UNIVERSIDAD EAFIT

IMPLEMENTACIÓN AUTOMATIZADA DE LFTAGS AWS

Susana Tilano Florez

CONTENIDO

1. Contextualización
2. Planteamiento del problema
3. Formulación
4. Solución

CONTEXTUALIZACIÓN – LA NUBE EN CIFRAS

1

Flexibilidad

Las organizaciones pueden **escalar** su infraestructura y servicios **a demanda**, según las necesidades de operación, seguridad y control.

2

Valor estratégico

Los proveedores aplican la **mejora continua** en sus servicios permitiendo contar con tecnología actualizada sin tener que dedicar gran cantidad de recursos TI a la gestion de infraestructura. Incremento de la **productividad**.

3

Eficiencia

Accesibilidad de las herramientas y datos desde prácticamente cualquier dispositivo conectado a Internet con **velocidad** de comercialización y redución de costos gracias al **pago por uso**.

CONTEXTUALIZACIÓN – LA NUBE EN CIFRAS

Distribución del mercado

AWS **62%**, Microsoft (48%) y Google (33%) son los proveedores de nube más grandes.

O'Reilly, 2021

Carga Operacional

El **70%** de las empresas ahora alojan más de la mitad de sus cargas de trabajo en la nube.

Palo Alto Networks, 2021

Seguridad de la información

El **92%** de las empresas intentarán implementar el privilegio mínimo en la nube.

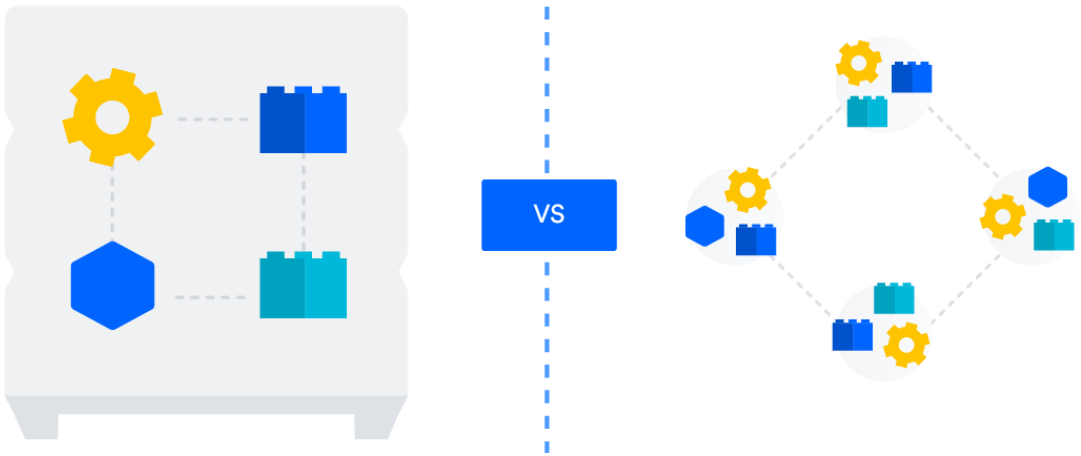
Ermetic, 2022

CONTEXTUALIZACIÓN – ARQUITECTURA

La arquitectura de la nube permite implementar estrategias para combinar componentes tecnológicos, coordinando recursos, compartiéndolos en redes y respondiendo ante necesidades empresariales específicas.

Monolith

Modelo tradicional de un programa de software que se compila como una unidad unificada y que es autónoma e independiente de otras aplicaciones.



Microservices

Modelo que se basa en una serie de servicios que se pueden implementar de forma independiente. Estos servicios tienen su propia lógica empresarial y base de datos con un objetivo específico.

