Project Report - CloudVault

Abstract

CloudVault is a cloud-based file storage software that allows users to store and access various types of files from any device. Hosted on AWS infrastructure, it offers a seamless and secure file management system. CloudVault incorporates encryption, secure authentication, and access control mechanisms to protect user data. With a focus on convenience, security and sustainability, CloudVault software offers a modern and efficient cloud storage solution for users.

1. Project Description

1.1 Project Overview

CloudVault is a web-based cloud storage platform designed to allow users to upload, organize and access files such as documents, images, and media content. The cloud storage is hosted on Amazon Web Services (AWS), leveraging technologies of AWS S3, DynamoDB and Cognito. The application is designed for scalability and high availability, the platform supports multi-web platform access and interactive UI.

The target audience of this application is who in need of an online storage service.

1.2 Objectives

- Create an application designed to allow for file transfer and organization.
- Design the UI to be modern in design and simple in use.
- Allow seamless file management through multiple platforms, i.e. Android, Apple, Windows etc.
- Use energy and cost-efficient services to provide file storage and access.
- Enable multiple file formats to be stored within the database.

1.3 Technologies

The following technologies are used to implement the CloudVault application.

Technology	Role	Source
Javascript	Implementation language for all application functionality.	https://www.javascript.com/
NodeJs	Backend server technology of application	https://nodejs.org/en

React	Frontend technology of application	http://legacy.reactjs.org/
AWS Cognito	User management technology, store application's user authentication information and authenticate methods	https://aws.amazon.com/cn/pm/cognito
AWS S3	Database technology, used to store application file data	https://aws.amazon.com/cn/pm/serv-s3
AWS DynamoDB	Database technology, used to store application file information metadata	https://aws.amazon.com/cn/dynamodb/

1.4 Application Usage

Both front-end and back-end applications can be run with the command `npm init` to install dependencies, then create a .env file that contains environment variables that are needed in this application, and finally run `npm start` to start projects.

When using the application, the user will be met with the login page, this will allow the user to enter their credentials so that they can access the database that stores the files specific to that account. Once the user is logged in, they will be presented with the main home page that shows the files currently stored in the database as well as other details regarding their files sure as file size and type.

(Please contact weny36@mcmaster.ca to request .env file environment information if you need)

2. Requirements

2.1 Product use cases

- Personal Storage:
    - Users can upload, store, and access personal documents, images, and media files. The cloud-based solution eliminates the risk of personal data loss due to device failure and provides seamless access across multiple devices.
- Business Collaboration:
    - Teams can share documents, collaborate on files, and manage access permissions. Businesses can streamline their workflow by integrating CloudVault with productivity tools and using shared storage for documents, reports, and media assets. Version control and audit logging ensure document integrity and traceability.
- Industrial microservice:
    - Workers can record the manufactory status by uploading the product information. CloudVault can be used as an industrial microservice to record information on

manufacturing production, such as manufacture date, material or any industrial information.

- Educational Use:
    - Students and educators can store the study material, assignments, and research papers. CloudVault provides an online shared folder for group projects, assignment submission portals, and integration with e-learning platforms. Educators can maintain version history and provide real-time feedback on student work.
- Media Storage:
    - Users can store high-quality images, videos and design files in cloud storage, and access media anywhere. CloudVault ensures seamless media playback and preview capabilities while maintaining high upload/download speeds for large files.

2.2 Functional requirements

R01	Users can register new accounts with multi-factor authentication(MFA).
R02	Secure file upload, storage, retrieval, and deletion.
R03	Real-time collaboration and version control for shared documents.
R04	File details are displayed on the home page.
R05	File sharing capabilities with customizable access control settings.
R06	File and folder organization features, including renaming, tagging, and sorting.

