
Herramienta Analítica para Detección de Anomalías HADA

Manual de Usuario

Introducción

¡Bienvenido(a)! 🎉 Este manual te guiará en el uso de la Herramienta Analítica para Detección de Anomalías **HADA** desarrollado para Contugas. El objetivo de esta herramienta es ayudarte a identificar de manera temprana comportamientos inusuales en el consumo de gas de los clientes industriales, optimizando la operación y reduciendo riesgos. ☀️



¿Qué es y que hace el sistema?

Este artefacto analítico es un sistema avanzado diseñado para monitorear continuamente el consumo de gas natural de los clientes industriales de Contugas. Utilizando técnicas de aprendizaje automático 🤖 no supervisado, el sistema analiza las variables operacionales clave –presión (bar) ⚙, temperatura (°C) 🔥 y volumen (m³) 📊– registradas cada hora para identificar patrones que se desvían del comportamiento normal 🚨.

Funciones Principales:

- Análisis Continuo: Procesa datos históricos (desde 2019) y datos que llegan en tiempo casi real desde los medidores.
- Detección de Anomalías: Identifica automáticamente puntos o secuencias de datos que son estadísticamente inusuales o atípicos para cada cliente o segmento.
- Clasificación de Criticidad: Asigna un nivel de criticidad (Baja, Media, Alta) a las anomalías detectadas, ayudando a priorizar la atención y respuesta.
- Visualización Centralizada: Presenta toda la información relevante (anomalías, tendencias, estado actual) en un tablero interactivo (dashboard) conectado a la base de datos.

base de datos analítica (AWS Athena), facilitando el monitoreo y la toma de decisiones.

El corazón 🌟 de este prototipo es el modelo **Random Cut Forest (RCF)** 🌳, un algoritmo robusto y eficiente, especialmente adecuado para detectar **anomalías** 📈 en datos de series temporales multivariadas como los de **Contugas**, sin necesidad de tener un historial previo de fallas etiquetadas. El modelo opera sobre la plataforma **AWS (Amazon Web Services)** utilizando **Amazon SageMaker**.

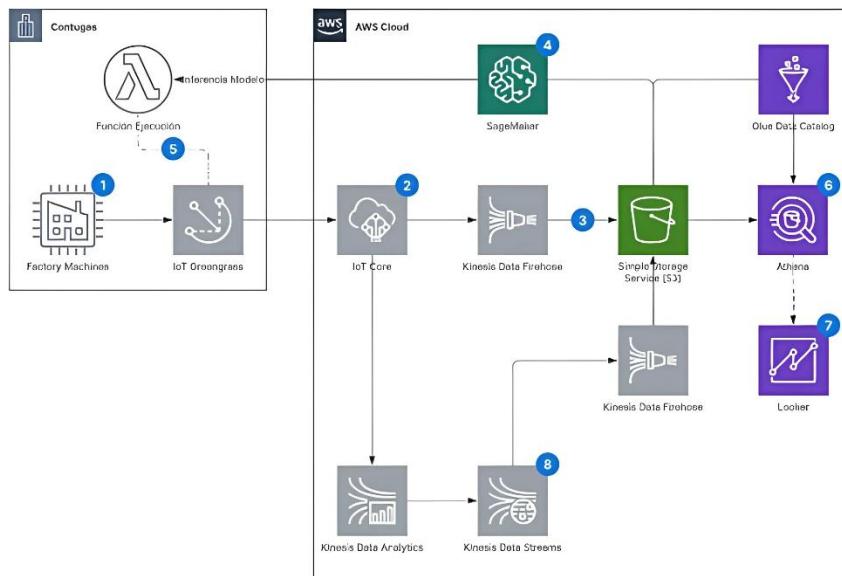


¿Cómo funciona el sistema?

El sistema opera mediante un flujo de trabajo automatizado eficiente y robusto, diseñado para transformar los datos crudos de los sensores en información accionable para la detección de anomalías. 🎛 Este proceso inicia con la **ingesta de datos** 🚲 horarios (presión, temperatura, volumen), seguido de un riguroso **preprocesamiento** ✂ que incluye imputación de faltantes, manejo de duplicados y estandarización. Posteriormente, los datos preparados alimentan el modelo **Random Cut Forest (RCF)** 🌳 en SageMaker, el cual asigna un score de anomalía. Estos scores son luego **clasificados por criticidad** ⚡ y almacenados en Athena para su **visualización** 📊 en el dashboard y la potencial generación de **alertas** 🚨.

1. Entrada de Datos: Los registros horarios de presión, temperatura y volumen por cada cliente industrial son ingeridos (idealmente vía AWS IoT Core y Kinesis Data Streams/Firehose, como se ve en el diagrama de backend).
2. Almacenamiento Inicial: Los datos crudos o semi-procesados se almacenan en un repositorio centralizado (Data Lake en Amazon S3).
3. Preprocesamiento: Antes del análisis, los datos pasan por una serie de transformaciones clave para asegurar su calidad y adecuación al modelo:
 - Imputación de Faltantes: Se rellenan valores ausentes (detectados como Sensor_Error) usando interpolación lineal para mantener la continuidad temporal.
 - Eliminación/Marcado de Duplicados: Se eliminan duplicados exactos y se marcan registros con lecturas múltiples en la misma hora (Doble_Lectura).