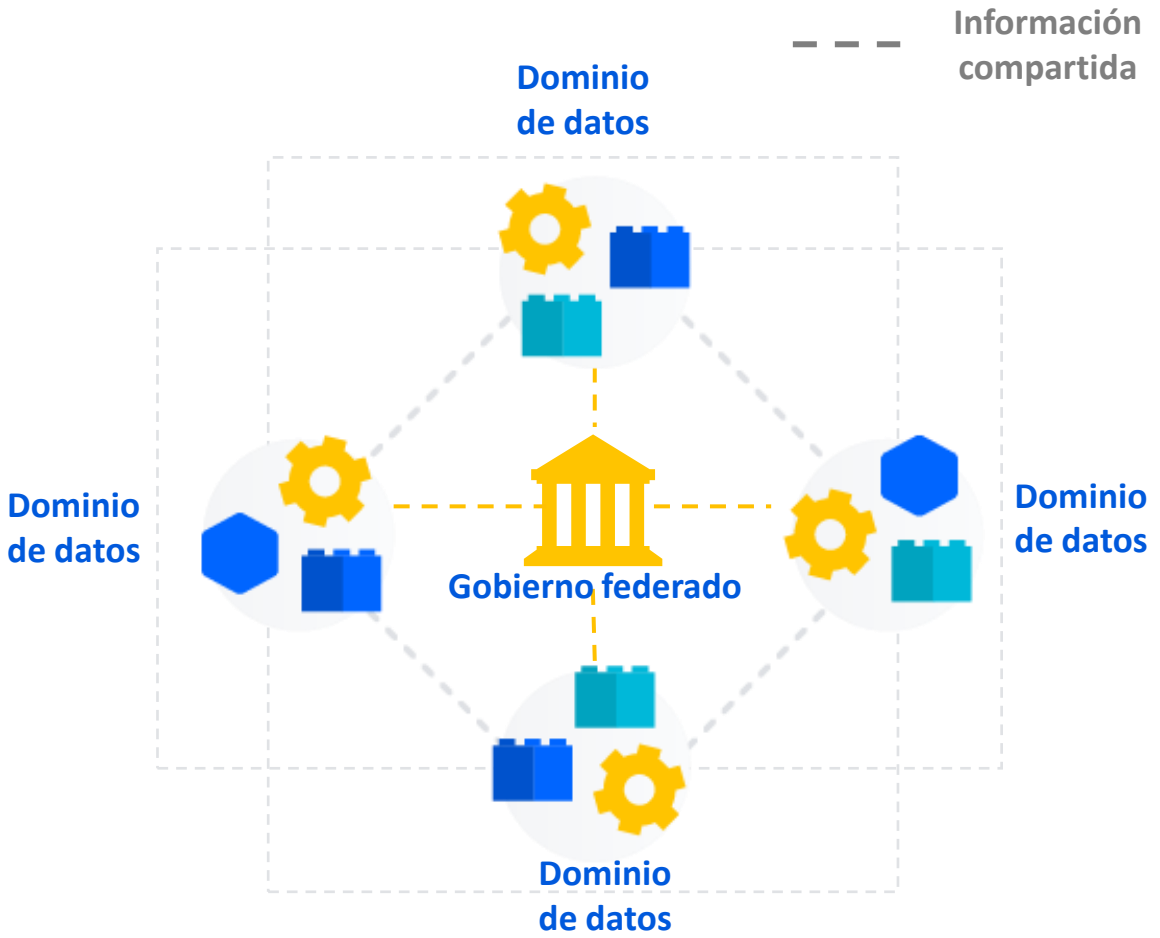
NETFLIX

Una de las primeras empresas destacadas en migrar de un monolito a una arquitectura de microservicios basada en la nube. Ganó el premio JAX Special Jury de 2015,

CONTEXTUALIZACIÓN – ARQUITECTURA

Uno de los marcos de arquitectura desarrollado en los últimos años se conoce como Data Mesh (Zhamak Dehghani), este asegura la descentralización y democratización de los datos, implementado un gobierno de datos transversal a las capacidades de cada línea de negocio. Cada una de estas mantiene el control sobre sus datos, quién, cómo y en qué formato se accede a estos.

> Una malla de datos aumenta la complejidad de la arquitectura, pero también aporta **eficacia** al mejorar el acceso a los datos, la **seguridad** y la **escalabilidad** [4]



Empresas que han implementado una arquitectura data mesh

CONTEXTUALIZACIÓN – IaC

La **infraestructura informática de nube** es la recopilación de elementos de hardware y software necesarios para hacer posible la informática de nube. Incluye capacidad de procesamiento, red y almacenamiento, así como una interfaz para que los usuarios accedan a sus recursos virtualizados.

La **infraestructura como código** (IaC) permite gestionar y preparar la infraestructura a través del código, en lugar de hacerlo mediante procesos manuales.

VENTAJAS

- Reducción de costos
- Aumento en la velocidad de implementación
- Disminución de la cantidad de errores
- Mayor uniformidad de la infraestructura
- Eliminación de los desajustes de configuración

HERRAMIENTAS

TERRAFORM

CHEF

ANSIBLE

CLOUD FORMATION

PLANTEAMIENTO DEL PROBLEMA

PLANTEAMIENTO DEL PROBLEMA

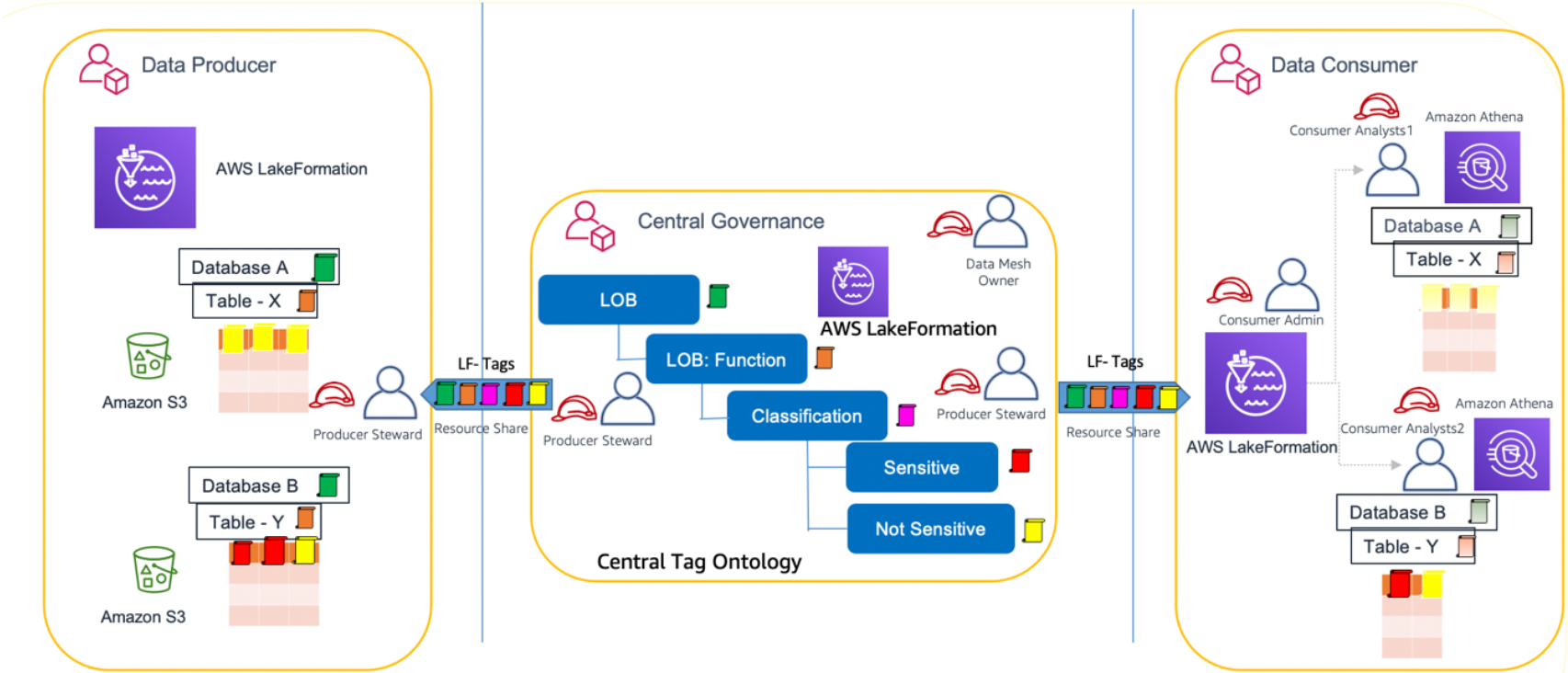
En Colombia, se establece la **ley 1266 de 2008**, esta indica las disposiciones de la protección de datos para las entidades financieras, por lo tanto los servicios y tecnologías que usan las personas o empresas, se ven relacionados en la intención de que su información sea protegida.

