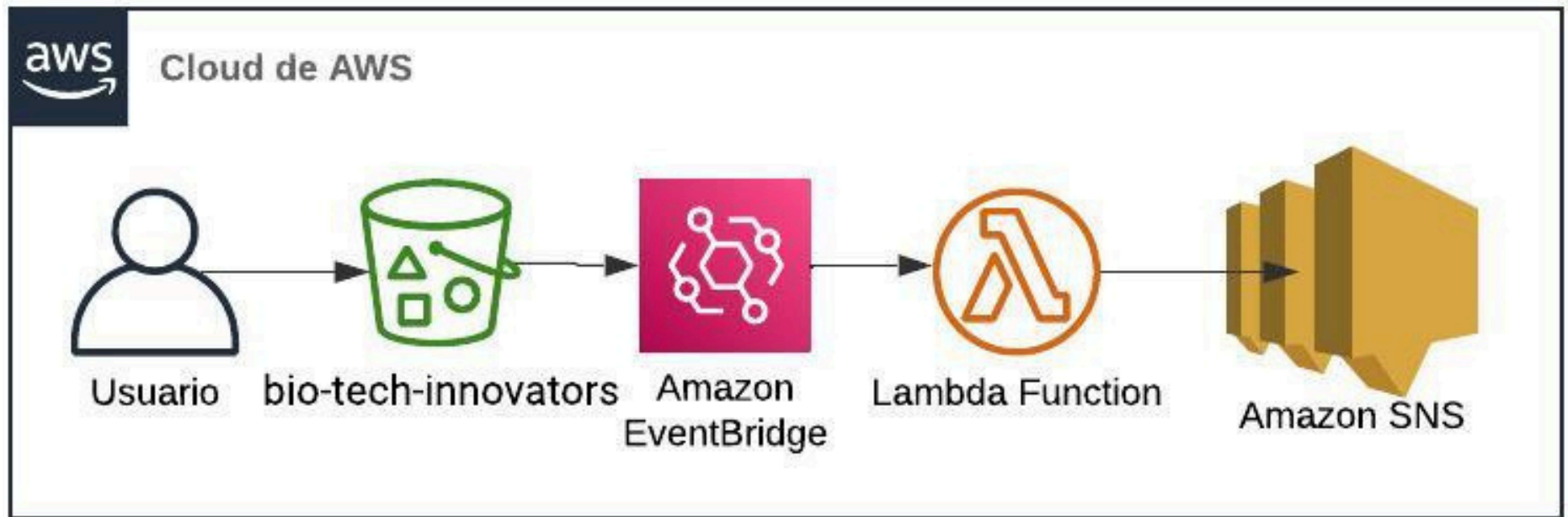


Optimización de la Gestión y Manejo de Datos en BioTechInnovators



Índice del Proyecto para BioTech Innovators

1. Introducción

- 1.1 Descripción de BioTech Innovators
 - *BioTech Innovators es una empresa ficticia dedicada a la investigación biotecnológica. Su propósito es ilustrar un caso práctico para la implementación de soluciones en la nube usando AWS.*
- 1.2 Problemáticas Identificadas
 - 1.2.1 Método de Carpetas Compartidas
 - 1.2.2 Compartir Archivos
 - 1.2.3 Respaldo de Archivos

2. Grupo de Trabajo y Políticas de Permisos

- 2.1 Equipo de Investigación
- 2.2 Equipo Clínico
- 2.3 Equipo de Calidad y Compliance

3. Estructura y Volumen de Información

- 3.1 Carpetas y Propósitos
 - 3.1.1 Datos de Investigación
 - 3.1.2 Resultados Clínicos
 - 3.1.3 Protocolos y Procedimientos
- 3.2 Volumen de Información

4. Propuesta de Solución

- 4.1 Cómo Amazon S3 Resuelve las Problemáticas
 - 4.1.1 Gestión de Archivos

- 4.1.2 Seguridad y Compartición de Archivos
- 4.1.3 Respaldo y Protección de Datos
- **4.2 Cómo S3 Intelligent-Tiering (S3 INT) optimiza costos:**
 - 4.2.1 Optimización de Costos con S3 Intelligent-Tiering (S3 INT)

