



PhUSE US Connect 2020

Automated Social Media Monitoring for Pharmacovigilance using Cloud Solutions



Bayer Inc.

Rohit Banga





Agenda

- // Introduction and Process Flow
- // Data Ingestion
- // Data Classification – Natural Language Processing
- // Data Engineering
- // Data Visualization
- // Looking into the Future



Introduction

Pharmacovigilance and Social Media Monitoring

Pharmacovigilance (PV) as the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem

Pharmacovigilance and Medical Device vigilance laws and regulations require Pharmaceutical companies to collect, analyze and report any suspected adverse event and/or quality issues that come to their knowledge about any products for human use

Social media is a promising source for new safety data and potential emergent safety signals.

Social Media data is closer to real-time occurrence of the event and it arises from direct user experience and can add to the information received from traditional post-marketing reporting methods



Exercise

We will live track tweets mentioning **PhUSEDrugA**

Please take out your mobile phones and start tweeting.

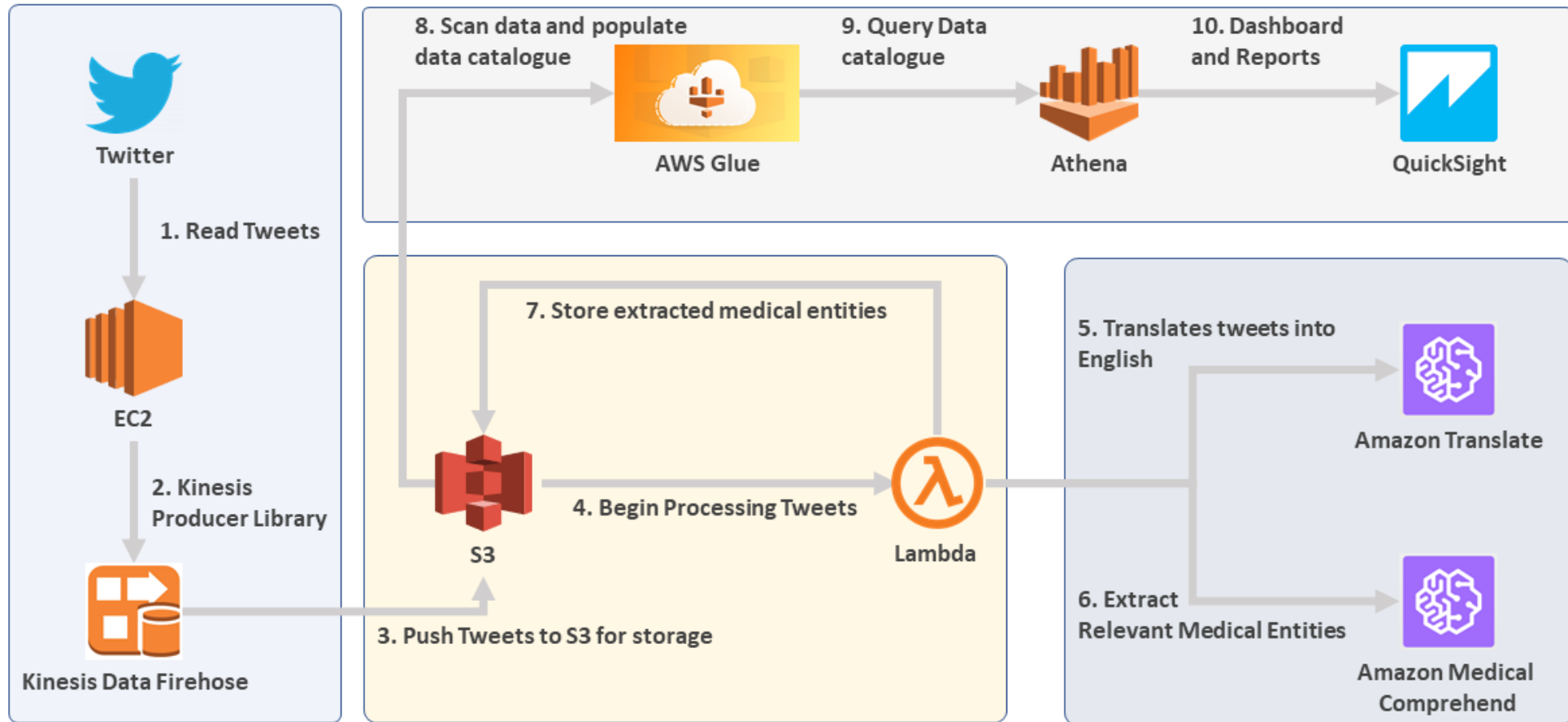
Example –

// I am a 45y old female. I got a headache after taking 25mg of PhUSEDrugA

// I have a stomach ache since last evening after taking PhUSEDrugA

Process Flow

Architecture Diagram





// Twitter will assess your use case and grant access to your app.