¿Cómo cumplir con la ley en la nube?

En la nube las **políticas basadas en roles** son importantes para segregar de manera correcta y establecer sus privilegios, permitiendo conocer quien realiza algún acceso y permitiendo llevar un control de la información a la que se accede.

PLANTEAMIENTO DEL PROBLEMA

Dentro de la arquitectura del Data Mesh, la implementación de **Lake Formation Tags** (LFTags) junto con el control de acceso basado en roles IAM, soportará que la democratización de la información se dé siguiendo los lineamientos de control, seguridad y cumplimiento de la normatividad vigente para la protección de datos, de manera escalable y simplificando la administración de los datos. Se muestra un ejemplo de como por medio de LFTags se controla el acceso a la información de las bases de datos A y B, con sus tablas X y Y.



https://aws.amazon.com/es/blogs/big-data/build-a-modern-data-architecture-and-data-mesh-pattern-at-scale-using-aws-lake-formation-tag-based-access-control/

FORMULACIÓN

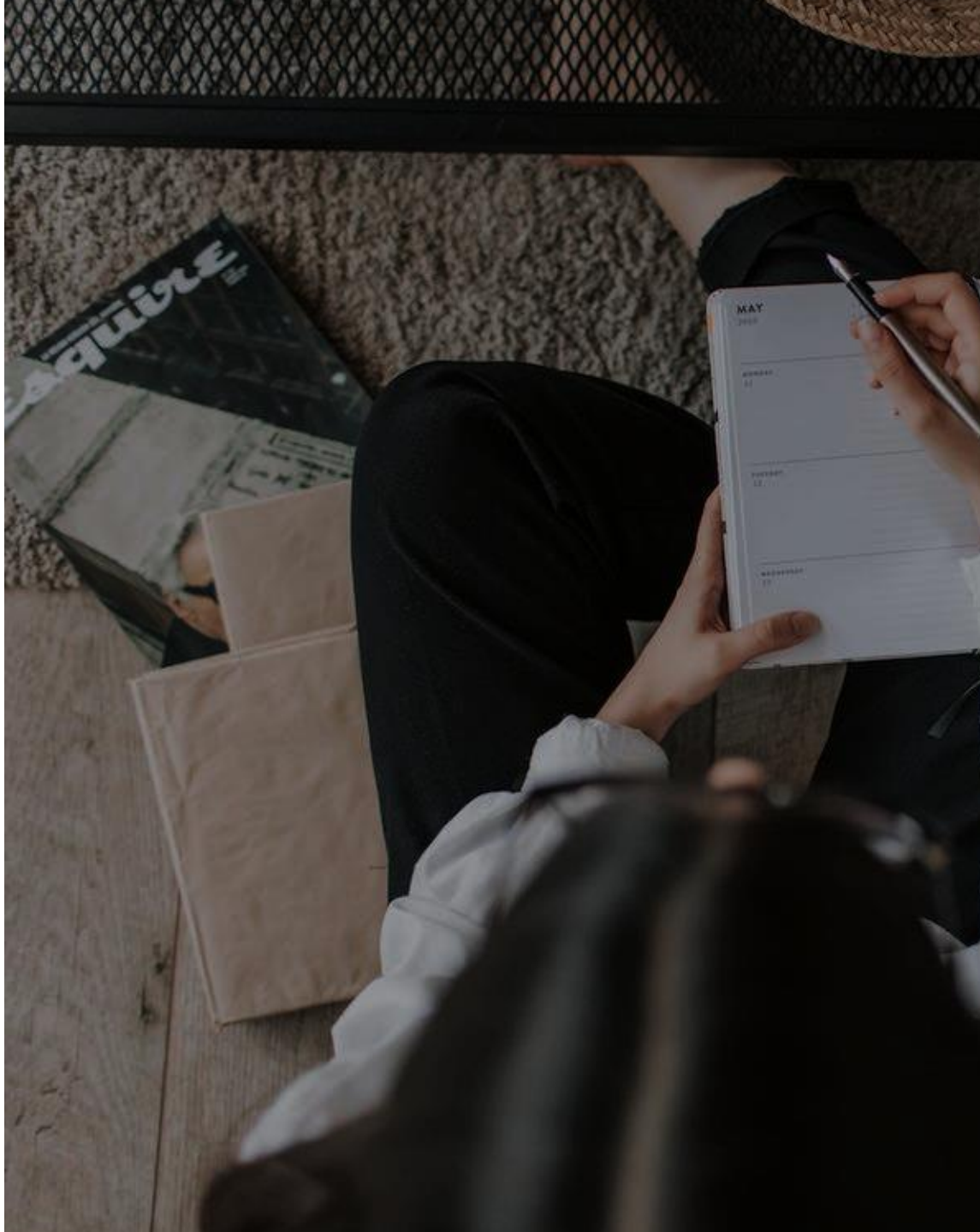
FORMULACIÓN - OBJETIVOS

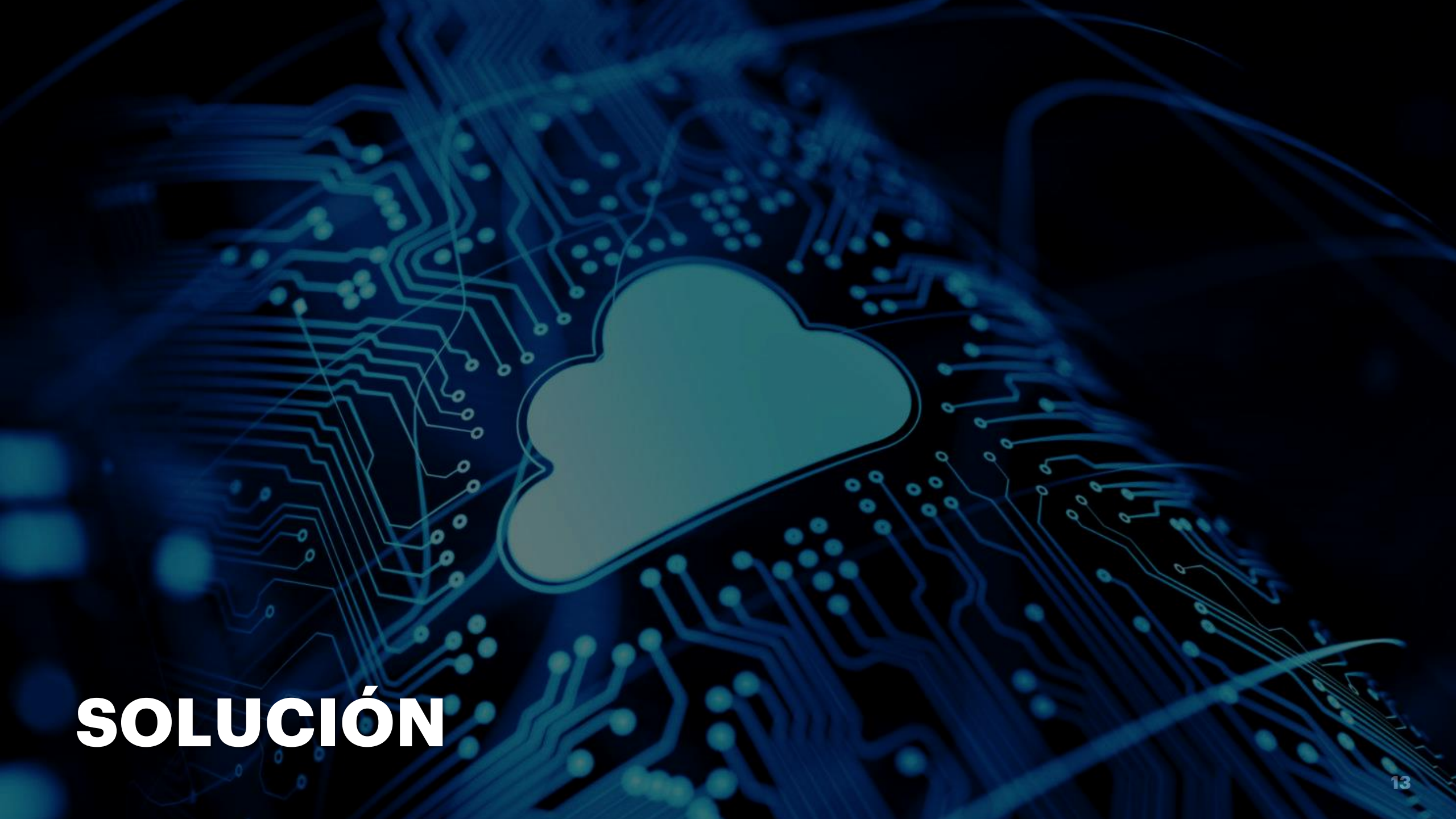