5. Solución Técnica

- 5.1 Creación de Grupos IAM
- 5.2 Plantillas Usuarios IAM
- 5.3 Creación de Políticas IAM
 - Política de Carga y Modificación en Carpeta de Investigación
 - Política de Carga y Modificación en Carpeta de Resultados Clínicos
 - Política de Carga y Modificación en Carpeta de Protocolos y Procedimiento
 - Política de Solo Lectura
- 5.4 Plantilla Lambda y SNS
- 5.5 Plantilla de S3 y EventBridge

6. Estimacion de costos

- 6.1 Acceso Frecuente (Frequent Access)
- 6.2 Acceso Infrecuente (Infrequent Access)
- 6.3 Acceso Instantáneo a Archivos Archivados (Archive Instant Access)
- 6.4 Acceso a Archivos Archivados (Archive Access)
- 6.5 Acceso a Archivos en Archivo Profundo (Deep Archive Access)
- 6.6 Otros Costos
- 6.7 Costo Total Mensual Aproximado
- 6.8 Costo Mensual Total Estimado

1. ¿Qué es y cuál es su problemática?

1.1 Descripción de BioTech Innovators

BioTech Innovators es una empresa científica de vanguardia dedicada a la investigación y desarrollo en biotecnología. Su misión es descubrir y desarrollar soluciones innovadoras que mejoren la salud y el bienestar a través de la biotecnología. La empresa trabaja en una amplia gama de proyectos, desde investigaciones básicas en biología molecular hasta el desarrollo de productos biotecnológicos aplicados en el ámbito clínico y farmacéutico. Actualmente, enfrentan desafíos significativos relacionados con la gestión de la información.

1.2 Problemáticas Identificadas

La empresa enfrenta varios problemas con su gestión de información. Primero, la creación de carpetas compartidas entre computadoras ha demostrado ser inviable debido a problemas de *sincronización y actualización*, lo que genera riesgos de versiones desactualizadas y pérdida de datos. Además, el uso del correo electrónico para compartir archivos es limitado, lo que causa ineficiencias en la colaboración y la organización, además de riesgos de seguridad al enviar información sensible. Por último, el almacenamiento de archivos importantes en computadoras individuales sin respaldo centralizado presenta riesgos de pérdida de datos y dependencia de dispositivos de almacenamiento extraíbles que requieren gestión manual.

- **Desglose de estas problemáticas**

1.2.1 Método de Carpetas Compartidas:

- La creación de carpetas compartidas entre computadoras no ha sido viable.
- Desafíos en la sincronización y actualización de archivos entre dispositivos.
- Riesgo de versiones desactualizadas y pérdida de datos.

1.2.2 Compartir Archivos:

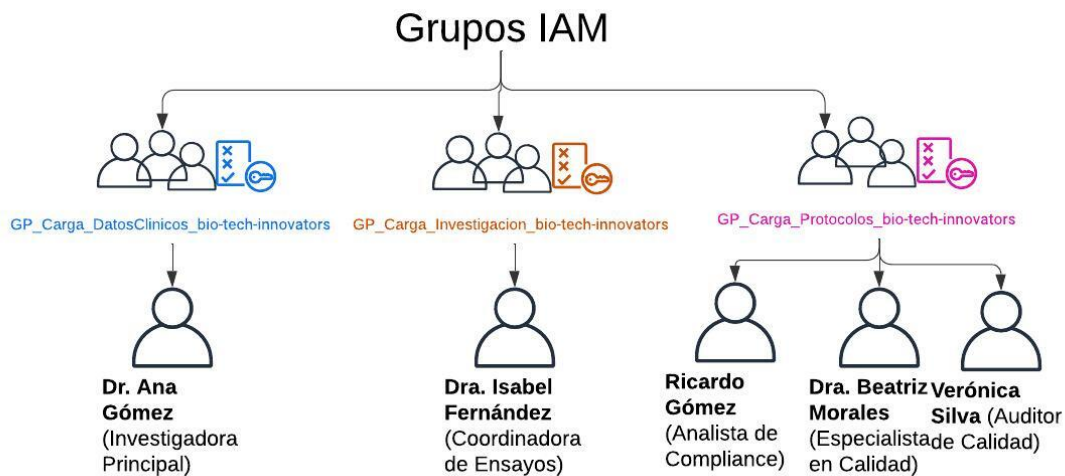
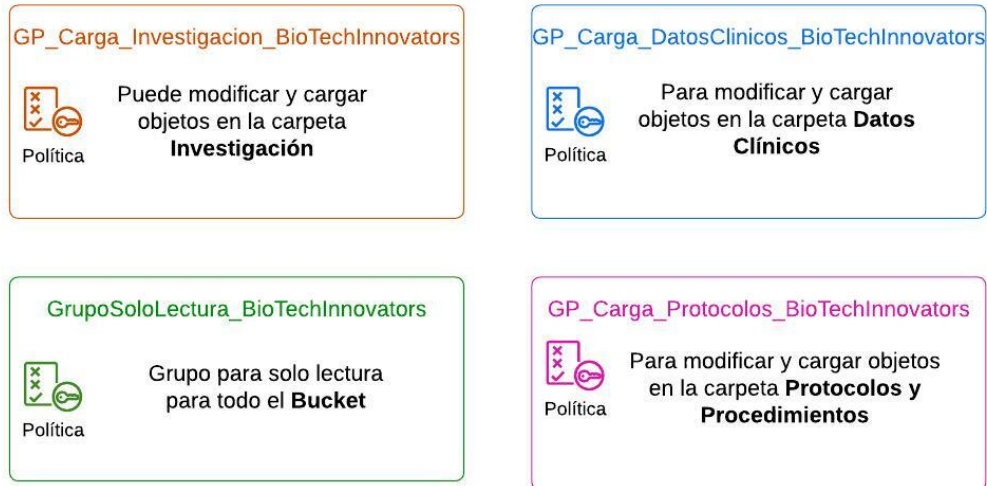
- Uso limitado del correo electrónico para compartir archivos.
- Ineficiencia en la colaboración y organización de documentos.
- Riesgos de seguridad al enviar información sensible por canales no seguros.