- Estandarización (Z-score): Las variables (Presión, Temperatura, Volumen) se escalan por segmento/cliente para que tengan media cero y desviación estándar uno, evitando que variables con magnitudes mayores dominen el análisis.
 - Shingling (Ventanas Temporales): Se crean vectores de características agrupando datos de ventanas consecutivas (24 horas) para capturar patrones diarios y anomalías contextuales.
4. Detección con RCF: Los vectores preprocesados son enviados al modelo RCF desplegado en Amazon SageMaker. El modelo asigna un "score de anomalía" a cada punto/ventana, indicando qué tan atípico es.
 5. Clasificación de Criticidad: Los scores de anomalía se comparan con umbrales estadísticos (ej. basados en la regla de 3 sigmas sobre datos históricos) para clasificar cada anomalía y/o alerta detectada como Baja, Media o Alta criticidad.
 6. Almacenamiento Analítico: Los resultados (datos originales + score + criticidad) se cargan en un Data Warehouse optimizado para consultas (AWS Athena).
 7. Visualización y Alerta: El dashboard (desarrollado en Power BI) se conecta a Athena para mostrar la información actualizada. Se pueden configurar alertas (vía correo, Slack, etc.) para notificar anomalías críticas.
 8. Validación Operativa: El sistema permite que los operadores revisen las anomalías detectadas y proporcionen retroalimentación (confirmar/descartar), la cual se almacena para futuras mejoras del modelo.





Instrucciones de uso

Una vez que ingreses al dashboard , encontrarás un entorno interactivo diseñado para simplificar la detección de anomalías. Podrás aplicar filtros por fecha, cliente y criticidad para personalizar tu vista. El dashboard te presentará gráficas de series temporales donde verás claramente resaltadas las desviaciones. También dispondrás de una tabla detallada para cada anomalía, con opciones para validarla o descartarla , y la facilidad de exportar tus hallazgos.

Requisitos del Usuario:

- Acceso autorizado a la plataforma de visualización Power BI
- Comprensión básica de las variables operativas (presión, temperatura, volumen) y su comportamiento esperado.
- Habilidad para interpretar gráficos de series temporales y dashboards interactivos.
- **NO** se requiere experiencia previa en machine learning o programación.

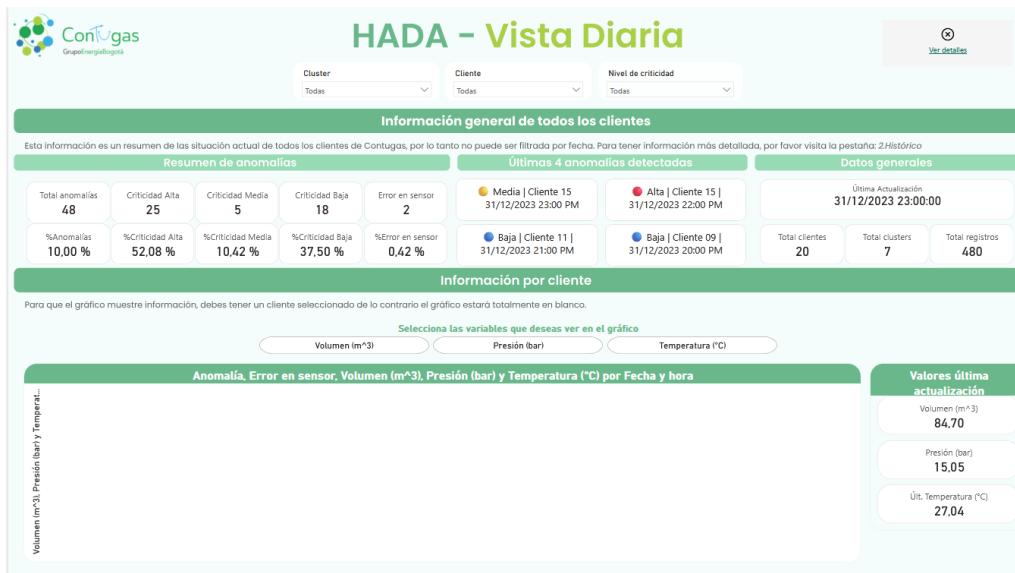
Flujo de Uso del Dashboard:

Al ingresar al [DASHBOARD](#) con tus credenciales , la página de **Inicio (Home)** te dará la bienvenida, ofreciéndote acceso directo a las diferentes secciones del 'Tablero' o a la 'Documentación' . Desde aquí, podrás navegar a:



The screenshot shows the 'HADA' (Herramienta Analítica para Detección de Anomalías) dashboard. At the top left is the Contegas logo. The main title 'HADA' is centered above the subtitle 'Herramienta Analítica para Detección de Anomalías'. Below the title, a message says 'Para dar inicio, selecciona el [enlace](#) de tu interés:'. There are two main sections: 'Tablero' and 'Documentación'. The 'Tablero' section contains three green buttons labeled '1. Diario', '2. Histórico', and '3. Mapa de calor'. The 'Documentación' section contains two green buttons labeled '1. Guía rápida de uso' and '2. Manual de uso'. At the bottom of the page, a footer note reads 'Para reportar inconvenientes, comunícate al correo electrónico: soporte@hada.com'.