Disminuir el tiempo y los procesos manuales involucrados en la implementación de la estrategia de control de acceso con Lake Formation Tags (**LFTags**) en la nube AWS para una importante plataforma financiera del país.

1. Mapear el proceso de automatización y las herramientas o requisitos para su ejecución con IaC.

2. Crear los recursos y pipeline para la correcta ejecución del proceso mapeado.

3. Corroborar la efectividad de la automatización sugerida.





SOLUCIÓN

SOLUCIÓN – MAPEO DEL PROCESO INICIAL

¿Cómo funciona manualmente el proceso en la consola de AWS?

1 Definir etiquetas

Crear una ontología de atributos o etiquetas LF, y decida a quién puede asignar los recursos de catálogo de datos.

2 Asignar etiquetas

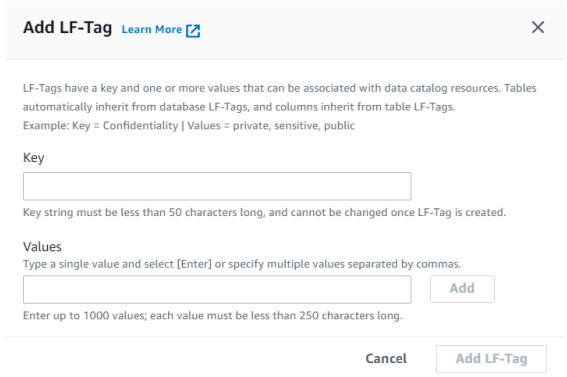
Combinaciones asociadas de LF-Etiquetas (clave y valor) específicas a bases de datos, tablas y columnas.

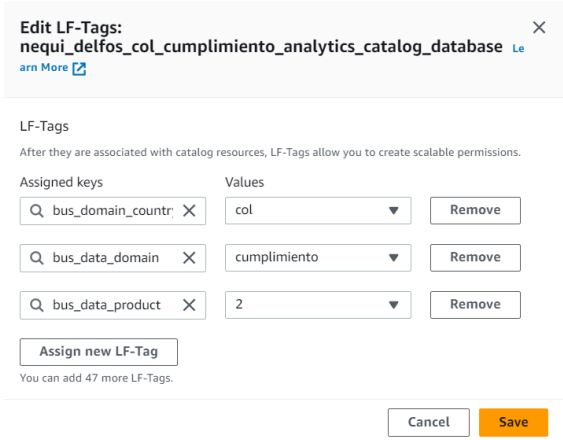
3 Conceder el acceso

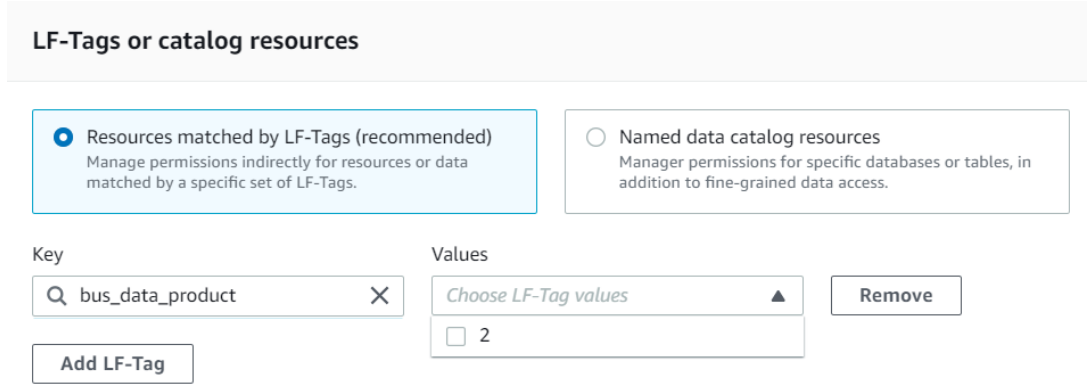
Definir permisos escalables que otorgan acceso al catálogo de recursos a través de combinaciones de etiquetas LF específicas.

