AWS State, Local, and Education Learning Days

Chicago

aws Learning Days
State, Local, and Education

Compliant Research Data Architecture and Data Sharing

Niris Okram (he/him)

Sr. Solutions Architect
AWS
niris@amazon.com

Ignatius Narchetty (he/him)

Solutions Architect
AWS
inarchet@amazon.com

aws Learning Days
State, Local, and Education

Why now?

Securing research data has never been more important.

Hacker Groups

Criminal gangs

Ransomware

IP theft

Nation states

Internal threats

Why now?

Securing research data has never been more important.
Research data has value and is an active target.

Ransom costs **Remediation costs** **Forensic costs**

Hacker groups

Criminal gangs

Ransomware

IP theft

Nation states

Internal threats

Reputational damage

Why now?

Responsible stewardship of research data is expected.

HIPAA

GDPR

PIPA

Data privacy

Responsible use

NIST / CUI

DFARS

CMMC

Why now?

Responsible stewardship of research data is expected.

Compliance defines specific responsibilities for research data.

Penalties and fines

data breach

HIPAA

GDPR

PIPA

Data privacy

Funding withheld

non-compliance/data exfiltration

NIST / CUI

DFARS

CMMC

Responsible use

Reputational damage

Research presents a unique challenge

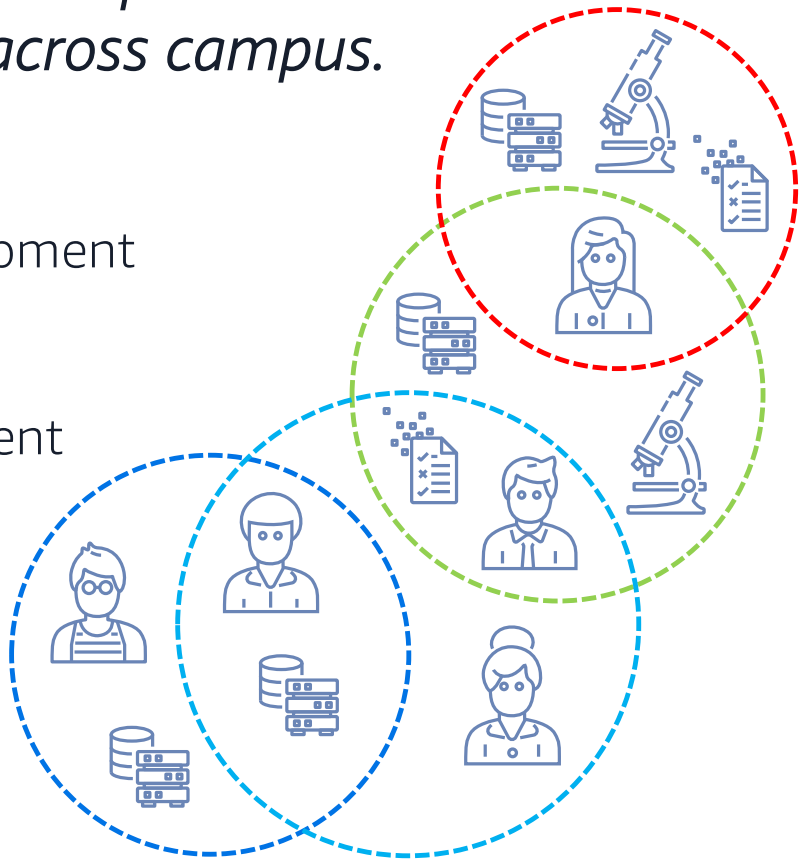
Research presents a unique challenge

Research is challenging to secure and make compliant because it often operates within and between islands across campus.

