

Hawai'i Cloud Innovation Summit 2023

# Democratizing Data Analytics

Leveraging AI and ML to empower individuals to make data-driven decisions and foster innovation

**Pranav Khadilkar**

Solutions Architect

AWS

**Kelli Gonsalves**

Data Scientist

DataHouse

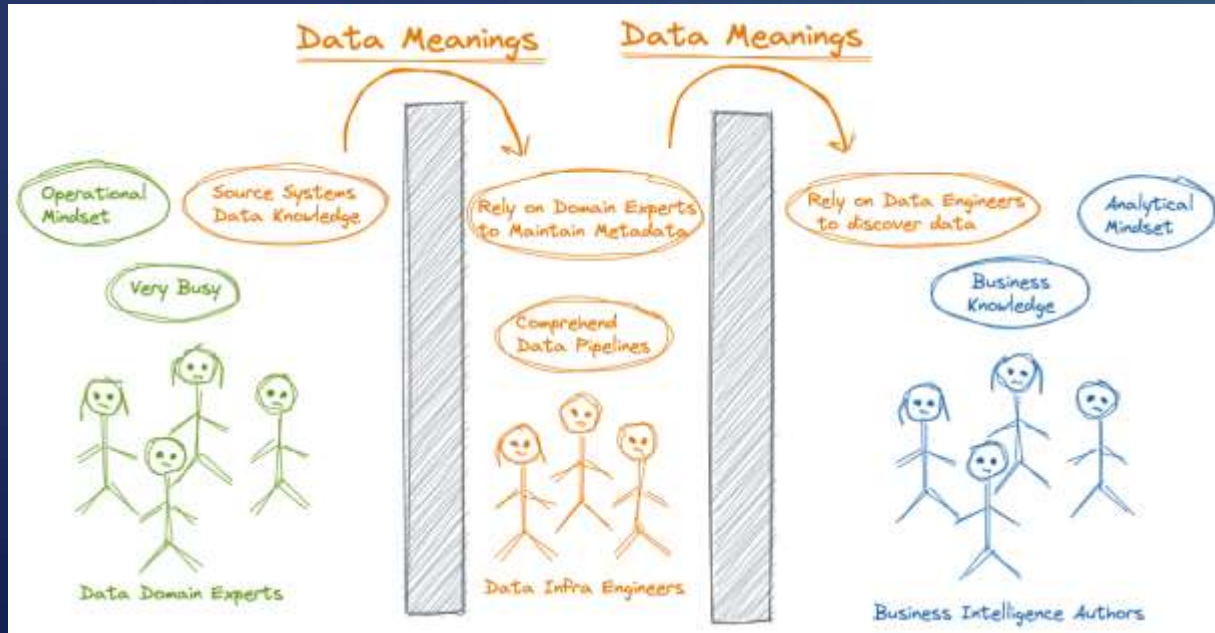
**Jeff Masumoto**

VP Solutions and Consulting Services

DataHouse



# Democratizing – Data and Analytics



Making data discoverable and available for right people at right time. To unlock data opportunity, it is important that right analytics tool is available to everyone when needed.

# Why democratize Data and Analytics

# What if I can democratize data and Analytics?

- Achieve your business outcomes



## Improve operational efficiency

Increase agility, run your business more efficiently, and provide a better experience for customers by moving data to the cloud



## Make more informed decisions

Inform business decision making with more meaningful insights by bringing together the full picture of data across your organization



## Accelerate innovation

Unlock opportunities that were either too difficult or impossible to do before by automating processes with AI and machine learning

By making 10% more data accessible, a typical Fortune 1000 company will see a **\$65 million increase in net income**<sup>1</sup>

<sup>1</sup> Dykes, "The Four Key Pillars To Fostering A Data-Driven Culture," [www.forbes.com/sites/brentdykes/2019/03/28/the-four-key-pillars-to-fostering-a-data-driven-culture/?sh=45d40bee7d90](http://www.forbes.com/sites/brentdykes/2019/03/28/the-four-key-pillars-to-fostering-a-data-driven-culture/?sh=45d40bee7d90)



# Key Challenges

# However, challenges are in the way



More data than ever is being generated



Data of all types is stored in silos across multiple data stores



Machine learning adoption is challenged by lack of skills and organizational inertia