1.2.3 Respaldo de Archivos:

- Almacenamiento de archivos importantes en computadoras individuales sin un respaldo centralizado.
- Riesgo de pérdida de datos debido a fallas técnicas o daños en los dispositivos.
- Dependencia de dispositivos de almacenamiento extraíbles para respaldos, con gestión manual y falta de actualización constante.
- Dificultad para compartir y mantener actualizados los respaldos con colegas ubicados en diferentes regiones del país.

2. Grupo de Trabajo y Políticas de Permisos

Grupos con permisos de modificación según carpeta



2.1 Equipo de Investigación

- **Funciones:** Realizar experimentos, analizar datos y desarrollar nuevas tecnologías.

- **Permisos en Carpeta de Datos de Investigación:** Lectura y descarga para todos; carga y modificación solo para investigadores principales.
- **Conformado por:**
 - **Dr. Ana Gómez** (Investigadora Principal)
 - **Laura Pérez** (Investigadora)
 - **Carlos Rodríguez** (Investigador)

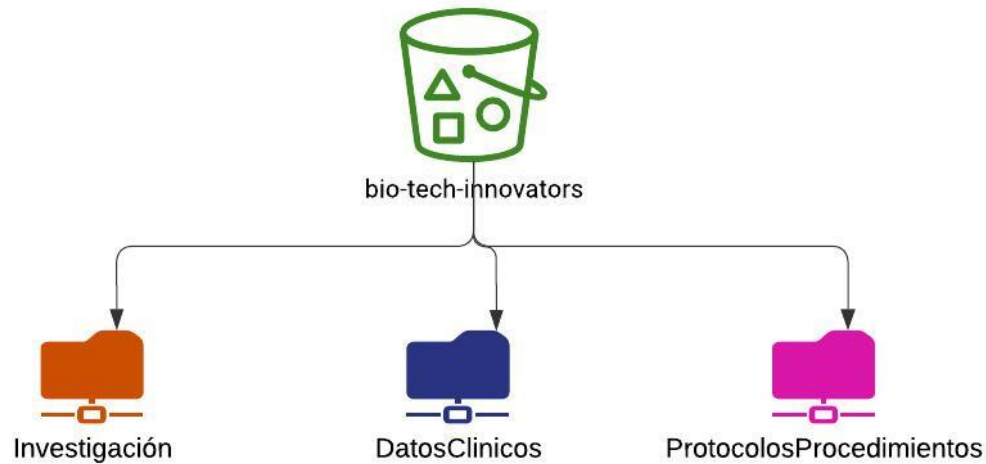
2.2 Equipo Clínico

- **Funciones:** Conducir ensayos clínicos, recolectar y analizar datos de pacientes.
- **Permisos en Carpeta de Resultados Clínicos:** Lectura y descarga para todos; carga y modificación solo para coordinadores de ensayos.
- **Conformado por:**
 - **Dra. Isabel Fernández** (Coordinadora de Ensayos)
 - **María López** (Especialista en Datos Clínicos)
 - **Pedro Sánchez** (Técnico Clínico)