Factors:

- Faculty/researcher procured and managed equipment

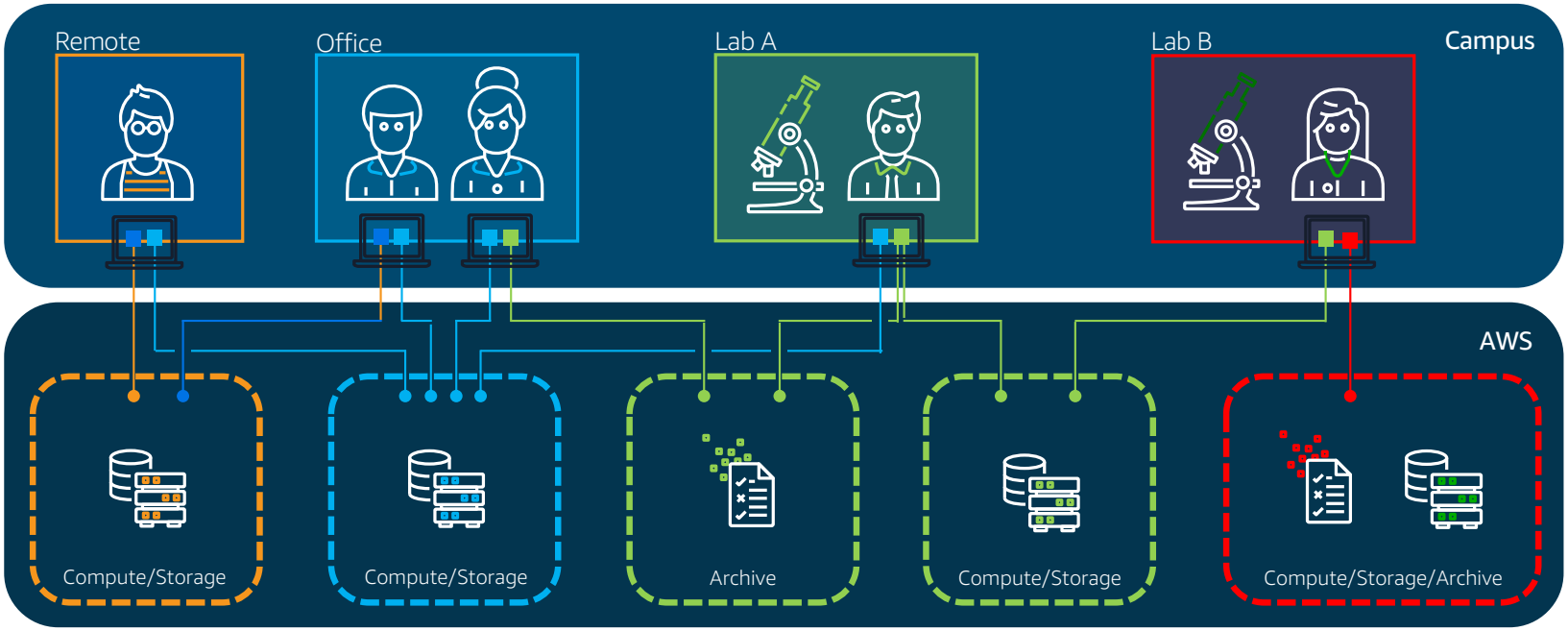
- Faculty/researcher/student population
  - Collaborative, distributed, mobile, and transient
  - Bring your own device (BYOD)



Why AWS?