2.3 Data requirements

- Support for various file formats

Name	Type	Extension
Text Plain	text/plain	.txt
Text Markdown	text/markdown	.md
Text HTML	text/html	.html
JSON data	application/json	.json
PNG	image/png	.png

JPEG	image/jpeg	.jpg
WebP	image/webp	.webp
GIF	image/gif	.gif

- File data is stored in AWS S3 buckets, providing scalable and reliable object storage with high durability and availability.
- Metadata such as filenames, user IDs, upload timestamps, tags, and access permissions is stored in AWS DynamoDB, enabling fast lookups and real-time data access.
- Each user is assigned a unique folder identified by a globally unique ID (UUID), ensuring organized and isolated storage spaces within the cloud infrastructure.
- Backup and recovery mechanisms are implemented to protect data integrity in the event of accidental deletion or service failure.
- **The data limit of uploading is 5MB**

2.4 Performance requirements

	Related	Performance requirements
R07	Login speed	Users should be able to log in within 1 second after entering the correct credentials.
R08	Authentication Email Delivery	When a new account is registered, the system should send a verification email to the user's email within 2 seconds. This prompt response helps streamline the onboarding process and reduce user drop-off.
R09	Home Page Load Time	The average time to load the main dashboard/home page should be no more than 1.5 seconds. This ensures users can quickly access their file list and perform operations without delay.
R10	Document Scrolling	When scrolling up or down through a multi-page document, the system should render the next page in under 1 second to maintain a smooth user experience.
R11	File Detail Page Reliability	The probability of failure on demand (POFOD) when loading the file detail page (e.g., preview, metadata, sharing settings) should be less than 1%. This ensures consistent access to file-specific information.
R12	File Uploading Performance	The average time to upload a standard file (e.g., 5MB document or image) should not exceed 2 seconds under normal network conditions. This supports quick file submission and reduces wait time.

R13	File Accessing Time	The system should enable users to open and preview files within 1 second on average. Fast access is crucial for productivity and supports real-time collaboration scenarios.
R14	Low Latency Transfers	All file uploads and downloads should occur with minimal latency and maintain high throughput, especially when handling media files or large documents.
R15	Scalability	The system must support an increasing number of concurrent users and growing data volumes without significant degradation in performance.
R16	High Availability	CloudVault infrastructure should ensure a 99.9% uptime Service Level Agreement (SLA), leveraging AWS availability zones for redundancy.
R17	Caching and Optimization	Frequently accessed files and metadata should be cached at the edge to accelerate retrieval times.
R18	Auto-Scaling and Load Balancing	Backend services should automatically scale based on demand, with load balancers distributing traffic evenly to prevent bottlenecks.

2.5 Security requirements

	Related	Requirements
R19	User Security	All user accounts must be protected by Multi-Factor Authentication (MFA) using AWS Cognito
R20		Passwords must be securely hashed and never stored in plaintext
R21		Session tokens should have expiration times and be securely stored and transmitted via HTTPS
R22		User roles and permissions must be enforced to restrict unauthorized access
R23	Application Security	All data transmission between client and server must be encrypted using HTTPS

R24		The application backend must validate all incoming data to prevent injection attacks. For example, SQL and XSS
R25		Authentication tokens must be checked on every user action that involves file access or modification
R26		Logging and monitoring must be implemented for authentication events and access patterns to detect anomalous behaviour
R27	File Security	All files must be encrypted at rest using AWS S3's built-in encryption (e.g., AES-256)
R28		Access to files must be governed by fine-grained permissions stored in DynamoDB
R29		Shared file links must be tokenized and expire after a set period or on demand by the owner
R30		Version control should preserve historical file integrity and prevent tampering with audit trails

2.6 Maintenance and support requirements

	Related	Requirements
R31	Monitoring and Alerts	The system should implement real-time monitoring of backend services (via AWS CloudWatch) with alerts for performance degradation, errors, or service downtime
R32	User Support	A basic support system should be in place, including user documentation, FAQs, and an email-based helpdesk to handle user issues or feedback
R33	Backup and Recovery	Daily automated backups of file data (S3) and metadata (DynamoDB) must be maintained. Recovery procedures should ensure restoration within 24 hours of a failure
R34	Dependency Management	Regularly audit and update third-party libraries and frameworks (e.g. Node.js, React) to the latest stable versions to ensure compatibility and security

R35	Scalability Reviews	Conduct periodic system reviews to assess performance under increased load and adjust auto-scaling rules or infrastructure as needed
R36	Access Management Audits	Review user access controls and permission settings quarterly to ensure appropriate access and compliance with data protection policies