2.3 Equipo de Calidad y Compliance

- **Funciones:** Implementar y revisar protocolos y procedimientos, asegurar conformidad con regulaciones.
- **Permisos en Carpeta de Protocolos y Procedimientos:** Lectura y descarga para todos; carga y modificación solo para el equipo de calidad y compliance.
- **Conformado por:**
 - **Dra. Beatriz Morales** (Especialista en Calidad)
 - **Ricardo Gómez** (Analista de Compliance)
 - **Verónica Silva** (Auditor de Calidad)
 - **Fernando López** (Consultor en Regulaciones)

3. Estructura y Volumen de Información



3.1 Carpetas y Propósitos:

3.1.1 Datos de Investigación:

- **Propósito:** Almacenar datos crudos y procesados de los experimentos científicos.
- **Permisos:** Todos los miembros del equipo pueden leer y descargar los datos. Solo los investigadores principales tienen permisos para cargar y modificar los archivos.

3.1.2 Resultados Clínicos:

- **Propósito:** Almacenar registros de pacientes, resultados de pruebas y reportes de análisis de ensayos clínicos.
- **Permisos:** Todos los miembros del equipo clínico pueden leer y descargar los datos. Solo los coordinadores de ensayos tienen permisos para cargar y modificar los archivos.

3.1.3 Protocolos y Procedimientos:

- **Propósito:** Almacenar protocolos de laboratorio, manuales y procedimientos operativos estándar.
- **Permisos:** Todos los empleados pueden leer y descargar los documentos. Solo el equipo de calidad y compliance tiene permisos para cargar y modificar los archivos.

3.2 Volumen de Información

BioTech Innovators maneja 50 GB de datos que se acceden con frecuencia durante los primeros 30 días y 20 GB de datos que se acceden de manera infrecuente entre los 30 y 90 días. A partir de este período, el resto de la información, que incluye 30 GB de datos, se accede ocasionalmente.

4. Propuesta

Como asesor recomiendo implementar una solución basada en Amazon S3 y IAM para abordar estos desafíos. Amazon S3 proporcionará almacenamiento escalable y seguro en la nube, centralizando archivos accesibles desde cualquier lugar, eliminando problemas de sincronización y versiones desactualizadas, y mejorando la colaboración y la seguridad en el acceso a archivos.

4.1 Cómo Amazon S3 y S3 INT nos ayudan a resolver estas problemáticas

4.1.1 Gestión de Archivos:

- **Amazon S3:** Proporciona almacenamiento escalable y seguro en la nube, centralizando archivos accesibles desde cualquier lugar.
- **Beneficios:** Elimina problemas de sincronización y versiones desactualizadas, mejorando la colaboración y la seguridad en el acceso a archivos.

4.1.2 Seguridad y Compartición de Archivos:

- **Amazon S3:** Permite configurar políticas de acceso y permisos granulares para compartir archivos de manera controlada y segura.
- **Beneficios:** Mejora la seguridad frente al uso de métodos inseguros como el correo electrónico.

4.1.3 Respaldo y Protección de Datos:

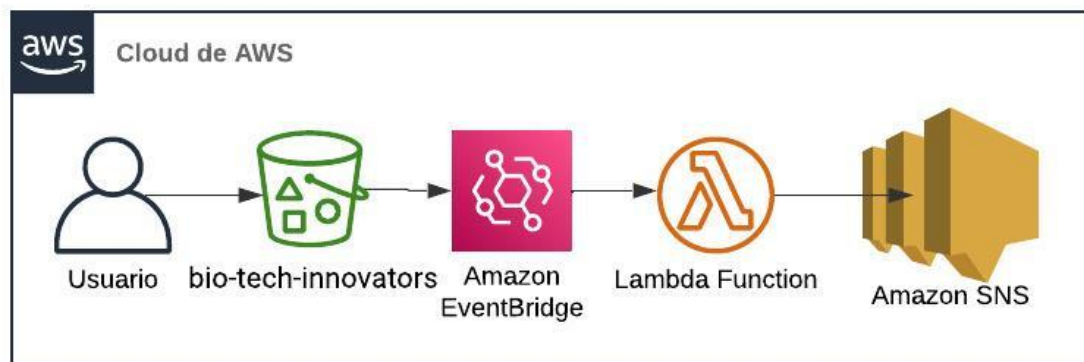
- **Amazon S3:** Ofrece redundancia y replicación de datos entre regiones, asegurando la protección contra pérdidas por fallos técnicos o desastres naturales.
- **Beneficios:** Elimina la necesidad de dispositivos de almacenamiento físico propensos a fallos, garantizando la disponibilidad y seguridad de los datos en todo momento.