// Consumer API Key and API Secret Key

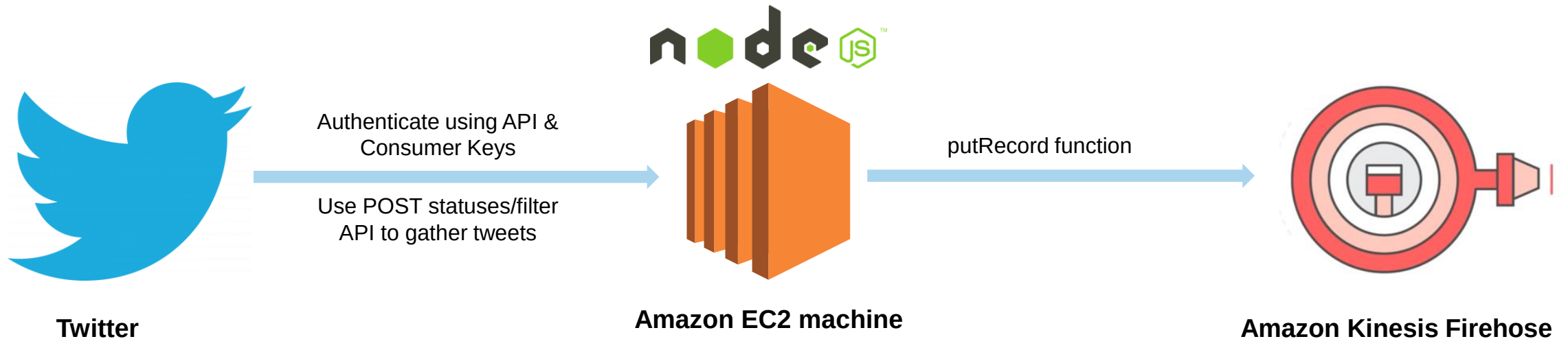
// Use POST statuses/filter API to filter realtime tweets**

[**https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter](https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter)

Data Ingestion – Ingest Tweets

Twitter Stream Producer Application

- // Create a twitter stream producer application
- // NodeJS application is deployed on an ubuntu virtual machine that is hosted on Amazon EC2
- // NodeJS app filters tweets matching “PhUSEDrugA” from twitter and pushes them into Kinesis Firehose





Data Ingestion – Store Tweets



Amazon Kinesis + S3

- // Kinesis firehose is amazon's service to prepare and load real-time data streams into data stores
- // Kinesis firehose streams tweets sent from the NodeJS app into Amazon S3 (Amazon Simple Storage Service) in near real-time for storage
- // S3 bucket will store the tweets as file object in JSON format



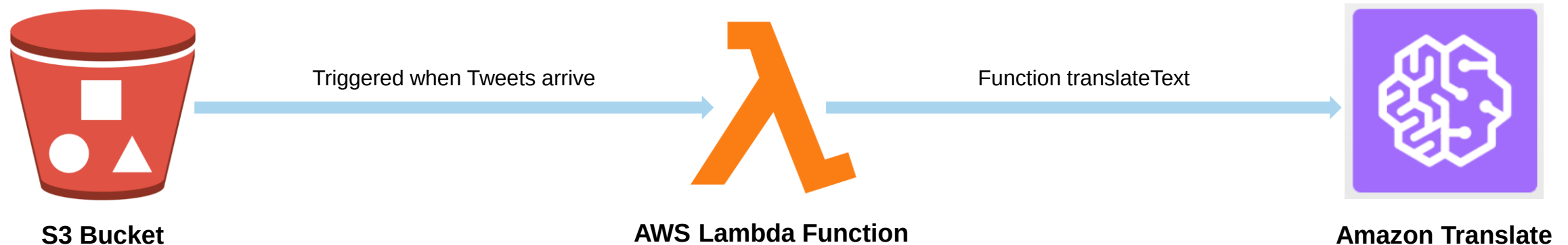


Data Classification – Translate tweets into english



Amazon Translate

- // Tweets from Spanish, German, French, Arabic and Portuguese language can be translated into English using Amazon Translate (a neural machine translation service)
- // Every file object stored in S3 triggers a Lambda function
- // AWS Lambda lets you run code (NodeJS in this example) without provisioning or managing servers
- // Lambda function reads tweets, uses the Translate API to translate them into english