2.7 Operation and environment requirements

	Related	Requirements
R37	Hosting Environment	CloudVault operates entirely within the AWS cloud ecosystem, leveraging services like S3 (file storage), DynamoDB (metadata storage), and Cognito (user authentication)
R38	Deployment Configuration	The application is deployed via a Node.js backend and React frontend, both of which can be run using npm start after setting appropriate environment variables
R39	Port Configuration	The default communication occurs over HTTP port 8080, configurable through the backend settings
R40	Cross-Platform Access	The application is accessible via modern web browsers on Windows, macOS, iOS and Android platforms, ensuring broad user accessibility
R41	Browser Requirements	Supports latest versions of Chrome, Safari, Edge with JavaScript enabled
R42	Environment Variables	Environment configuration (e.g., API keys, AWS resource links) must be stored securely in a .env file and never hardcoded in source code
R43	Uptime and Reliability	CloudVault must maintain a minimum 99.9% uptime SLA by using AWS Availability Zones and failover strategies

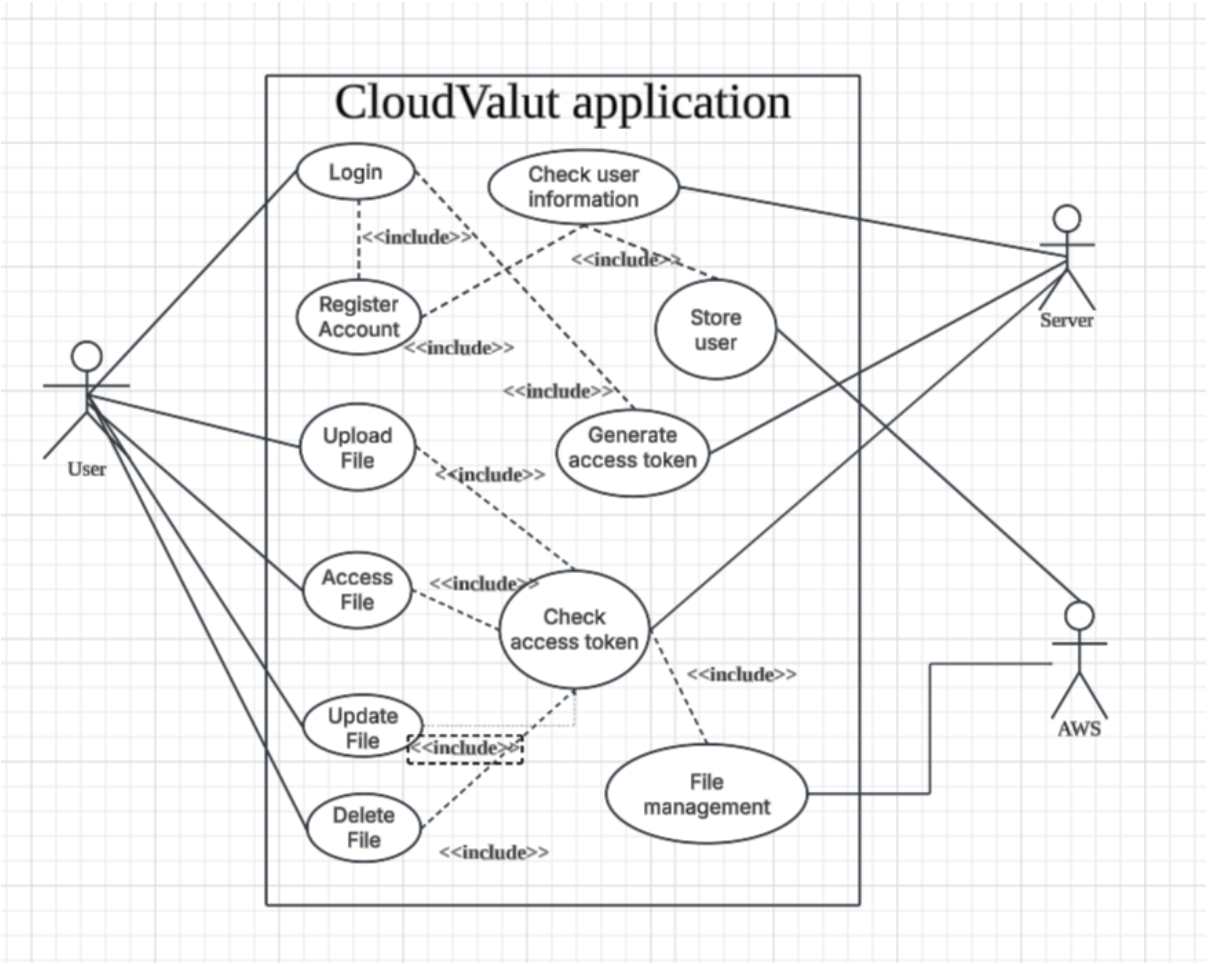
3. Design

3.1 Scenario View

Concerns: Understanding the central functionality of the system

Stakeholders: All stakeholders, but particularly the end user