Data security, privacy, and compliance regulations are increasingly important



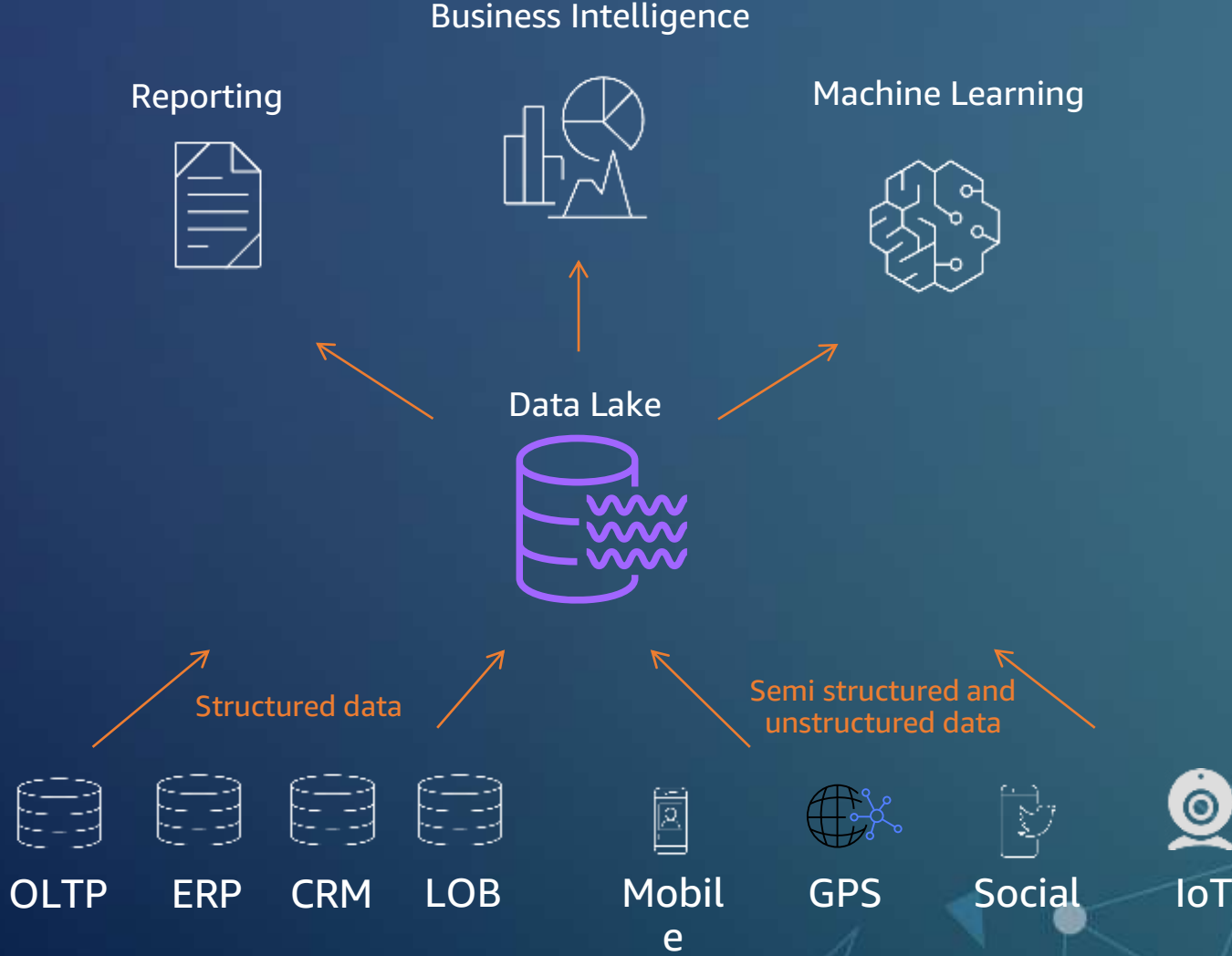
# Data silos



# How to Solve these Challenges

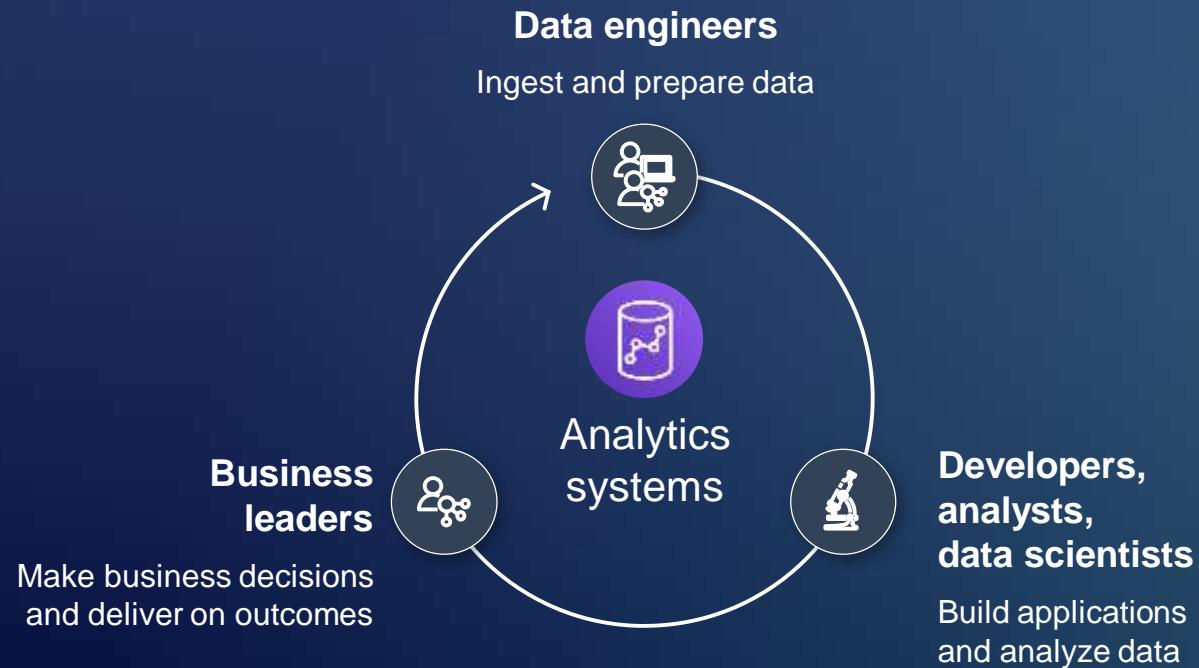


# Data Lakes | Democratize Data



# Easy analytics for everyone | Democratize Analytics

- Focus on getting from data to insights in seconds



## Automatic provisioning and scaling

Automatically provisions and scales the underlying compute resources to deliver high performance for demanding and unpredictable workloads

## Visualize your data

Single, visual interface for querying data to improve productivity through one-click visual analytics, collaboration, version control, and scheduling

## Bypass administrative tasks

Take advantage of automated provisioning, backup, patching, tuning, and monitoring in Amazon Redshift

# Lake House Architecture



SCALABLE DATA LAKES

PURPOSE-BUILT DATA SERVICES

SEAMLESS DATA MOVEMENT

UNIFIED GOVERNANCE

PERFORMANT AND COST-EFFECTIVE

# Scalable Data Lakes



Variety of sources  
and data types



Data volume and velocity



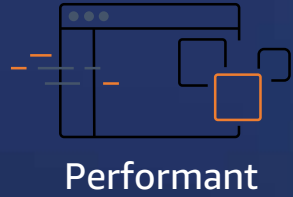
Cost Effective



A data lake is a **centralized repository** that allows you to store all your **structured and unstructured** data at any scale



# Seamless Data Movement



## Inside out

*Example:*

Move only the last 24 hours of application clickstream data from the data lake to a data warehouse for daily reporting.

## Outside in

*Example:*

Move query results for regional sales of products from the data warehouse into the data lake to run product recommendation algorithms against a larger dataset using ML.

## Around the perimeter

*Example:*

Copy the product catalog data from a database to a search service to make it easier to look through their product catalog and offload the search queries from the database.

# Purpose Built Services | Performant and Cost-Effective



Performance



Scalability



Cost Effective



- ❖ Decouple storage from compute
- ❖ Prefer Serverless over Managed over Self-hosted
- ❖ Pay as you go
- ❖ Compress and partition
- ❖ Purpose-built service for higher performance
- ❖ Data lake for lower cost

# Unified Governance



Secure and  
Compliance



Manageable



- ❖ Governance as an Enabler
- ❖ Flexible Granular Access Control
- ❖ Auditability
- ❖ Security In The Cloud vs Security Of The Cloud

