IA APLICADA A LA PRODUCCIÓN PESQUERA

UNIDAD 5

El viaje del dato

Arquitectura de datos y Big Data, del barco al tablero



Damian Giacone

Material del curso · UTN FRCH, Puerto Madryn

IA APLICADA A LA PRODUCCIÓN PESQUERA · UNIDAD 5

El viaje del dato

Arquitectura de datos y Big Data, del barco al tablero

Damian Giacone

www.damiangiacone.com

Material del curso **IA aplicada a la Producción Pesquera**
UTN Facultad Regional Chubut, Puerto Madryn
Octubre de 2026

Antes de empezar

Este libro acompaña la Unidad 5 del curso *IA aplicada a la Producción Pesquera*. Está escrito para gente de la pesca: patrones, jefes de planta, técnicos, calidad, administración. No hace falta saber de informática.

La idea que lo ordena es una sola: **un dato hace el mismo viaje que un lote de merluza**. Sale del agua, pasa por la bodega, se descarga, se recibe en planta, se clasifica, se procesa, se envasa, se etiqueta, se guarda en cámara y se despacha. Si ya sabés cómo se recorre un lance de punta a punta, ya conocés el esqueleto de lo que sigue. Lo único que cambia es qué miramos mientras el lote avanza.

Cómo está armado

Cada capítulo arranca con una situación que conocés del barco o de la planta. Recién después aparece el nombre técnico, y siempre atado a esa situación. Si una palabra te suena rara, no pasa nada: la primera vez que aparece va en **negrita** y al final hay un glosario con todas, dichas en idioma pesquero.

Vas a encontrar dos tipos de recuadros:

> **IDEA CLAVE**
>
> Lo que conviene que te quede después de leer el capítulo.

PARÁ Y PENSÁ

Una pregunta para responder con tu propio barco, tu propia planta o tu propio puesto. Sirven para armar el mini-entregable del final.

La pregunta que guía el libro

En el curso usamos un asistente llamado **Fishy**, conectado a datos de ejemplo de una planta pesquera. Le podés preguntar cosas en lenguaje de todos los días, como qué pasó con el rendimiento de la merluza o cómo vienen los kilos por embarcación.

El capítulo 1 arranca con una pregunta que Fishy no puede responder entera. El libro termina volviendo a esa misma pregunta, con una idea clara de qué hace falta para que sí pueda.

PARA QUIÉN ES

Para quien va a trabajar con asistentes, modelos y tableros en su empresa y quiere entender de dónde sale lo que esas herramientas usan. Las unidades que vienen, de modelos predictivos, visión artificial, agentes y tableros, dependen de lo que se explica acá.

CAPÍTULO 1

Una pregunta sin respuesta entera

Empecemos por lo que pasa hoy. Imaginá que le preguntás a Fishy lo que se preguntaría cualquier jefe de planta un lunes a la mañana:

```
> ¿Por qué bajó el rendimiento de la merluza esta semana?

Bajó, sí. Los kilos de producto terminado respecto de los kilos capturados están por debajo de la semana anterior. La causa no la puedo confirmar: no tengo la temperatura de cámara de ese lote ni el control de calidad de esa jornada.
```

Respuesta de ejemplo, escrita para ilustrar la situación. No es una salida real de Fishy.

Fijate lo que pasó. El asistente no se equivocó ni inventó nada. Hizo lo que podía con lo que tenía: sabía cuántos kilos entraron y cuántos salieron, pero no tenía la temperatura de la cámara ni el control de calidad de ese día. Esas dos cosas existen. Alguien las midió y alguien las anotó. Simplemente no están donde Fishy puede verlas.

IDEA CLAVE

Los datos existen. El problema es dónde están y si se pueden juntar. Una herramienta de inteligencia artificial responde tan bien como lo permitan los datos que alcanza a ver.