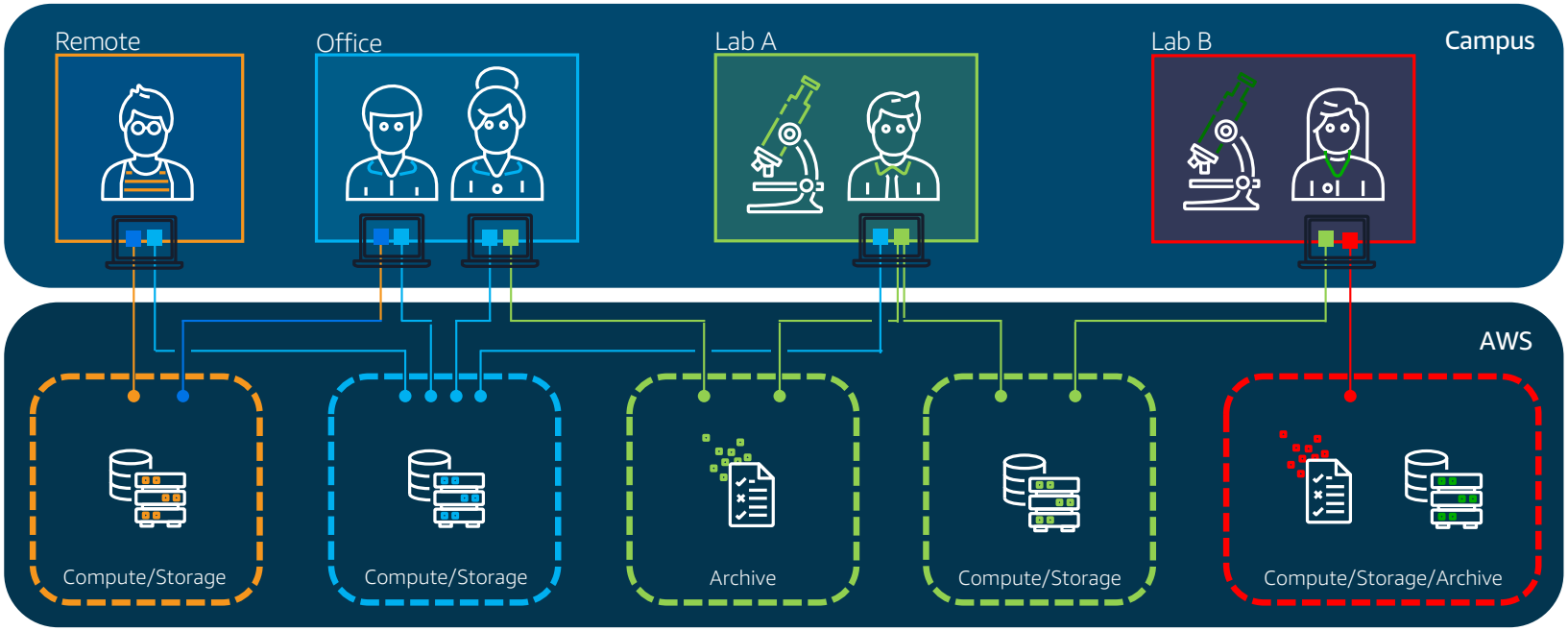
Why AWS?

Enables deployment of repeatable research environments that help institutions achieve their security and compliance goals



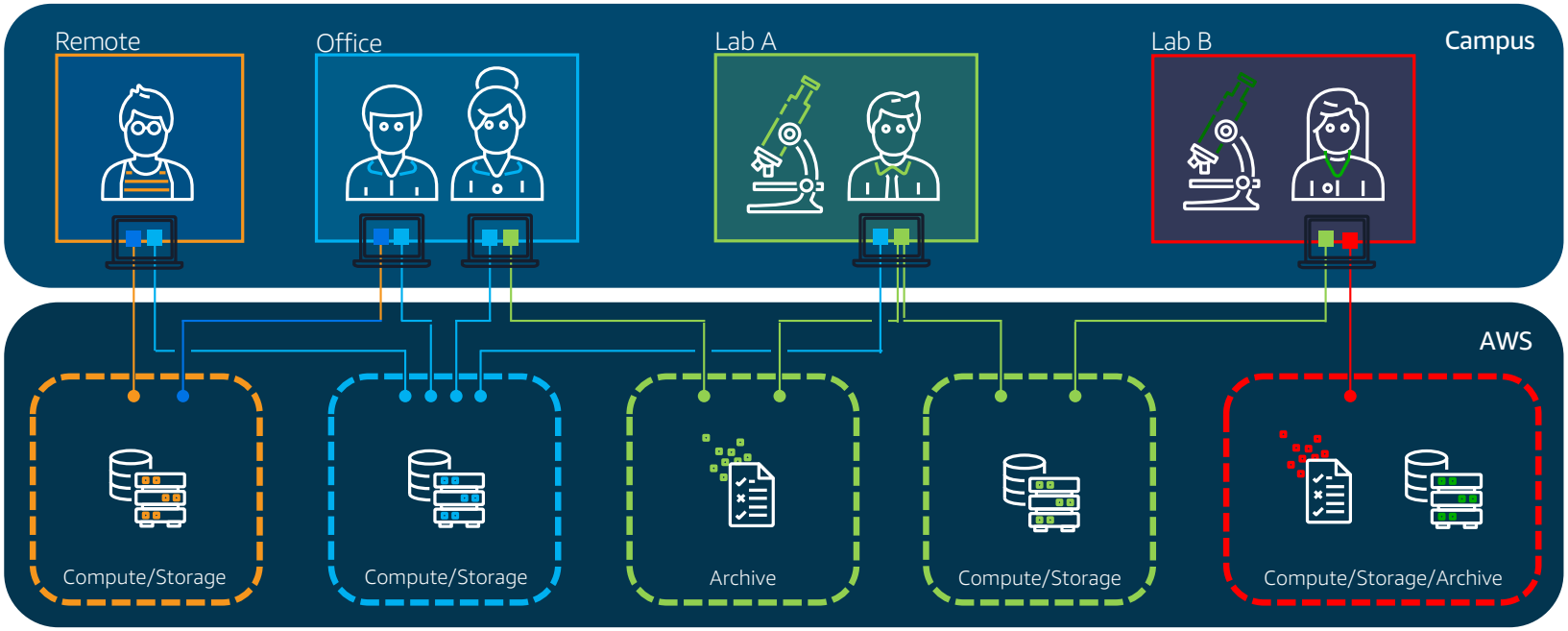
Why AWS?

