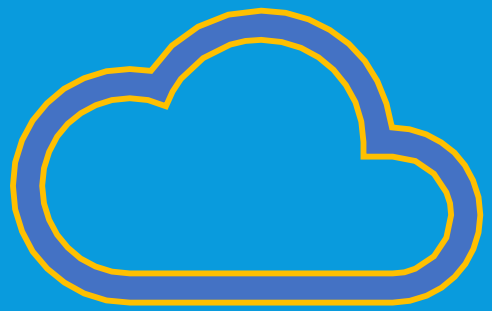


CLOUD PROJECT – TECHNOSPARTAN

PROJECT IDEA

The severity of COVID-19 is not acknowledged by a lot of people which leads to incorrect information and decisions; to be regretted later. During this holiday season many are meeting and travelling across the world ignoring the consequences.

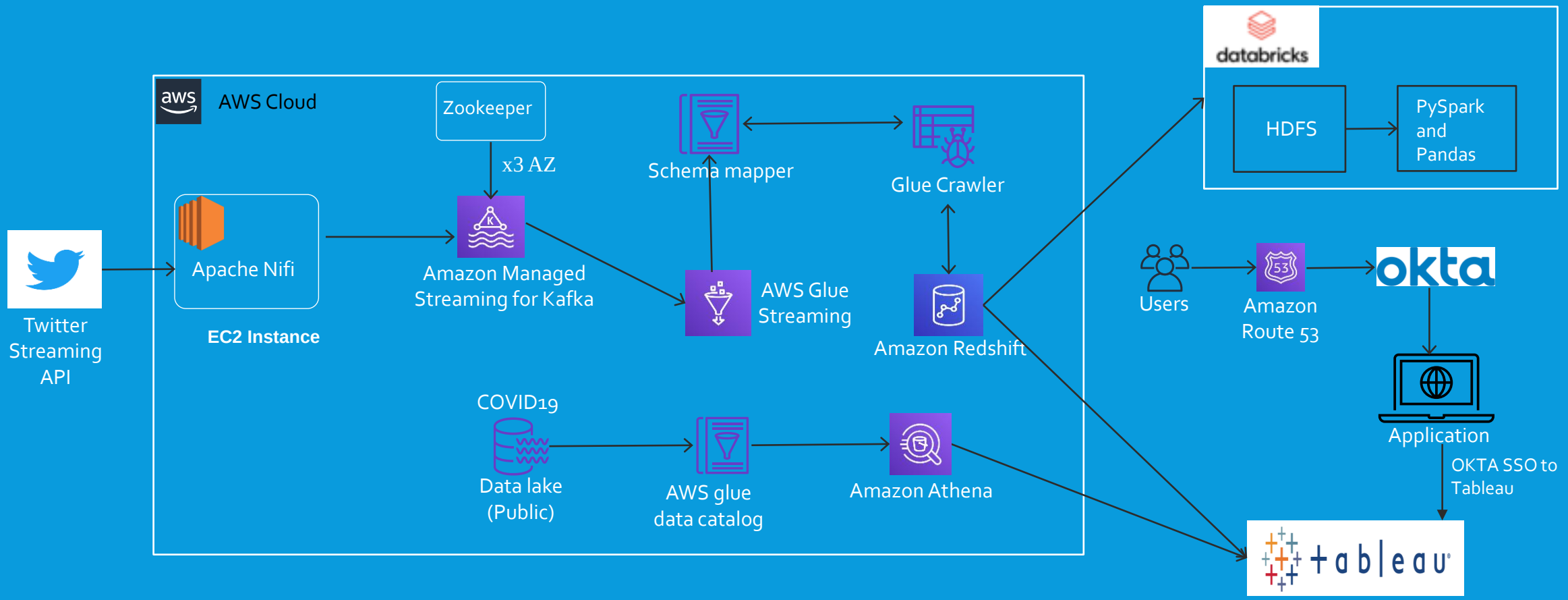
The data available today is humongous for any individual to manage and comprehend on their own.

Our project provides insights to people on the global rise of COVID-19 cases along with information on where people are meeting or planning to meetup. With our project, we want to help people understand that COVID-19 should be taken seriously. People should be aware of the locations where the cases are high, and for the same location, they can see whether people are meeting there.