Modelling techniques: UML Use Case Diagram



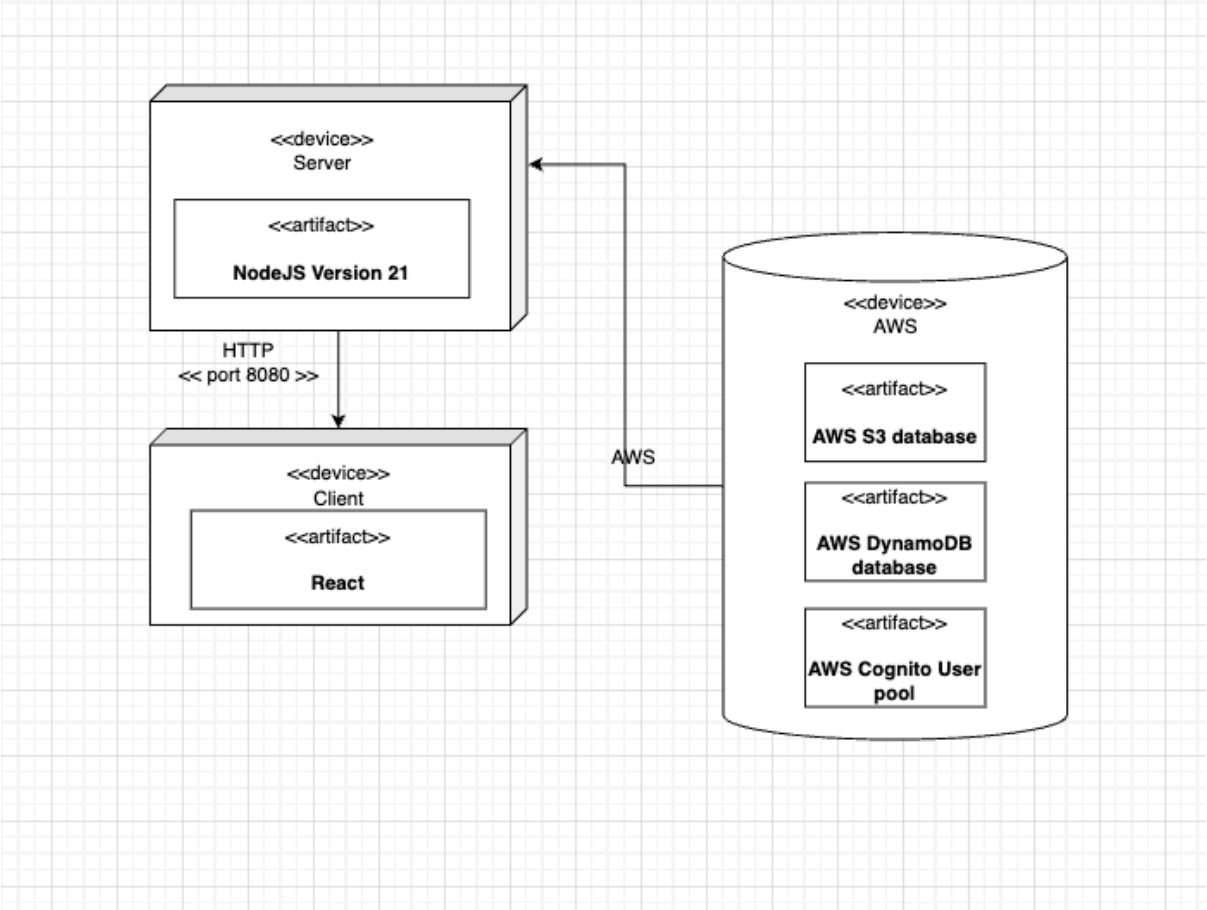
3.2 Physical View

Concerns: Mapping of software to hardware, communication protocols and modules related to Communication

Stakeholders: software architect, software developers

Modelling techniques: UML Deployment Diagram

The console application communicates with the server over HTTP port 8080, though this is configurable via the backend server.js file if the port needs to be changed in the future. The AWS S3 database and AWS DynamoDB is technically hosted by AWS.



3.3 Development View

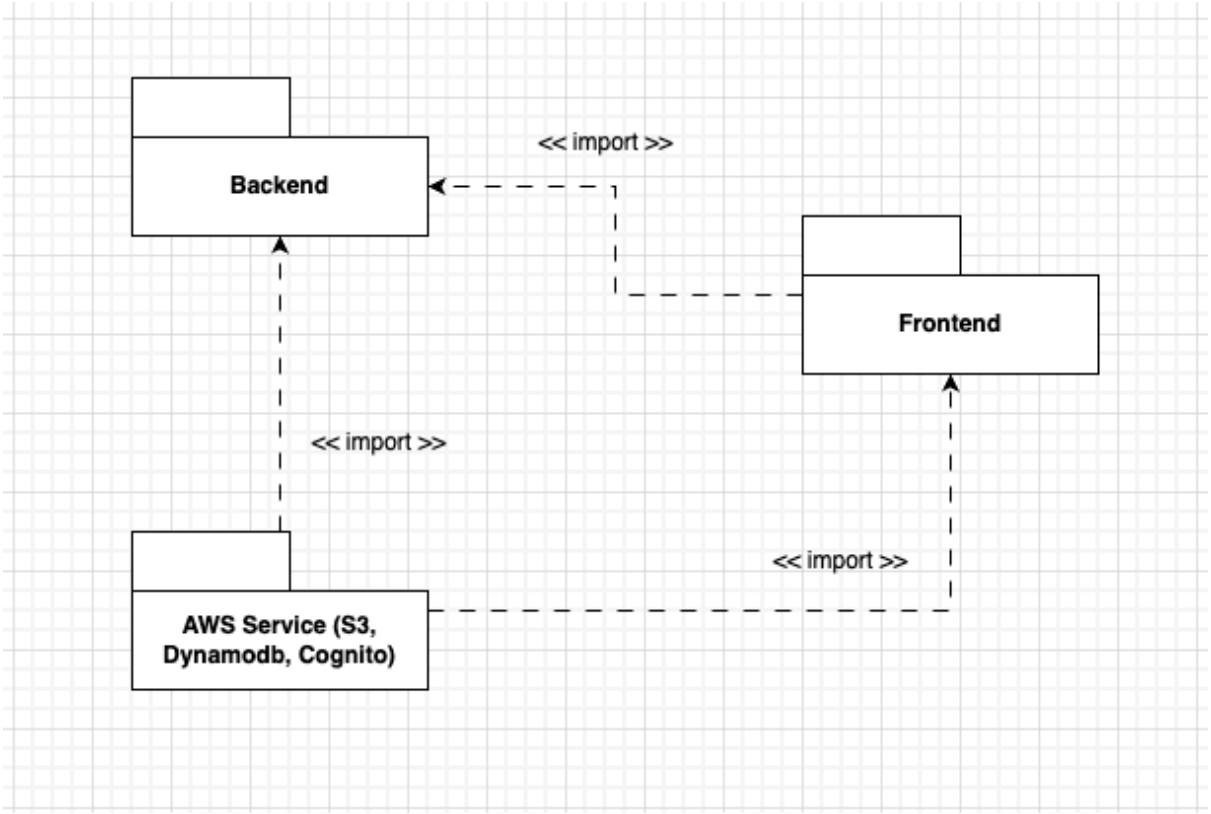
Concerns: Organization of software modules

