

La Idea Principal: Dos Formas de “Cocinar” los Datos

Imagina que ya tienes tus **Productos de Datos** (los ingredientes de alta calidad) y tu **Domain Data Store (DDS)** (tu cocina profesional). Ahora, ¿qué vas a cocinar?

El capítulo explica que hay dos grandes “recetarios” o formas de generar valor:

1. **Business Intelligence (BI)**: Es el arte de **mirar hacia atrás y hacia el presente** para entender el negocio. Su objetivo es responder a las preguntas “**¿Qué pasó?**” y “**¿Por qué pasó?**”. Es como el panel de control de un coche: te muestra la velocidad, la gasolina que queda y si el motor se está sobrecalentando. Te da la información para que *tú* tomes una decisión.
2. **Analítica Avanzada (MLOps)**: Es el arte de **mirar hacia el futuro** para predecir y automatizar decisiones. Su objetivo es responder a las preguntas “**¿Qué pasará?**” y “**¿Qué deberíamos hacer?**”. Siguiendo la analogía, es el sistema de piloto automático del coche: no solo te muestra la información, sino que **toma el control del volante** para llevarte a tu destino.

Ambos son cruciales, pero resuelven problemas diferentes y tienen complejidades distintas. Ahora, vamos a profundizar en los conceptos que mencionaste.

Business Intelligence (BI) en Detalle

El objetivo del BI es hacer que los datos sean comprensibles y accesibles para que los humanos puedan tomar mejores decisiones. Los dos conceptos que mencionas son las herramientas clave para lograrlo.

1. Capas Semánticas (Semantic Layers)

- **El Problema que Resuelven:** Los datos en un DDS están en tablas con nombres técnicos (`fact_sales`, `dim_customer`) y requieren `SQL` y `JOINS` para ser consultados. Un gerente de marketing no sabe (ni le interesa) hacer eso.
- **¿Qué son?:** Una “**capa de traducción**” que se pone encima de los datos técnicos. Oculta la complejidad y presenta los datos en términos de negocio que todos entienden.
 - En lugar de ver la tabla `fact_sales`, el usuario ve un objeto llamado “**Ventas**”.
 - En lugar de escribir `SUM(sale_amount)`, el usuario arrastra un campo llamado “**Ingresos Totales**”.

- En lugar de escribir un JOIN complejo, el usuario simplemente selecciona “Ingresos Totales” y “Nombre del Cliente”, y la capa semántica sabe cómo unir las tablas por detrás.
- **La Analogía:** Es como el menú de un restaurante. No ves la complejidad de la cocina, solo ves “Lasaña Boloñesa” y haces clic para pedirla. La capa semántica es el menú de tus datos.
- **Relación con Strengtholt:** Este es un componente crucial dentro de un **DDS Blueprint** para un caso de uso de *reporting*. Es una de las “herramientas” que un dominio consumidor utiliza para generar valor.

2. Autoservicio (Self-Service)

- **El Problema que Resuelve:** El cuello de botella tradicional donde el negocio tiene que pedirle cada nuevo reporte al equipo de TI y esperar semanas.
- **¿Qué es?: Empoderar** a los usuarios de negocio (analistas, gerentes) para que puedan explorar los datos y responder a sus propias preguntas sin depender de un intermediario técnico. Las capas semánticas son un habilitador clave del autoservicio.
- **La Distinción Clave de Strengtholt:** Para que el autoservicio no se convierta en un caos, es vital distinguir entre dos tipos de actividades:
 - **Datos Gestionados (Managed Data):** Son los reportes y *dashboards* oficiales, críticos para el negocio (ej: el reporte de ventas trimestral). Son estables, automatizados y están bajo un gobierno estricto. Son como los platos fijos del menú del restaurante.
 - **Datos de Autoservicio (Self-service Data):** Son análisis *ad-hoc* y exploraciones puntuales para responder una pregunta específica (ej: “¿Por qué cayeron las ventas en la región norte el martes pasado?”). Son temporales por naturaleza. Strengtholt sugiere que para esto se creen “workspaces” o entornos de “sandbox” temporales. Son como el chef experimentando con un nuevo plato que quizás nunca llegue al menú.

Analítica Avanzada (MLOps) en Detalle

Aquí pasamos de entender el pasado a predecir el futuro.

- **El Reto:** Un modelo de *Machine Learning* (ML) en el *notebook* de un científico de datos es como un prototipo de motor de F1 construido con Lego. Funciona en un entorno ideal. **MLOps** es el conjunto de prácticas y herramientas para tomar ese prototipo y convertirlo en un motor real, instalarlo en un coche de carreras, ponerlo en la pista, y asegurarse de que siga funcionando vuelta tras vuelta sin explotar.
- **¿Qué es MLOps?:** Es el **marco de trabajo para gestionar el ciclo de vida completo de un modelo de ML en producción**. Es la unión de *Machine Learning* (ML) y *DevOps* (Operations). Su objetivo es hacer que el despliegue y mantenimiento de modelos de ML sea **repetible, confiable y escalable**.
- **¿Por qué es diferente de DevOps?:** El software tradicional es estático. Una vez que lo despliegas, no cambia. Los modelos de ML, en cambio, se “degradan” con el tiempo porque el mundo cambia (*data drift*). Por lo tanto, deben ser **constantemente reentrenados, recalibrados y redesplegados**. MLOps gestiona este ciclo de vida dinámico.

Fases de MLOps

1. **Fase de Exploración:** El “laboratorio”. Aquí el científico de datos juega con los datos en un entorno seguro (un *sandbox* o *workspace* dentro de un DDS), prueba diferentes algoritmos y valida si es posible construir un modelo que resuelva el problema de negocio. El resultado es un prototipo.
2. **Fase de Operacionalización:** La “fábrica”. Esta es la fase más difícil. Aquí se toma el prototipo y se **automatiza todo el proceso en un pipeline**. Este *pipeline* se encarga de:
 - Extraer automáticamente los datos nuevos.
 - Procesarlos y crear las *features*.
 - Reentrenar el modelo.
 - Evaluar el nuevo modelo.
 - Si el nuevo modelo es mejor, desplegarlo automáticamente en producción.
3. **Fase de Monitoreo:** La “telemetría”. Una vez que el modelo está en producción, se monitorea constantemente su rendimiento. ¿Sus predicciones siguen siendo precisas? ¿El *data drift* lo está afectando? Si el rendimiento cae por debajo de un

umbral, se dispara una alerta y, a menudo, se activa automáticamente el *pipeline* de reentrenamiento.

Relación con Strongholt: El proceso de MLOps es la forma **gestionada y escalable** de construir y operar las soluciones de “analítica avanzada” dentro de un **DDS Blueprint** para *Machine Learning*. Es el corazón operativo de los casos de uso más complejos en el lado del consumidor.