

ARQUITECTURA DE PRODUCTOS DE DATOS

P R O Y E C T O F I N A L

22 de mayo de 2023

Arquitectura de la solución



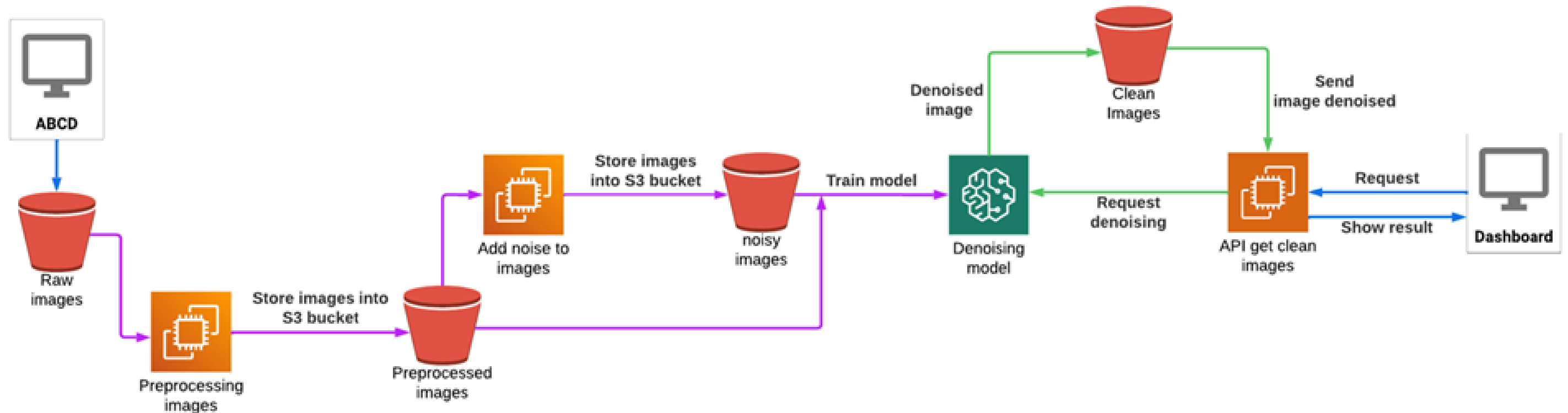
Alberto
Fuentes
Chavarría

Brandon
Hernández
Troncoso

Christopher
Chávez
Jiménez

David
Damián
Arbeu

Denoising Ray-x



La imagen anterior representa la arquitectura del producto de datos propuesto, en la cual se pueden identificar todas las fases y elementos que conforman la solución desarrollada. Esta arquitectura busca cumplir las necesidades y objetivos planteados, además de esclarecer cada proceso del producto de forma independiente sin dejar de lado su papel dentro del producto completo. Adicionalmente, en la imagen se puede observar que para cada proceso se tiene asociada una herramienta de AWS, lo que favorece el entendimiento del papel que tiene cada proceso en la solución completa. A continuación se detalla cada uno de los elementos mostrados:

ABCD Solutions:

- En primer lugar, el equipo de ABCD solutions es el encargado de **recolectar y almacenar las imágenes de radiografías** a utilizar.
- En seguida, con ayuda de **Amazon S3**, se almacenan las imágenes recolectadas para su debido tratamiento y procesamiento.
- Para el procesamiento de las imágenes es necesario la ejecución de un conjunto de códigos, sin embargo, también se requiere de la disponibilidad de equipos de cómputo con ciertas características para soportar todas las acciones realizadas a las imágenes. Por lo tanto, el procesamiento debe ser realizado con ayuda de instancias **Amazon EC2** que cumplan con la memoria y espacio en disco suficiente para soportar dichas ejecuciones.
- El resultado del procesamiento de las imágenes debe ser almacenado nuevamente con ayuda de **Amazon S3**.
- Como el objetivo del proyecto es mejorar la calidad de radiografías, es necesario simular cuál es el ruido esperado, por lo que se requiere nuevamente de la ejecución de otro conjunto de códigos. Para llevar a cabo este procedimiento, se puede aprovechar la ventaja de la eficiencia computacional de la instancia **Amazon EC2** creada anteriormente para la ejecución de dichos códigos.
- El resultado de la simulación de las imágenes con ruido debe ser almacenado nuevamente con ayuda de **Amazon S3**.
- Una vez que ya se cuenta con los conjuntos de imágenes con ruido y sin ruido, es necesario el desarrollo de un modelo de aprendizaje automático para la reconstrucción de las imágenes sin ruido. En este proceso resulta conveniente utilizar el elemento **Amazon Sagemaker** para entrenar y desplegar el modelo.

Usuario final:

- Una vez que el modelo entrenado se encuentra disponible para su uso, resulta pertinente especificar la arquitectura en la que estará involucrada el usuario. En este caso, la interacción del usuario con la solución se lleva a cabo mediante un dashboard que muestra el resultado obtenido. Este dashboard está desarrollado en una **Shiny App**.
- Una vez que el usuario realiza una petición para mejorar la calidad de la imagen, esta petición se procesa con la ayuda de una API que recibe peticiones on-demand, es decir que está disponible para su uso las 24 horas. Para soportar esta acción, se requiere nuevamente de una instancia de Amazon EC2 que contenga la solución en ejecución constante.
- La API es la encargada de conectarse directamente con el modelo entrenado, por lo que la petición de imagen se envía al **punto de conexión del modelo en Sagemaker** para generar el resultado de la imagen reconstruida.
- El resultado de la reconstrucción de la imagen debe ser almacenado nuevamente con ayuda de **Amazon S3**.
- Una vez que se almacenó el resultado, la API es la encargada de obtenerlo para visualizarlo en el dashboard como resultado final.