Our web application provides following visualizations –

- Data extracted from Twitter filtered over keywords like meetup, meet
- Real-time COVID-19 new cases, new deaths
- Predicted Deaths

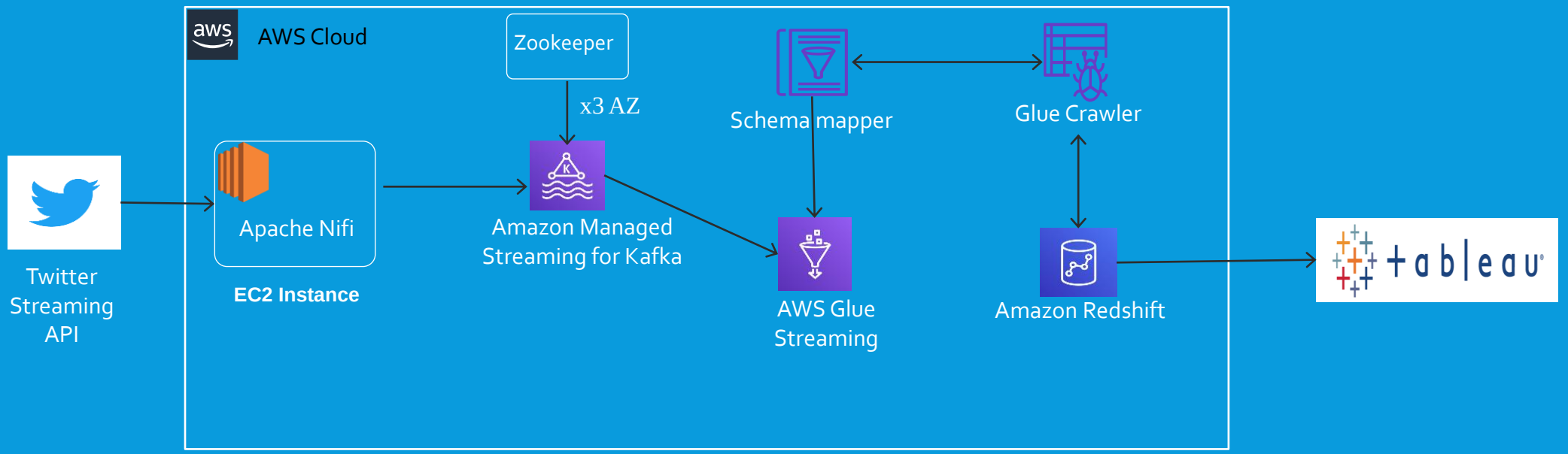
PROJECT ARCHITECTURE DIAGRAM



FEATURES

- Authentication and Authorization using OKTA.
- Single Sign-On to Tableau via OKTA.
- Real-time Data visualization on Tableau.
- Big data transformation, routing and visualization using Apache NIFI.
- Distributed coordination service using Zookeeper.
- Analytical data storage using Amazon Redshift
- Queuing mechanism using Amazon Managed Streaming for Apache Kafka.
- ETL jobs and data integration using AWS Glue
- Application has high availability and low latency.

COMPONENT 1 ARCHITECTURE DIAGRAM



Twitter Flow

COMPONENT 1 IMPLEMENTATION

- Twitter streaming API produces approximately ~3000 tweets per minute.

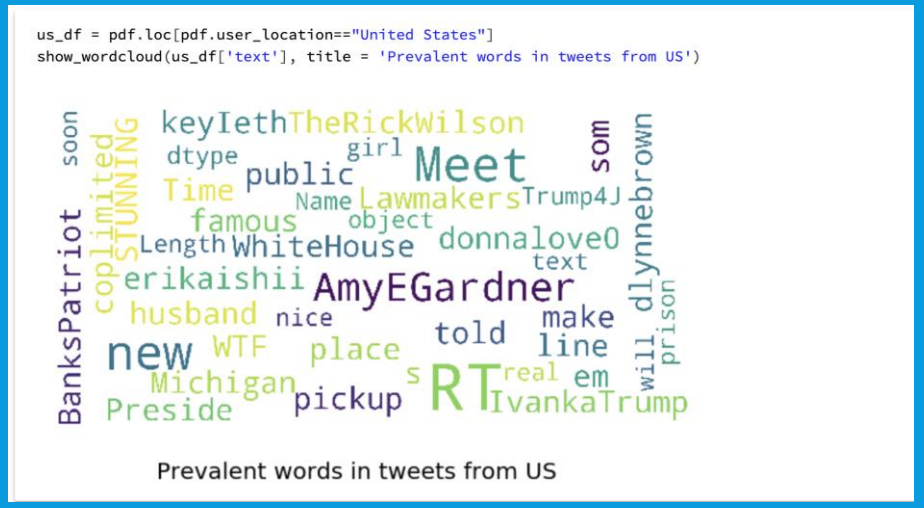
Back-end:

- ❖ To manage this streaming data we have created a Kafka topic in an EC2 instance.
- ❖ The big data workflow is viewed and managed using an Apache Nifi server running in the EC2 instance.
- ❖ The data extracted from the Kafka topic is transformed and loaded to Redshift using AWS Glue.