# Lake House Architecture on AWS



# Lake House architecture on AWS



Scalable data lakes

Purpose-built data services

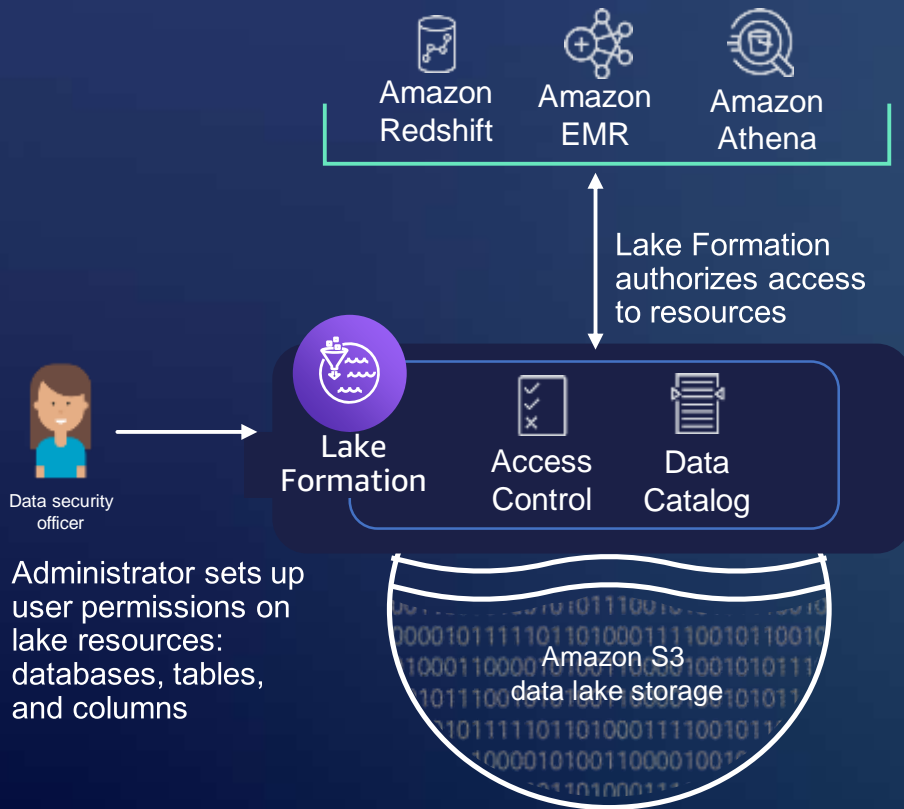
Seamless data movement

Unified governance

Performant and cost-effective

# Simplify security management with Lake Formation

Centrally define security, governance, and auditing policies in one place



Centrally define security, governance, and auditing

Policies are consistently enforced

Integrated with security, storage, analytics, and machine learning services

Permissions on databases, tables, and columns

# The AWS AI/ML stack

BROADEST AND MOST COMPLETE SET OF MACHINE LEARNING CAPABILITIES

## AI SERVICES

|                    |  |               |  |   |  |   |
|--------------------|--|---------------|--|---|--|---|
| <b>SPECIALIZED</b> | <b>BUSINESS PROCESSES</b>  | <b>SEARCH</b> | <b>CONVERSATION</b>  | <b>CODE + DEVOPS</b>  | <b>INDUSTRIAL</b>  | <b>HEALTH</b>   |
|                    | Amazon Personalize<br>Amazon Forecast<br>Amazon Fraud Detector<br>Amazon Lookout for Metrics | Amazon Kendra | Amazon Lex<br>Amazon Transcribe Call Analytics<br>Contact Lens<br>Voice ID | Amazon CodeGuru<br>Amazon CodeWhisperer<br>Amazon DevOps Guru | Amazon Monitron<br>Amazon Lookout for Equipment<br>Amazon Lookout for Vision | Amazon HealthLake<br>Amazon Comprehend Medical<br>Amazon Transcribe Medical<br>Amazon Omics |

|                      |  |             |                                       |                            |   |
|----------------------|--|-------------|---------------------------------------|----------------------------|---|
| <b>Generative AI</b> | <b>TEXT/VISION</b>                                     | <b>CORE</b> | <b>TEXT</b>                           | <b>SPEECH</b>              | <b>VISION</b>   |
|                      | Amazon Bedrock Foundation Models<br>Amazon Bedrock API |             | Amazon Translate<br>Amazon Comprehend | Amazon Polly<br>Transcribe | Amazon Textract<br>Rekognition<br>Amazon AWS Panorama |

## AMAZON SAGEMAKER

|   |                               |                                   |   |
|---|-------------------------------|-----------------------------------|---|
| <b>CANVAS</b><br>No-code ML for business analysts | <b>STUDIO LAB</b><br>Learn ML | <b>GROUND TRUTH</b><br>Label data | <b>JUMPSTART</b><br>Foundation Models Solutions |
|---|-------------------------------|-----------------------------------|---|

|                                |               |                      |   |  |                                 |  |  |
|--------------------------------|---------------|----------------------|---|--|---------------------------------|--|--|
| Prepare data<br>Store features | Geospatial ML | Build with notebooks | <b>STUDIO IDE</b>                               |  | Train models<br>Tune parameters | Deploy in production<br>Manage and monitor | <b>EDGE MANAGER</b><br>Manage edge devices |
|                                |               |                      | ----- CI/CD   GOVERNANCE   RESPONSIBLE ML ----- |  |                                 |  |  |

