

Tecnológico de Monterrey
Proyecto Integrador (TC5035)
Avance 2: Casos de Uso



Daniel Guzmán Ávila A00781387
Gabriel Alejandro Amezcua Baltazar A01795173
Luis Miguel Farfán Lara A01360268

Proyecto Integrador – Avance 2

Objetivo

En la entrega anterior se realizó un análisis exploratorio limitado de las imágenes disponibles en un dataset de estantería y se generaron algunas transformaciones limitadas.

En la entrega actual se definió dejar por escrito los casos de uso de la solución para la cual se desarrollará el proyecto integrador.

Antecedentes

En el sector retail, asegurar el cumplimiento de planogramas (distribución planificada de productos en estanterías) es crucial para maximizar ventas y mantener la imagen de marca. (PlanoHero, s.f.) Los impulsores (merchandisers) se encargan de surtir y organizar productos de alta rotación (bebidas, lácteos, etc.) siguiendo planogramas predefinidos en supermercados y tiendas. Existe una relación directa entre la rentabilidad de las categorías y el cumplimiento de un planograma que asegure la gestión eficiente de inventario (Arenas, 2024)

Propuesta de solución

Este informe describe cómo una solución de inteligencia artificial generativa basada en el modelo multimodal **IBM WatsonX Granite Vision que puede** automatizar la verificación de estas estanterías mediante análisis de imágenes. Esta propuesta abarca los casos de uso dependiendo de los roles en el flujo de trabajo y una propuesta técnica de arquitectura que describa de forma general el funcionamiento de dicha solución.

IBM WatsonX Granite Vision es un modelo fundacional imagen a texto especializado en visión por computadora. Se trata de un modelo multimodal (imagen + texto) capaz de ingerir imágenes junto con texto y producir salidas en lenguaje natural (IBM, s.f.). El modelo Vision disponible en Watson X es un modelo fundacional que fue construido específicamente para el análisis de fotografías, documentos, gráficos, diagramas etc. Actualmente el modelo con arquitectura decoder se encuentra en la versión 3.2.2, con un tamaño de 2 mil millones de parámetros y una ventana de contexto global de 131,072 token (entrada y salida). El modelo está entrenado con textos en Inglés, alemán, español, francés, japonés, portugués, árabe, checo, italiano, coreano, holandés y chino (IBM, s.f.)

IBM (s.f.) indica que sus modelos multimodales pueden realizar “image-to-text generation”, útil para “visual question answering” o chatear con imágenes subidas

(responder preguntas sobre una imagen), interpretación de gráficos, entre otras funciones. Esto significa que podemos preguntarle al modelo, por ejemplo: “¿Esta imagen de la góndola cumple con el planograma esperado?” y obtener una respuesta fundamentada.

Casos de Uso

A continuación se describen los casos de uso para una solución en WatsonX que apoye a la gestión de punto de venta con base en el análisis de imágenes. Existen cuatro casos de uso principales:

1. **Gestión del punto de venta por parte del impulsador (merchandiser):** Donde el impulsador toma la fotografía inicial del punto de venta, acomoda los artículos de acuerdo con el planograma y toma una fotografía con el trabajo terminado.
2. **Supervisión por parte de un coordinador comercial:** Donde el supervisor recibe y revisa informes de la solución con base en las fotografías recibidas y analizadas y utiliza la información como un insumo para su gestión.
3. **Reportes de cumplimiento realizados por un auditor que verifica el trabajo de los impulsadores.** : Donde el supervisor recibe y analiza informes de un auditor externo que da seguimiento al trabajo de los impulsadores, estableciendo una medida de control de calidad cruzada.
4. **Mantenedor:** Donde el supervisor obtiene del negocio instrucciones de la estrategia comercial e instruye a la solución para tomar en cuenta nuevos productos, promociones temporales etc.

El proceso de negocio “should be” debería ser el siguiente:

1. **Fotografía inicial del estado de la estantería.**
 - **Responsable:** Impulsador.
 - Al llegar al punto de venta, el impulsador utiliza la aplicación móvil nativa para tomar una fotografía inicial de la estantería de productos. Esta imagen inicial documenta el estado antes de cualquier intervención, mostrando cómo se encontraba el exhibidor al momento de la llegada. Los metadatos de la foto (fecha, hora, ubicación GPS si disponible) se capturan automáticamente para su registro. Este paso sirve para tener un punto de comparación y evidencia del antes.

