Nimbus Technology - UGRADS

Team:
Itreau Bigsby, Matthew Cocchi
Richard Deen, Benjamin George

Mentor:
Austin Sanders

Sponsor:
Daniel Boros

Problem Statement

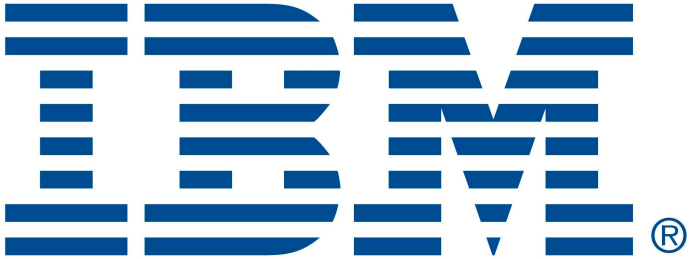
Cloud Storage Services

- Businesses have various forms of data that need storing, such as customer history, market performance, etc.
- Many businesses are moving to cloud data storage solutions, rather than company-owned servers.
- Most cloud services offer only cloud storage, not data management, which is cumbersome.

Problem Statement (cont'd)

IBM Spectrum Protect

- Businesses purchase storage through vendors such as AWS.
- Those businesses have storage needs ranging from less than a terabyte (a thousand gigabytes) to several petabytes (millions of gigabytes).
- IBM provides tools and services to its client businesses for managing their cloud storage.



Problem Statement (cont'd)

Costs of Cloud Storage

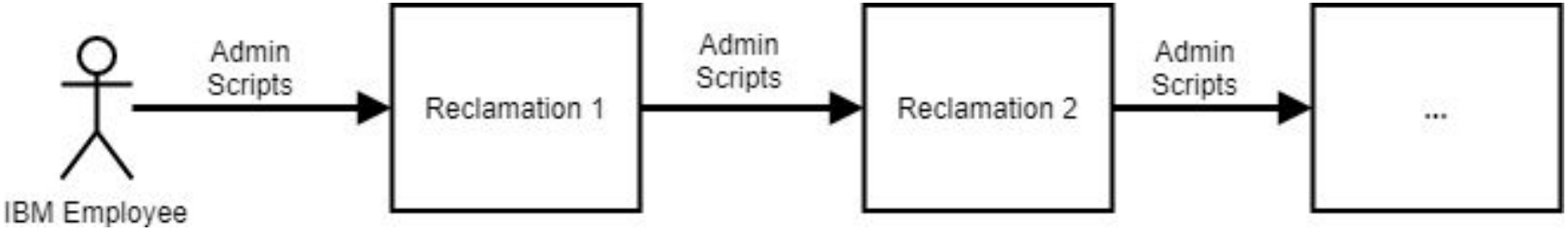
	Standard Storage
First 50 TB / month	$0.026 per GB
Next 450 TB / month	$0.025 per GB
Over 500 TB / month	$0.024 per GB

PUT, COPY, or POST Requests	$0.01 per 1,000 requests
GET and all other Requests	$0.01 per 10,000 requests

Reclamation

1. Identify Expired Chunks
2. Reclaim Space
3. Reformat Data

Problem Statement (cont'd)