Offers a wide range of flexible, on-demand infrastructure that enables and evolves with researcher demand



Why AWS?

Gives researchers and institutions the flexibility to meet research, security, and compliance needs



Why AWS?

Provides a wide range of services enabling institutions to create solutions to meet their security and compliance requirements

Identity & access management	Detection	Infrastructure protection	Data protection	Incident response
IAM	Security Hub	Firewall Manager	Macie	Detective
AWS SSO	GuardDuty	Shield	AWS KMS	CloudEndure DR
Organizations	Amazon Inspector	AWS WAF	CloudHSM	AWS Config Rules
Directory Service	CloudWatch	Amazon VPC	ACM	Lambda
Amazon Cognito	AWS Config	AWS PrivateLink	Secrets Manager	
AWS RAM	CloudTrail	Systems Manager	AWS VPN	
	VPC Flow Logs		Server-Side Encryption	

Why AWS?

Elevates your institution's research capabilities along with its security and compliance posture

Inherit global security and compliance controls

Scale with superior visibility and control

Highest standards for privacy and data security

Automate and reduce risk with deeply integrated services

Largest community of security partners and solutions

How?

You need a landing zone

- A secure, scalable, multi-account AWS environment based on AWS best practices
- A starting point for net new development and experimentation
- A starting point for migrating applications
- An environment that allows for iteration and extension over time

Landing zone elements

Secure and compliant

Meets the organization's security and auditing requirements

Scalable and resilient

Ready to support highly available and scalable workloads

Adaptable and flexible

Configurable to support evolving mission requirements

The **Landing Zone Accelerator on AWS** is an open-source software solution that accelerates the implementation of a customer's technical security controls and infrastructure foundation on AWS

Landing Zone Accelerator benefits

How AWS delivers

Example: secure and compliant landing zone

UCSD Health Secure Research Cloud (HSRC) for HIPAA compliance

Drivers

- Prevent removal of research data assets and inappropriate third-party data transfers.
  (IRB vs. policy and legal compliance)
- Prevent proliferation of unmanaged cloud accounts.
  (and gain visibility to monitor activity, data types, workloads, and potential risks)
- Prevent ransomware and research data on mobile devices as a breach source
  (unmanaged, unprotected, or misconfigured devices)

Partnered with AWS, UCSD Health IS security, institutes, and research groups early

- Compliance is more than technical controls: BAA, governance, and policy

How AWS delivers

Example: secure and compliant landing zone

UCSD Health Secure Research Cloud (HSRC) for HIPAA compliance

Solution goals

- Access controls – technical policies and procedures allowing only authorized persons to access electronic protected health information (ePHI)
- Audit controls – hardware, software, and/or procedural mechanisms to record and examine access and other activity
- Integrity controls – policies, procedures, and measures to ensure and confirm ePHI is not improperly altered or destroyed
- Transmission security – technical security measures guarding against unauthorized access to ePHI transmitted over a network

How AWS delivers

Example: deployment of a research workload

UCSD Health Virtual Research Desktop (VRD) – within UCSD HSRC (saw earlier)