Esto no le pasa solo a Fishy. Le pasa a cualquier persona de la planta cuando quiere cruzar el parte del barco con el control de calidad, o la temperatura de cámara con el rendimiento de la línea. Para responder una pregunta que atraviesa el recorrido, hay que juntar datos que se generaron en lugares distintos, por personas distintas, en formatos distintos.

De eso trata este libro: de cómo se organiza el recorrido de los datos para que esa pregunta se pueda contestar entera.

PARÁ Y PENSÁ

Pensá en una pregunta que te hayan hecho alguna vez en el trabajo y que te llevó un buen rato contestar, porque tuviste que juntar información de varios lados. Anotala. ¿De cuántos lugares distintos tuviste que sacar datos?

CAPÍTULO 2

El recorrido de un lance de merluza

Antes de hablar de datos, hablemos de pescado. Seguí con la cabeza un lance de merluza desde que sube a bordo hasta que sale en un camión. Estos son los pasos, con lo que se mide o se anota en cada uno:

#	**Paso**	**Dato que se genera**
1	Captura	Posición, hora, kilos
2	Bodega	Temperatura
3	Descarga	Peso
4	Recepción	Temperatura, control visual
5	Clasificado	Especie, calibre
6	Línea de proceso	Rendimiento, merma
7	Envasado	Peso neto
8	Etiquetado	Lote, fecha
9	Cámara	Temperatura, humedad
10	Despacho	Destino, camión



Cada etapa del recorrido y los datos que genera. Ilustración.

La lista no tiene nada nuevo para vos: es el recorrido que ya conocés. Lo que sí es nuevo es la tercera columna. En cada punto del viaje, alguien anota, mide o fotografía algo. Un lote de merluza no pasa por ninguno de esos pasos sin dejar un rastro de información.

IDEA CLAVE

Un lote de merluza y su información hacen el mismo viaje. El pescado avanza físicamente de un paso al siguiente. Los datos, en cambio, se quedan donde se generaron, y ahí empieza el problema que vemos en el próximo capítulo.

Este recorrido va a ser nuestro hilo. Cada vez que aparezca un concepto nuevo, vamos a volver a preguntarnos en qué punto del lance estamos.

PARÁ Y PENSÁ

Escribí el recorrido de tu barco o de tu planta con tus propios pasos. Puede tener más o menos de diez. Al lado de cada uno anotá qué dato se genera y quién lo anota.

CAPÍTULO 3

¿Dónde queda el dato?

En el capítulo 2 vimos que cada paso del recorrido genera un dato. Ahora preguntémonos dónde queda ese dato cuando el lote sigue su camino.

El pescado viaja. El dato, la mayoría de las veces, no. Se queda donde nació, y casi siempre en uno de tres lugares:

Dónde queda	**Ejemplo**	**Quién lo puede ver**
El **bolsillo del capitán**	El parte de pesca anotado a mano, la posición que sólo sabe el patrón	El patrón
El **cajón**	La planilla de la balanza, el registro de temperatura impreso	Quien tiene la llave o está en ese puesto
La **carpeta**	El Excel del jefe de turno, el control de calidad en una planilla propia	Quien sabe en qué carpeta está y cómo se llama el archivo

Ninguno de los tres es un error de nadie. Son la forma natural en que trabaja cada puesto. El efecto, sin embargo, es que el dato queda aislado. A eso se le llama **silo de datos**: información guardada en un lugar cerrado, que sólo entiende y alcanza quien la generó.

IDEA CLAVE

Un silo es el equivalente a una bodega con la puerta trabada: el pescado está adentro, pero nadie más puede sacarlo. El dato existe, pero no sirve para responder preguntas que cruzan el recorrido.



Cada etapa guarda su dato en su propia cámara cerrada: nadie ve lo que hay en las otras. Ilustración.



Islas de información: cada área registra y guarda sus datos por su cuenta. Ilustración.

Tres tipos de dato

No todos los datos son iguales, y eso importa cuando hay que juntarlos.