4 Retirar el acceso

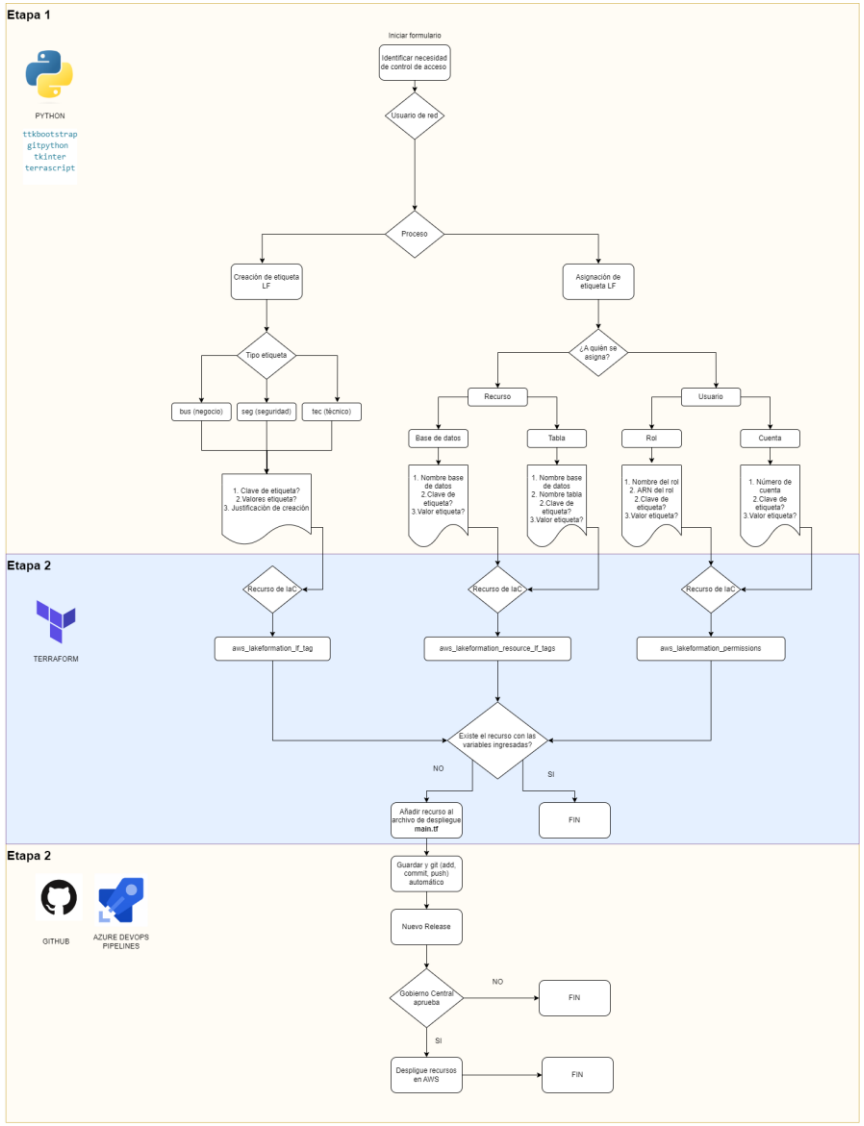
Revocar el acceso directo a los recursos que son reemplazados por LF-Tag.







SOLUCIÓN – MAPEO DEL PROCESO NUEVO



Se ha estructurado la automatización del proceso en 3 etapas.

Etapa1: Creación del formulario, en el cual el encargado del proceso (Data Steward) ingresará las variables requeridas según sea necesario. Las variables a tener en cuenta son:

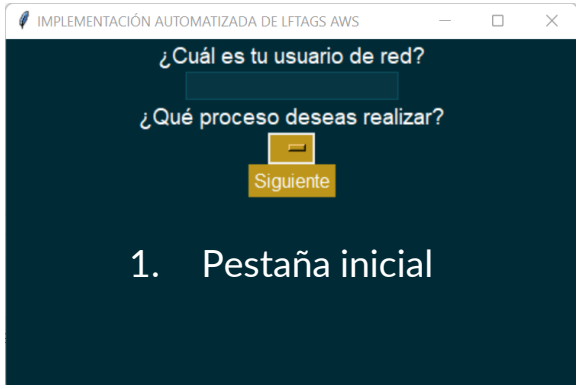
Variable	Tipo
Nombre base de datos	str
Nombre tabla	str
Nombre persona/rol	str
ARN persona/rol	str
# Cuenta AWS	int
Valor/clave etiquetas	str

Etapa2: En esta etapa se genera el script de IaC de lo solicitado en la etapa 1 con el formulario.

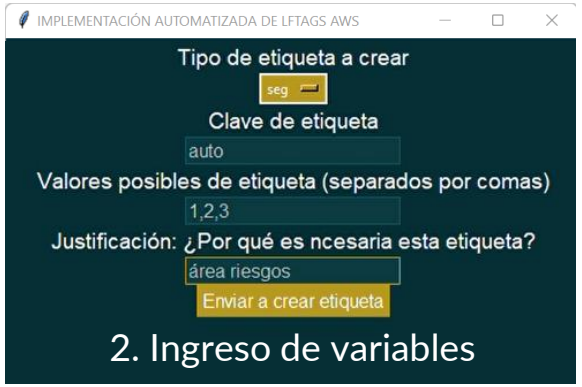
Etapa3: Se guarda el script solicitado añadido en la etapa 2. Se conecta el repositorio con el pipeline de despliegue en la cuenta AWS y se termina la automatización del proceso.