Researcher enablement

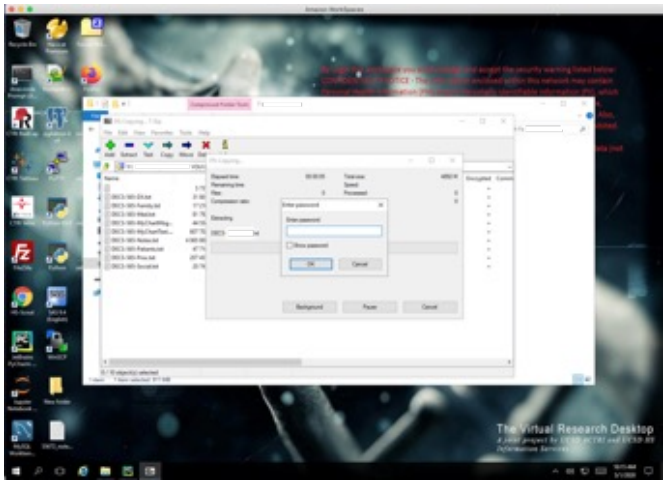
- A solution that balances security and privacy while still providing a quality user experience
- Access to ePHI via UCSD Data Extraction Concierge Service (DECS) and VRDs
  (data extracted from clinical data warehouse by DECS and placed into investigator's VRD "secure" folder)
- Hardened Amazon WorkSpaces Windows 10 virtual machines
  - Runs within UCSD HSRC and approved by UCSD Health CISO for ePHI
  - Provisioned with: SPSS, R/RStudio, Python/PyCharm, Java 8, and others
  - With approval, access to internal databases

How AWS delivers

Example: deployment of a research workload

UCSD Health Virtual Research Desktop (VRD) – within UCSD HSRC (saw earlier)

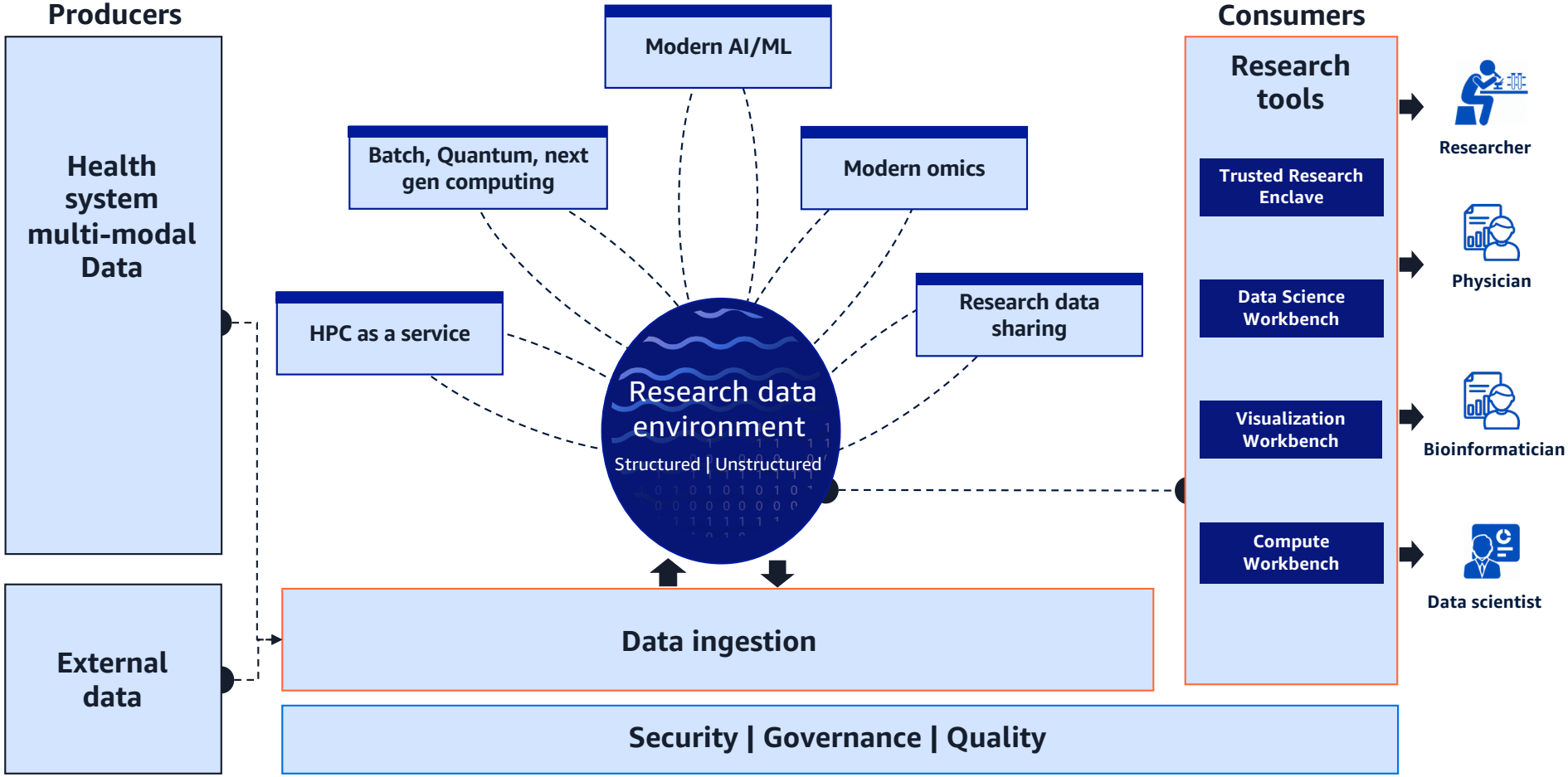
View of remote desktop from investigator's computer