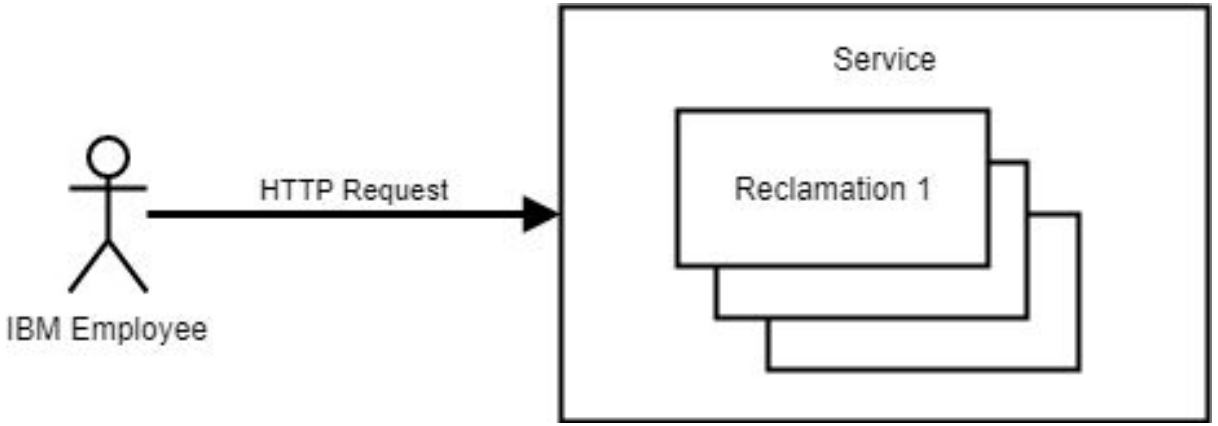


Problems

- Slow, cumbersome: New set of scripts for each container.
  - Millions, for the largest of IBM's clients ➔ difficult, if not entirely infeasible.

- Error prone: Scripts made by hand, potential errors at each step.

Solution Overview: Automation



Solutions

- Fast, easy: Send HTTP request for each container, let the service do the rest.
  - Millions of containers now feasible.

- Less error prone: Automate procedure of steps performs reclamations consistently.

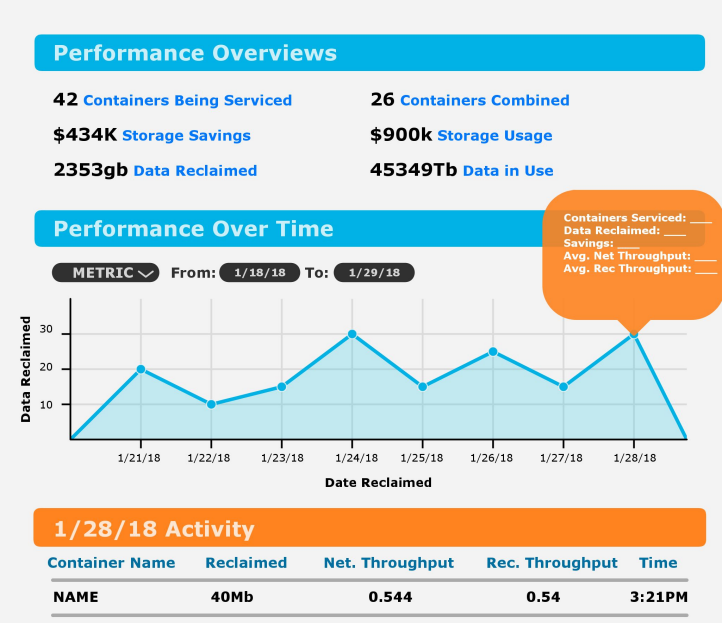
Solution Overview: Statistics Display

Statistics and Metrics

- Frontend web display that shows variety of useful metrics:
  - Data storage savings
  - Monetary Savings
  - Fragmentation (expired data) percentage

- Data displayed is based on all reclamations performed for an IBM customer's data.

- User can select to display data over a given range of dates.



Requirements and Specifications

Requirements Acquisition

- Weekly meetings held with our client, Dan Boros.
  - Occasionally joined by a frontend/UI developer, Jeff Placer.
- Review and refine specifications of desired software.

Key Requirements

- Reliability: Maintain IBM's customer data protection.
- Cost Effectiveness: Ensure monetary savings.
- Performance: Handle hundreds, possibly thousands of reclamations simultaneously.