## ML FRAMEWORKS & INFRASTRUCTURE

|                                   |            |       |       |                |              |              |      |
|-----------------------------------|------------|-------|-------|----------------|--------------|--------------|------|
| PyTorch, Apache MXNet, TensorFlow | Amazon EC2 | CPU's | GPU's | AWS Inferentia | AWS Trainium | Habana Gaudi | FPGA |
|-----------------------------------|------------|-------|-------|----------------|--------------|--------------|------|

# Pacific Island Health Officers' Association (PIHOA) Customer Use Case



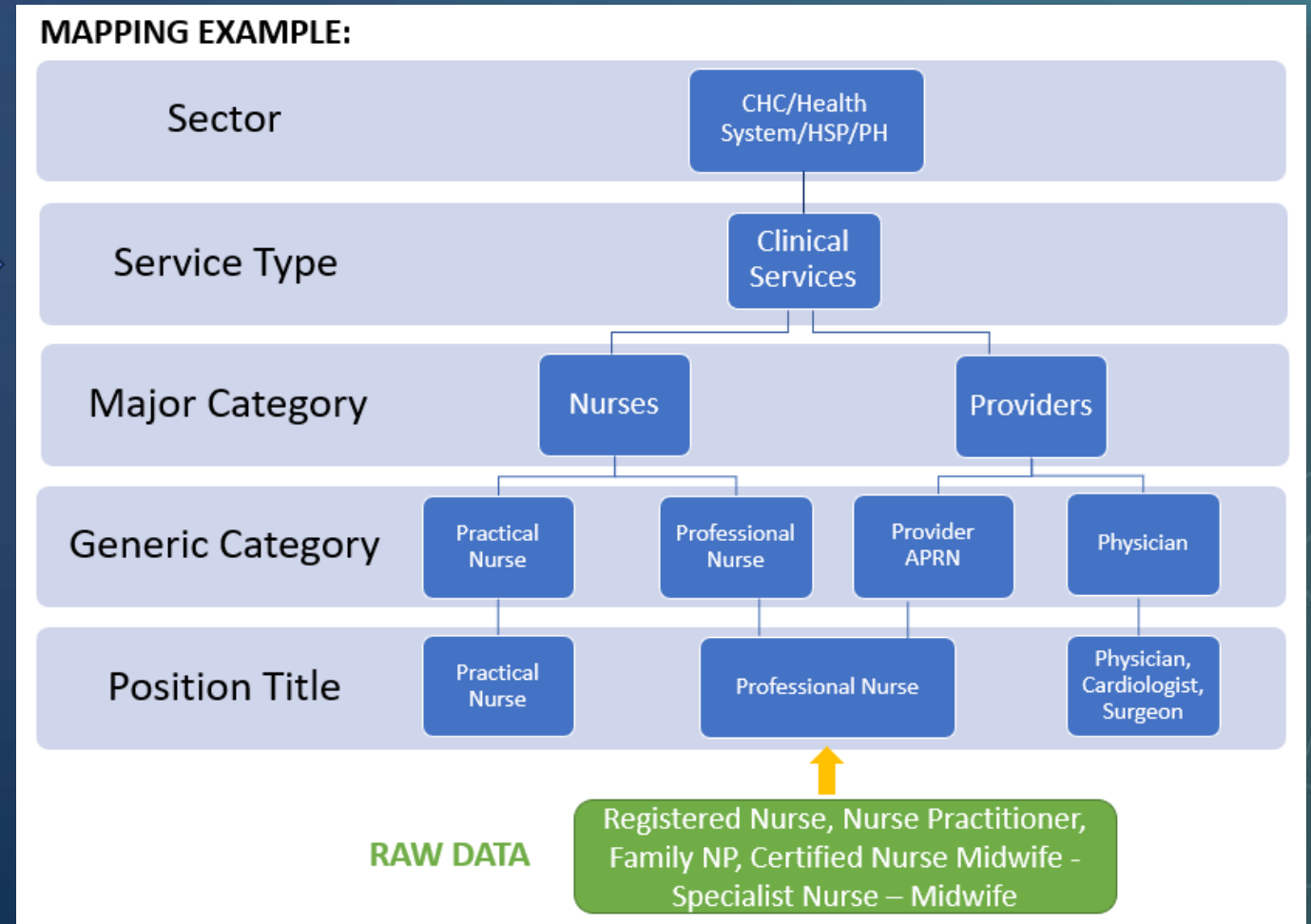
# Pacific Island Health Officers Association

- Non-profit representing the collective interest of the USAPI
- Mission is to improve the health and well-being of USAPI communities
- PIHOA Board priorities include the need to gain a better understanding of the USAPI healthcare workforce
- Human Resources for Health (HRH) Enumeration framework developed with stakeholder input
- Data collection to date: Palau, CNMI, RMI, Am. Samoa, and FSM: Kosrae, Pohnpei, & Chuuk.