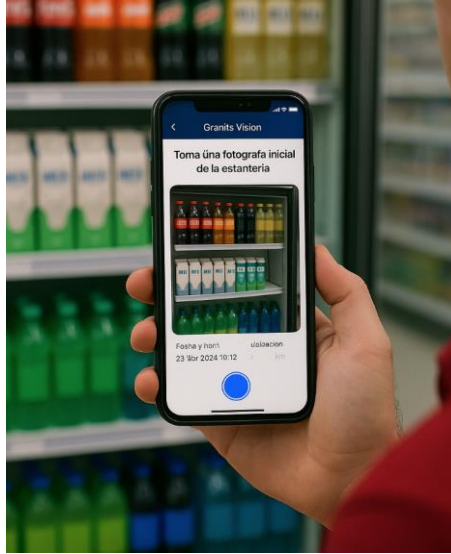


Imagen1. Mockup de la solución caso de uso para el impulsador en el punto de venta. Fuente: Elaboración propia (asistido con 4o).

2. Acomodo de productos según planograma.

- **Responsable:** Impulsador.
- Tras la foto inicial, el impulsador procede a acomodar los productos conforme al planograma establecido por la compañía. El planograma define la colocación óptima de cada producto (por ejemplo, qué marcas van en cada nivel de la repisa, cuántas facings de cada producto, etc.). El impulsador reorganiza o repone mercancía hasta cumplir con esta distribución ideal. La aplicación móvil puede mostrar el planograma de referencia para asistir al impulsador.

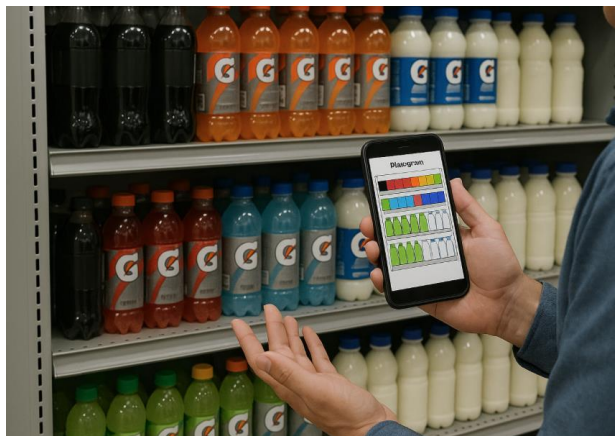


Imagen 2. Mockup de la solución caso de uso para el impulsador en el punto de venta. Fuente: Elaboración propia (asistido con 4o).

3. Fotografía posterior y evaluación automática.

- **Responsable:** Impulsador.
- Completado el acomodo, el impulsador toma una fotografía posterior de la estantería ya organizada. Esta imagen después es enviada al sistema backend para ser evaluada automáticamente contra el planograma esperado. Aquí entra en acción el modelo de IA generativa WatsonX Granite Vision, el cual analizará la imagen para identificar la disposición de los productos, sus posiciones y etiquetas/marcas visibles
- El modelo comparará esta información visual con el planograma objetivo (proporcionado al modelo como contexto o descripciones esperadas) y generará un juicio de valor sobre el acomodo. En otras palabras, Granite Vision emitirá una evaluación en lenguaje natural indicando el nivel de cumplimiento del planograma y la calidad del acomodo.
- Dado que Granite Vision es un modelo texto a partir de imagen, puede describir con detalle la escena y responder preguntas sobre ella. La evaluación generada se almacena junto con la imagen y la metadata para informes posteriores.

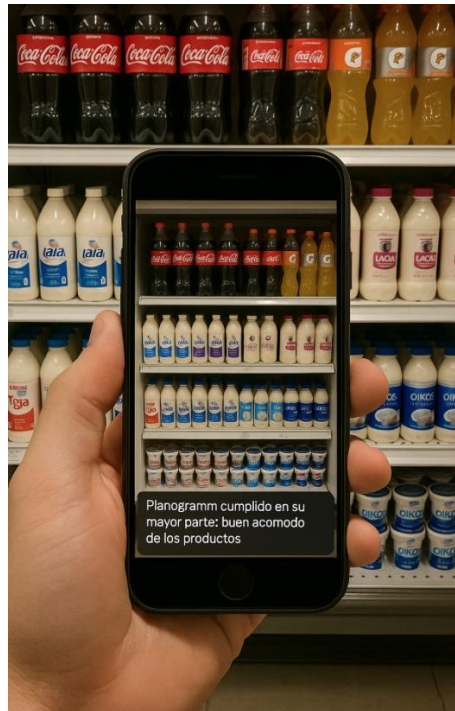


Imagen 3. Mockup de la solución caso de uso para el impulsador en el punto de venta para evaluación del acomodo. Fuente: Elaboración propia (asistido con 4o).

