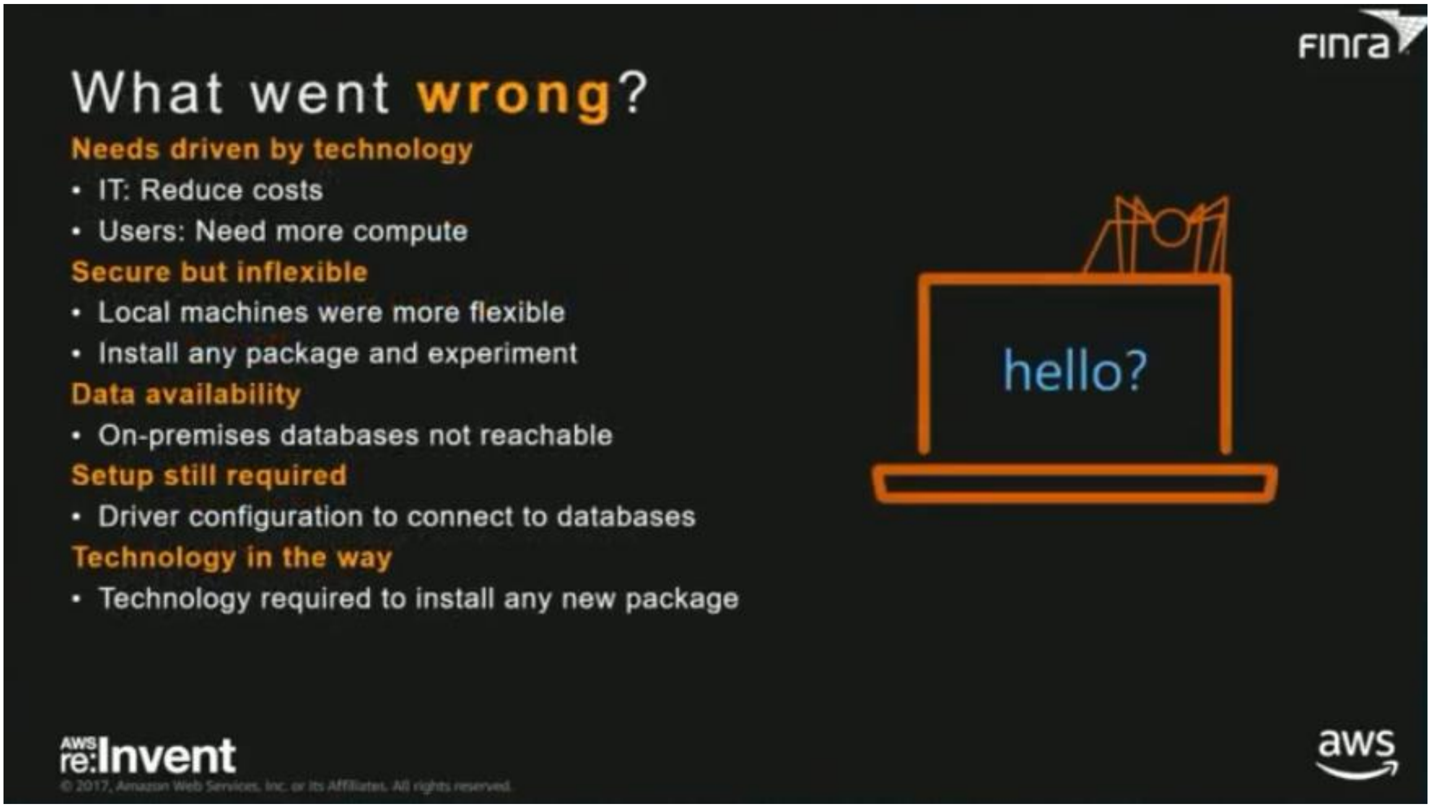


Security & productivity collide

Users were fine if they only needed what the system provided.
Users, unfortunately, may NEED to add packages/libraries.

How did one add new packages/libraries with V1?

- **The official way**
  1. Put in a request to Technology
  2. Technology downloads, builds, and bundles into next release
  3. Available when a release deploys
- **The unofficial way**
  1. Download package to local machine
  2. Upload to cloud
  3. Build and install to instance



What went wrong?