Vista Diaria : Aquí encontrarás un resumen rápido del estado actual. Verás las últimas anomalías detectadas, un conteo general y los valores más recientes de las variables. Podrás usar filtros  por Cliente, Cluster y Nivel de Criticidad para enfocar tu revisión del día.



HADA - Vista Diaria

Información general de todos los clientes

Resumen de anomalías

Total anomalías	Criticidad Alta	Criticidad Media	Criticidad Baja	Error en sensor
48	25	5	18	2
%Anomalías	%Criticidad Alta	%Criticidad Media	%Criticidad Baja	%Error en sensor
10,00 %	52,08 %	10,42 %	37,50 %	0,42 %

Últimas 4 anomalías detectadas

Categoría	Detalle	Fecha
Media Cliente 15	Media Cliente 15	31/12/2023 23:00 PM
Baja Cliente 15	Baja Cliente 15	31/12/2023 22:00 PM
Baja Cliente 11	Baja Cliente 11	31/12/2023 21:00 PM
Baja Cliente 09	Baja Cliente 09	31/12/2023 20:00 PM

Datos generales

Total clientes	Total clusters	Total registros
20	7	480

Última Actualización: 31/12/2023 23:00:00

Información por cliente

Para que el gráfico muestre información, debes tener un cliente seleccionado de lo contrario el gráfico estará totalmente en blanco.

Seleccióna las variables que deseas ver en el gráfico

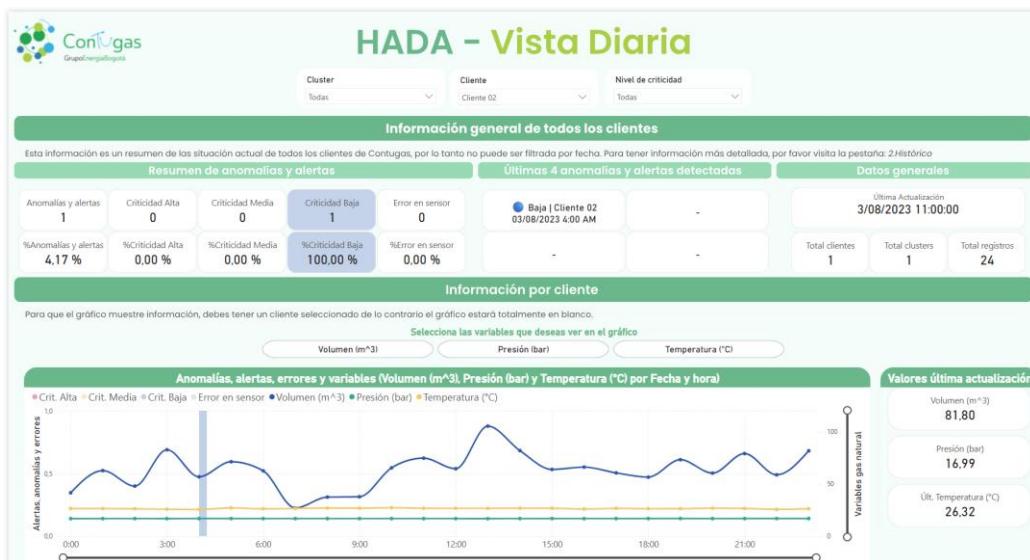
Volumen (m³) Presión (bar) Temperatura (°C)

Anomalía, Error en sensor, Volumen (m³), Presión (bar) y Temperatura (°C) por Fecha y hora

Valores última actualización

Volumen (m ³)	Presión (bar)	Ult. Temperatura (°C)
84,70	15,05	27,04

Para ver el gráfico dinámico debes seleccionar un cliente específico en el filtro superior:



HADA - Vista Diaria

Información general de todos los clientes

Resumen de anomalías y alertas

Anomalías y alertas	Criticidad Alta	Criticidad Media	Criticidad Baja	Error en sensor
1	0	0	1	0
%Anomalías y alertas	%Criticidad Alta	%Criticidad Media	%Criticidad Baja	%Error en sensor
4,17 %	0,00 %	0,00 %	100,00 %	0,00 %

Últimas 4 anomalías y alertas detectadas

Categoría	Detalle	Fecha
Baja Cliente 02	Baja Cliente 02	03/08/2023 6:00 AM

Datos generales

Total clientes	Total clusters	Total registros
1	1	24

Última Actualización: 3/08/2023 11:00:00

Información por cliente

Para que el gráfico muestre información, debes tener un cliente seleccionado de lo contrario el gráfico estará totalmente en blanco.