4. Informe diario al supervisor.

- **Responsable:** Supervisor comercial
- Al finalizar la jornada, el supervisor comercial recibe un informe consolidado de todas las visitas realizadas por los impulsores bajo su cargo. Este informe diario incluye:
 - Tiendas visitadas: una lista o tabla de los puntos de venta atendidos en el día, identificados por nombre o código.
 - Fecha y hora: el momento de cada visita, obtenido de los metadatos de las fotografías (timestamp que asegura evidencia temporal de la ejecución) y ubicación del punto de venta.
 - Evaluación de cumplimiento del planograma: para cada visita, se incluye el resultado del análisis de Granite Vision sobre la foto del después. Esto puede presentarse como una calificación (por ejemplo, cumplimiento 95%), un semáforo (verde/amarillo/rojo) y/o un breve comentario generativo del modelo resaltando hallazgos (p.ej., "Estantería ordenada correctamente, solo falta reponer 2 facings de producto X en el estante superior").
 - La información deberá ser puesta a disposición de otras áreas como planificación de ventas (SOP) y marketing para hacer transparente la gestión del punto de venta.

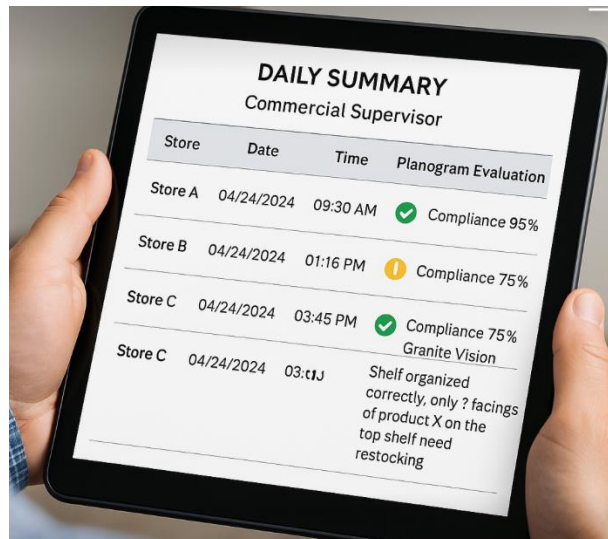


Imagen 4. Mockup de la solución caso de uso para el supervisor comercial. Fuente: Elaboración propia (asistido con 4o)

5. Revisión en reunión comercial.

- **Responsables:** Supervisor comercial e impulsores.

- Los informes generados por el sistema se pueden usar como un insumo más para las reuniones periódicas que realiza el equipo de venta todas las semanas y durante el cierre de mes. De tal forma que el equipo comercial pueda decidir con esa información cambios en políticas de descuento, promociones y administración de categorías.

6. Informes de auditoría.

- **Responsable:** Auditor
- En algunas organizaciones existe un rol de auditoría que visita el punto de venta para asegurar que el impulsador o merchandiser está visitando el punto de venta. Mediante una versión de la misma aplicación móvil el auditor toma fotografías de la estantería en momentos aleatorios, independientemente de las visitas de los impulsadores. Estas imágenes son enviadas al sistema para obtener una evaluación independiente del cumplimiento del planograma en ese momento. Estos datos de auditoría pueden confrontarse con los reportes de los impulsadores, sirviendo como mecanismo de control de calidad cruzado.

7. Mantenedor:

- **Responsable:** Supervisor comercial y canal de venta
- Los planogramas no suelen estar estáticos en el largo plazo. Con frecuencia se introducen o retiran nuevos productos y se habilitan promociones temporales. El objetivo de contar con un mantenedor es proporcionar al modelo instrucciones frescas que demuestren el cambio en las políticas comerciales de la organización al respecto de productos y promociones.



Imagen 5. Mockup de la pantalla del mantenedor. Fuente: Elaboración propia (asistido con 4o)

A continuación, se detalla la arquitectura técnica recomendada para implementar esta solución de punta a punta, integrando la aplicación móvil, la nube de IBM y el modelo WatsonX Granite Vision. La solución integra el patrón recomendado por IBM considerado como mejor práctica para la implementación de soluciones de IA generativa mediante WatsonX (IBM Cloud, s.f.)