Tipo	Cómo se ve	Ejemplo pesquero
Datos en **tabla**	Filas y columnas, números y fechas	Kilos por lance, temperatura por hora, peso neto por caja
Datos en **texto libre**	Frases, notas, partes escritos	El parte de pesca manuscrito, un mail del observador, una observación del control de calidad
Datos en **imagen, audio o video**	Archivos que hay que mirar o escuchar	Una foto del lote en recepción, la cámara de la línea, un audio del patrón

Las tablas son las más fáciles de juntar. Los textos y las imágenes tienen muchísima información, pero hay que prepararlos antes de poder cruzarlos con el resto. Más adelante vemos cómo.

Las fuentes de datos del recorrido

Cuando hablamos de **fuentes de datos** hablamos, simplemente, de los lugares donde nace la información. Agrupadas por zona del recorrido:

En el mar

Fuente	Qué aporta	Tipo
GPS del barco	Posición, rumbo, velocidad	Tabla
Sonda y sensores de red	Profundidad, lo que detecta la sonda, apertura de red	Tabla
Parte de pesca	Lances, horas, kilos estimados por especie	Texto libre o tabla, según cómo se anote
Sensores de bodega	Temperatura	Tabla
Observador a bordo	Observaciones de captura y descarte	Texto libre
Consumo de combustible	Litros por jornada o por lance	Tabla

En el puerto y la planta

Fuente	Qué aporta	Tipo
Balanza de descarga	Peso por cajón o por viaje	Tabla
Control de calidad en recepción	Temperatura, aspecto, frescura	Tabla y texto libre
Clasificado	Especie, calibre	Tabla, a veces imagen
Línea de proceso	Rendimiento, merma	Tabla
Envasado y etiquetado	Peso neto, lote, fecha	Tabla
Cámara frigorífica	Temperatura y humedad en el tiempo	Tabla
Despacho	Destino, camión, remito	Tabla y documentos

En el mercado y el entorno

Fuente	Qué aporta	Tipo
Precios de mercado	Cuánto se paga cada producto	Tabla
Clima y condiciones del mar	Viento, oleaje, temperatura del agua	Tabla
Cuotas y permisos	Cuánto se puede capturar y dónde	Documentos
Certificaciones y auditorías	Requisitos que hay que poder demostrar	Documentos
Clientes y pedidos	Qué se vendió, a quién, cuándo	Tabla

Esta lista no es cerrada. Cada empresa tiene las suyas, y seguramente te encuentres con alguna que acá no figura.

PARÁ Y PENSÁ

Tomá el recorrido que armaste en el capítulo 2. Para cada paso, marcá dónde queda hoy el dato: bolsillo, cajón, carpeta o un sistema. Después elegí dos fuentes de datos que nunca se hayan cruzado y preguntate: *¿podrían ponerlas juntas en cinco minutos?* Si la respuesta es no, ya encontraste un silo.

CAPÍTULO 4

Big Data: el concepto

Antes de hablar de dónde se guardan los datos y cómo se preparan, conviene frenar en una palabra que seguramente escuchaste: **Big Data**. Suena a algo enorme y lejano, pero es una idea simple, y entenderla ayuda a ver por qué hace falta todo lo que viene después.

Big Data es una idea, no un lugar

Big Data no es un programa, ni una base de datos, ni una máquina que se compra. Es un **concepto**: el nombre que se le da a la situación en que los datos son tantos, llegan tan rápido y son tan variados que las herramientas de siempre, como la planilla de Excel, dejan de alcanzar para guardarlos, ordenarlos y usarlos.

Pensalo con el recorrido del capítulo 2. Un barco, una balanza y una planilla de calidad generan pocos datos y se manejan bien a mano. Ahora sumá los sensores de las bodegas, el GPS que informa todo el tiempo, las cámaras de la línea, las cámaras frigoríficas y los partes de todos los barcos, durante años. Ahí ya no alcanza la carpeta: hace falta método. A esa situación se la llama Big Data.

