Isabelle Gruijs, Lukas Hartansky, Martin Micko, Tushar Shingrani

Project walkthrough

Project S3

Scrum 5

Contents

[1. Project description 1](#_Toc178797083)

[1.1 Stakeholders & Client 2](#_Toc178797084)

[1.2 Client Requirements 3](#_Toc178797085)

[1.3 Product Backlog 3](#_Toc178797086)

[1.4 Functional Requirements 3](#_Toc178797087)

[2. Project Goal 5](#_Toc178797088)

[3. User experience 5](#_Toc178797089)

[3.1 Use cases & User stories 5](#_Toc178797090)

**Document History**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Date | Author | Group | Summary |
| 0.1 | 25-9-2024 | Isabelle Gruijs | SCRUM 5 | Started agile working |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## Project description

**Dropbox-like File Storage and Sharing Service**

This project aims to build a scalable, cloud-based file storage and sharing solution like Dropbox, with hybrid cloud integration. Users can upload, manage, and share files securely, with some data stored on-premises and the rest in the cloud.

**Key Components and Challenges:**

1. **File Storage & Management**:  
   Store user files in Amazon S3, with options for encrypted storage (using S3’s server-side encryption or client-side encryption). Users can share files using pre-signed URLs or role-based access control (RBAC) to restrict access to specific users or groups. On-premises file storage can be used for highly sensitive data with secure synchronization to AWS using AWS Storage Gateway.
2. **Data Synchronization**:  
   Implement a system where files can be seamlessly synced between on-premises storage and AWS S3. AWS DataSync or Storage Gateway can be used for efficient data movement, ensuring minimal latency for critical files stored on-premises.
3. **Versioning and Backup**:  
   Use S3’s built-in versioning feature to track file changes and enable recovery of previous versions. Set up automatic backups using AWS Backup for cloud-stored data and a similar solution for on-premises storage.
4. **Hybrid Cloud Implementation**:  
   Provide local file access using an on-premises file system like NFS or SMB, integrated with AWS Storage Gateway, allowing files to be cached locally but stored securely in the cloud for global access and disaster recovery.
5. **User Authentication & Permissions**:  
   Utilize AWS IAM for managing user permissions in the cloud. On-premises Active Directory or LDAP servers can be integrated for user authentication, with AWS Managed Microsoft AD as a potential bridge between on-premises and cloud-based user authentication. Make sure your solution doesn't imply keeping access keys on the users’ hard drives
6. **File Sharing and Collaboration**:  
   Build a sharing module that allows users to invite others to collaborate on files. Use pre-signed URLs with expiration for time-bound file sharing or OAuth2 for secure, long-term collaboration access.
7. **Security and Compliance**:  
   Implement stringent security measures such as file encryption, multi-factor authentication (MFA), and logging using AWS CloudTrail. On-premises systems would adhere to local compliance needs, while cloud infrastructure could follow standards like HIPAA, GDPR, or SOC 2.
8. **Scalability and Cost Optimization**:  
   Use S3 storage classes (Standard, Intelligent-Tiering, or Glacier) to optimize costs depending on file access patterns. Enable lifecycle policies to archive infrequently accessed files to cheaper storage automatically. On-premises storage can be used for high-frequency, low-latency access.
9. **API for External Integrations**:  
   Develop an API that allows other apps or services to upload, manage, or retrieve files programmatically, making the service extensible for various business or academic use cases. AWS API Gateway and Lambda can be used for managing API requests efficiently.

### Stakeholders & Work method

This project includes stakeholders and a direct client, Andrius Kuprys.

We are working with the method of SCRUM. Scrum is a lightweight framework that helps people, teams and organizations generate value through adaptive solutions for complex problems. In a nutshell, Scrum requires a Scrum Master to foster an environment where:

1. A Product Owner orders the work for a complex problem into a Product Backlog.

2. The Scrum Team turns a selection of the work into an Increment of value during a Sprint.

3. The Scrum Team and its stakeholders inspect the results and adjust for the next Sprint.

4. Repeat Scrum is simple.

Try it as is and determine if its philosophy, theory, and structure help to achieve goals and create value. The Scrum framework is purposefully incomplete, only defining the parts required to implement Scrum theory.

Scrum is built upon by the collective intelligence of the people using it. Rather than provide people with detailed instructions, the rules of Scrum guide their relationships and interactions. Various processes, techniques and methods can be employed within the framework. Scrum wraps around existing practices or renders them unnecessary. Scrum makes visible the relative efficacy of current management, environment, and work techniques, so that improvements can be made.