Seleccióna las variables que deseas ver en el gráfico

Volumen (m³) Presión (bar) Temperatura (°C)

Anomalías, alertas, errores y variables (Volumen (m³), Presión (bar) y Temperatura (°C) por Fecha y hora)

Variables: Crit. Alta, Crit. Media, Crit. Baja, Error en sensor, Volumen (m³), Presión (bar), Temperatura (°C)

Gráfico: Alertas, anomalías y errores vs. Tiempo (0:00 a 21:00)

Valores últimas actualizaciones

Volumen (m ³)	Presión (bar)	Ult. Temperatura (°C)
81,80	16,99	26,32

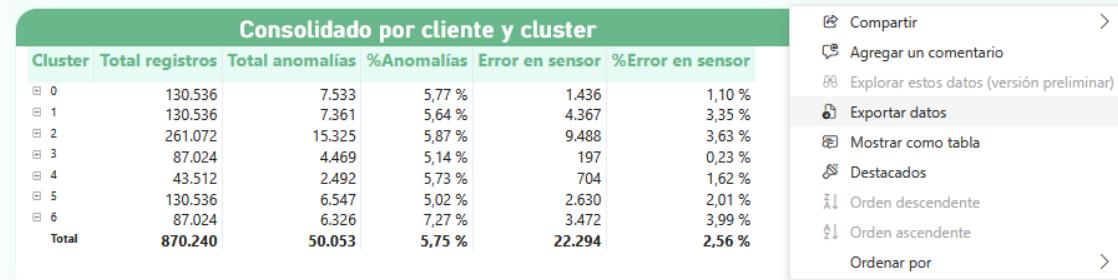
Vista Histórica ✎: Esta sección te permite explorar en detalle los datos pasados. Dispondrás de filtros más amplios (rangos de fecha, día de la semana, cliente, cluster, criticidad, etc.) para analizar las tendencias de Presión, Temperatura y Volumen a lo largo del tiempo en gráficos interactivos, así como tablas consolidadas de anomalías por cliente y cluster. Aquí también encontrarás una tabla con el **Detalle** 📊 de cada registro individual.



Mapa de Calor 🔥: Esta vista te ayudará a identificar patrones de anomalías según la hora del día. Podrás filtrar los datos y seleccionar qué variable quieras ver en las filas (Cliente, Cluster o Día de la semana) para descubrir en qué momentos se concentran más las desviaciones, ya sea en número total o porcentaje.



Descarga de Resultados 📁: Si necesitas compartir o analizar los datos fuera del dashboard, busca la opción para exportar la tabla de detalles o un resumen en formato Excel (CSV).



The dashboard displays a table with the following data:

Cluster	Total registros	Total anomalías	%Anomalías	Error en sensor	%Error en sensor
0	130.536	7.533	5,77 %	1.436	1,10 %
1	130.536	7.361	5,64 %	4.367	3,35 %
2	261.072	15.325	5,87 %	9.488	3,63 %
3	87.024	4.469	5,14 %	197	0,23 %
4	43.512	2.492	5,73 %	704	1,62 %
5	130.536	6.547	5,02 %	2.630	2,01 %
6	87.024	6.326	7,27 %	3.472	3,99 %
Total	870.240	50.053	5,75 %	22.294	2,56 %

A context menu is open on the right side of the table, listing options such as Compartir (Share), Agregar un comentario (Add comment), Explorar estos datos (versión preliminar) (Explore these data (preliminary version)), Exportar datos (Export data), Mostrar como tabla (Show as table), Destacados (Highlights), Orden descendente (Descending order), Orden ascendente (Ascending order), and Ordenar por (Sort by).



Ventajas, limitaciones y advertencias

Al usar este sistema, te beneficiarás de una detección temprana de problemas sin necesidad de datos etiquetados, y de un dashboard amigable 🖥️. Sin embargo, ten en cuenta que el sistema identifica anomalías, pero no su causa raíz automáticamente, y su precisión óptima se alcanza con tu validación continua ✅. Es crucial asegurar la calidad de los datos 💧 y recordar que las alertas críticas siempre deben ser revisadas por personal experto.

Ventajas

- Detección Temprana: Permite identificar problemas potenciales antes de que se conviertan en emergencias costosas.
- No Supervisado: Funciona sin necesidad de datos históricos etiquetados de fallas (que suelen ser escasos).
- Adaptabilidad: El modelo RCF puede adaptarse a los patrones cambiantes y específicos de cada cliente o segmento.
- Contexto Temporal: El uso de "shingling" permite detectar anomalías basadas en el patrón diario completo, no solo en puntos aislados.
- Escalabilidad: La arquitectura basada en AWS permite manejar grandes volúmenes de datos y escalar según las necesidades.
- Visualización Intuitiva: El dashboard está diseñado para ser fácilmente comprensible por el personal operativo.