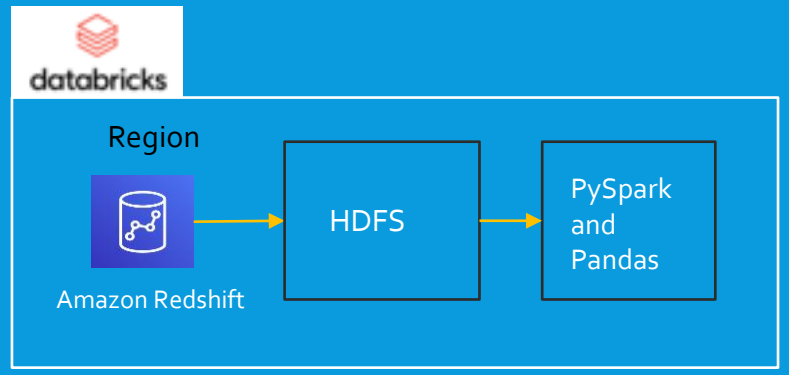
Visualization:

- ❖ Data extracted from Twitter filtered over keywords like meetup, meet
- ❖ Displayed using Tableau.

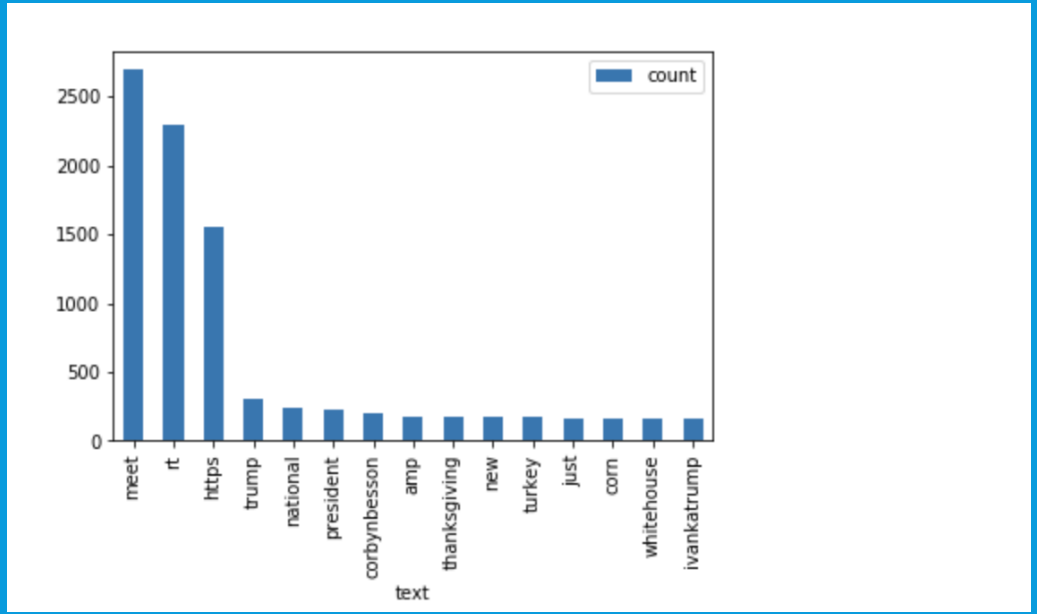
Twitter stream for one minute



HOW TO KNOW PEOPLE ARE MEETING OFFLINE/ONLINE?

