Building Real World Evidence on Cloud in Practice

Naoki Mashiko
Senior Solution Architect
Healthcare and Life Sciences (HCLS) Solutions
Amazon Web Services Japan K.K.

2019.10.24
Goal & Take Away

• The changing healthcare landscape requires healthcare industry companies to be data-driven organizations.

• Cloud has the enabling technologies and Quick Starts to become such organizations.

• What you want from RWE is not to have tools, but to bring out outcomes.

• Try it first anyway! In the cloud, this can be achieved with low risk.
Why Cloud?
What is the value of the cloud?
Most companies had electric generation capabilities on-site as differentiating factor

https://www.informationweek.com/software/information-management/the-cloud-electric-generator-analogy/d/d-id/1075830
The emergence of power plants and power grids is a **Paradigm Shift**
When you need it, you can use it anytime with a low price

Electric Power

Power Plant

Power Grid

Factory

IT

Data Center

Internet / Dedicate Network

Office
The same transformation as electricity is in the IT world!

- Initial investment
  - Surplus / shortage risk
  - Fixed Cost

- No initial investment required
  - Pay as you go
  - Variable costs
The essence of AWS = Building Blocks
You can combine and rapidly build services
Provide over 165 services
Customer of Healthcare & Life Sciences in Japan
Real World Evidence on AWS
Why is the industry investing in Real World Evidence?

**Sustainability**
Creating pressure on the healthcare system to produce better overall outcomes

**Reimbursement**
Payers utilizing new data sources to redefine value based payment models and formulary preference

**Healthcare Data is Exploding**
44 fold increase from 2009, growing to 35ZB¹ fueling new insights and raising bar for proving outcomes
Clinical trials activity is increasing across industry

Number of registered studies over time and some significant events (as of January 10, 2019)

- Number of registered studies
- Significant events

Source: https://ClinicalTrials.gov

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark
New requirements break the traditional approach

Customers need to:
- Capture and store new non-relational data at EB scale
- Secure and combine data from new and existing sources
- Do new types of analysis (ML, big data & real-time)

Traditional approach:
- DW is optimized for relational data at PB scale
- Data exists in silos, ETL does not scale at EB data volumes
- Operational and ad hoc on relational only
Traditionally, analytics look like this

- Relational data
- TBs–PBs scale
- Schema defined prior to data load
- Operational reporting and ad hoc
- Large initial CAPEX + $10K–$50K/TB/Year
Data Lake

- Collect
- Store
- Process/Analyze
- Consume
Building a Data Lake on AWS

COLLECT

COLLECT

STORE

STORE

CONSUME

CONSUME

PROCESS / ANALYZE

PROCESS / ANALYZE

Protect and Secure
Use entitlements to ensure data is secure and users' identities are verified

Protect and Secure
Use entitlements to ensure data is secure and users' identities are verified

Data Ingestion
Get your data into S3 quickly and securely

Data Ingestion
Get your data into S3 quickly and securely

Catalog & Search
Access and search metadata

Catalog & Search
Access and search metadata

Access & User Interface
Give your users easy and secure access

Access & User Interface
Give your users easy and secure access

Processing & Analytics
Use of predictive and prescriptive analytics to gain better understanding

Processing & Analytics
Use of predictive and prescriptive analytics to gain better understanding

Kinesis
Direct Connect
Snowball
Database Migration Service

Kinesis
Direct Connect
Snowball
Database Migration Service

IoT

IoT

Identity & Access Management
Security Token Service
CloudWatch
CloudTrail
Key Management Service

Identity & Access Management
Security Token Service
CloudWatch
CloudTrail
Key Management Service

Athena Query Service
QuickSight
EMR
Redshift
Lambda
Batch
Glue
SageMaker

$300/TB/Year

$300/TB/Year
Data Lakes extend the traditional approach

- Relational and non-relational data
- TBs–EBs scale
- Schema defined during analysis
- Diverse analytical engines to gain insights
- Designed for low cost storage and analytics
Functions that enable Data Exchange