Las cinco V

Para describirlo se usan cinco palabras que empiezan con V. Las llevamos a la pesca:

Las 5 V	Qué significa	En la pesca
Volumen	Mucha cantidad de datos	Años de lances, lecturas de temperatura por hora, miles de fotos de la línea
Velocidad	Los datos llegan rápido y seguido	El GPS informa posición todo el tiempo; el sensor de cámara mide a cada minuto
Variedad	Formatos muy distintos	Tablas, partes manuscritos, fotos, audios, documentos
Veracidad	**Cuánto se puede confiar en el dato**	**Un peso mal anotado, una hora mal puesta, un sensor descalibrado**
Valor	**Convertir los datos en información útil para decidir**	**Que la balanza, la temperatura de cámara y el parte terminen explicando por qué cambió un rendimiento**

VERACIDAD Y VALOR NO VIENEN EN EL DATO

Volumen, velocidad y variedad describen el dato tal como llega. **Veracidad y valor**, no: ningún dato llega confiable ni útil por sí solo.

Se **logran** con un proceso que las asegura: controles de calidad, un recorrido ordenado del dato (ETL o ELT, capítulo 7) y reglas de gobernanza (capítulo 8).

Por eso más adelante dedicamos capítulos enteros a cómo se prepara el dato y a cómo se cuida su confianza.

El verdadero problema del Big Data no es la cantidad: es la combinación de fuentes que hablan distinto y llegan a distinto ritmo. Ilustración.

La imagen muestra la idea central. Un dato enorme pero ordenado, en una sola tabla y con un solo formato, se maneja con herramientas comunes: no es un problema de Big Data. El problema aparece cuando se juntan muchas fuentes que hablan distinto y llegan a distinto ritmo.

IDEA CLAVE

Big Data es el nombre del problema, no de la solución: hay datos de más, de muchas clases y llegando sin parar. Todo lo que sigue en el libro, la bodega, la recepción, el procesamiento y la cámara, es la manera de ordenar ese problema.

Con esta idea en la cabeza, ya podemos pasar a lo concreto: dónde se guardan todos esos datos mientras llegan.

PARÁ Y PENSÁ

De las cinco V, ¿cuál es la que más te complica hoy en tu trabajo? ¿Te sobran datos, te llegan muy rápido, vienen en formatos incompatibles o no te podés fiar de ellos?

CAPÍTULO 5

La bodega: el data lake

Pensá en la bodega de un barco. Lo que sube en cada lance se estiba como viene: distintas especies, distintos tamaños, distintas horas. Nadie lo ordena en alta mar. Lo importante es que entre, que se conserve y que no se pierda nada.

En el mundo de los datos existe exactamente ese lugar: el **data lake**, o lago de datos. Es un depósito grande donde los datos se guardan **tal como llegan**, sin ordenarlos ni corregirlos, pero **con una etiqueta mínima** que dice de dónde vienen. Entra la tabla de la balanza, entra el parte escaneado, entra la foto de la cámara.

> En el barco La bodega recibe todo lo que se capturó, en el estado en que salió del agua.
>
> En los datos El data lake recibe todo lo que se generó, en el formato en que se generó.



La red del barco y la red del dato: se captura de todo y recién después se separa y se clasifica. Ilustración.

Para qué sirve guardar todo

Porque hoy no sabés qué pregunta te van a hacer mañana. Si descartás un dato por parecer inútil, y dentro de seis meses hace falta para entender un problema de rendimiento, ya no está. La bodega sirve para eso: conservar la materia prima completa.

Todo, pero con una etiqueta mínima

Guardar todo no quiere decir guardar sin anotar. En el barco, cada cajón sale del lance con un papel mínimo: la fecha, la zona y el número de lance. Todavía no se sabe qué calibre tiene ni cómo va a salir en la línea, pero se sabe de dónde viene. Con los datos pasa igual: cada uno entra al lago con su **etiqueta de origen**.