SOLUCIÓN – MAPEO DEL PROCESO ETAPA 1

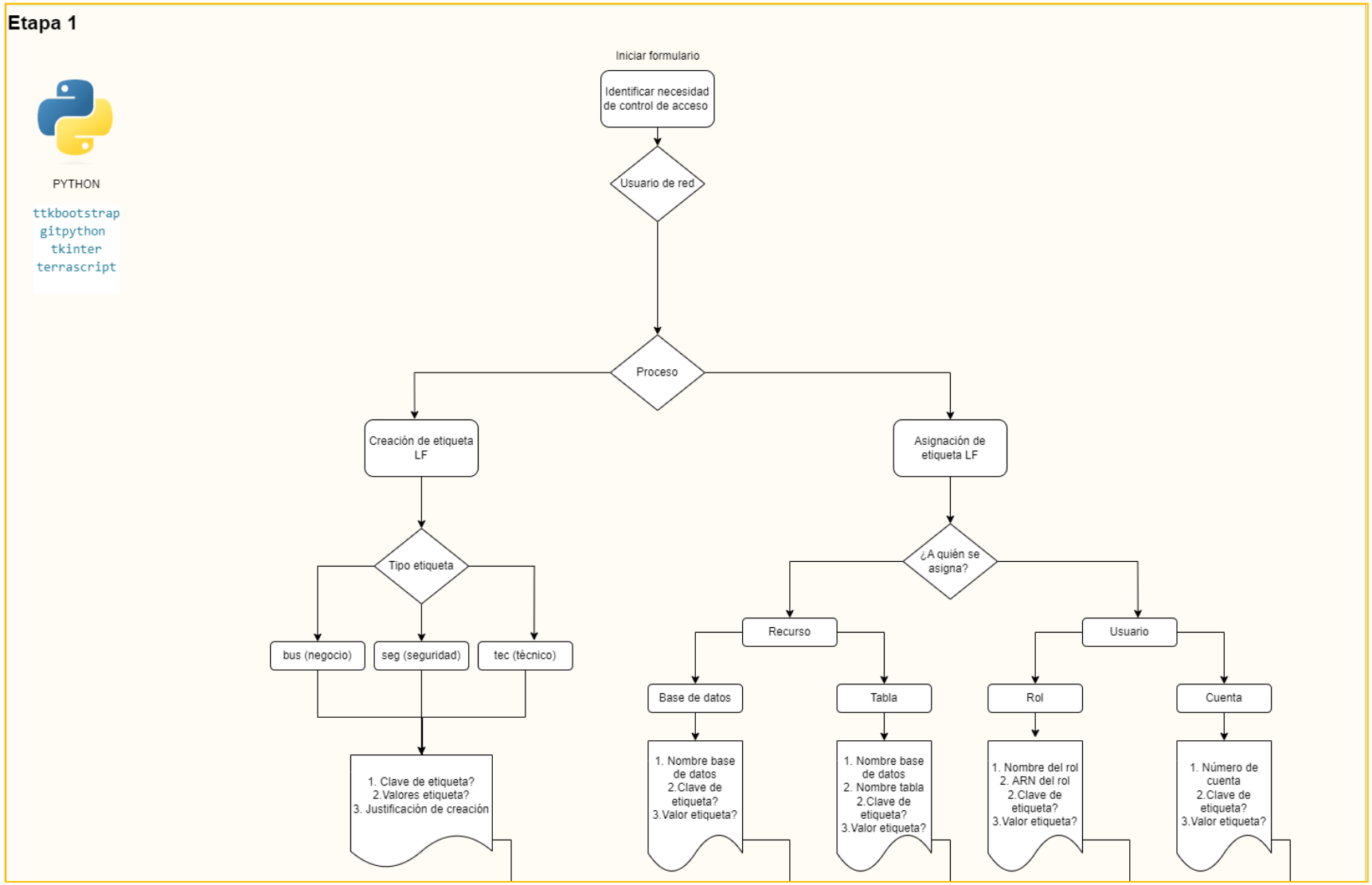
Para la creación del formulario se ha utilizado Python. A continuación se muestra un ejemplo de la visualización para la creación de una etiqueta LF.