1. Aplicación Móvil Nativa (Front-end)

El cliente móvil será una aplicación nativa (iOS/Android) instalada en los dispositivos de los impulsores y auditores. Sus funciones incluyen:

Captura de Imágenes: Interfaz de cámara para tomar fotos de las estanterías. La app puede guiar al usuario para tomar la foto con el ángulo y distancia adecuados, asegurando que toda la góndola sea visible.

Gestión de Sesiones y Datos: Autenticación de usuarios (impulsador, auditor) y selección de tienda o ubicación actual. La app puede obtener la lista de tiendas asignadas al impulsador ese día. Cada foto se etiqueta con la identificación de tienda, ubicación en la tienda (ej. "pasillo lácteos") y hora.

Envío de Datos al Backend: Tras tomar una foto, la aplicación envía la imagen junto con sus metadatos (ID de usuario, tienda, marca temporal, posiblemente coordenadas GPS) al servidor backend mediante una API REST segura. Se puede emplear HTTPS para todas las comunicaciones. Las imágenes podrían comprimirse o redimensionarse en el para optimizar la transferencia, manteniendo suficiente calidad para el análisis.

2. Backend en la Nube (Servidor de Aplicación y Almacenamiento)

El backend actúa como cerebro central coordinando el flujo de información entre la app, el modelo de IA y los sistemas de informes. Este backend residirá en la nube (IBM Cloud) con los siguientes subcomponentes potenciales:

- 1) **API REST:** Como web-service para que la app móvil pueda subir datos como las imágenes y usuarios del aplicativo. De igual forma puede solicitar reportes (en el caso del supervisor comercial)
- 2) **Almacenamiento:** Las imágenes subidas se almacenan en un repositorio centralizado como (IBM CloudStorage) con un indicador único asociado a la visita a esa tienda, estante , fecha, y hora específica.
- 3) **Base de datos:** Una base de datos con información de las imágenes, catálogo de usuarios, tiendas visitables, días y horarios de visita (si aplica) , planogramas esperados y resultados de las evaluaciones.
- 4) **Integración con WatsonX / Granite Vision:** El backend se conecta con el servicio IBM WatsonX.ai donde reside el modelo Granite Vision. Granite Vision puede consumirse como SaaS en IBM Cloud sin necesidad de desplegar infraestructura propia. (IBM watsonx.ai and watsonx.governance. (s. f)

- a. Procesamiento de la respuesta: Una vez que WatsonX Granite Vision devuelve la respuesta (texto generado), el backend la procesa. Esta respuesta contendría la evaluación del acomodo.
 - b. El resultado final (ej. 90% cumplimiento, comentario "Falta producto X") se guarda en la BD ligado a la visita.
- 5) **Generación de informes:** Por último, el backend cuenta con un proceso (por ejemplo, un job diario o bajo demanda) para compilar el informe del supervisor. El informe puede tomar forma de un dashboard web, un PDF enviado por correo, o simplemente una vista en la propia app para el perfil de supervisor.

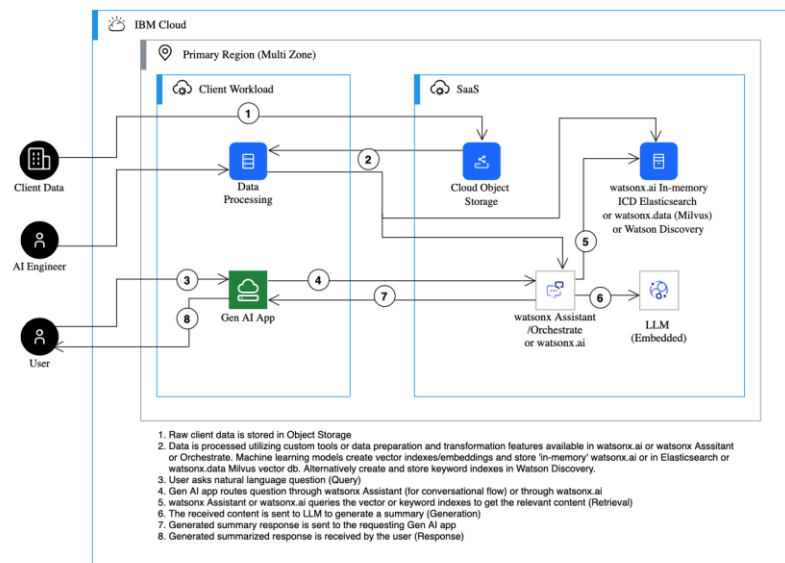


Figura1. Patrón de Implementación para Soluciones con IA de IBM (Fuente: IBM Cloud s.f.)