4.2 Cómo S3 Intelligent-Tiering (S3 INT) optimiza costos:

4.2.1 Optimización Automática de Costos:

- **S3 Intelligent-Tiering:** Clasifica automáticamente los datos según su frecuencia de acceso, moviéndolos a capas de almacenamiento más económicas cuando son menos accedidos.
- **Beneficios:** Reduce los costos de almacenamiento sin comprometer la accesibilidad inmediata de los datos, optimizando automáticamente los costos de almacenamiento basado en el uso real de los archivos.

5. Solución Técnica



• 5.1 Creación de Grupos IAM

- Estos grupos tendrán políticas específicas para un bucket S3

```
AWSTemplateFormatVersion: '2010-09-09'
```

```
Description: Creación de grupos IAM que tendrán políticas es
```

```
Resources:
```

```
# Grupo para solo lectura
```

```
GrupoSoloLectura:
```

```
  Type: AWS::IAM::Group
```

```
  Properties:
```

```
    GroupName: GrupoSoloLectura_BioTechInnovators
```

```
  Metadata:
```

```
    Description: "Este grupo tiene los permisos de solo le
```

```
# Grupo para Carga y Modificación en la Carpeta Investigac:
```

```
GrupoCargaInvestigacion:
```

```
  Type: AWS::IAM::Group
```

```
  Properties:
```

```
    GroupName: GP_Carga_Investigacion_BioTechInnovators
```

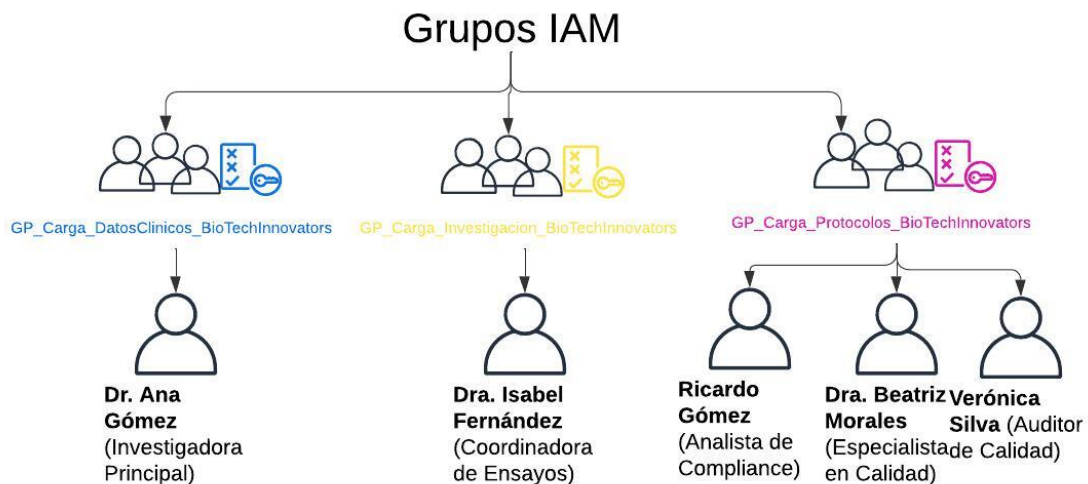
```
  Metadata:
```

```
    Description: "Este grupo tiene los permisos para carga
```

```
# Grupo para Carga y Modificación en la Carpeta Datos Clín:
GrupoCargaDatosClinicos:
  Type: AWS::IAM::Group
  Properties:
    GroupName: GP_Carga_DatosClinicos_BioTechInnovators
  Metadata:
    Description: "Este grupo tiene los permisos para carga y modificación en la carpeta Datos Clínicos"

# Grupo para Carga y Modificación en la Carpeta Protocolos
GrupoCargaProtocolos:
  Type: AWS::IAM::Group
  Properties:
    GroupName: GP_Carga_Protocolos_BioTechInnovators
  Metadata:
    Description: "Este grupo tiene los permisos para carga y modificación en la carpeta Protocolos"
```