- **Data Exchange**
  - **Internet**:
    - Transfer Object File
    - SFTP
    - MQTT
    - AWS Transfer for SFTP
    - AWS IoT
  - **Internet-VPN**:
    - Amazon VPC
    - Site-to-Site VPN
    - AWS Client VPN
    - NFS/iSCSI/SMB
    - Amazon FSx
    - AWS Storage Gateway
    - AWS Direct Connect
    - AWS Private Link
    - Amazon VPC Peering
    - AWS Transit Gateway
  - **Dedicate Network**:
    - AWS Direct Connect
    - NFS/iSCSI/SMB
    - Amazon FSx
    - AWS Storage Gateway
    - AWS Private Link
    - Amazon VPC Peering
    - AWS Transit Gateway
  - **Between AWS Services**:
    - AWS Private Link
    - Amazon VPC Peering
    - AWS Transit Gateway
  - **Hardware moving**:
    - AWS Snowball
  - **Storage**:
    - Amazon EC2 AMI
    - Amazon Redshift
    - Amazon RDS
  - **Via AWS Services**:
    - Share Snapshot
  - **Share Snapshot**:
    - Amazon EC2 AMI
    - Amazon Redshift
    - Amazon RDS

What you can do with other companies:

- Transfer File via HTTPS/CLI/SDK
- Transfer File via SFTP
- Bi-directional Communicate via MQTT
- Connect between office and office via VPN
- Connect between office and client via VPN
- Connect between on-premise via Dedicate Network
- Share storage between on-premise and AWS
- HTTP endpoint service with cross account
- L3 Communication with specific account
- Full-mesh L3 connectivity
- Moving data from on-premise to AWS
- Share your VM to other account
- Share you DWH to other account
- Share your Database to other account
Data Exchange Platform

Hospital A

AWS Account #A
- VPC
- EC2 (VM)
- De-identify

Clinic B

AWS Account #B
- VPC
- EC2 (VM)
- De-identify

Company C

AWS Account #C
- VPC
- Redshift (DWH)
- De-identify

Third Party AWS Account
- VPC
- Pre-process Platform
- De-identify

Common Data
- Amazon Simple Storage Service (S3)
- Processed Data

Use

Provide

Data Exchange Platform
Provide data (MDV’s online delivery of Real World Data (RWD))

MDV’s AWS env.

Gather RWD

MDV

Hospital 1
HP 2
HP 3

Doctors, Patients

DB

MDV’s AWS env.

(1) Delivery w/ S3

RWD Delivery (Fee-based)
Amazon S3

(2) Delivery w/ Redshift

Amazon Redshift

RWD Delivery through Redshift Snapshot (Fee-based)

BioPharma A’s AWS env.

Tableau on EC2

Amazon QuickSight

Researchers

Medical Representatives

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
Use Case: Takeda Pharmaceutical Company / Data Hub

- AWS is adopted as a data analysis infrastructure.
- Provides data transparency, data-driven decision making, analytical standardization, and automation.

[Links]
- https://www.youtube.com/watch?v=chDWhbAfFp8
uMotif drives research and increases data quality

“...We had 5,000 people register to join the clinical study in the first seven days, and we scaled the AWS infrastructure accordingly, no problem. In total, 13,500 people participated.”

Bruce Hellman,
Chief Executive Officer

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark
Customer Problems

1.2 Billion unstructured clinical documents created per year

Critical information “trapped” in these documents

Difficult to extract insights
Mr. Smith is a 63-year-old gentleman with coronary artery disease and hypertension. CURRENT MEDICATIONS: taking a dose of LIPITOR 20 mg once daily.

aws comprehend-medical detect-entities --region us-east-1 --text "<Insert Text Here>"
Extract and visualize clinical entities using Amazon Comprehend Medical

1) Upload Clinical notes to Amazon S3
2) Use Comprehend Medical API to extract various clinical entities
3) Extracted entities file is parsed and insert into Amazon DynamoDB table
4) DynamoDB has a stream attached to it. This stream is parsed using an AWS Lambda that is triggered by stream event
5) Lambda function inserts the records into Amazon Elasticsearch Service
6) Kibana dashboard visualize the clinical entities.

