PKWARE

Data Privacy and Protection at Global Biotech

Introduction

This PKWARE customer is a biotechnology company based in California and is dedicated to scientific research and development for medicines to treat people with serious and life-threatening diseases. The company employs almost 15,000 employees and achieves nearly $2 billion in annual revenue.

The Challenge

1. Different customer projects were using a variety of data types. They needed a solution that offered flexible policy creation to address finding different kinds of sensitive data ranging from PHI to HIPAA data. They had 17 different classifiers. Every data set would need a new policy.

2. The customer was setting up a data lake into which they were going to ingest data from a variety of third-party data sets coming primarily in JSON and CSV files.

3. PKWARE needed to work with AWS as well as with a designated solution provider. The PS Consultant was stationed on site with the customer. The proof of concept (POC) was partially completed with AWS. The next step was to have the additional solution provider independently test PKWARE APIs and automation. The final steps would include fine-tuning the rules and creating a de-identified data lake with fully protected data to be consumed by the customer's analytics team.

About PKWARE and AWS

PKWARE offers the only data discovery and protection solution that locates and secures sensitive data to minimize organizational risks and costs, regardless of device or environment. Our ultra-efficient, scalable software is simple to use on a broad range of data types and repositories, enabling precise, automated visibility and control of personal data, even in the fastest-moving, most complex IT environments. With more than 1,200 customers, including many of the world's largest financial institutions, retailers, healthcare organizations, and government agencies, PKWARE continues to innovate as an award-winning global leader in data discovery, security, and compliance. To learn more, visit PKWARE.com

For Amazon Web Services (AWS) customers, it offers industry-first support for sensitive data stored in Amazon Simple Storage Service (Amazon S3) and accessed via Amazon Elastic MapReduce (Amazon EMR) for all your data analytics, compliance, and information governance purposes. Amazon Web Services provides a broad platform of secure, cost-effective, high performance cloud services that help you collect, store, and analyze Big Data workloads.

Key Requirements

1. Performance was key. They also had strict time restrictions for data ingestion.

2. A hands-off, automated solution was extremely important.

3. Accuracy in minimizing false positives in an unstructured format was also crucial.

The POC Process

Step 1:
a. Leveraged certain out-of-the-box elements from the PKWARE platform. Then used the PKWARE policy builder to build custom rules (elements) for their unique, specific elements: patient IDs and serial numbers.

b. Fine-tuned rules to capture the nuances of their environment, resulting in no false positives and no false negatives.

Step 2: Created an automated flow. Leveraged PKWARE APIs and AWS Lambda functions.

Step 3: Leveraged EMR at scale to achieve detection and masking for a specific window of time.

AWS Cloud

Data Profile   Antivirus

Source

Source

Source

Staging Area   Safe Zone

S3
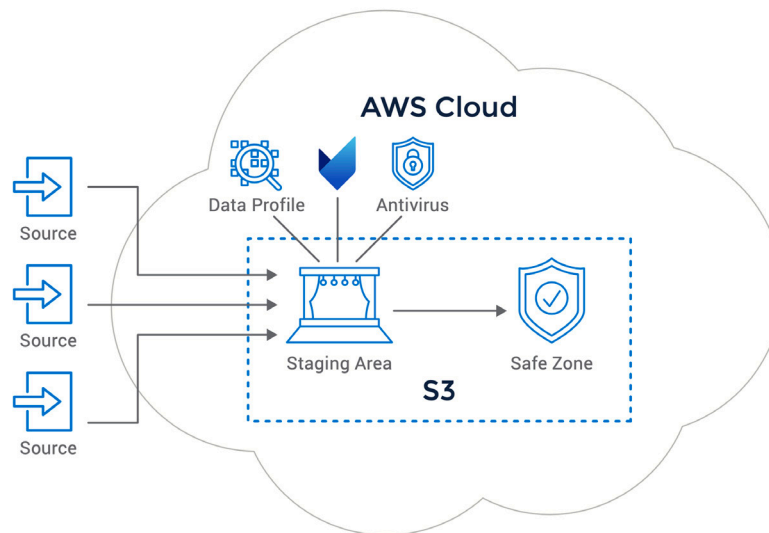
The Benefits

Fully Automated Solution
Processes that had been manual, slow, and inaccurate became rapid and accurate.

Full De-identification
Every element of sensitive data was identified and masked as required for compliance.

Safe Zone for Analytics
Fully protected data now flows into the customer's safe zone data lake. Analytics can deliver full value safely.

PKWARE