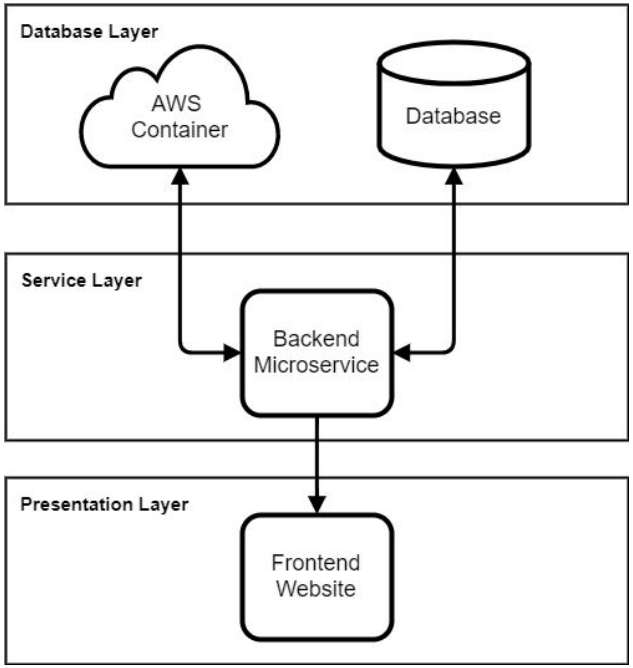
Implementation Overview

Layered Architecture

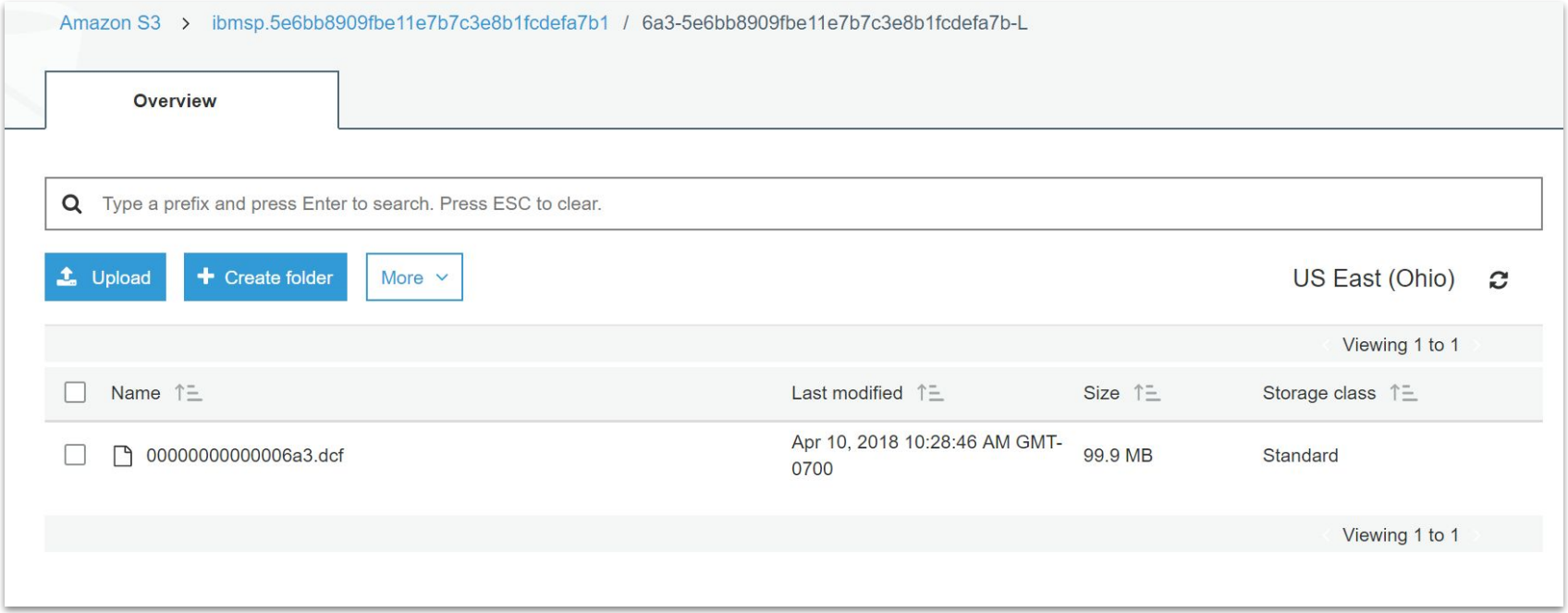
- Database Layer: Where data is stored.
- Service Layer: Where data is altered.
- Presentation Layer: Where data is shown.

Use Case

- IBM employee sends HTTP request.
- Backend fetches container file from AWS.
- Backend reclaims, reformats container.
- Backend records statistics in database.
- Backend uploads container to AWS.
- Frontend displays statistics.