Extract and visualize clinical entities using Amazon Comprehend Medical

You can see the extracted entities from the notes. We get the attributes like Category, Type and also a confidence score.

You will see the dashboard with visualizations generated from the extracted entities.
Architecture: SNS Analytics using Comprehend Medical Demo

- AWS Cloud
- AWS Lambda
- Amazon Kinesis Data Firehose
- Amazon Simple Storage Service (S3)
- Amazon Athena
- Amazon QuickSight
- Crawl Twitter
- Translation ja into en
- detect_entities by Comprehend Medical
- User

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Demo - Comprehend Medical

- Twitter (in this demo, # anticancer drug side effects, #hay fever) is crawled regularly, and natural language processing specialized in the medical field is performed at Amazon Translate and Amazon Comprehend Medical.

- Analyze and visualize results with Amazon Athena, Amazon QuickSight.
Demo: Twitter analysis using Comprehend Medical
Create data science environments on AWS for health analysis using OHDSI

※ The Observational Health Data Sciences and Informatics (OHDSI) program and community are working toward this goal by producing data standards and open-source solutions to store and analyze observational health data.

1) CloudFormation build this architecture automatically in about 30min.
2) Load data as OMOP format.
3) Build DB for Atlas application.
4) Use ATLAS for visualization
5) Execute prediction using R-Studio and Jupyter notebook.

The following screenshot is just one example of the population health analysis that is possible with the OHDSI tools. This Atlas visualization shows the prevalence of various drugs within the given population of people. This information helps researchers and clinicians discover trends and make informed decisions about patient health.
Mobile data capture with AWS
Clinical trials transformation: mobile data capture (HIPAA eligible)

1. Collect data
Collect real-time, streaming data from medical devices and personal wearables

2. Store data
Store raw data on Amazon S3 for future analysis

3. Data processing—fast lane
Process and move actionable KPIs in Amazon DynamoDB (real-time, sliding window basis)

4. Data processing—batch
Extract, transform, and load (ETL) data using AWS Glue and move to Amazon S3. Load data in Amazon Redshift for detailed analysis

5. Visualize and act on data
Leverage Amazon QuickSight (or other BI tools) for data visualization. Provide real-time feedback via emails and text messages using Amazon SNS
It will take just 30 min!

https://aws.amazon.com/jp/quickstart/architecture/sas-viya/
Security / Compliance
 Considerations for Using AWS Products in GxP Systems

AWS製品をGxP関連システムにおいて使用する際の考慮事項

2018 年 1 月
Reduce risk.
Move fast. OR Stay secure.
Move fast. AND Stay secure.
Access a deep set of cloud security tools

**Networking**
- Virtual Private Cloud
- Isolated cloud resources
- Web Application Firewall
- Filter Malicious Web Traffic
- Shield
- DDoS protection
- Certificate Manager
- Provision, manage, and deploy SSL/TSL certificates

**Encryption**
- Key Management Service
- Manage creation and control of encryption keys
- CloudHSM
- Hardware-based key storage
- Server-Side Encryption
- Flexible data encryption options

**Identity & Management**
- IAM
  - Manage user access and encryption keys
- SAML Federation
  - SAML 2.0 support to allow on-prem identity integration
- Directory Service
  - Host and manage Microsoft Active Directory
- Organizations
  - Manage settings for multiple accounts

**Compliance**
- Service Catalog
  - Create and use standardized products
- Config
  - Track resource inventory and changes
- CloudTrail
  - Track user activity and API usage
- CloudWatch
  - Monitor resources and applications
- Inspector
  - Analyze application security
- Macie
  - Discover, Classify & Protect data
Goal & Take Away

• The changing healthcare landscape requires healthcare industry companies to be data-driven organizations.

• Cloud has the enabling technologies and Quick Starts to become such organizations.

• What you want from RWE is not to have tools, but to bring out outcomes.

• Try it first anyway! In the cloud, this can be achieved with low risk.
Thank you!