

ASU DECISION THEATRE CHALLENGE

Frequently Asked Questions (FAQ)

1. How long will it take to migrate on-premises to the cloud?

Depends if we are going to do a lift and shift or change services that may provide better performance, scalability, and cost management.

2. Do we have to move data in a specific order?

No, ideally though the services that have the most scale need, as defined below would be moved first.

Datapod – Getting external data sources. (high scale needs)

Message Broker – Queue messages from the data source.

Spark Cluster – Processing from the message broker. Does the ML and DA.

Data is moved ES – Saves the raw data and the analyzed data. (high scale needs)

Manager UI (GUI) – See the dashboard and see the data becoming knowledge. Custom application built for displaying UI/UX

3. How do we prioritize what data to move first?

The data that has the most need to scale as identified in “Question 2”.

4. How do we ensure migration doesn't impact current customer use?

Migrate the application to AWS and setup a new URL to point to the website. Conduct testing and when site is working as expected perform DNS cutover.

5. How will we know the eventual cost of our cloud usage vs. our onsite solution?

AWS Simple monthly calculator will be created once the services to be used are finalized. Once the solution is running we can use cost explorer to get the exact amount of spend.

6. How quickly will I start being able to access my data in the cloud post migration?

Right away.

7. How quickly can I set up AI/ML functionality in my new environment to start running my algorithms?

We can instantly get started as the trained models already exist. The following 3 models below are currently in use.

Model 1 – Sentiment classifier.

Model 2 – Spam classifier. (can remove based on the tone of the tweet)

Model 3 – Categorize tweets in different classes.

In phase one we will deploy the current models to the cloud and in phase two we can work on training new models in AWS sagemaker.

8. How do we automate these AI/ML algorithms to increase productivity?

We need the ability to retrain models based on trends changing in the marketplace. Retraining could be done via Sagemaker

9. What data can I input? In what format?

Twitter API. Writing python code currently to pull from other data sets.

10. What data limitations exist?

Hard limit: What services are we legally allowed to scrape.

Soft limit: Not having the python code created to scrape that source.

11. Can I pull in multiple sources from social media? (e.g. – Twitter, Facebook, Reddit, etc.)

See “question 11”.

12. How long will it take to get results from a query?

- Depends on the amount of the data
- Depends single instance vs multiple (serial vs parallel)
- Depends on the format of the data.

13. How quickly will the tool be able to scale?

Within minutes.

14. How many users can run simultaneous searches against our data?

Unlimited users can be supported with a solution designed for scale.

15. How can we get content closer to our end-users, to have better performance and scalability?

CloudFront, Multiple regions, and Route 53.

16. How does the tool handle confidential information?

Need to develop full RBAC controls to separate data between customers. So customer X does not see customer Y's model or find out who the other customer is.

17. What is the log-in/identify access management process to access data? How do we assign users rights?

AWS will be configured using best practices to enable logging at the AWS level. Need to enable application logging. Can use cloud services to help pull these logs and certificate them.

18. How will users get notifications of real time events?

Need to create a system to notify users via their preferred contact method.

19. What's the emergency plan to make PULSE resilient in the event of an outage?

We will use multiple AZ's to make it more resilient.

20. How many AZ's do we need to ensure we have enough redundancy?

At least two.

21. What legal, technical or policy issues limit access to information?

Need to create RBAC to comply with NDA. Need to create RBAC to fully segregate user information.

22. How do we ensure compliance with GDPR?

Talk to the ASU legal team.

23. How do we ensure a smooth user interface and connectivity between the front end application and backend infrastructure?

- Have the backend be scalability.
- Look at catching common requests.
- Need to look at current data sets for NoSQL or RDBMS recommendation.

24. Where will the demand be coming from?

Anywhere in the world.

25. Will I have mobile access to my data? From my Desktop?

Customers will be able to access the tool from any recently updated web browser. At a later time apps for smartphones can be developed.

26. How will admin users label XXX & also learn real-time?

Have a view where someone can look at data that has been identified as “not relevant” and be able to mark it as “relevant” to have it included in the calculations.

27. What makes PULSE better/different?

- Its ability to integrate multiple data sets.
- Highly configurable and customizable data sets and views in real time.

###