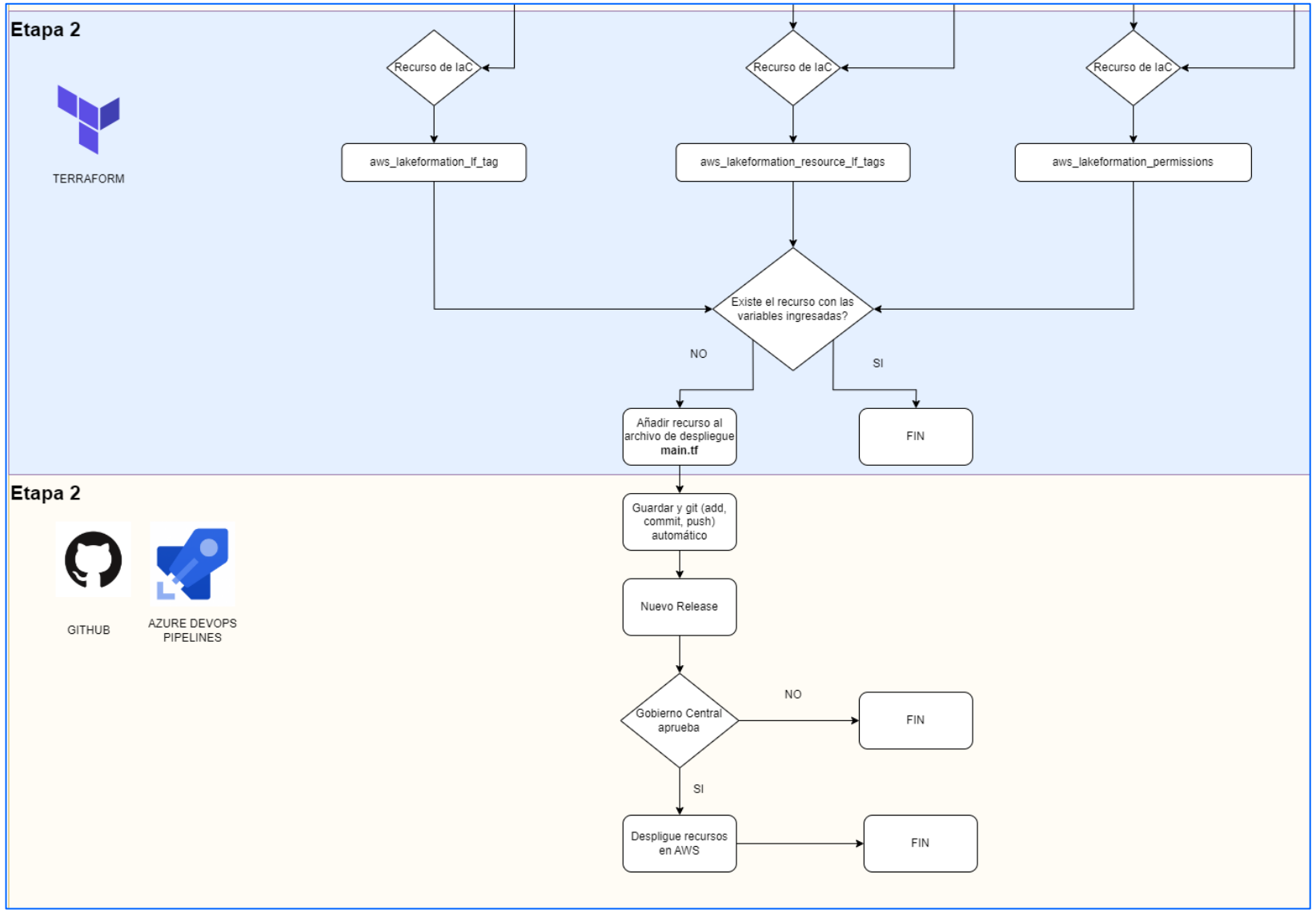
1. Pestaña inicial



2. Ingreso de variables



SOLUCIÓN – MAPEO DEL PROCESO ETAPA 2-3



Se almacenan las variables obtenidas en el formulario en un archivo “respuestas.py”. Este es leído y de acuerdo al proceso de etiquetas a realizar se genera el propio script de IaC con Terraform en el archivo “mainLFTag.tf”.

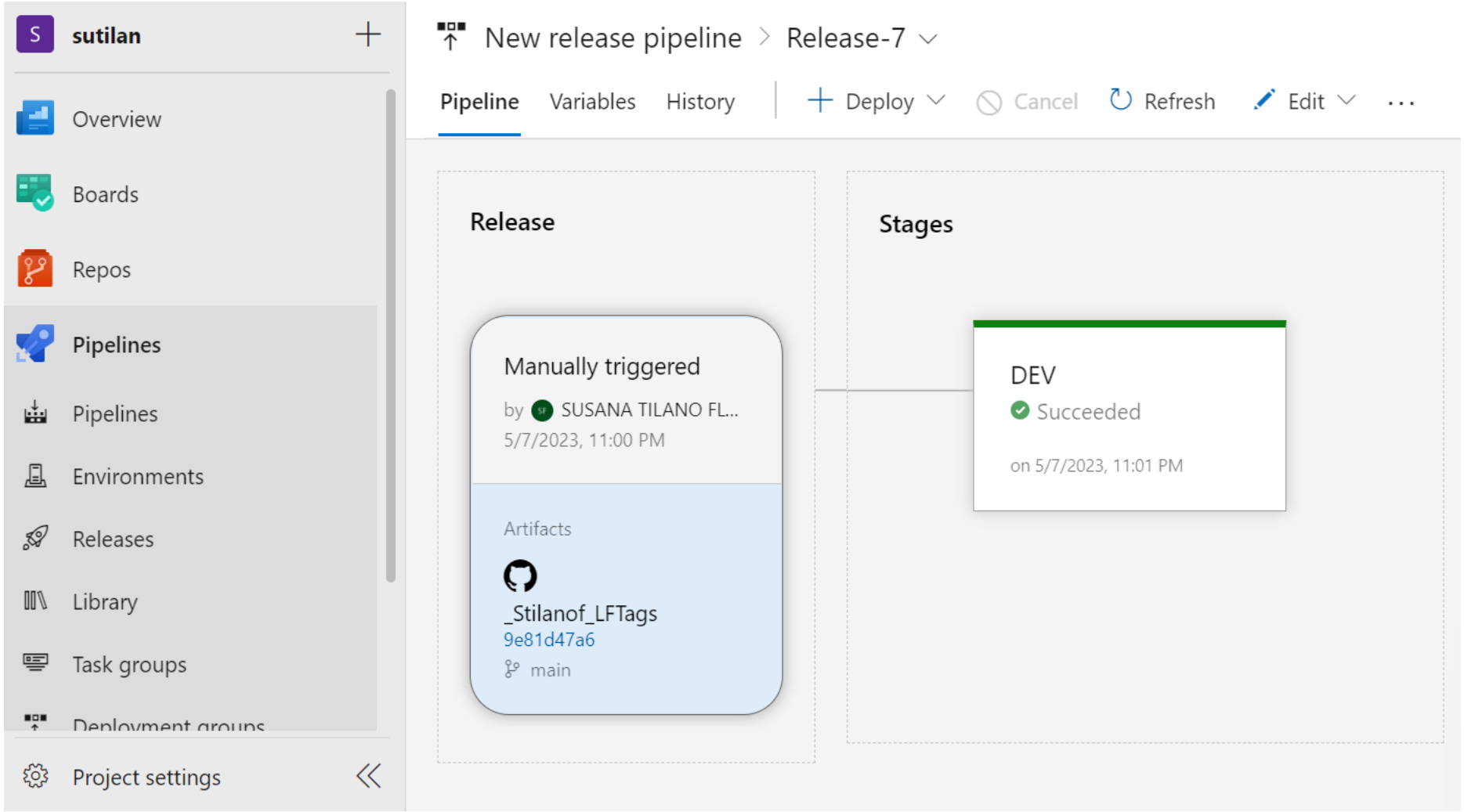
Se cuenta con un repositorio en Github para el proyecto con todos los archivos mencionados, el cual se actualiza al llenar el formulario.