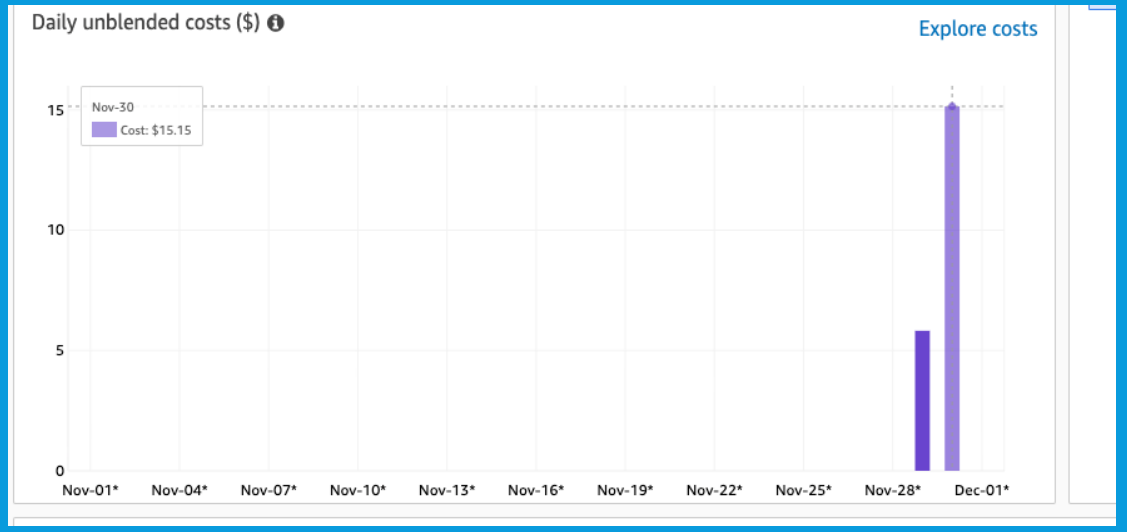


- More than 2500 people per minute are meeting offline.
- Around 1500 people per minute are meeting online



Why this Architecture ?

- AWS MSK – Helps to integrate Kafka as AWS service.
- Tableau – To help visualize and eliminate noise in the twitter data.
- Databricks – Community edition is free to use along with the clusters ☺
- Redshift – Large data warehouse with easy-to-use business intelligence tools.



Total cost $15 per day

Cost ?

- AWS EC2 - $ 0.384 per hour

☐	KafkaClientInstance	i-0d28d6ceb8464f76a	Running	m5.2xlarge

- Apache Nifi open source
- Databricks community edition
- AWS MSK - $0.0456 per hour

	Cluster name	Status	Creation time	Apache Kafka version	EC2 instance type	Brokers per Availability Zone	Availability Zones
○	MSKCluster	Active	November 29, 2020, 10:52:38 AM PST	2.2.1	kafka.t3.small	1	3

- AWS Glue – Free
- AWS Redshift – 2 months (free tier)
- Tableau Desktop – (Student 1-year free subscription)

COMPONENT 3


aws
AWS Cloud
COVID19
Data lake
(Public)
AWS glue
data catalog
Amazon Athena
tableau

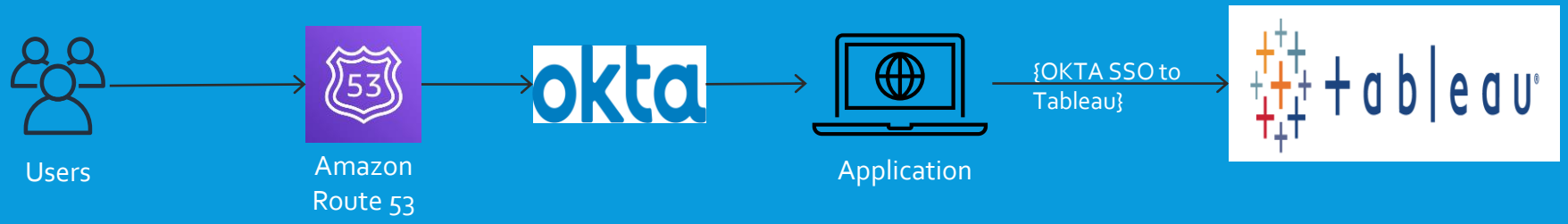
COMPONENT 3 IMPLEMENTATION

- Utilized publicly available COVID-19 data on AWS called "COVID-19 Data Lake".
- Multiple trusted sources are updating this S3 bucket like NY Times, John Hopkins etc.
- By using AWS Glue and Amazon Athena we are directly querying the S3 data lake for data analysis.
- Tableau provides connectivity to Amazon Athena which we have used to generate Map views from the COVID-19 data.

Data Visualization:

- Global view of new cases and new deaths
- COVID-19 positive cases over the time
- Predicted Deaths

COMPONENT 4


Users
Amazon
Route 53
okta
Application
{OKTA SSO to
Tableau}
tableau

COMPONENT 4 IMPLEMENTATION

- We chose React JS for our front-end application development.
- For Authentication and Authorization, we have used Okta.
- Okta also provides SSO for Tableau.
- For user information sharing between Okta and Tableau enabled SCIM feature.
- Using Okta with Tableau eliminates user's requirement to have Tableau license to view visualizations.

MONETIZATION AND FUTURE

- There isn't much social media integration with COVID-19 data.
- Use the API in messaging platforms and alert the users to the COVID-19 rate in the meetup locations.

Future

- Add extensions to personal and professional communication and social media applications such as Slack, Teams, Facebook messenger etc. to alert on the COVID-19 risks in a mentioned location based on analytical data.
- Use NLP to extract the real time tweet locations and intent from the social media text.
- With funding, we could use other AWS technologies and improve the efficiency of big data pipeline.

Thank You ☺