Stakeholders: software developers, manager

Modelling techniques: UML Class Diagram

The UML package diagram below demonstrates the organization of project and code management. The backend module contains the logical operations, creating a user, user sign-in, checking access

token and file management operations. The frontend module contains the rendering of the webpage and user token access operations including render file list, file detail page, generate and user access token. The AWS S3 service is used to store file data, the Dynamodb service is used to store file metadata and Cognito is used to manage the user pool of applications.



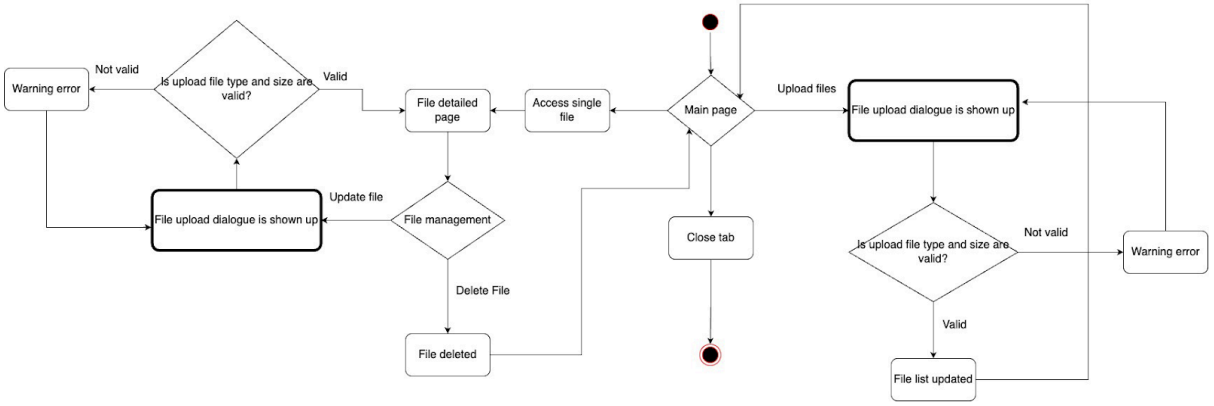
3.4 Logical View

Concerns: Functional requirements

Stakeholders: End user, software architect

Modelling technique: UML Activity Diagram

The use cases are realized via a series of actions that allow users to upload new files or select files to access/edit/delete. All of these options are implemented on the main page. The sequence of options is captured in the image below