# HCW Key Challenges

- Manual data collection & cleaning processes
- No standardization in job categories and job titles across jurisdictions
- Limited resources to help with data collection, mapping, and validation
- Guidance needed regarding tooling and metrics



# HCW Solution

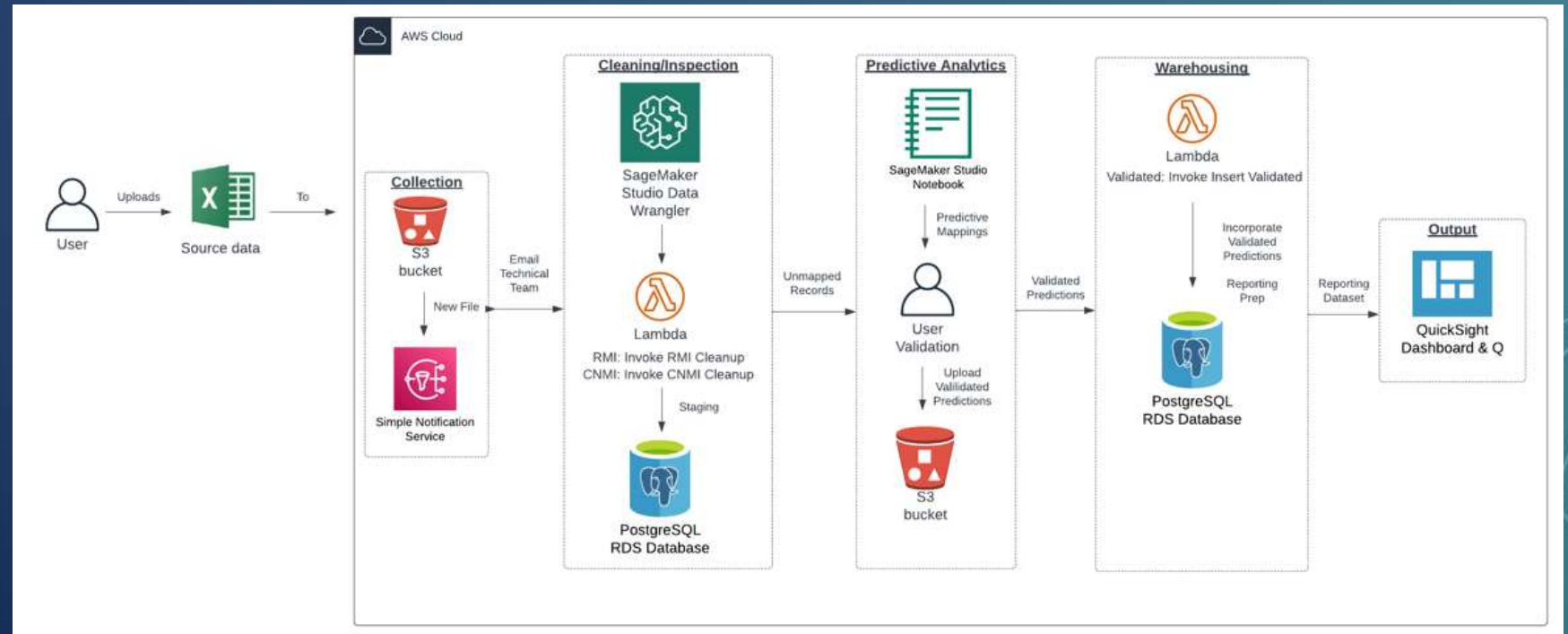
## Develop efficient data workflows:

- ✓ Automate data collection and cleaning (as much as possible)
- ✓ Programmatic mappings using AI/Machine Learning predictions
- ✓ Consolidate data seamlessly into an organized database structure
- ✓ Utilize intuitive reporting tool to turn data into information