Etiqueta mínima	Qué dice	En el cajón
Fuente	De dónde viene: la balanza, el GPS, el parte, la cámara	El barco y el lance
Cuándo entró	Fecha y hora en que llegó al lago	La fecha de captura
Quién lo cargó	La persona o el sistema que lo cargó	Quién anotó el cajón
Formato	Tabla, texto, imagen, audio	Si es merluza entera o cajón de calamar

Esta etiqueta es distinta de la que aparece más adelante en la planta. La de la bodega es mínima y la pone quien carga el dato; la de la planta, con lote, calibre y controles, se agrega recién al transformarlo. Pero sin la primera no hay segunda: no se puede etiquetar bien lo que no se sabe de dónde vino.

Una bodega sin esa etiqueta tiene un problema conocido: nadie sabe qué hay adentro ni de qué lance es cada cosa. Un lago al que se tira de todo, sin anotar de dónde viene cada dato, se convierte en lo que en la jerga se llama un **pantano de datos**: está todo, pero no se encuentra nada.

IDEA CLAVE

El data lake es la bodega del dato: guarda todo, tal como llega y con su etiqueta de origen. Es indispensable, pero no alcanza. Nadie le sirve un filete al cliente directamente desde la bodega.

PARÁ Y PENSÁ

¿Qué datos de tu empresa se tiran, se borran o se pisan hoy porque “ya no hacen falta”? ¿Alguna vez te hubiera servido tenerlos?

CAPÍTULO 6

De la bodega a la cámara

Entre la bodega del barco y la cámara frigorífica donde se guarda el producto terminado hay un camino con varias estaciones. Cada una tiene su equivalente en el mundo de los datos. Vamos de a una.

1. La descarga

> En la planta Se baja el pescado de la bodega y se pesa.
>
> En los datos Se **extraen** los datos del lake para empezar a trabajarlos.

2. La recepción en planta: el staging

El pescado llega en cajones y se deja un tiempo en recepción, antes de entrar a la línea. Ese lugar es una zona de paso: se almacena de forma temporal, se controla la temperatura y se vacía a medida que el lote avanza.

> En la planta Cajones en recepción, temporalmente almacenados.
>
> En los datos El **staging** es una copia intermedia de los datos extraídos. Se guarda ahí mientras se prepara y se vacía una vez que la carga termina.

3. El clasificado

> En la planta Se separa por especie y por calibre.
>
> En los datos Se **clasifican** los registros: se separan por tipo, se descartan los que están incompletos o repetidos y se corrigen los que están mal escritos.

4. La línea de proceso

> En la planta Se filetea, se limpia, se corta. Aparecen rendimiento y merma.
>
> En los datos Se **procesan** los datos: se calculan totales, rendimientos, se cruzan fuentes y se arman los indicadores.

5. El envasado

> En la planta Se arma la caja o la bolsa con el peso neto.
>
> En los datos Se **empaquetan** los datos en el formato final que va a usar el warehouse: columnas ordenadas, mismas unidades, mismas fechas.