Limitaciones

- Explicabilidad Limitada: Si bien RCF detecta anomalías, no explica automáticamente la causa raíz del problema (requiere análisis humano o modelos adicionales).
- Sensibilidad a Cambios: Cambios operativos drásticos y legítimos (ej. inicio/fin de un nuevo proceso industrial en un cliente) podrían ser marcados inicialmente como anomalías hasta que el modelo se adapte.
- Dependencia de Umbrales: La clasificación de criticidad depende de umbrales estadísticos que pueden requerir ajustes periódicos.
- Mejora con Feedback: La precisión óptima se alcanza cuando los operadores validan continuamente las alertas, ayudando a refinar los umbrales y el modelo.

Advertencias

- Calidad de Datos: La precisión del sistema depende fundamentalmente de la calidad, completitud y continuidad de los datos de entrada de los sensores. Errores o faltantes prolongados pueden afectar la detección.
- Necesidad de Revisión: El sistema es una herramienta de apoyo a la decisión. Las anomalías, especialmente las críticas, siempre deben ser revisadas por personal calificado antes de tomar acciones operativas mayores.
- Reentrenamiento Periódico: Aunque RCF se adapta, si se observan cambios estructurales significativos y persistentes en el comportamiento del consumo, puede ser necesario revisar o reentrenar partes del modelo para mantener su relevancia.



Casos de uso y ejemplos

Para que veas el sistema en acción, aquí te presentamos escenarios comunes. Podrás realizar tu revisión operativa diaria 🕒, investigando anomalías críticas 💡 que el sistema te notifique. También te facilitará generar informes mensuales 📈 sobre los clientes con mayor incidencia o validar el comportamiento de un medidor después de un mantenimiento 🔧, todo desde una interfaz centralizada.

Caso de Uso / Descripción	Acción Esperada desde el Dashboard
Identificación de Patrones Horarios de Anomalías	Un analista quiere saber si las anomalías en un grupo de clientes (o un cluster específico) se concentran en ciertas horas del día, lo que podría indicar problemas recurrentes en turnos específicos o con equipos que operan en esos horarios.
Ánalisis Comparativo del Desempeño Histórico entre Clientes	Se necesita comparar cómo ha evolucionado el volumen de consumo y las anomalías de dos clientes industriales clave durante el último trimestre para entender si un cambio operativo en uno afectó su estabilidad.
Verificación Rápida del Estado General de la Red (Vista Diaria)	Al iniciar el turno, un operador necesita una visión general y rápida de la situación actual de todos los clientes para priorizar tareas.
Evaluación de la Confiabilidad del Modelo Post-Ajuste	Después de realizar ajustes en los umbrales o parámetros del modelo RCF, el equipo técnico quiere verificar si la precisión general y la tasa de falsos positivos han mejorado.
Análisis Detallado de un "Error en Sensor" Específico	La "Vista Diaria" muestra un "Error en sensor". El operador necesita investigar qué cliente y qué lecturas específicas están involucradas para coordinar una revisión del medidor.



Detalles útiles

Para que aproveches al máximo el sistema, considera estos puntos:

- 🔒 El dashboard es ágil y tu acceso es seguro 🔑.
- 🌐 Tus búsquedas específicas son privadas.
- 🌐 Accede desde cualquier dispositivo con navegador.
- ⏱ La información se actualiza constantemente para darte datos frescos.
- 📁 Exporta fácilmente tus hallazgos para informes.
- ✅ Tu validación de alertas mejora la precisión del sistema.
- 🔎 Usa los filtros para análisis rápidos y enfocados.
- 🚦 Los colores te guían visualmente sobre la criticidad de las anomalías.
- 🤖 No necesitas ser experto en IA para interpretar los resultados.
- ❓ Y si tienes dudas o necesitas ayuda, el equipo de soporte está para asistirte.



Preguntas frecuentes (FAQ)

Sabemos que pueden surgir dudas al utilizar una nueva herramienta. Por eso, hemos recopilado aquí algunas de las consultas más comunes 🤷‍♂️ con sus respectivas respuestas, para que puedas sacar el máximo provecho del sistema de detección de anomalías desde el primer momento y resolver rápidamente cualquier inquietud que tengas. ¡Explora y aclara tus dudas!

❓ ¿Dónde puedo comunicarme si tengo algún inconveniente?

Puedes comunicarte con alguno de los miembros del equipo responsable del desarrollo de la herramienta, enviando un correo a:

- ✉️ soporte@hada.com
- ✉️ j.padillas@uniandes.edu.co
- ✉️ ek.diaz@uniandes.edu.co
- ✉️ a.quinoneso@uniandes.edu.co
- ✉️ m.gonzalezc@uniandes.edu.co

❓ ¿Con qué frecuencia se actualizan los datos y las alertas en el dashboard? 🕒

R: El sistema está diseñado para procesar datos y generar alertas en tiempo casi real. La visualización en el dashboard se actualizará con una alta frecuencia (cada hora según la configuración de ingesta de datos desde los sensores) para reflejar el estado operativo más reciente.

❓ ¿Qué debo hacer si encuentro una anomalía marcada como "Alta Criticidad"? 🚨

Una anomalía de "Alta Criticidad" requiere atención prioritaria. Debes verificarla inmediatamente en el dashboard, analizar las variables asociadas (Presión, Temperatura, Volumen) para entender el contexto, y si la consideras válida, escalar la situación al equipo técnico u operativo correspondiente según los protocolos establecidos por Contugas para este tipo de eventos.