Needs driven by technology

- IT: Reduce costs
- Users: Need more compute

Secure but inflexible

- Local machines were more flexible
- Install any package and experiment

Data availability

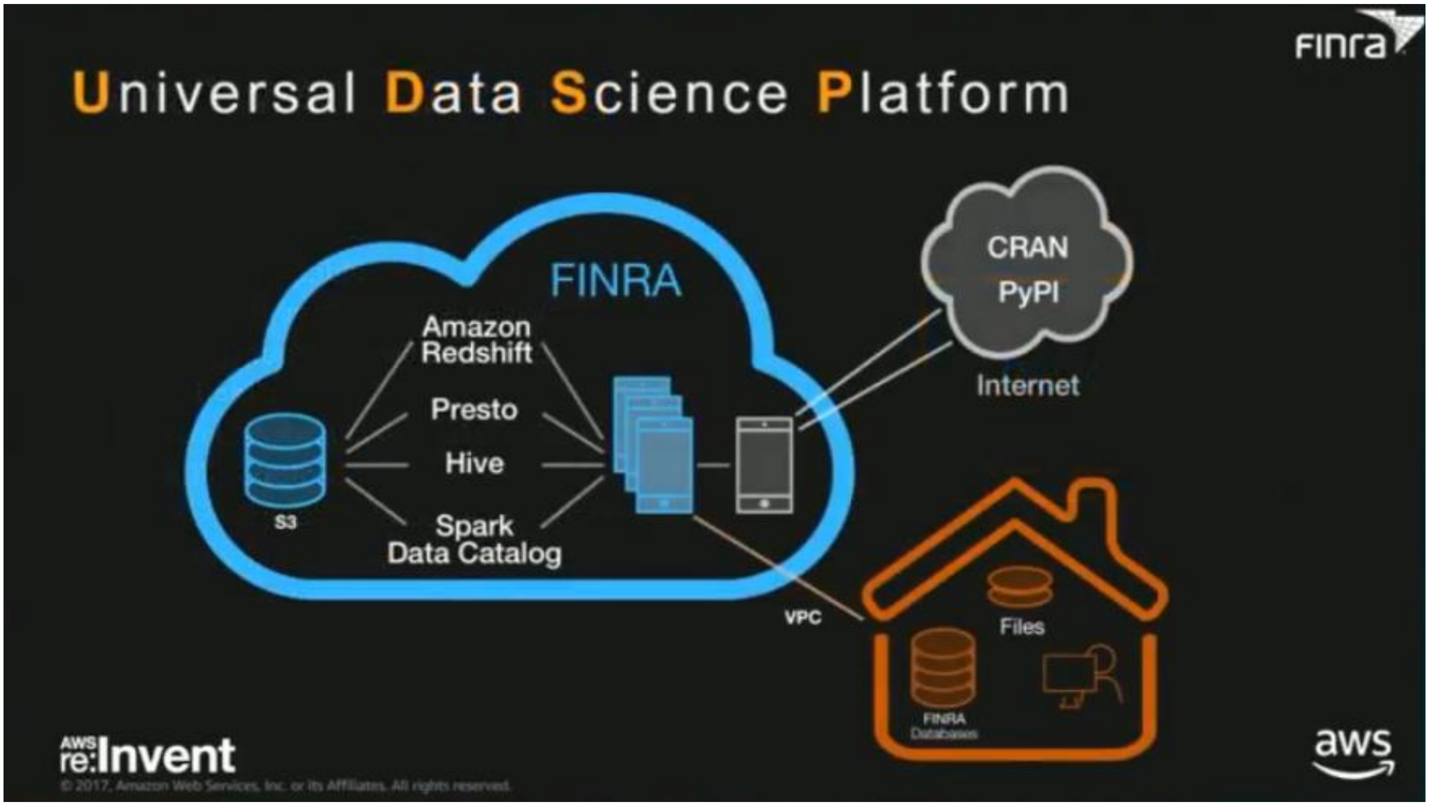
- On-premises databases not reachable

Setup still required

- Driver configuration to connect to databases

Technology in the way

- Technology required to install any new package



Universal Data Science Platform