6. El etiquetado

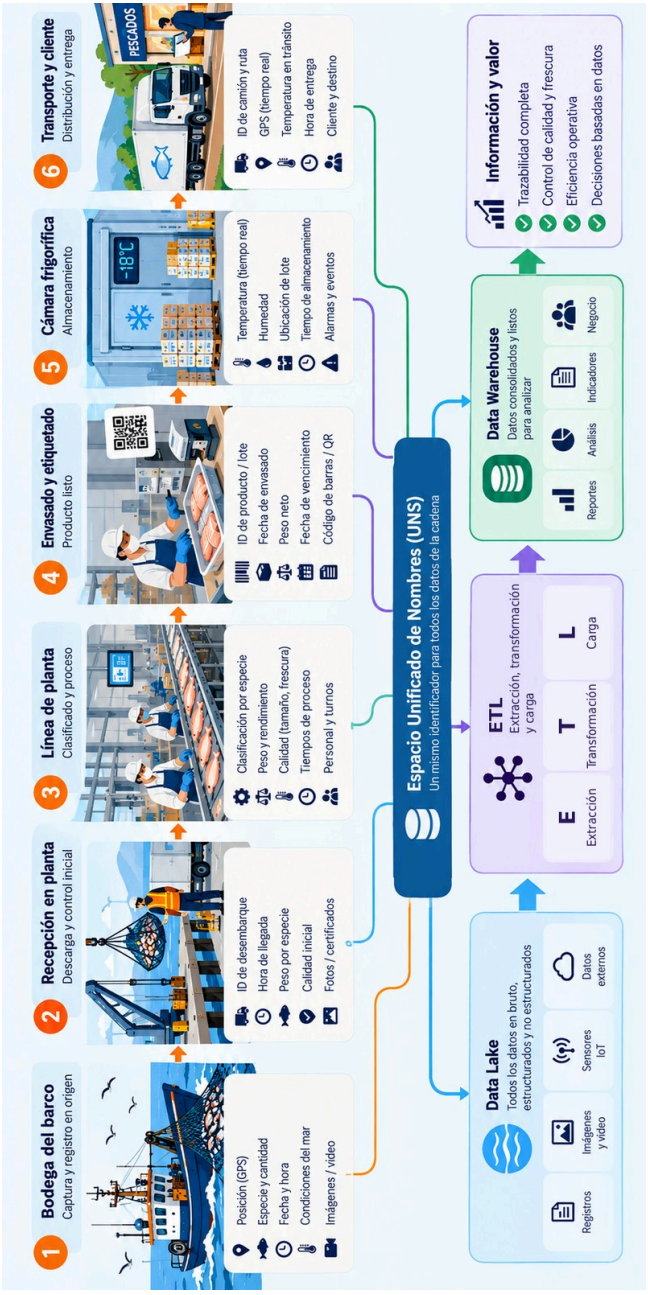
> En la planta Cada caja lleva lote, fecha y origen.
>
> En los datos Cada dato lleva su **etiqueta**: de qué lance viene, quién lo cargó, cuándo y bajo qué criterio.

7. La cámara frigorífica: el data warehouse

> En la planta El producto terminado se estiba ordenado, por lote, listo para despachar.
>
> En los datos El **data warehouse** guarda los datos ya limpios, ordenados y etiquetados, listos para consultar.

Lake y warehouse, lado a lado

	Data lake (bodega)	**Data warehouse (cámara)**
Qué guarda	Todo, como llegó	Sólo lo ya preparado
Orden	Ninguno, pero cada dato lleva su etiqueta de origen	Estricto, por lote y por fecha
Para qué sirve	Conservar la materia prima	Responder preguntas rápido y sin dudas
Quién lo usa	Quien prepara los datos	Quien consulta, analiza o arma tableros



Las estaciones del camino y los datos de cada una. Ilustración.

IDEA CLAVE

El camino de la bodega a la cámara es siempre el mismo: descargar, recibir, clasificar, procesar, envasar, etiquetar y estibar. Si lo sabés hacer con pescado, ya lo sabés hacer con datos.

PARÁ Y PENSÁ

En tu empresa, ¿en qué estación del camino se pierde más información o aparecen más errores? ¿Es antes del clasificado, o recién cuando el dato ya está “envasado”?

ETL y las cinco etapas del dato

El camino que acabamos de recorrer tiene nombre técnico: **ETL**. Son las iniciales de tres palabras en inglés. Mirá cómo se corresponden con lo que ya conocés.