¿ Si descarto una anomalía como "Falso Positivo", ¿el sistema aprenderá de mi feedback?

¡Sí! Tu validación es muy importante. Aunque este prototipo inicial con RCF es no supervisado, la retroalimentación que proporcionas al marcar anomalías como "Válidas" o "Descartadas" se registra. Esta información es crucial y se utilizará en futuras iteraciones para reajustar umbrales, refinar el modelo o incluso entrenar modelos supervisados que mejoren la precisión y reduzcan los falsos positivos con el tiempo.

¿ ¿Puedo ver el historial de anomalías de un cliente específico de semanas o meses anteriores? 

¡Sí! El dashboard te permite filtrar los datos por rangos de fecha. Simplemente selecciona el cliente que te interesa y ajusta las fechas en los filtros para explorar su comportamiento y las anomalías detectadas en períodos anteriores, facilitando análisis retrospectivos o la identificación de patrones recurrentes.

⚙️ Manual Técnico ⚙️

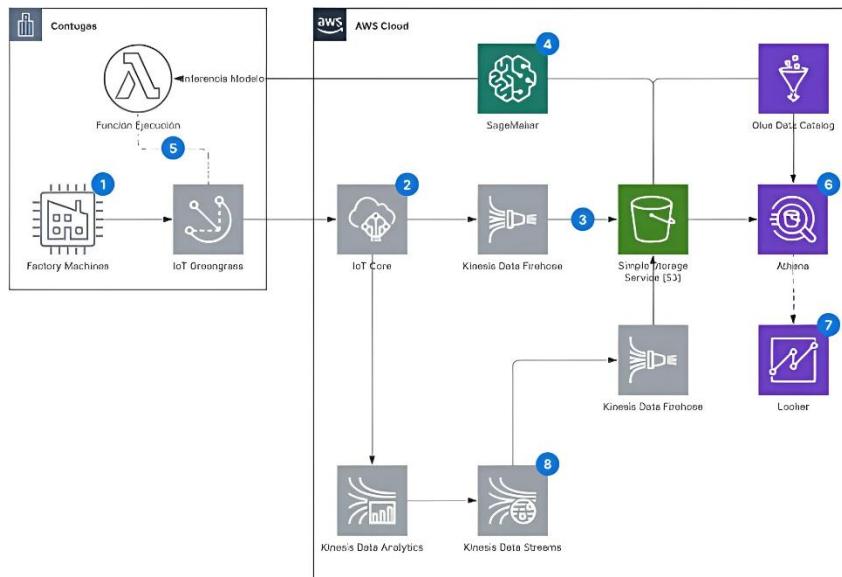
Implementación y Arquitectura del Sistema HADA

Este documento técnico describe la arquitectura de backend, el flujo de datos, la configuración del modelo de detección de anomalías y las interfaces de consumo para el sistema **HADA** (Herramienta Analítica para Detección de Anomalías) de Contugas. Está dirigido a personal técnico encargado del mantenimiento, despliegue y evolución de la solución.



Arquitectura del Modelo y Flujo de Datos en AWS

El sistema **HADA** se ha diseñado sobre la plataforma Amazon Web Services (AWS) para garantizar escalabilidad, robustez y eficiencia en el procesamiento de datos de series temporales y la ejecución de modelos de machine learning. El flujo principal es el siguiente.



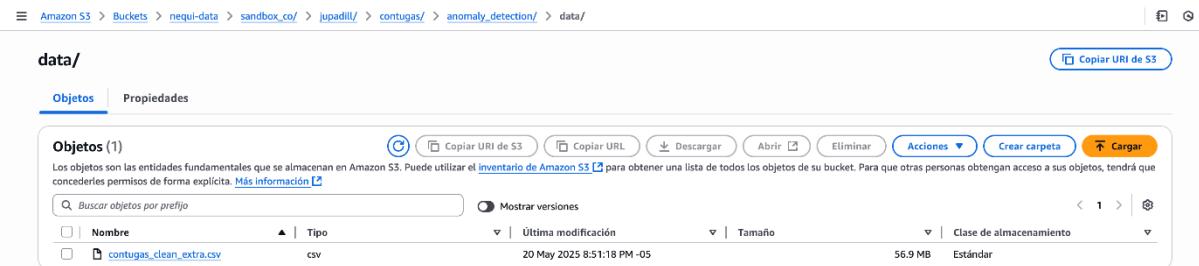
Ingesta de Datos desde Sensores (LoT) 🏠 (#1, #2):

- Los datos de Presión, Temperatura y Volumen, generados horariamente por los medidores en las Factory Machines, son idealmente capturados por **AWS IoT Core**

- **AWS IoT Core** actúa como el punto de entrada seguro y escalable para los datos de los dispositivos.