• 5.2 Plantillas Usuarios IAM



AWSTemplateFormatVersion: '2010-09-09'

Description: Crear usuarios IAM y asignarlos a grupos para B:

Resources:

Usuarios IAM

UsuarioAnaGomez:

Type: AWS::IAM::User

Properties:

UserName: AnaGomez

Groups:

- !Ref GrupoSoloLectura
- !Ref GrupoCargaInvestigacion

UsuarioJavierMartinez:

Type: AWS::IAM::User

Properties:

UserName: JavierMartinez

Groups:

- !Ref GrupoSoloLectura
- !Ref GrupoCargaInvestigacion

UsuarioLauraPerez:

Type: AWS::IAM::User

Properties:

UserName: LauraPerez

Groups:

- !Ref GrupoSoloLectura
- !Ref GrupoCargaInvestigacion

UsuarioCarlosRodriguez:

Type: AWS::IAM::User

Properties:

UserName: CarlosRodriguez

Groups:

- !Ref GrupoSoloLectura

- !Ref GrupoCargaInvestigacion

UsuarioIsabelFernandez:

Type: AWS::IAM::User

Properties:

UserName: IsabelFernandez

Groups:

- !Ref GrupoSoloLectura
- !Ref GrupoCargaDatosClinicos

UsuarioLuisRamirez:

Type: AWS::IAM::User

Properties:

UserName: LuisRamirez

Groups:

- !Ref GrupoSoloLectura
- !Ref GrupoCargaDatosClinicos

UsuarioMariaLopez:

Type: AWS::IAM::User

Properties:

UserName: MariaLopez

Groups:

- !Ref GrupoSoloLectura
- !Ref GrupoCargaDatosClinicos

UsuarioPedroSanchez:

Type: AWS::IAM::User

Properties:

UserName: PedroSanchez

Groups:

- !Ref GrupoSoloLectura
- !Ref GrupoCargaDatosClinicos

UsuarioBeatrizMorales:

Type: AWS::IAM::User

```
Properties:
  UserName: BeatrizMorales
  Groups:
    - !Ref GrupoSoloLectura
    - !Ref GrupoCargaProtocolos
```

```
UsuarioRicardoGomez:
  Type: AWS::IAM::User
  Properties:
    UserName: RicardoGomez
    Groups:
      - !Ref GrupoSoloLectura
      - !Ref GrupoCargaProtocolos
```

```
UsuarioVeronicaSilva:
  Type: AWS::IAM::User
  Properties:
    UserName: VeronicaSilva
    Groups:
      - !Ref GrupoSoloLectura
      - !Ref GrupoCargaProtocolos
```

```
UsuarioFernandoLopez:
  Type: AWS::IAM::User
  Properties:
    UserName: FernandoLopez
    Groups:
      - !Ref GrupoSoloLectura
      - !Ref GrupoCargaProtocolos
```

• 5.3 Creación de Políticas IAM

- **Estas políticas otorgaran permisos a las carpetas Especificadas**

```
AWSTemplateFormatVersion: '2010-09-09'
Description: Crear políticas IAM con permisos para modificar
```


Parameters:

NombreBucket:

Description: Nombre del bucket S3

Type: String

Resources:

Política para modificar y cargar objetos en la carpeta Investigación

PoliticaCargarInvestigacion:

Type: AWS::IAM::Policy

Properties:

PolicyName: Politica_carga_Investigacion_\${NombreBucket}

PolicyDocument:

Version: '2012-10-17'

Statement:

- Effect: Allow

Action:

- s3:ListBucket
- s3:GetBucketLocation

Resource:

- !Sub arn:aws:s3:::\${NombreBucket}

- Effect: Allow

Action:

- s3:GetObject
- s3:GetObjectVersion
- s3:PutObject

Resource:

- !Sub arn:aws:s3:::\${NombreBucket}/Investigacion

Groups:

- Ref: GP_Carga_Investigacion_BioTechInnovators

Política para modificar y cargar objetos en la carpeta Datos Clínicos

PoliticaCargarDatosClinicos:

Type: AWS::IAM::Policy

Properties:

PolicyName: Politica_carga_DatosClinicos_\${NombreBucket}

```

PolicyDocument:
  Version: '2012-10-17'
  Statement:
    - Effect: Allow
      Action:
        - s3:ListBucket
        - s3:GetBucketLocation
      Resource:
        - !Sub arn:aws:s3:::${NombreBucket}
    - Effect: Allow
      Action:
        - s3:GetObject
        - s3:GetObjectVersion
        - s3:PutObject
      Resource:
        - !Sub arn:aws:s3:::${NombreBucket}/DatosClinicos
  Groups:
    - Ref: GP_Carga_DatosClinicos_BioTechInnovators