A diagram of a scrum

Description automatically generated

### Client Requirements

*First Meeting*

The client has brought up that the functionality is greater than the design process of the product we are going to deliver, each sprint there will be a MVP presented that is relevant to the client. Other client requirements can be found in the workbook from Fonty’s.

### Product Backlog

For the product backlog we use Trello to divide the tasks and make it as efficient as possible.

### Functional Requirements

#### Main functionality/ Goal of website and the cloud

The website's primary objective is to provide a secure interface that allows users to upload and download files to and from AWS S3 buckets, Aswell as monitoring costs and logs. The system must ensure secure file transfer, access control, and data integrity while interacting with AWS S3 storage in a cost-efficient manner.

#### Functional Requirements

|  |  |  |
| --- | --- | --- |
| Requirements: | Description: | Priority: |
| The system must provide secure user authentication | Users must create an account and log in to gain access to the file upload and download functionalities. | High |
| The system must enforce role-based access control (RBAC) to determine who can upload, download, or view files. | Different user roles (e.g., admin, regular user) will have different permissions for accessing files. | High |
| Users must be able to upload files securely to specific AWS S3 buckets. | Provide a user-friendly interface where users can select files from their local device and upload them to AWS S3. | High |
| The system must support multiple file types for upload. | images, PDFs, documents, and videos. | Medium |
| The system should limit the size of each file upload | Administrators should be able to configure the maximum file size allowed for upload (e.g., 10MB). | Medium |
| The system should display upload progress and success/failure messages to the user. | Show real-time progress indicators and messages for successful uploads or errors. | Medium |
| Files must be automatically encrypted using server-side encryption when uploaded to AWS S3 | (e.g., AES-256 encryption). | High |
| Users must be able to download files securely from AWS S3 buckets. | After login, users can view a list of files they have access to and download them to their local device. | High |
| The system should log all download actions for auditing purposes. | Maintain a log of when files were downloaded, by whom, and from which IP address. | Medium |
| Users should be able to view a list of uploaded files with details. | such as file name, size, upload date, and uploader. | Medium |
| The system should support file versioning in AWS S3. | If the same file is uploaded multiple times, previous versions should be retained in the S3 bucket (versioning enabled). | High |
| Users should be able to delete files securely from the S3 bucket, with access restrictions in place. |  | Medium |
| The system must support SSL/TLS encryption for all data transfers between the website and AWS S3. | Files should be encrypted both in transit and at rest (e.g., SSL in transit, S3 server-side encryption at rest). | High |
| Users should receive notifications (e.g., email, on-screen) when file uploads or downloads are successful or encounter errors. |  | Low |
| The system must log all critical operations, including uploads, downloads, and file deletions, for auditing purposes. | Use AWS CloudTrail/CloudWatch and server-side logging to track user actions. | High |
| MYSQL Database | to track user information, file metadata, and access logs. | High |
| Some Automation to create Infrastructure | Ansible to create infrastructure and docker for containerization | High |
| Frontend and Backend security | Public Frontend and private Backend in separate subnets/VPCs. | High |
| Demonstrate the upgrade of HTTP to HTTS connection without any changes to EC2 instances (use of Certification Manager and load balancer) |  | High |
| Cost management System | Have alerts and a monitoring system for all costs | High |
|  |  |  |

## 2. Project Goal

The goal for the project is to deliver a platform with a GUI where users can upload and download files securely, as well as effectively and productively use in their environment. For this to happen there will be a infrastructure required to handle the workload and ensure that the platform will be secure so it can handle sensitive data as well.

## 3. User experience

### 3.1 Use cases & User stories

Below are twelve descriptions for user stories, and three use cases for the endgoal for the project.

#### User Stories

1. File Upload & Storage

* **User Story**

*As a user, I want to upload files to cloud storage securely, so that I can access my files from any device.*

* + **Acceptance Criteria**
    - Users can drag and drop files into the application to upload.
    - Files are uploaded securely to Amazon S3 with encryption (client-side or server-side).
    - The system displays progress indicators during uploads.
    - Upon completion, users receive a confirmation notification.

2. File Sharing with Pre-Signed URLs

* **User Story**

*As a user, I want to share files with others using pre-signed URLs, so that I can securely share time-bound access without giving permanent permissions.*

* + **Acceptance Criteria**
    - Users can generate pre-signed URLs with customizable expiration times.
    - URLs grant access to specific files only, without requiring the receiver to log in.
    - Expired URLs are automatically invalidated after the set time.