# HCW Data Analytics Workflow Using AWS

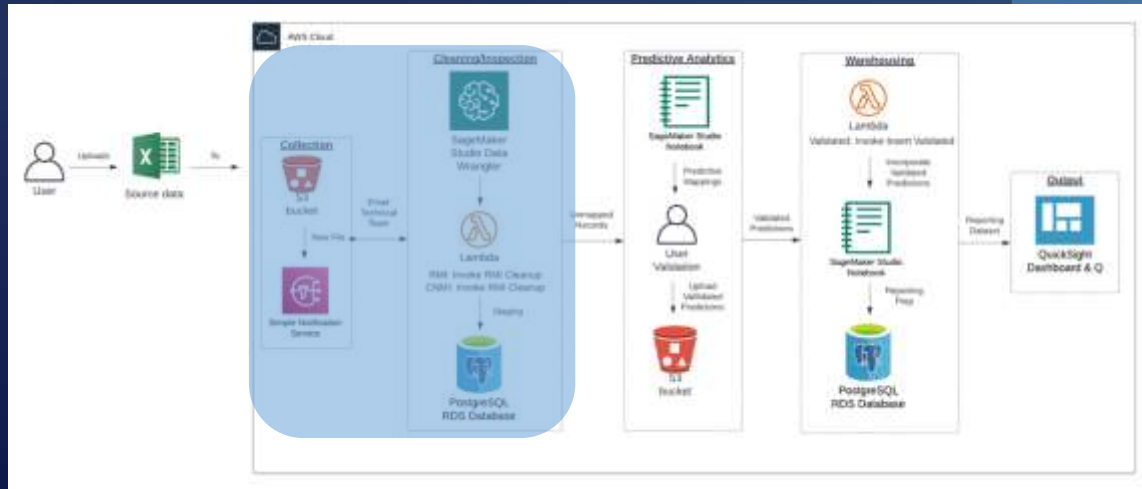
- S3
- Simple Notification Service (SNS)
- Lambda
- RDS
- SageMaker Data Wrangler & Notebooks
- QuickSight & Q





# HCW Data Analytics Workflow Using AWS

## Data Collection & Cleaning/Inspection



User uploads source data to S3 bucket

Technical team notified via SNS email

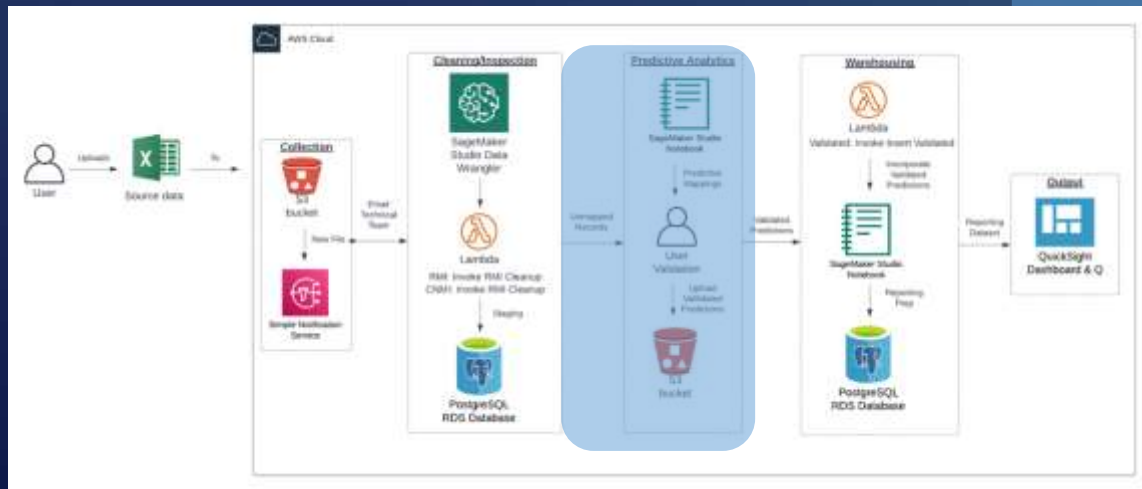
Data Wrangler used to inspect data

Clean-up lambda executed based on uploaded filename

Data loaded into RDS staging schema

# HCW Data Analytics Workflow Using AWS

## Predictive Analytics



Unmapped records filtered out to perform predictive analysis

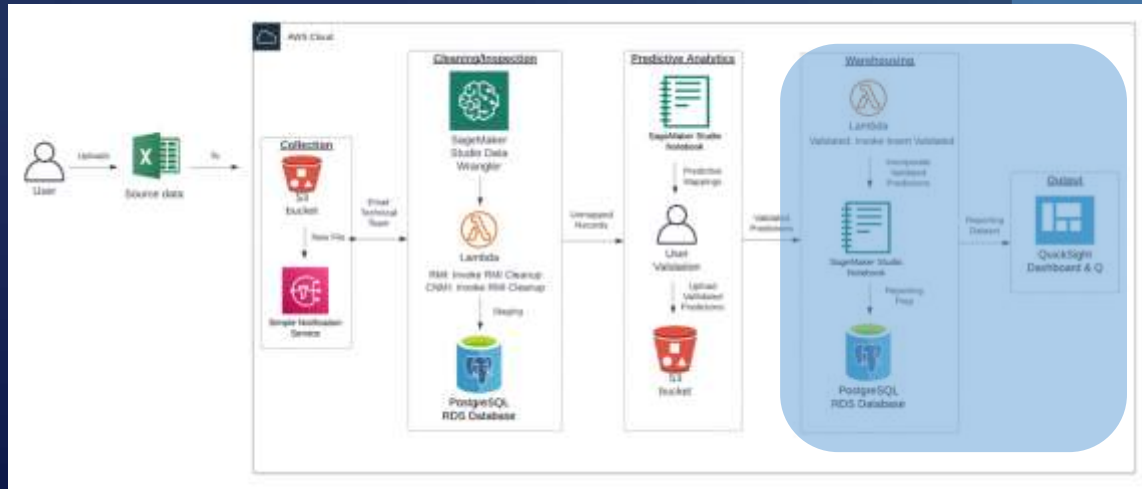
SageMaker notebooks used to produce Python scripts for ML