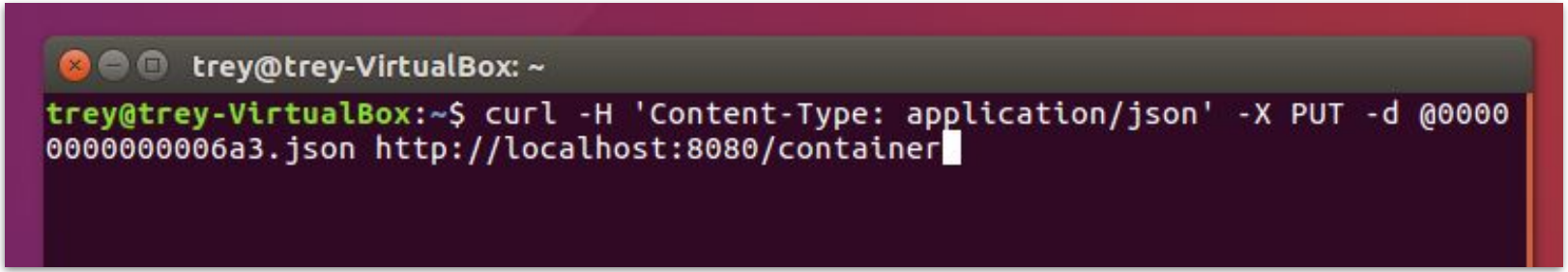


Demo: Backend



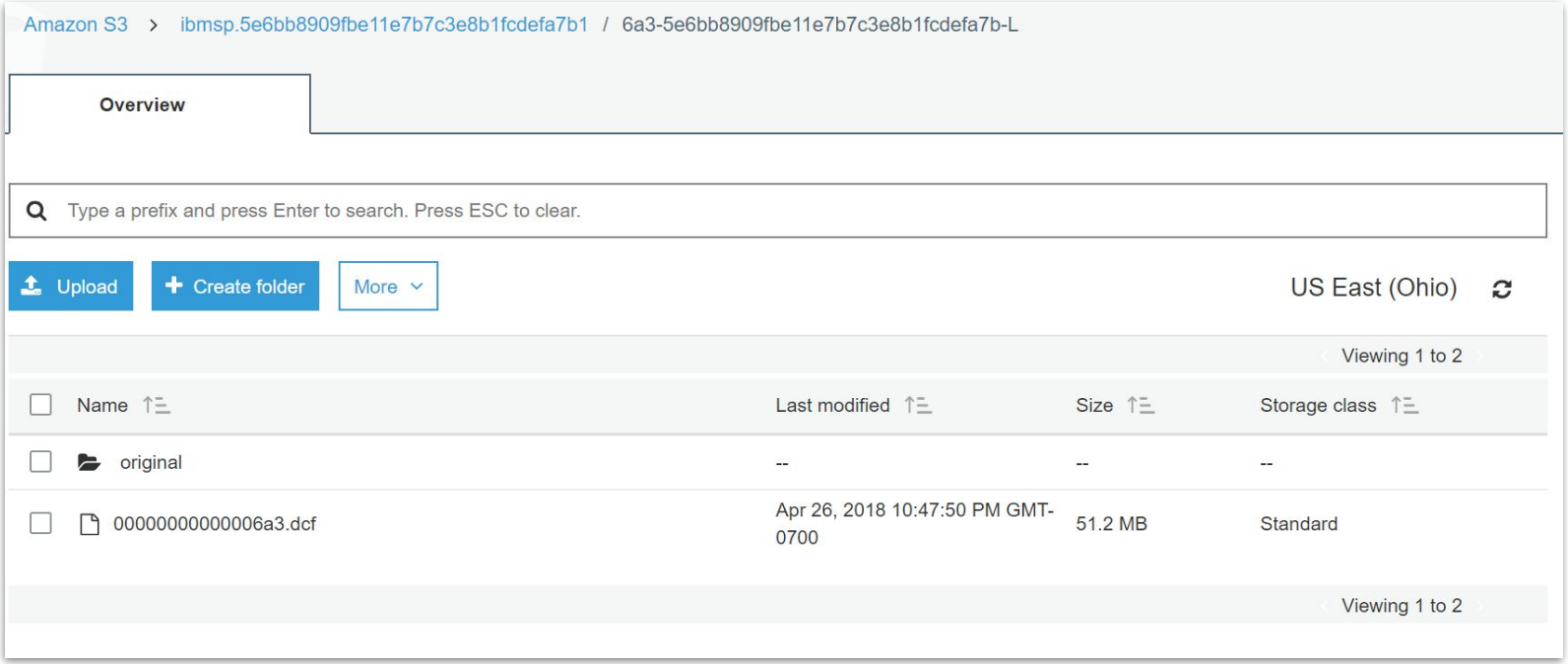
1. Identify Container

Demo: Backend (cont'd)