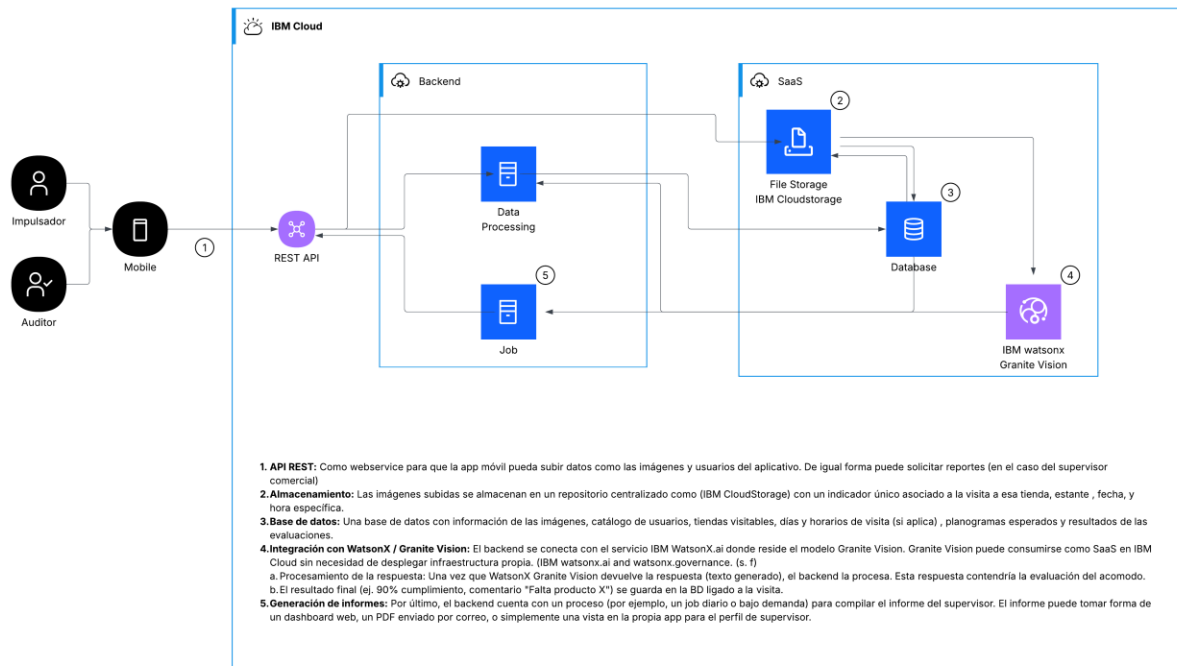


Figura2. Diagrama de arquitectura.

Bibliografía

1. ¿Qué es un planograma? (n.d.). PlanoHero.com.
<https://planohero.com/es/blog/what-is-a-planogram/>
2. Arenas Palacio, M. (2024, diciembre 15). Sigamos Aprendiendo de #Retail, Ahora el Turno es para los "#Planogramas" en los Supermercados y/o Tiendas; Facilitan ó No la Vida al Cliente? [Publicación en LinkedIn]. LinkedIn.
https://www.linkedin.com/posts/mauricioarenas_retail-planogramas-nielseni-activity-7266105693575823360-8ycV
3. IBM Corporation. (s. f.). IBM foundation models | IBM watsonx.
<https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/fm-models-ibm.html?context=wx#:~:text=Granite%20Vision%203>
4. IBM watsonx as a Service. (s. f.).
<https://www.ibm.com/docs/en/watsonx/saas?topic=lab-chatting-documents-images>
5. IBM Cloud Docs (s.f.) Gen AI Pattern for Watsonx on IBM Cloud
<https://cloud.ibm.com/docs/pattern-genai-rag?topic=pattern-genai-rag-genai-pattern#:~:text=AI%20holds%20the%20promise%20to,and%20governing%20Gen%20AI%20applications>

6. IBM watsonx.ai and watsonx.governance. (s. f.). Using the API to work with foundation models <https://www.ibm.com/docs/en/watsonx/w-and-w/1.1.x?topic=solutions-rest-api>