FINRA

Amazon Redshift

Presto

Hive

Spark Data Catalog

S3

CRAN

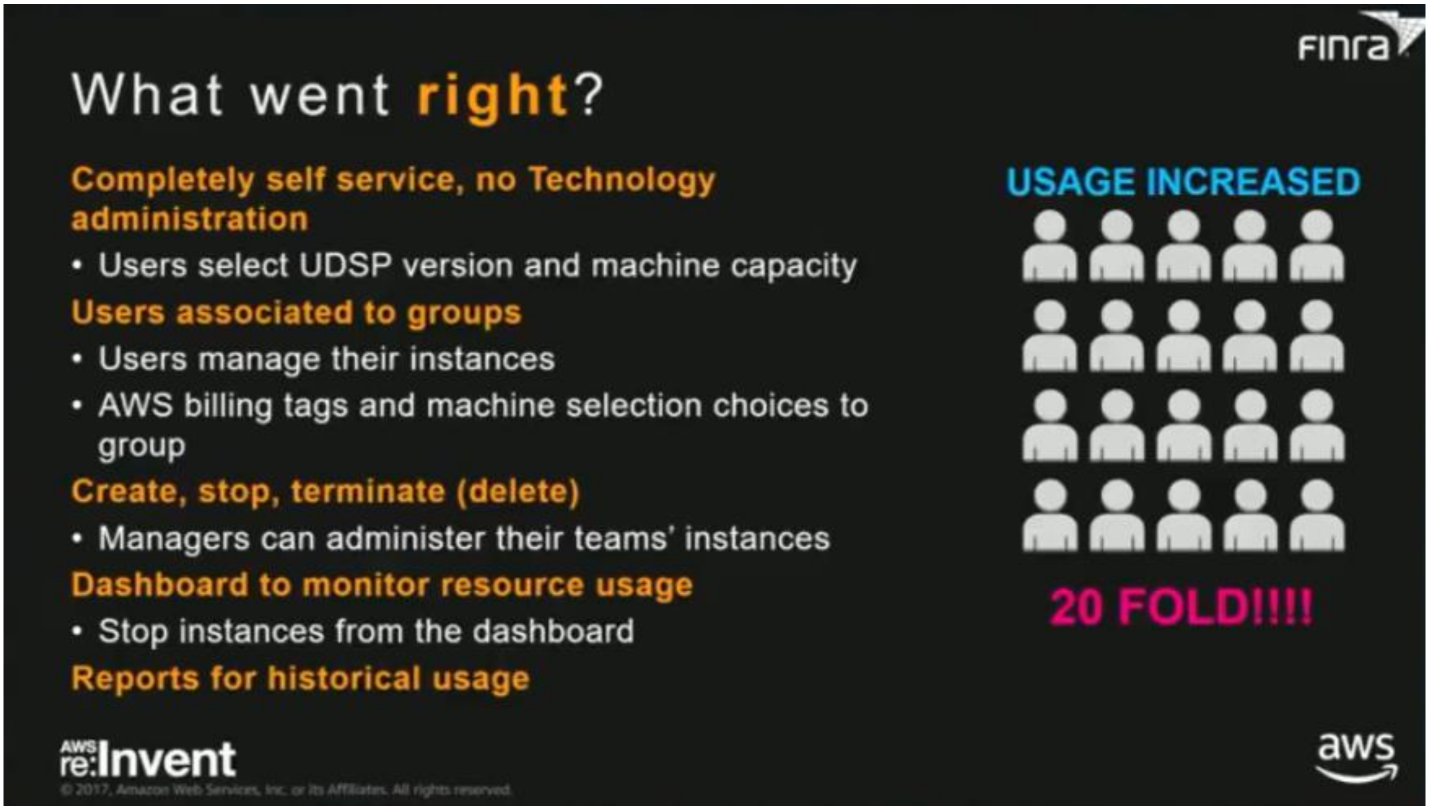
PyPI

Internet

VPC

Files

FINRA Databases



What went right?

Completely self service, no Technology administration

- Users select UDSP version and machine capacity

Users associated to groups

- Users manage their instances
- AWS billing tags and machine selection choices to group

Create, stop, terminate (delete)