2. Send Layout

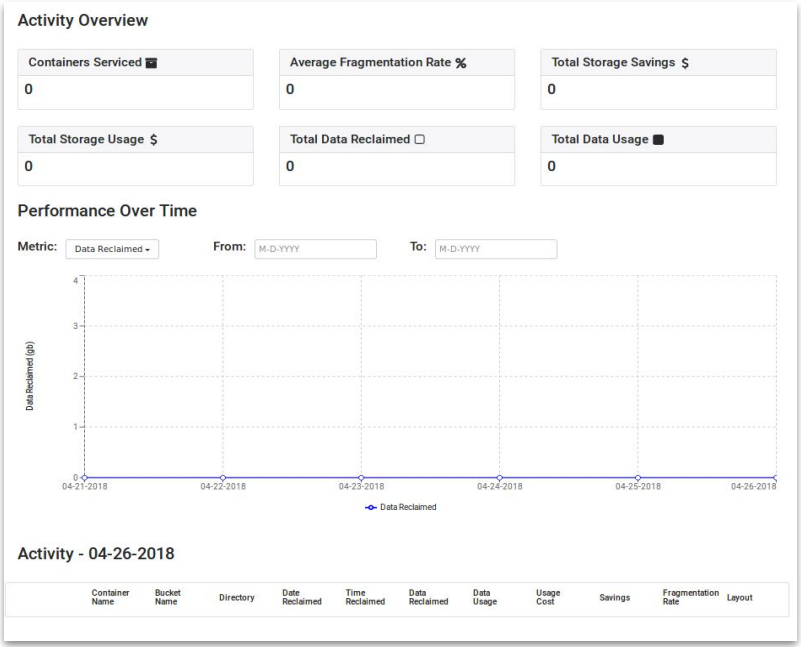
Demo: Backend (cont'd)