Procesamiento y Almacenamiento en el Data Lake (#2, #3):

- Desde IoT Core, los datos fluyen hacia **Amazon Kinesis Data Firehose**. Kinesis Firehose se encarga de la transformación básica y la carga confiable de los datos en el Data Lake.
- El Data Lake está implementado en **Amazon Simple Storage Service (S3)** (#3). Los datos se almacenan en formatos optimizados para análisis y entrenamiento de modelos.



The screenshot shows the AWS S3 console interface. The URL in the address bar is: [Amazon S3 > Buckets > nequi-data > sandbox_co/ > jupadill/ > contugas/ > anomaly_detection/ > data/](#). The page title is "data/". There is a "Copiar URI de S3" button. Below it, there are tabs for "Objetos" and "Propiedades", with "Objetos" selected. A table lists one object: "contugas_clean_extra.csv". The table columns include Nombre, Tipo, Última modificación, Tamaño, and Clase de almacenamiento. The object details show: Nombre: contugas_clean_extra.csv, Tipo: csv, Última modificación: 20 May 2025 8:51:18 PM -05, Tamaño: 56.9 MB, Clase de almacenamiento: Estándar. There are also buttons for Copiar URI de S3, Descargar, Abrir, Eliminar, Crear carpeta, and Cargar.

Catalogación de Datos (#3, #6):

- Una vez en S3, los metadatos de los conjuntos de datos son rastreados y catalogados por **AWS Glue Data Catalog** (#6). Esto permite que los datos sean fácilmente descubribles y accesibles por otros servicios de AWS, incluyendo Athena y SageMaker.

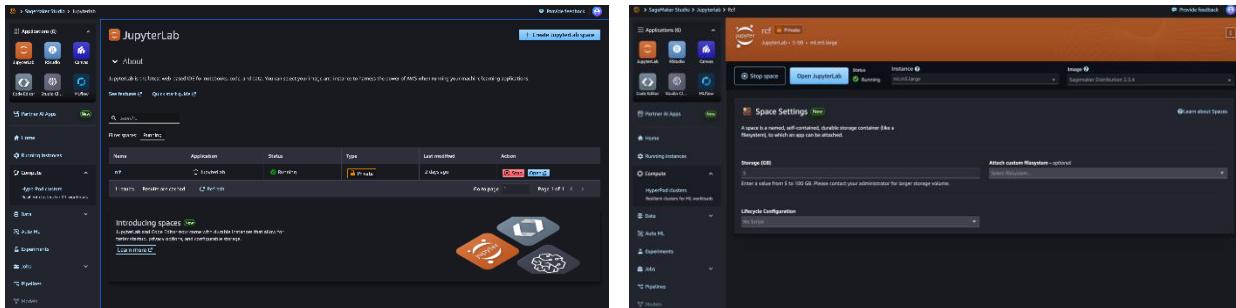


The screenshot shows the AWS Glue Data Catalog interface. The main section is titled "Datos". It includes fields for "Origen de datos" (set to "AwsDataCatalog"), "Catálogo" (set to "Ningún elemento"), and "Base de datos" (set to "analytics_co"). Below this is a "Tablas y vistas" section with a "Crear" button and a table showing "Tablas (178)" and "Vistas (0)". There is also a search bar for "Filtrar tablas y vistas".

Entrenamiento y Despliegue del Modelo de Detección de Anomalías (RCF) 🧠 (#4):

- **Amazon SageMaker (#4)** es el servicio central para el desarrollo, entrenamiento y despliegue del modelo **Random Cut Forest (RCF)** 🌳.

- Los datos preprocesados (limpios, imputados, escalados) desde S3 se utilizan para entrenar el modelo **RCF** 🌲.
- En **SageMaker Studio** se gestionan los entornos de desarrollo (JupyterLab).



- Una vez entrenado, el modelo **RCF** 🌲 se despliega como un **endpoint de SageMaker** para inferencia en tiempo real o por lotes.

Inferencia y Almacenamiento de Resultados 🔎:

- Los nuevos datos (o datos históricos para reprocesamiento) son enviados al endpoint de **RCF** 🌲. El modelo retorna un "score de anomalía" para cada registro.
- Estos scores, junto con los datos originales y la clasificación de criticidad (derivada de los scores), se almacenan nuevamente en S3.