- Managers can administer their teams' instances

Dashboard to monitor resource usage

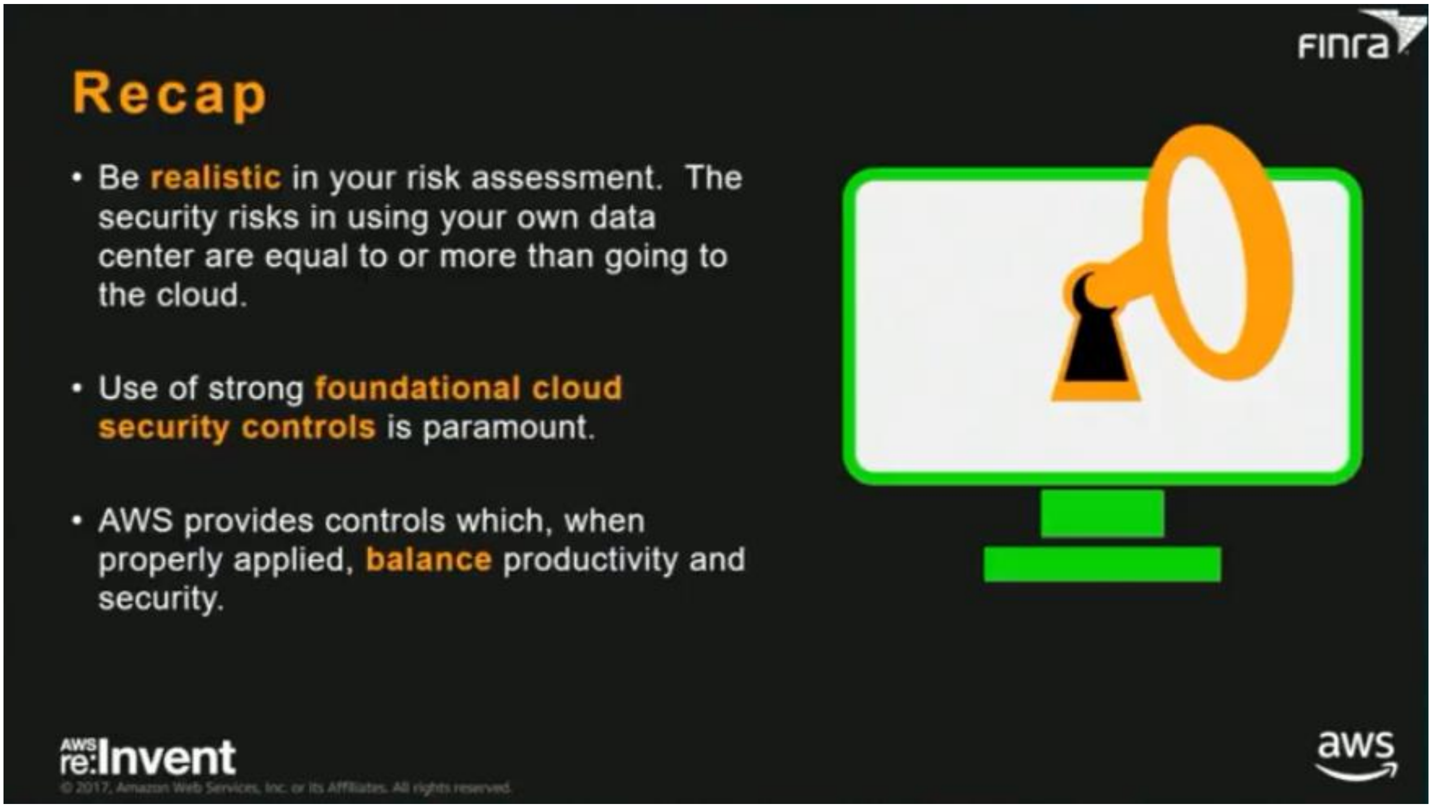
- Stop instances from the dashboard

Reports for historical usage

USAGE INCREASED

20 FOLD!!!!

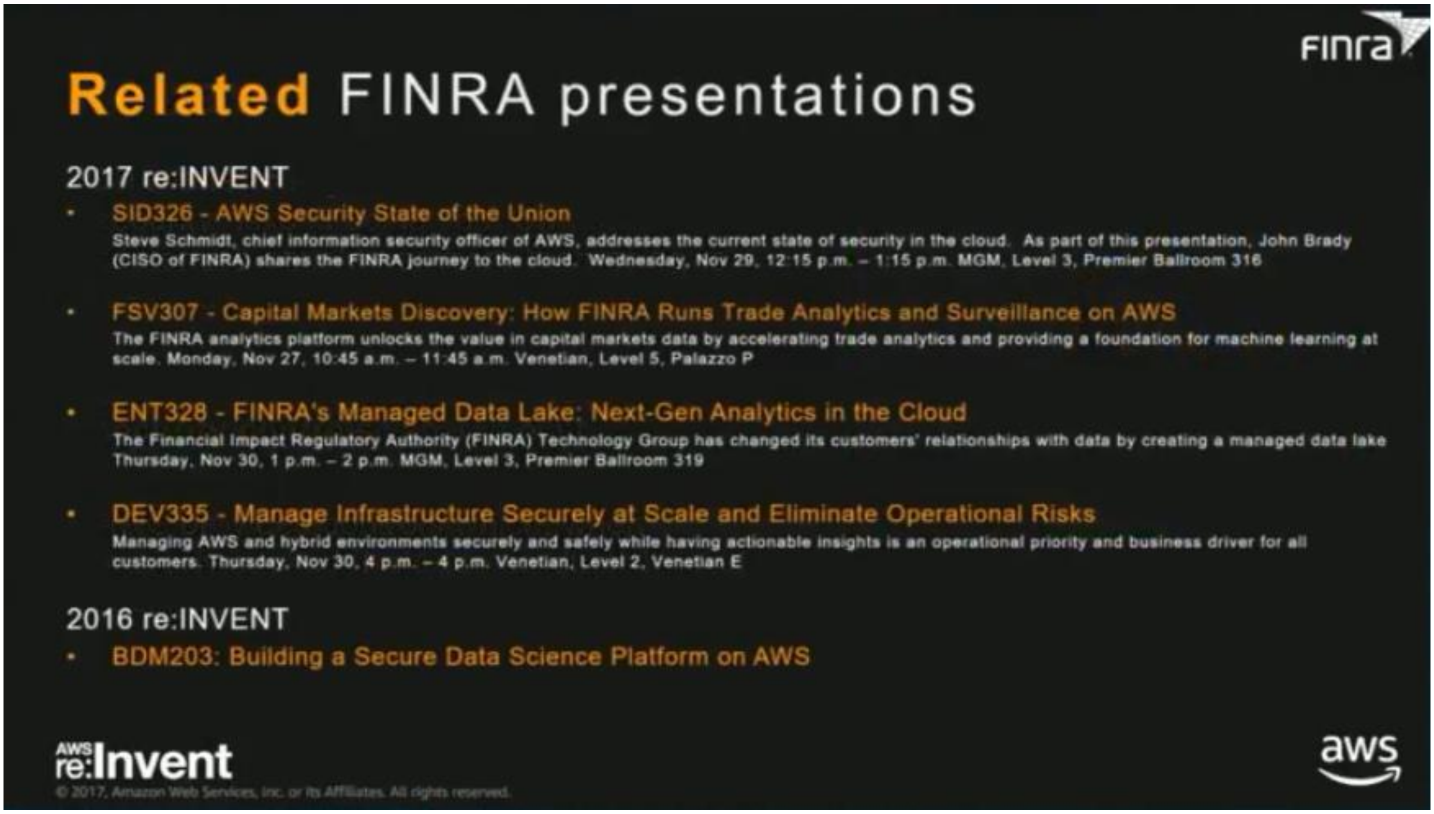
re:Invent



Recap

- Be **realistic** in your risk assessment. The security risks in using your own data center are equal to or more than going to the cloud.
- Use of strong **foundational cloud security controls** is paramount.
- AWS provides controls which, when properly applied, **balance** productivity and security.

re:Invent



Related FINRA presentations

2017 re:INVENT

- **SID326 - AWS Security State of the Union**
  Steve Schmidt, chief information security officer of AWS, addresses the current state of security in the cloud. As part of this presentation, John Brady (CISO of FINRA) shares the FINRA journey to the cloud. Wednesday, Nov 29, 12:15 p.m. – 1:15 p.m. MGM, Level 3, Premier Ballroom 316
- **FSV307 - Capital Markets Discovery: How FINRA Runs Trade Analytics and Surveillance on AWS**
  The FINRA analytics platform unlocks the value in capital markets data by accelerating trade analytics and providing a foundation for machine learning at scale. Monday, Nov 27, 10:45 a.m. – 11:45 a.m. Venetian, Level 5, Palazzo P
- **ENT328 - FINRA's Managed Data Lake: Next-Gen Analytics in the Cloud**
  The Financial Impact Regulatory Authority (FINRA) Technology Group has changed its customers' relationships with data by creating a managed data lake Thursday, Nov 30, 1 p.m. – 2 p.m. MGM, Level 3, Premier Ballroom 319
- **DEV335 - Manage Infrastructure Securely at Scale and Eliminate Operational Risks**
  Managing AWS and hybrid environments securely and safely while having actionable insights is an operational priority and business driver for all customers. Thursday, Nov 30, 4 p.m. – 4 p.m. Venetian, Level 2, Venetian E

2016 re:INVENT

- **BDM203: Building a Secure Data Science Platform on AWS**