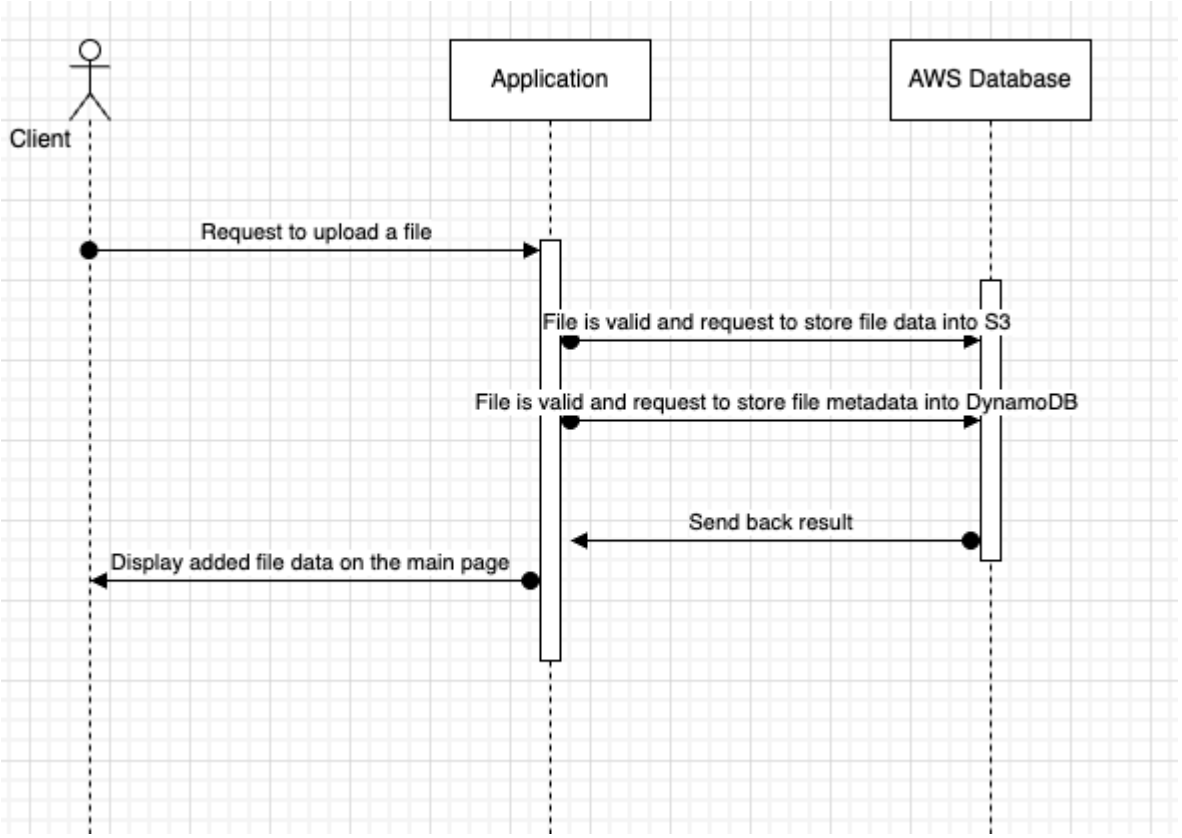
3.5 Process View

Concerns: Runtime communication

Stakeholders: software architect

Modelling technique: UML Sequence Diagram

The most important runtime communication that occurs are between the application and the AWS database. In particular, when the user uploads a file, the file is first checked by a file checking program to check the size and type of file, if it is a valid file, then its data will be uploaded into AWS S3 and the file information will be uploaded into AWS DynamoDB.



3.6 Database Schema

AWS S3 and DynamoDB are key-value databases which stores data as keys associated with values. One possible value is a hash, which is itself a set of keys and values. The database schema used in CloudVault is described in the table below.

Key	Value
fileId	The unique file ID generated by nanoid()
ownerId	The ownerId that generated by AWS Cognito
created	The file created date in format `yyyy-mm-dd hh:mm:ss`
updated	The file updated date in format `yyyy-mm-dd hh:mm:ss`
type	The file types (txt, md, html, json, png, jpg, webp. gif)
size	The size of the file, counted in Byte

4. Test cases

To ensure CloudVault meets its functional and non-functional requirements, we conducted the following test cases:

1. Verified the function of authentication mechanisms by testing valid and invalid login, and signup attempts.
2. Confirmed successful upload and preview of all supported file types (txt, jpg, pdf, JSON, etc). Tests included drag-and-drop uploads, mobile uploads, and multi-file batch uploads.
3. Tested multiple users with different accounts to validate file access restriction. Confirmed that the user could not access files they did not own.
4. Recovery and backup test
5. UI/UX testing, to ensure the application can be accessed on multiple browsers (Chrome, Firefox and Safari) and any devices (Windows, MacOS, Android, iOS) to ensure interface consistency and responsiveness.
6. Ran penetration tests to check for common vulnerabilities such as XSS, CSRF

5. Issues

During development and testing, we encountered several issues:

1. Loading larger images led to delays or UI lag. This is due to the network speed and file size issue, will try another way to minimize the size of the preview file.
2. UI elements such as buttons and modals are rendered differently across browsers. Resolved by using consistent styling libraries and vendor-specific fixes.
3. An incorrect redirect URI in the AWS Cognito configuration caused failed MFA setups. Fixed by correcting the callback URL and validating all auth flows.
4. Occasionally, uploaded files displayed incorrect metadata due to DynamoDB write conflicts. Solved by implementing retry logic and transaction-based writes.