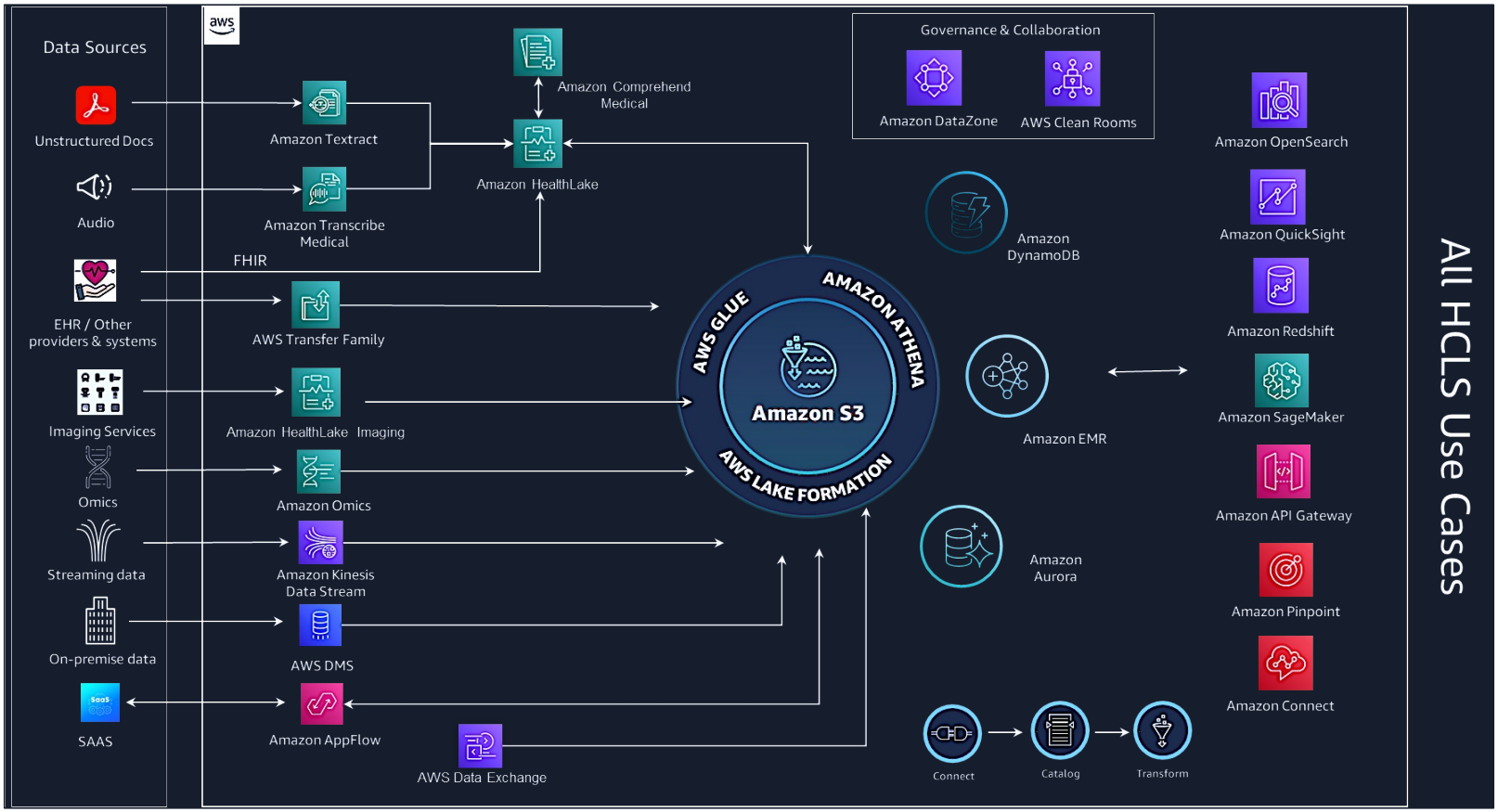
VRD desktop with application running

What about data?

Research for health on AWS

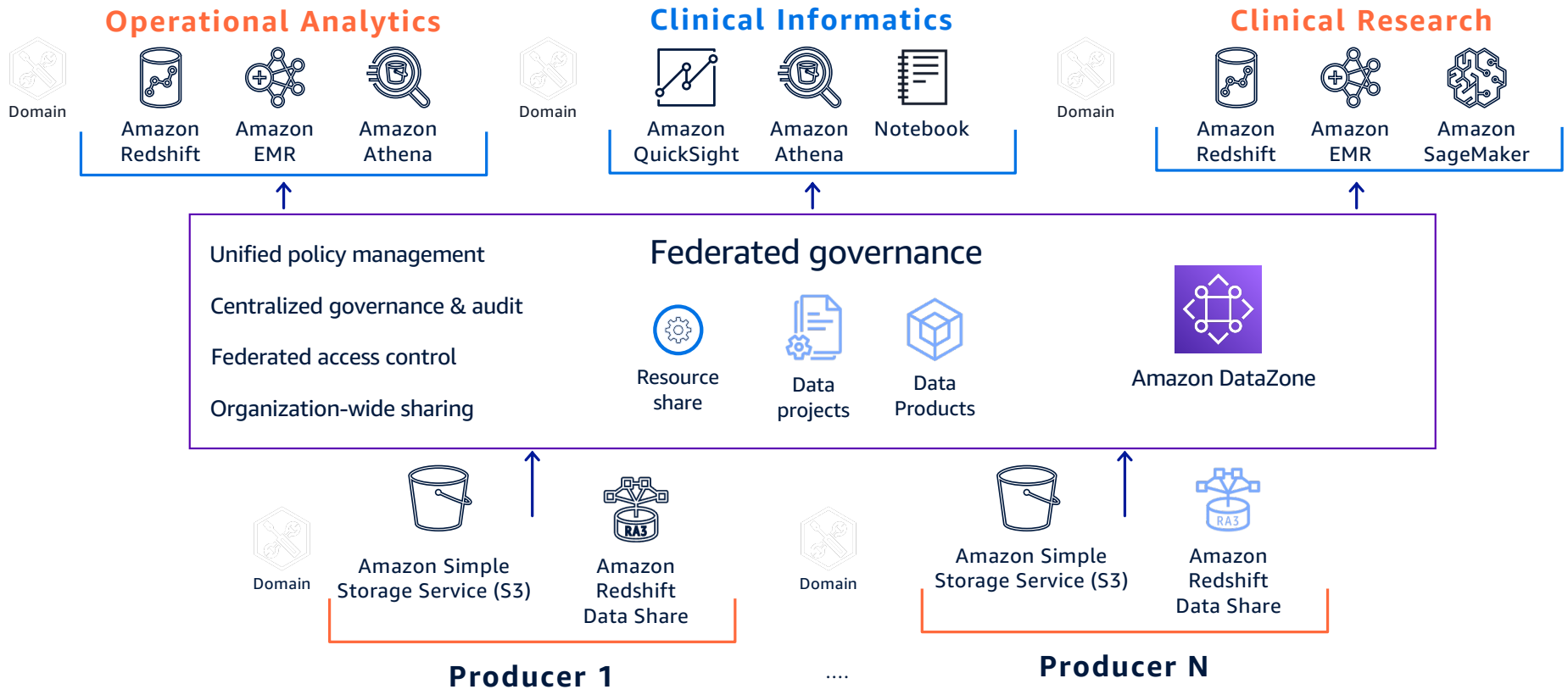


Modern Health Data Platform on AWS



Modern Health Data Mesh Architecture

Decentralized, lightweight federated governance across domain-oriented data systems to drive governed sharing