Ver repositorio aquí (https://github.com/Stilanof/LFTags)

Para hacer la ejecución del archivo “mainLFTag.tf” se realiza la conexión con Azure Devops por medio su herramienta de pipelines. El despliegue se activa al recibir un push del repositorio y debe ser aprobado por el revisor.

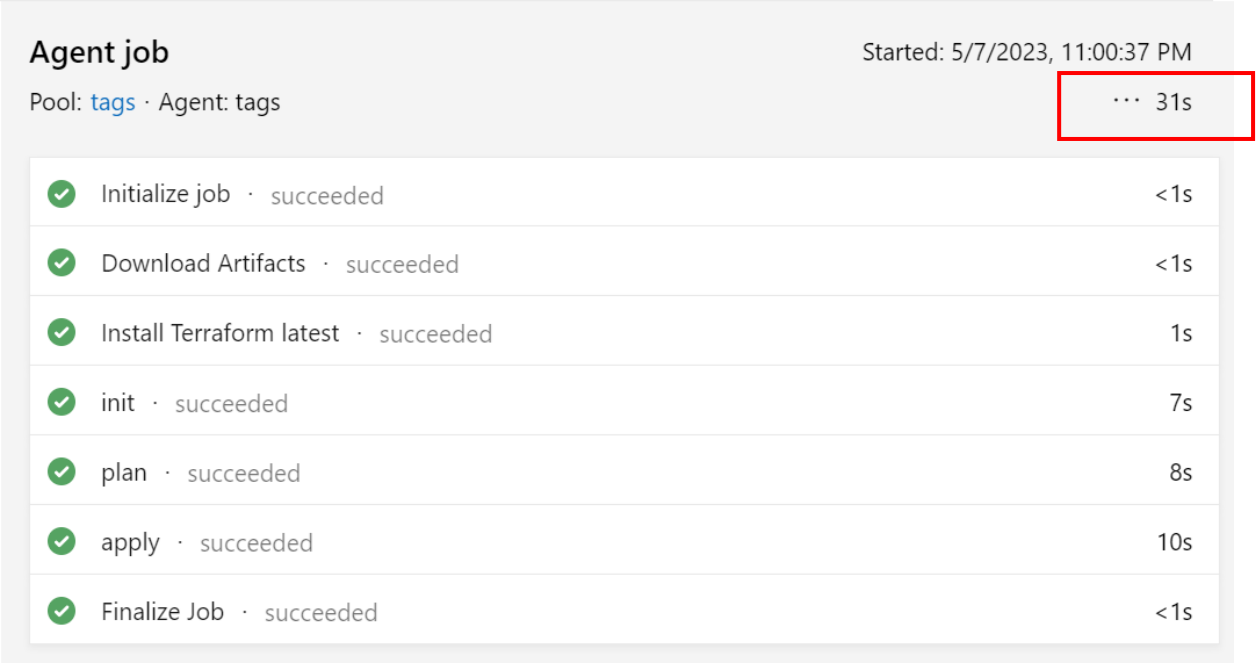
SOLUCIÓN – DESPLIEGUE ETAPA 3

En la consola de Azure Pipelines Releases se puede ver gráficamente la conexión con el repositorio GitHub y el despliegue exitoso.



SOLUCIÓN – DESPLIEGUE ETAPA 3

Si se ingresa al stage DEV se lista las tareas realizadas en el despliegue. Podemos ver que el tiempo de ejecución (31s) reduciéndose significativamente al proceso manual.



En la consola AWS es posible comprobar el despliegue, siguiendo el ejemplo de creación de etiqueta se corrobora que la etiqueta seg_auto efectivamente se creó de manera correcta.

SOLUCIÓN – CONCLUSIONES

La automatización de la IaC puede proporcionar beneficios como:

1. **Consistencia:** Garantiza que la infraestructura se configure de manera consistente y de acuerdo a reglas de negocio en todos los entornos, lo que reduce los errores y aumenta la eficiencia.

2. **Velocidad:** Ayuda a reducir el tiempo necesario para crear y configurar la infraestructura, lo que permite una entrega más rápida los recursos y servicios. Mayor escalabilidad

3. **Versionado:** Facilita el control de versiones de la infraestructura para la fácil reversión de cambios, rastrear y auditar los cambios.

4. **Ahorro de costos:** Reduce los costos asociados con la administración manual de la infraestructura y minimiza los errores que podrían resultar en gastos innecesarios.

REFERENCIAS BIBLIOGRÁFICAS

1. *¿Cómo será la seguridad en el Cloud en este 2022? Descúbrelo aquí*. (n.d.). RedesZone. Retrieved May 8, 2023, from https://www.redeszone.net/noticias/redes/tendencias-nube seguridad-2022/
2. *¿Qué es la infraestructura como código - Infrastructure as Code?* (n.d.). Www.redhat.com. Retrieved May 8, 2023, from https://www.redhat.com/es/topics/automation/what-is-infrastructure-as-code-iac
3. Atlassian. (n.d.). *Comparación entre la arquitectura monolítica y la arquitectura de microservicios*. Atlassian. Retrieved May 8, 2023, from https://www.atlassian.com/es/microservices/microservices-architecture/microservices-vs-monolith
4. *Beneficios de la computación en la nube | IBM*. (n.d.). Www.ibm.com. Retrieved May 8, 2023, from https://www.ibm.com/es-es/topics/cloud-computing-benefits
5. *Terraform by HashiCorp*. (n.d.). Terraform by HashiCorp. https://www.terraform.io
6. *¿Qué es una malla de datos? - Explicación de la malla de datos - AWS.* (n.d.). Amazon Web Services, Inc. Retrieved April 18, 2023, from https://aws.amazon.com/es/what-is/data-mesh/?nc1=h_ls
7. Alvarez, J. (2019). *LAS NECESIDADES DE LA SEGURIDAD EN LA NUBE.* http://repository.unipiloto.edu.co/bitstream/handle/20.500.12277/5978/00005161.pdf?sequence=1&isAllowed=y
8. *What is Cloud Computing Infrastructure? | VMware Glossary*. (2022, March 16). VMware. https://www.vmware.com/es/topics/glossary/content/cloud-computing-infrastructure.html