```

Política para modificar y cargar objetos en la carpeta Políticas

PoliticaCargarProtocolos:

Type: AWS::IAM::Policy

Properties:

PolicyName: Politica_carga_Protocolos_\${NombreBucket}

PolicyDocument:

Version: '2012-10-17'

Statement:

```

- Effect: Allow
  Action:
    - s3:ListBucket
    - s3:GetBucketLocation
  Resource:
    - !Sub arn:aws:s3:::${NombreBucket}
- Effect: Allow
  Action:
    - s3:GetObject

```

```

        - s3:GetObjectVersion
        - s3:PutObject
    Resource:
        - !Sub arn:aws:s3:::${NombreBucket}/Protocolos/
Groups:
    - Ref: GP_Carga_Protocolos_BioTechInnovators

# Política para acceso de solo lectura al bucket S3
PoliticaSoloLectura:
    Type: AWS::IAM::Policy
    Properties:
        PolicyName: Politica_SoloLectura_${NombreBucket}
        PolicyDocument:
            Version: '2012-10-17'
            Statement:
                - Effect: Allow
                  Action:
                    - s3:ListBucket
                    - s3:GetBucketLocation
                  Resource: arn:aws:s3:::${NombreBucket}
                - Effect: Allow
                  Action:
                    - s3:GetObject
                    - s3:GetObjectVersion
                  Resource: arn:aws:s3:::${NombreBucket}/*
Groups:
    - !Ref GrupoSoloLectura_BioTechInnovators

```

• 5.4 Plantilla Lambda y SNS

Arquitectura de la Solución

1. **EventBridge:** Detecta la carga de nuevos datos en el bucket S3 de datos crudos y activa la función Lambda.
2. **AWS Lambda:** La función Lambda procesa el evento de S3, extrayendo el nombre del bucket y del objeto cargado, y envía una notificación a través

de SNS.

3. **Amazon SNS:** Envía las notificaciones a los suscriptores mediante email.

```
AWSTemplateFormatVersion: '2010-09-09'
Description: Stack para crear una función Lambda que puede pu

Parameters:
  ARNTopicSNS:
    Type: String
    Description: ARN del tema que accionará la función lambda
  LambdaRegion:
    Type: String
    Description: Región en la que se desplegará la función L

Resources:
  MyLambdaExecutionRole:
    Type: AWS::IAM::Role
    Properties:
      AssumeRolePolicyDocument:
        Version: '2012-10-17'
        Statement:
          - Effect: Allow
            Principal:
              Service:
                - lambda.amazonaws.com
            Action: 'sts:AssumeRole'
      Policies:
        - PolicyName: Lambda-SNSPublishPolicy
          PolicyDocument:
            Version: '2012-10-17'
            Statement:
              - Effect: Allow
                Action: sns:Publish
                Resource: !Ref ARNTopicSNS
```

```

MyLambdaFunction:
  Type: AWS::Lambda::Function
  Properties:
    Environment:
      Variables:
        ARNTopicSNS: !Ref ARNTopicSNS
        LambdaRegion: !Ref AWS::Region
    Code:
      ZipFile: |
        import json
        import boto3

        def lambda_handler(event, context):
            # Inicializar el cliente de SNS
            sns_client = boto3.client('sns')

            # Extraer la información del evento
            bucket_name = event['Records'][0]['s3']
            object_key = event['Records'][0]['s3']

            # Construir el mensaje
            cuerpo = f"Se ha cargado un nuevo objeto en el bucket {bucket_name} con la clave {object_key}"

            # Publicar el mensaje en el tópico SNS
            sns_client.publish(
                TopicArn='arn:aws:sns:us-west-2::123456789012:my-topic',
                Subject='Notificación de carga de objeto',
                Message=cuerpo)

            return {
                'statusCode': 200,
                'body': json.dumps('Notificación enviada')
            }

    Handler: index.lambda_handler
    Role: !GetAtt MyLambdaExecutionRole.Arn

```

```
Runtime: python3.8
Timeout: 120
```

Outputs:

```
MyLambdaFunctionName:
  Description: Nombre de la función Lambda
  Value: !Ref MyLambdaFunction
```