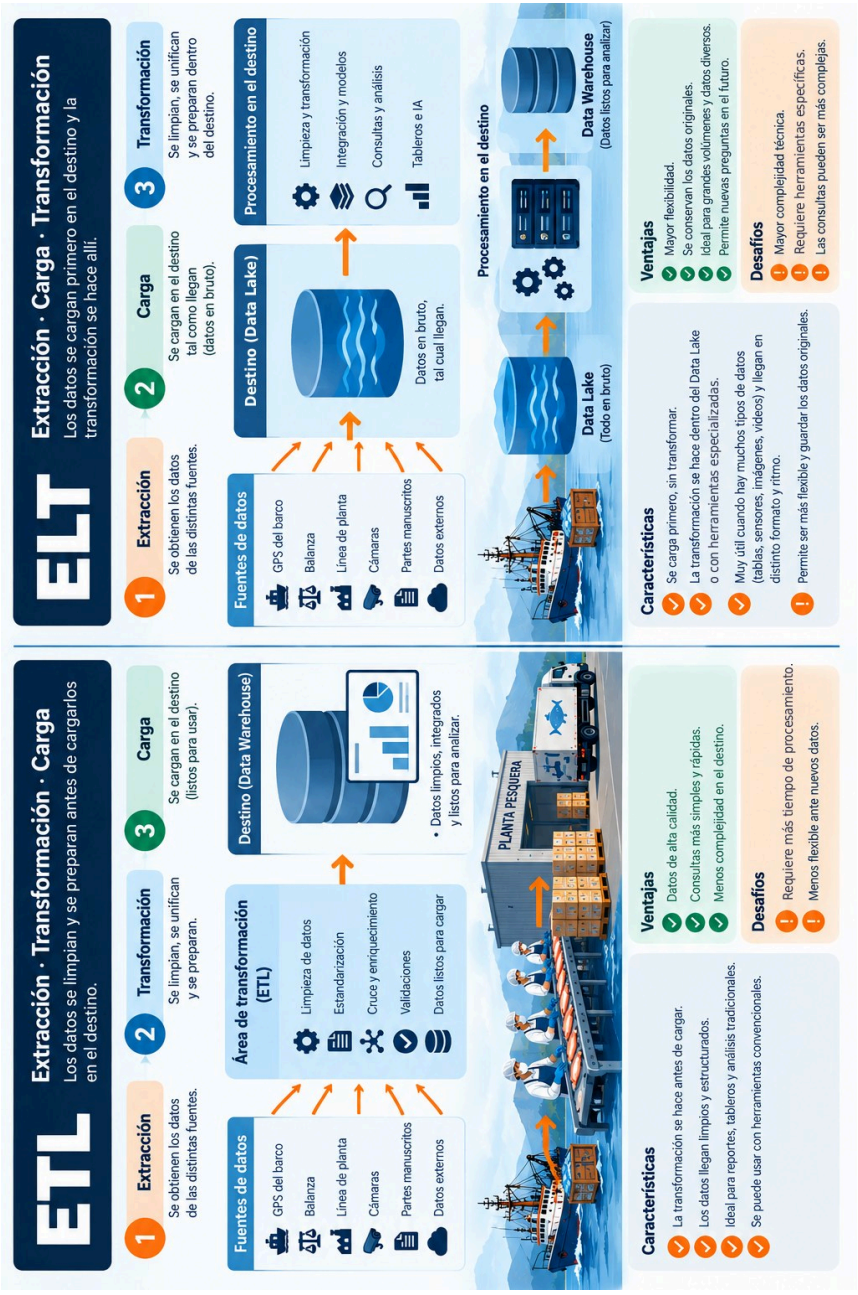
Letra	Palabra	En la planta	En los datos
E	Extracción	Descarga de la bodega	Sacar los datos del lake
T	Transformación	Clasificado, línea de proceso, envasado y etiquetado	Limpiar, ordenar, cruzar y etiquetar
L	Carga	Estiba en cámara	Guardar los datos listos en el warehouse

Fijate que la T es la parte más larga y la que más trabajo lleva, igual que en la planta. Es donde un dato en bruto se convierte en un dato que sirve.

ETL y ELT: dos órdenes posibles

El ETL clásico transforma **antes** de cargar: el dato se prepara en la línea de planta y recién entonces se estiba en la cámara. Hay otra forma, cada vez más usada, que invierte el orden: **ELT**. Se extrae, se **carga** primero y se transforma **después**, adentro del propio almacén.