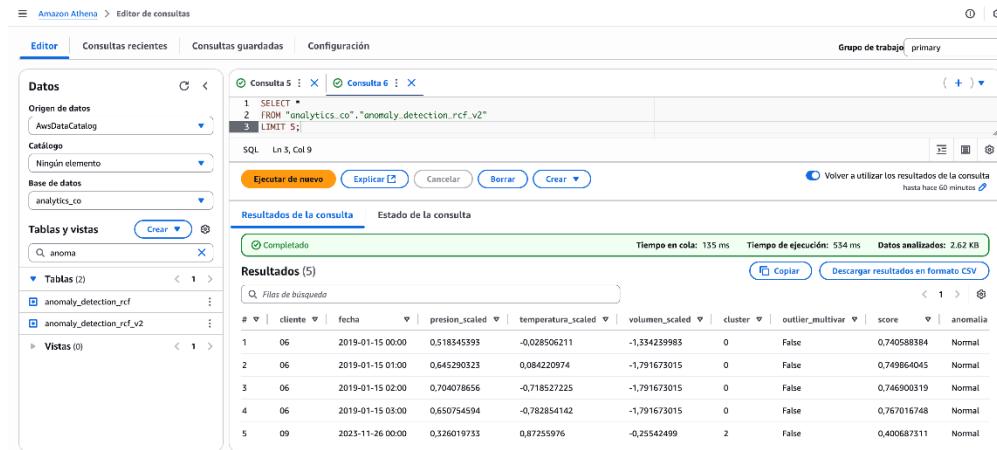


Conexión y Consulta de Datos con Power BI vía Amazon Athena

El dashboard de visualización **HADA** consume los datos procesados y los resultados de las anomalías directamente desde el Data Lake a través de **Amazon Athena (#6)**.

Amazon Athena como Motor de Consultas SQL:

- Athena es un servicio de consultas interactivas que facilita el análisis de datos en Amazon S3 utilizando SQL estándar.
- Utiliza el **AWS Glue Data Catalog** para obtener los esquemas de las tablas que residen en S3.



The screenshot shows the Amazon Athena Editor interface. On the left, the 'Datos' sidebar lists the 'Origen de datos' as 'AwsDataCatalog', 'Catálogo' as 'Ningún elemento', 'Base de datos' as 'analytics_co', and 'Tablas y vistas' containing 'anoma' and 'anomaly_detection_rcf' (with 'anomaly_detection_rcf_v2' listed under it). The main area displays 'Consulta 6' with the following SQL code:

```

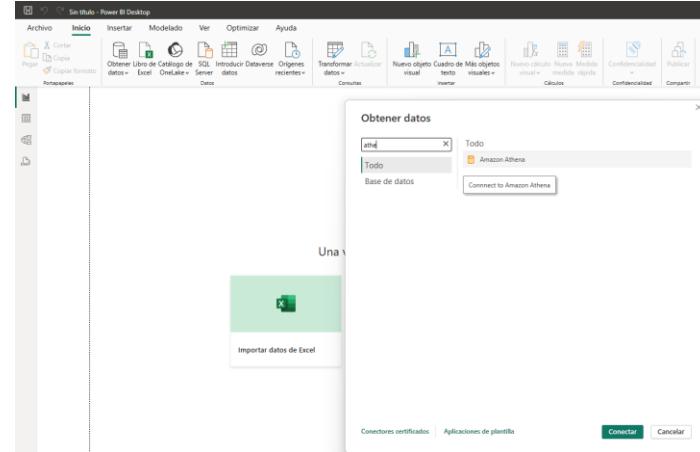
1 SELECT *
2 FROM "analytics_co"."anomaly_detection_rcf_v2"
3 LIMIT 5;
    
```

Below the code, there are buttons for 'Ejecutar de nuevo', 'Explicar', 'Cancelar', 'Borrar', and 'Crear'. A status bar at the bottom indicates: 'Tiempo en cola: 135 ms', 'Tiempo de ejecución: 534 ms', and 'Datos analizados: 2.62 KB'. The 'Resultados de la consulta' tab shows the results of the executed query, which include columns: '#', 'cliente', 'fecha', 'presion_scaled', 'temperatura_scaled', 'volumen_scaled', 'cluster', 'outlier_multivar', 'score', and 'anomalia'. The results show five rows of data.

- Los resultados de la consulta muestran columnas como cliente, fecha, presion_scaled, volumen_scaled, cluster, outlier_multivar, score y anomalia.

Conexión de Power BI a Athena:

- Power BI se conecta a Amazon Athena utilizando su conector nativo o un conector ODBC/JDBC.
- Se configura el origen de datos en Power BI especificando la región de AWS, el catálogo de datos, la base de datos y el grupo de trabajo de Athena.



- Una vez establecida la conexión, se pueden importar las tablas (o vistas creadas en Athena para optimizar) directamente en Power BI o usar DirectQuery para consultar los datos en tiempo real.

-
- Las visualizaciones del dashboard HADA (Vista Diaria, Histórica, Mapa de Calor, Desempeño) se construyen utilizando las tablas y campos expuestos por Athena.



Disposición de Archivos y Artefactos en GitHub

Para facilitar la colaboración, el versionamiento y la reproducibilidad del proyecto, todos los artefactos de código y configuración relevantes se gestionan a través de un repositorio en [GitHub](#).

Clonación y Uso:

- El repositorio (privado o compartido según las políticas de Contugas) puede ser clonado localmente o en una instancia de SageMaker Studio.
- `git clone https://github.com/ek-diaz/MIAD_final_project.git`
- Siga las instrucciones del README.md para configurar el entorno, instalar dependencias y ejecutar los scripts o cuadernos.

Versionamiento:

Se recomienda seguir prácticas estándar de control de versiones (commits frecuentes, ramas para nuevas funcionalidades, pull requests para revisiones) para mantener un historial claro de los cambios y facilitar la colaboración.