- **5.5 Plantilla de S3 y EventBridge**

```
AWSTemplateFormatVersion: '2010-09-09'
Description: "Plantilla para crear una regla de EventBridge (
```

Parameters:

```
BucketName:
  Description: "Nombre de bucket S3 existente"
  Type: String
```

```
LambdaFunctionName:
  Description: "Nombre de la función Lambda existente"
  Type: String
```

Resources:

```
# Creación de la regla de EventBridge para activar la funci
S3EventBridgeRule:
  Type: AWS::Events::Rule
  Properties:
    Description: "Regla para activar la función Lambda en
    EventPattern:
      source:
        - "aws.s3"
      detail-type:
        - "Object Created"
      detail:
        bucket:
```

```

        name:
          - !Ref BucketName
    State: "ENABLED"
    Targets:
      - Arn: !Sub "arn:aws:lambda:${AWS::Region}:${AWS::Account}:function:${LambdaFunctionName}"
        Id: "TargetFunctionV1"

# Permisos para que EventBridge invoque la función Lambda
LambdaInvokePermission:
  Type: AWS::Lambda::Permission
  Properties:
    Action: "lambda:InvokeFunction"
    FunctionName: !Sub "${LambdaFunctionName}"
    Principal: "events.amazonaws.com"
    SourceArn: !GetAtt S3EventBridgeRule.Arn
      - S3EventBridgeRule
      - Arn

Outputs:
  # Salida con el ARN de la regla de EventBridge
  RuleARN:
    Description: "ARN de la regla de EventBridge"
    Value: !GetAtt S3EventBridgeRule.Arn

```

6. Estimación de costos

Basado en el manejo de datos de BioTech Innovators y los niveles de acceso descritos, la estimación de costos sería la siguiente:

- **6.1 Acceso Frecuente (Frequent Access):**
 - **Descripción:** Datos que se acceden regularmente, como archivos de proyectos activos.
 - **Almacenamiento:** 50 GB
 - **Costo:** \$2.025 USD
- **6.2 Acceso Infrecuente (Infrequent Access):**
 - **Descripción:** Datos accedidos esporádicamente, como informes de proyectos pasados.
 - **Almacenamiento:** 20 GB
 - **Costo:** \$0.442 USD
- **6.3 Acceso Instantáneo a Archivos Archivados (Archive Instant Access):**
 - **Descripción:** Datos accedidos ocasionalmente, como información histórica relevante.
 - **Almacenamiento:** 13 GB
 - **Costo:** \$0.1079 USD
- **6.4 Acceso a Archivos Archivados (Archive Access):**
 - **Descripción:** Estimación de datos accedidos raramente.
 - **Almacenamiento:** 12 GB
 - **Costo:** \$0.092 USD
- **6.5 Acceso a Archivos en Archivo Profundo (Deep Archive Access):**
 - **Descripción:** Estimación de datos accedidos muy raramente, para auditorías o referencias muy infrecuentes.
 - **Almacenamiento:** 5 GB

- **Costo:** \$0.0161 USD
- **6.6 Otros Costos:**
 - **Monitoreo y Automatización de Objetos:** \$0.0085 USD
 - **Solicitudes PUT:** \$0.028 USD
 - **Solicitudes GET:** \$0.0031 USD
 - **Solicitudes de Ciclo de Vida:** \$0.003 USD
 - **Datos Seleccionados Retornados:** \$0.035 USD
 - **Datos Escaneados:** \$0.20 USD
- **6.7 Costo Total Mensual Aproximado:**
 - **Almacenamiento:** \$2.025 (Acceso Frecuente) + \$0.442 (Acceso Infrecuente) + \$0.1079 (Acceso Instantáneo) + \$0.092 (Acceso Archivado) + \$0.0161 (Archivo Profundo) = \$2.683 USD
 - **Otros Costos:** \$0.0085 (Monitoreo) + \$0.028 (Solicitudes PUT) + \$0.0031 (Solicitudes GET) + \$0.003 (Ciclo de Vida) + \$0.035 (Datos Seleccionados) + \$0.20 (Datos Escaneados) = \$0.2776 USD
- **6.8 Costo Mensual Total Estimado:**
 - **\$2.683 + \$0.2776 = \$2.9606 USD**

Este cálculo proporciona una estimación aproximada del costo mensual de almacenamiento en Amazon S3 para los diferentes niveles de acceso según las necesidades de BioTech Innovators.