Predictions sent to user for validation

User uploads validated predictions to S3 bucket

# HCW Data Analytics Workflow Using AWS

## Warehousing & Output



Lambda executed based on validated file upload

Lambda used to prep data to create final dataset

Final reporting dataset written to RDS

Reporting dataset visualized by QuickSight & Q

# HCW Data Analytics Workflow Using AWS

## *Live Demonstration*



# HCW File Samples

| Employee Name | Social Security | Gender | Date of Birth | Ethnicity   | Marital Status | Date Hired | Seniority | Employee Group                | Organization Unit        | Location |
|---------------|-----------------|--------|---------------|-------------|----------------|------------|-----------|-------------------------------|--------------------------|----------|
| 1             |                 | Male   |               | MARSHALLESE | Married        | 06/02/2022 | 0.3 Year  | Majuro Regular Emp. Below 18k | MDHHS-Majuro Maintenance | Majuro   |

**Commonwealth Healthcare Corporation**  
**Active Personnel as of PP#9 (05/2022)**

Legend

- Hospital & Support Services
- Public Health & Support Services
- Non-Hospital/ Non-PH Support Services

| Check Route | Check Route Description      |
|-------------|------------------------------|
| 1867A       | General Public Health        |
| 1862        | CHC - Data Processing        |
| 1801        | Chief Executive Office       |
| 1863        | CHC -Business Office/Billing |
| 1863        | CHC -Business Office/Billing |
| 1863        | CHC -Business Office/Billing |
| 1863        | CHC -Business Office/Billing |
| 1863        | CHC -Business Office/Billing |
| 1863        | CHC -Business Office/Billing |
| 1863        | CHC -Business Office/Billing |
| 1801        | Chief Executive Office       |
| 1801        | Chief Executive Office       |
| 1801        | Chief Executive Office       |
| 1801        | Chief Executive Office       |

Active Personnel- PP#9

**Bureau Health Administration and Support Services**

| SSNo | EmployeeName | Degree                    | Department | Position |
|------|--------------|---------------------------|------------|----------|
|      |              | Two Year Degree / Diploma |            |          |
|      |              | Bachelors Degree          |            |          |
|      |              | College Credit            |            |          |
|      |              | Bachelors Degree          |            |          |
|      |              | College Credit            |            |          |
|      |              | College Credit            |            |          |
|      |              | College Credit            |            |          |
|      |              | Bachelors Degree          |            |          |

**Bio-Med**

| SSNo | EmployeeName | Degree                    |
|------|--------------|---------------------------|
|      |              | Two Year Degree / Diploma |
|      |              | High School               |
|      |              | High School               |
|      |              | A5 Degree                 |
|      |              | A5 Degree                 |

**JURISDICTION AND CHC (OR OTHER ENTITY) NAME : Kagman Community Health Center, Inc.**

**INDIVIDUAL DATA ELEMENTS (to be collected for each worker):**  
 (This info is to be collected from either HR or Department Head - depending on the site)

**Identifier and work HX info**

Name  
 Employee number  
 DOB  
 Gender: Male, Female or Other Female

**Current Position Info**

Division Medical  
 Program TICH  
 Position TITLE (currently in) Medical Assistant

Date of hire in the organization  
 Date started in current position  
 FTE status on a 0-1.0 scale 1  
 Permanent or Contractor Permanent

**Education and Credentialing Info**

Highest educ degree relevant to position High School  
 If prof lic required: is it current? N/A

K0046 K0025 K0014 K0024 K0003 K0056 K0028 K0042 K0005 K0001

Please submit your **feedback**. Thank you!



<https://www.pulse.aws/survey/NE4IZMCR>

- Pranav Khadilkar | Solutions Architect | AWS
- Email: pranavom@amazon.com
  
- Kelli Gonsalves | Data Scientist | DataHouse
- Email: kelli\_gonsalves@datahouse.com
  
- Jeff Masumoto | VP | DataHouse
- Email: jeff\_masumoto@datahouse.com

# Hawai'i Cloud Innovation Summit 2023

# Thank you!