6. Ethics and sustainability considerations

CloudVault is developed with a strong emphasis on ethical responsibility and environmental sustainability. As a cloud-based solution that handles sensitive user data and operates on internet infrastructure, we took careful measures to ensure responsible design and long-term social impact.

6.1 Data Privacy and Ethical Use

CloudVault ensures end-to-end encryption of user data both in transit and at rest, preventing unauthorized access. We implemented secure authentication, including multi-factor authentication

(MFA), to protect user accounts. The service hosted by AWS also ensures the security of data transmission.
CloudVault will not collect or access user data for any reason. The platform complies with major data protection regulations, including GDPR and CCPA. We maintain transparency in data handling by allowing users to view and control access to their stored content.

6.2 Environmental Sustainability

CloudVault is hosted entirely on AWS which is committed to achieving 100% renewable energy usage in its data center.

CloudVault also contributes to reducing paper consumption and physical storage needs by promoting paperless workflows.

6.3 Social Responsibility and Accessibility

CloudVault is designed to be accessible on all major platforms, including desktops, tablets, and smartphones. The interface is intuitive and easy to use, even for non-technical users, supporting digital inclusivity. We offer a freemium model, allowing users from all financial backgrounds to access core features for free.

6.4 Ethical Development Practices

All development was done following fair labour practices and team collaboration based on transparency and respect. Consumers will be notified in advance of what the application and database requires for proper installation and usage based on consumer requests. This is to ensure that any practices that are presented during development and installation of the consumers specifications are both made aware to both parties, ensuring that all actions are following protocol.

7. Commercialization

CloudVault targets both individual and business users who require a secure and efficient storage solution. Unlike conventional cloud storage platforms, CloudVault focuses on a sleek, interactive, and aesthetically pleasing design with high-end security measures. CloudVault will provide 5-10 GB of initial storage space, allowing the user to experience what CloudVault has to offer. Further benefits would be provided if the user decides to upgrade their plans to suit more of their needs. Further support, benefits and storage increase will be provided depending on the consumer's plan.

7.1 Business opportunity

With the rising demand for secure, easy-to-use cloud solutions among schools and small businesses, CloudVault can address a growing market that seeks affordable and simple online storage services. Most applications today involve the usage of cloud storage for their platform, but the solution that CloudVault provides is a more affordable and simplified application which can provide ease of use for the consumer.

7.2 Target Market

Personal use

- CloudVault provides a platform for individual users who seek a reliable and user-friendly platform for managing their personal digital data. Many users struggle with limited storage space on their local devices or expensive online storage solutions. By using CloudVault, they can securely store personal documents, photos and other media files online.

Education facility

- CloudVault supports educational institutions such as schools, colleges, and universities by providing a centralized platform for digital content management. Instructors can upload lecture notes, assignments, syllabi, and multimedia learning materials, while students can submit homework and collaborate on group projects.

Business environment

- In a corporate setting, CloudVault enhances productivity by offering a secure and organized file management system for internal teams. Employees can upload reports, presentations, and working documents while maintaining strict control over who can view or edit files.

Industrial environment

- In manufacturing or industrial settings, CloudVault can be used to store and manage large volumes of production data, such as equipment manuals, inspection checklists, quality control documents, or real-time reports from the production floor.

7.3 Sales Strategy

In the early sales stage, CloudVault can be introduced to schools, startups and creators through partnerships and group discounts to attract users.

A freemium model with 5GB-10GB free storage.

Basic Plan: $5/month (100GB, priority support).
Team Plan: $15/month/user (collaborative tools, audit logs).
Enterprise Plan: Custom pricing, API integration, dedicated support.

8. Conclusion

CloudVault aspires to be a revolutionary cloud storage solution that combines security, sustainability, and user experience. With a robust validation plan, clear commercialization strategy, and ethical business approach, CloudVault is poised to make cloud storage more accessible, efficient, and environmentally friendly. The application ensures a balance between convenience, security, and scalability, catering to both personal and professional users seeking a modern cloud storage platform.