3. Role-Based Access Control (RBAC)

* **User Story**

*As a user, I want to control access to my files using role-based permissions, so that I can restrict access to specific users or groups.*

* + **Acceptance Criteria**
    - Users can assign roles (e.g., viewer, editor, admin) to individuals or groups for each file or folder.
    - Permissions are enforced both on-premise and in the cloud using AWS IAM.
    - Users can modify permissions at any time, and changes are reflected instantly.

4. On-Premise File Synchronization

* **User Story**

*As a system admin, I want to sync on-premise files with AWS S3, so that sensitive data can be stored locally but also backed up in the cloud.*

* + **Acceptance Criteria**
    - AWS Storage Gateway is configured to automatically sync changes between on-premise and cloud storage.
    - Users experience minimal latency when accessing high-priority files stored on-premise.
    - Changes to files are synced in real-time, with no manual intervention needed.

5. File Versioning & Recovery

* **User Story**

*As a user, I want to access previous versions of my files, so that I can recover older versions if needed.*

* + **Acceptance Criteria**
    - Each file version is stored using S3’s versioning feature.
    - Users can view and restore previous versions directly from the file interface.
    - Automatic backups are performed using AWS Backup for cloud files and a similar solution for on-premise storage.

6. Local File Caching for Fast Access

* **User Story**

*As a user, I want to access frequently used files quickly via local storage, so that I can work with minimal delays.*

* + **Acceptance Criteria**
    - Frequently accessed files are cached on-premise using AWS Storage Gateway.
    - Users can access cached files with low latency through local file systems like NFS or SMB.
    - Files not accessed recently are offloaded to S3 and can be accessed when needed.

7. User Authentication via AWS IAM & Active Directory

* **User Story**

*As an admin, I want to integrate on-premise Active Directory with AWS IAM, so that users can authenticate securely across both environments without duplicating credentials.*

* + **Acceptance Criteria**
    - Users authenticate using existing on-premise Active Directory credentials.
    - AWS Managed Microsoft AD bridges on-premise and cloud authentication.
    - Multi-factor authentication (MFA) is enforced for additional security.
    - No sensitive access keys are stored on users’ local machines.

8. File Sharing & Collaboration with OAuth2

* **User Story**

*As a user, I want to invite collaborators to work on files, so that we can edit and manage files together securely.*

* + **Acceptance Criteria**
    - Users can send invitations for collaboration via email.
    - Collaborators log in using OAuth2, without needing to create new accounts.
    - Role-based permissions (viewer, editor) can be set for each collaborator.
    - Collaboration activities (file edits, comments) are logged for transparency.

9. File Encryption & Security

* **User Story**

*As a security officer, I want all files to be encrypted in transit and at rest, so that sensitive data is protected from unauthorized access.*

* + **Acceptance Criteria**
    - All files are encrypted using S3’s server-side encryption (SSE-S3 or SSE-KMS) or client-side encryption (CSE).
    - Files are transmitted securely using SSL/TLS during uploads and downloads.
    - Access to files is logged using AWS CloudTrail for auditing purposes.
    - On-premise files adhere to the same encryption and security standards.

10. Lifecycle Management for Cost Optimization

* **User Story**

*As a user, I want infrequently accessed files to be automatically archived, so that I can reduce storage costs without losing access to my data.*

* + **Acceptance Criteria**
    - Files not accessed in a specified period are moved to S3 Glacier automatically.
    - Users are notified when files are archived and can restore them if needed.
    - Lifecycle policies are customizable by folder or file type.

11. API for External Integrations

* **User Story**

*As a developer, I want to use an API to programmatically upload and manage files, so that I can integrate the service into my own application.*

* + **Acceptance Criteria**
    - The API allows for file uploads, downloads, and metadata management.
    - API requests are authenticated using OAuth2 or API keys, depending on the use case.
    - AWS API Gateway and Lambda are used to efficiently handle requests, with rate limits in place.