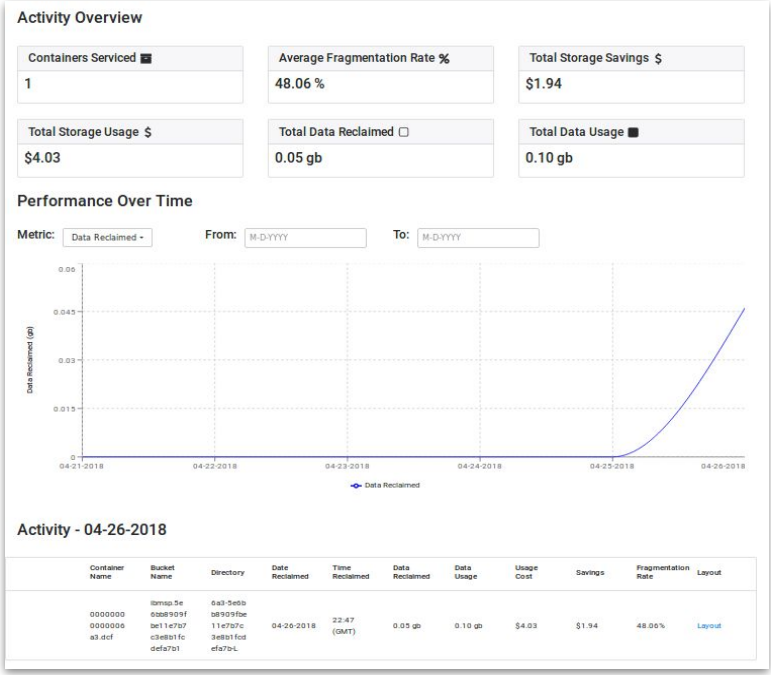
3. Reformat Container & Layout

Demo: Frontend

Before Reclamation



After Reclamation

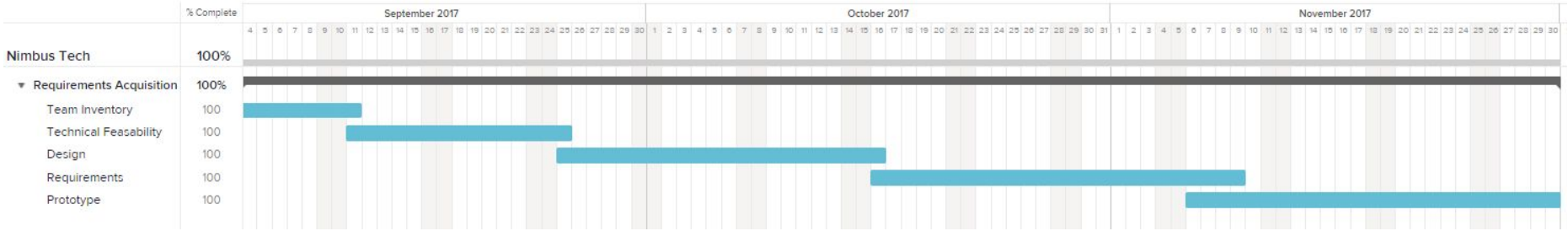


Challenges and Resolutions

Challenges	Resolutions
Multithreaded efficiency ● Memory Management ● Processing Time	Multithreaded efficiency ● Input/output streams to disk ● Queue for requests
Chart re-rendering frequency ● Too often: Unable to observe changes ● Too infrequent: Not getting useful metrics	Chart re-rendering frequency ● Use “activity” metrics to determine frequency of updates

Schedule: Requirements Acquisition

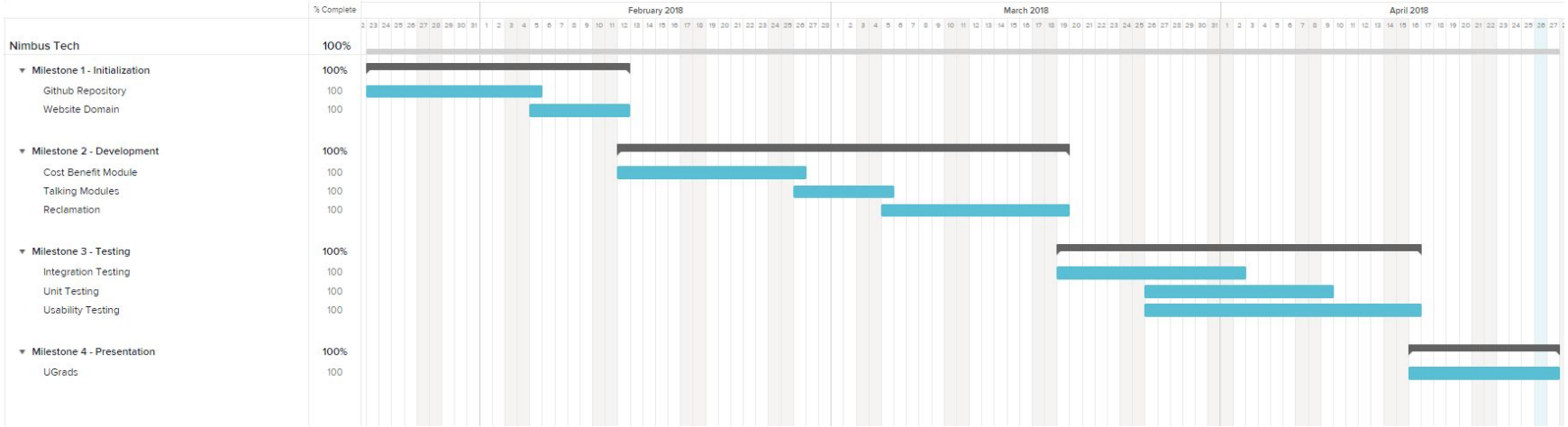
Nimbus Tech Schedule



Complete: Incomplete:

Schedule: Development and Testing

Nimbus Tech Schedule



Complete: Incomplete:

Software Testing

Unit Testing

- A dozen modules, many functions with wide range of inputs.
- Heavy unit testing: around 100 tests to verify all ranges of inputs for all functions.

Integration Testing

- Four major components, all need to work in tandem.
- Moderate integration testing, focus on backend module interactions.

Usability Testing

- Two phases of testing:
  - Categorical Acceptance: Match categories of displayed content to colors.
  - Live Usability: Gauge user's ability to intuitively navigate frontend.

Future Work

Expanding our Product

- Custom library of HTTP responses.
- SHA1 encryption checking to verify integrity of data.
- Batch reclamation via file with names of multiple containers.
- Automatic frontend re-rendering as reclamations occur.
- Ability for user to adjust scales of frontend charts.

Conclusion

- Cloud storage is costly, upwards of seven figures for the biggest consumers.
- Automated service to reclaim cloud data storage, saving businesses thousands.
- Worked closely with Dan Boros at IBM to acquire the specifications.
- Service is reliable, secure, and cost-effective.
- Our product eliminates vast man hours of work for IBM employees, making large-scale reclamations not just possible, but easy.