Research for health

Secure, self-service research from Research IT to Researcher

Research Data Platform	Modern Omics	HPC as a Service	ML for Research	Trusted Research Enclave	Next Gen Research Compute	Data Sharing and Federation
Singular OMOP data platform for all research data	High scale, high performance genomics cloud services	Centralized large scale Batch and HPC research on demand clusters	One platform for all ML and data science needs	Secure and isolated research enclaves for PIs	Massive compute scale with latest generation compute	Secure data sharing, and data cleanroom
Research use cases						
Multi-modal research data, de-identified data, research data meshes, genomics storage, imaging storage	Secondary Analysis, Tertiary Analysis, Genomics workflows, native Genomics CLI support	Research HPC clusters, SLURM, Genomics, Massive Batch computing	Deep learning, machine learning Imaging AI, AI assisted annotation, PyTorch, TensorFlow, CNN, DNN	Enclaves for researcher workbenches for data, compute, data science, and visualization	Nextgen NVIDIA GPUs, FPGAs, ARM, Intel, AMD, and Quantum	NIH DMS 2023 sharing, research consortia, federated learning, federated queries

Amazon DataZone

UNLOCK THE POWER OF ALL DATA FOR ALL USERS WITH TRUSTED AUTONOMY

Data producers	Amazon DataZone	Data consumers
Teams who want to share data	Team who runs the data marketplace	Teams who want to use data

Amazon DataZone

UNLOCK DATA ACROSS ORGANIZATIONAL BOUNDARIES WITH BUILT-IN GOVERNANCE

Manage **organization-wide governance** in one place

Catalog your data with business context

Simplify access to analytics for everyone in your organization

Solve specific business use cases through **data projects**

AWS Clean Rooms helps organizations collaborate on datasets without sharing underlying data

Multi-party collaborations

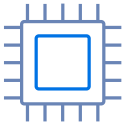
Collaborate with up to five parties in a single collaboration; extract insights from multiple companies

No AWS data movement

Use Amazon S3 data with direct permissioning and no AWS data movement

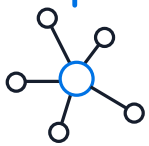
Query controls and enforcement

Configure analysis rules to restrict the type of analysis allowed on your data



Cryptographic computing

Pre-encrypt data so that it is encrypted at all times, including during query execution



Programmatic access

Automate and integrate functionality into existing workflows and products; create white-labeled clean room offering

CHOP accelerates pediatric research using AWS-powered data resource

Challenge

As medical researchers generate more and more clinical data, they're faced with the challenge of storing and organizing that data so that researchers can access, study, and cross-reference it to facilitate medical breakthroughs.

Benefits

CHOP provided the research community with access to genomic and associated clinical data and increased KFDRC's collaborative potential.

CHOP stored 26 billion occurrences of 215 million unique genomic variants from 5,000 participants, while meeting the FHIR industry standard

Solution

CHOP built the Gabriella Miller Kids First Data Resource Center (KFDRC), a data source that brings genomics, clinical and imaging data as an open resource for researchers to focus on discovers in pediatric cancer and structural birth defects.

> All of our system is currently built on AWS. . . . We went from zero to managing a few petabytes of genomic data within a year using this setup."
>
> Allison Heath
> Director of Data Technology and Innovation, Center for Data-Driven Discovery in Biomedicine

aws

Q & A

Thank you!

Niris Okram (he/him)
Sr. Solutions Architect
AWS
niris@amazon.com

Ignatius Narchetty (he/him)

Solutions Architect
AWS
inarchet@amazon.com

Track: **Data and Analytics**

Session: **Compliant Research Data Architecture and Data Sharing**

aws Learning Days
State, Local, and Education