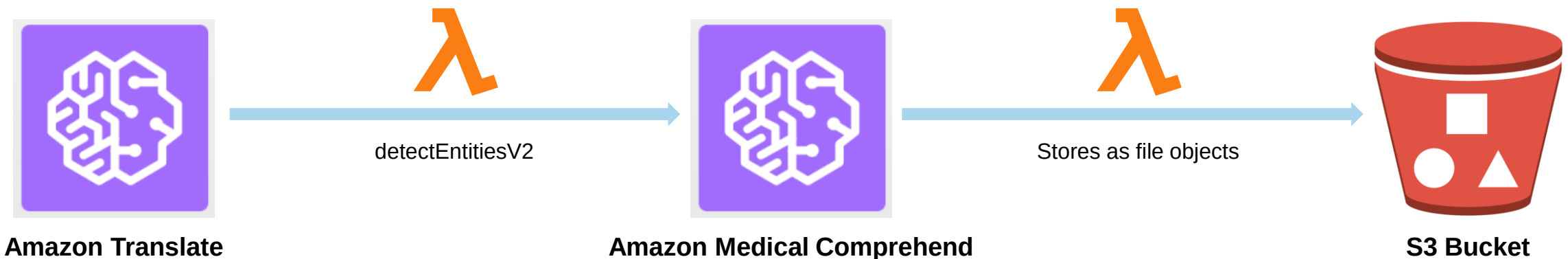


Data Classification – Natural Language Processing



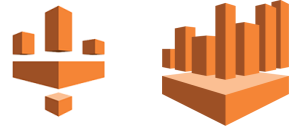
Amazon Medical Comprehend

- // Meaningful clinical information from unstructured tweet data can be extracted with the help of Amazon Medical Comprehend
- // Comprehend is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text.
- // Use custom classification models and plug them in Amazon Comprehend or use Amazon Medical Comprehend to extract clinically relevant information.
- // Lambda function passes translated text into Amazon Medical Comprehend. Clinically relevant data is extracted and stored back to S3 as file objects in JSON format.



Data Engineering

AWS Glue + Athena



- // AWS Glue can extract, transform and load data from S3 and build a data warehouse. It can automatically discover the data structures of tweets, translated text and clinically relevant entities in our S3 bucket.
- // AWS Glue can crawl S3 regularly and create/update tables in a Data Catalog.
- // Amazon Athena is used to query Amazon S3 data using the data catalog created by AWS Glue.



Data Visualization

Amazon QuickSight

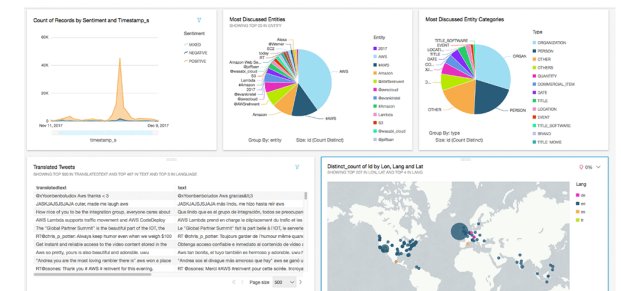
- // Amazon QuickSight is used to build interactive dashboards and reports that connects seamlessly with Athena tables
- // QuickSight has “Super-fast Parallel In-memory Calculation Engine” (SPICE), which features in-memory optimized calculation for data, and is designed for quick and up-to-date analysis.



AWS Athena



Amazon QuickSight



Dashboard & Reports



Exercise Results

**Live Social Media Monitoring
of PhUSEDrugA**



Looking into the Future

- // Data obtained from Social Media Monitoring is unstructured and obtained via uncontrolled and ungoverned processes in a non-regulated environment and is neither driven by data quality standards nor by specific business area orientation
- // However, social media feedback is too valuable to ignore.
- // The application essentially demonstrated real-time comprehension and analysis of unstructured data at scale.
 - // FDA uses Real-world data (RWD) and real-world evidence (RWE) to monitor post market safety and adverse events and to make regulatory decisions. **
 - // The health care community is using RWD to support coverage decisions and to develop guidelines and decision support tools for use in clinical practice **
 - // Medical product developers are using RWD and RWE to support clinical trial designs (e.g., large simple trials, pragmatic clinical trials) and observational studies to generate innovative, new treatment approaches.

[**https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence](https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence)



Questions?

Thank you!



Bye-Bye