12. Compliance with Security Standards (HIPAA, GDPR, SOC 2)

* **User Story**

*As a compliance officer, I want the file storage solution to meet industry standards such as HIPAA, GDPR, and SOC 2, so that I can ensure the service is secure and legally compliant.*

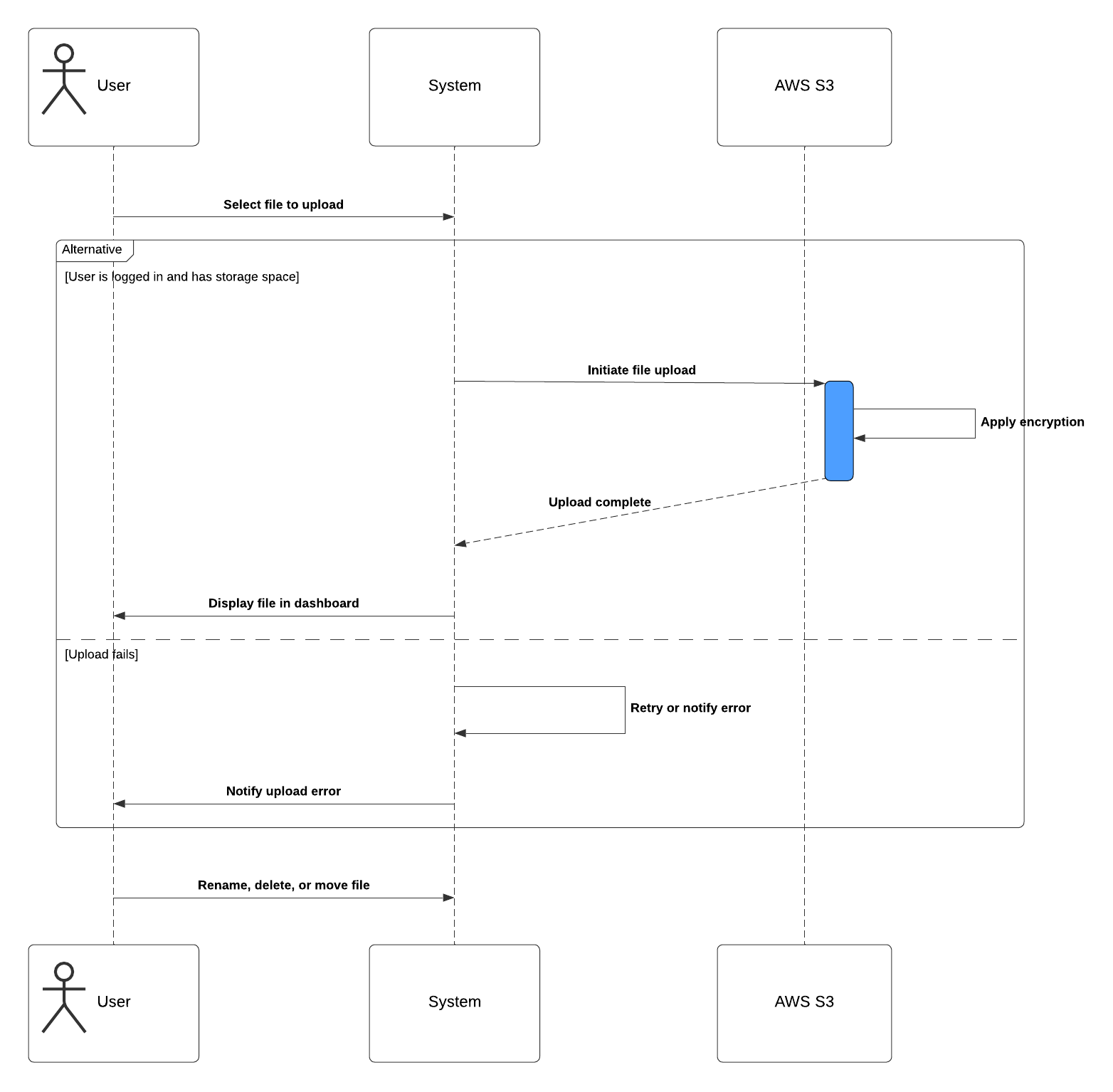
* + **Acceptance Criteria**
    - Files in the cloud are stored in AWS regions that comply with data residency regulations (e.g., GDPR).
    - Auditing and logging features (AWS CloudTrail) are enabled for tracking access to sensitive files.
    - The system supports data encryption, multi-factor authentication, and secure file-sharing features to meet compliance requirements.
    - Regular security audits and vulnerability scans are conducted to maintain compliance.

#### Use Cases

Use Case 1

**Upload and Manage Files**

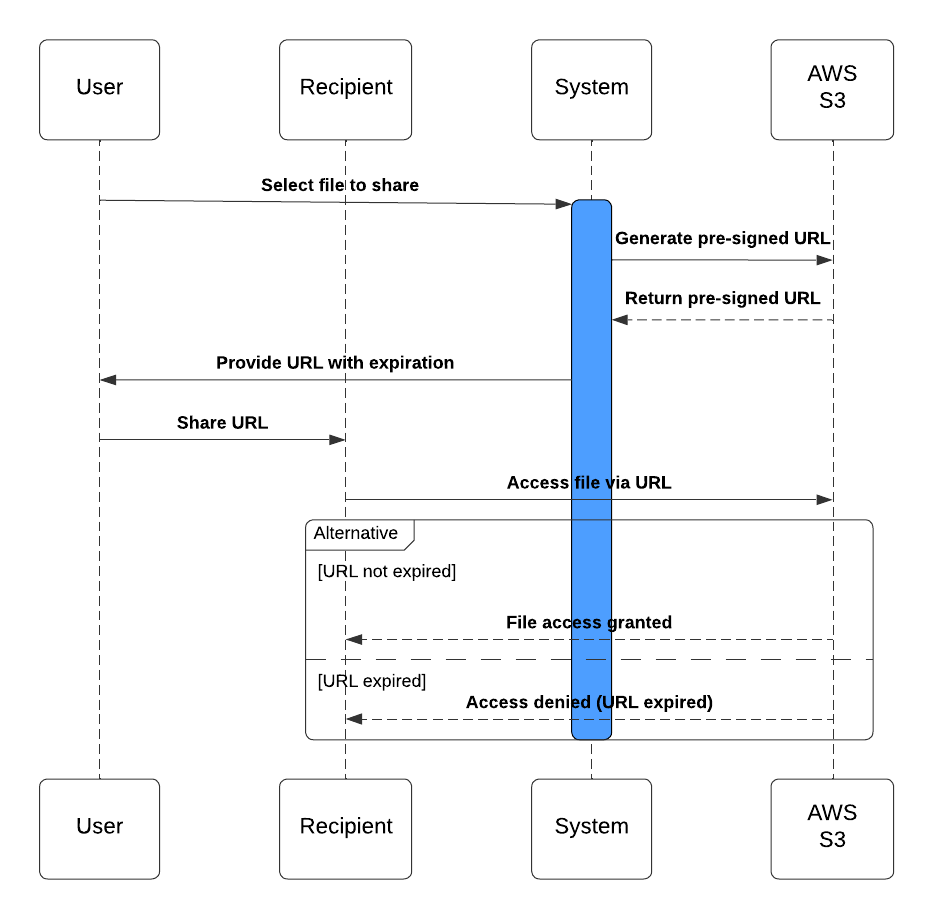
* **Name:** Upload and Manage Files
* **Actors:** User, AWS S3
* **Preconditions:** User is logged into the system and has available storage space.
* **Main Flow**
  1. User selects the file(s) to upload from their device.
  2. The system initiates the upload and shows progress indicators.
  3. The file is uploaded to AWS S3, with encryption applied.
  4. Once the upload is complete, the file appears in the user's file management dashboard.
  5. User can rename, delete, or move the file into folders.
* **Alternative Flow**
  1. If the upload fails, the system retries or notifies the user of the error.
* **Postconditions:** The file is securely stored in AWS S3 and accessible via the user's dashboard.
* **Trigger:** User initiates a file upload.

****

Use Case 2

**Share File with Pre-Signed URL**

* **Name:** Share File with Pre-Signed URL
* **Actors:** User, AWS S3
* **Preconditions:** User is logged into the system and owns the file to be shared.
* **Main Flow**
  1. User selects a file to share.
  2. The system generates a pre-signed URL with an expiration time.
  3. User copies the URL and shares it with the recipient.
  4. The recipient accesses the file via the URL.
* **Alternative Flow**
  1. If the URL expires, the recipient cannot access the file and receives an error message.
* **Postconditions:** The file is shared securely via a time-limited URL.
* **Trigger:** User selects a file and requests to share it.

****

Use Case 3

**Sync On-Premise Files with Cloud Storage**

* **Name:** Sync On-Premise Files with Cloud Storage
* **Actors:** System Admin, AWS Storage Gateway
* **Preconditions:** On-premise file system is configured with AWS Storage Gateway.
* **Main Flow**
  1. Admin configures file folders for synchronization with AWS S3.
  2. The system automatically syncs files between on-premise storage and the cloud.
  3. Changes made to files on-premise are reflected in the cloud, and vice versa.
  4. Users can access synced files from both local storage and cloud storage.
* **Alternative Flow**
  1. If the connection to the cloud fails, changes are queued for syncing when the connection is restored.
* **Postconditions:** Files are synced between on-premise storage and AWS S3, ensuring data availability and backup.
* **Trigger:** Admin config