	ETL	**ELT**
Orden	Extraer, transformar, cargar	Extraer, cargar, transformar
Dónde se transforma	En un paso intermedio, antes de llegar al destino (la línea de planta)	Dentro del destino, una vez que el dato ya está cargado
En la planta	Se clasifica, procesa, envasa y etiqueta, y recién después se estiba en cámara	Se estiba todo tal como llega en una cámara grande, y se clasifica y procesa ahí adentro cuando hace falta
Cuándo conviene	Cuando se sabe de antemano cómo tiene que quedar el dato y se quiere guardar sólo lo ya limpio	Cuando hay muchos datos o todavía no se sabe qué preguntas se van a hacer, y se prefiere conservar todo y decidir después



ETL y ELT lado a lado: la diferencia es el orden en que se transforma el dato. Ilustración.

No es una cuestión de cuál es mejor. Son dos maneras de organizar el mismo trabajo, y muchas empresas usan las dos según el tipo de dato. Lo que no cambia es que las tres tareas, extraer, transformar y cargar, siempre hay que hacerlas.

Cuándo usar ETL y cuándo ELT

Conviene ETL cuando…	Conviene ELT cuando…
Sabés cómo tiene que quedar el dato. Las preguntas, los indicadores y el formato final están definidos. En la planta: un producto con especificación cerrada, que se clasifica, envasa y etiqueta antes de estibarlo.	**No sabés qué preguntas vas a hacer.** Guardás todo y decidís después. En la planta: estibar todo en una cámara grande y procesar a pedido.
Querés guardar sólo lo ya limpio. El warehouse queda chico, ordenado y barato de consultar.	**Hay muchos datos o muy variados:** imágenes, textos y sensores que no entran fácil en un formato único.
Hay datos sensibles que no deberían guardarse tal cual: se anonimizan o se descartan antes de que entren al destino.	**Necesitás conservar el dato original** para auditar, reprocesar con otro criterio o entrenar modelos.
La calidad se exige antes del destino: si algo no cumple, no entra. Es el control de calidad como puerta de entrada.	**El destino tiene mucha capacidad de cómputo,** como en las plataformas modernas en la nube, y transformar adentro sale barato y rápido.
El destino tiene poca capacidad de procesamiento: la transformación se hace afuera, en un paso intermedio.	**Querés cargar rápido,** sin esperar a que termine la transformación para tener los datos disponibles.



Cuándo conviene ETL y cuándo ELT, con el criterio práctico para decidir. Ilustración.

CRITERIO PRÁCTICO

Si el dato ya tiene un uso claro y reglas fijas, ETL. Si todavía estás explorando, o el dato original vale por sí mismo, ELT. En la práctica muchas empresas usan los dos: ELT para lo exploratorio, con el lake como bodega, y ETL para lo que alimenta los tableros de siempre.

Dos advertencias. Primero, **ELT no es "sin transformar"**: la transformación existe igual, sólo que ocurre después de cargar. Si nadie la hace, el lake termina en un pantano de datos. Segundo, **la elección no es sólo técnica**: también pesa quién es el responsable del dato, si hay requisitos legales o de calidad previos a guardarlo y cuánto sabe hacer el equipo que lo mantiene.

Las cinco etapas

Otra forma de mirar el mismo recorrido es agrupar todo en cinco etapas, que van a ser el esqueleto del resto del curso:

Etapa	Qué es	En el recorrido del lance
1. Captación	Lugar y momento donde nace el dato	Captura, bodega, balanza, sensores
2. Transporte	Cómo viaja el dato de un lugar a otro	Del barco a tierra, de la planta a la oficina
3. Almacenamiento	Dónde descansa: lake, staging, warehouse	Bodega, recepción, cámara
4. Procesamiento	Se clasifica, procesa, envasa y etiqueta	Clasificado, línea, envasado, etiquetado
5. Salida	Lo que le llega a quien decide	Despacho: tableros, informes, asistentes como Fishy

Esta división es la que usamos en el curso para ordenar el trabajo. No es la única posible, y en otros libros vas a ver variantes, pero sirve para no perderse.

PARÁ Y PENSÁ

Pensá en un indicador que mires seguido, por ejemplo el rendimiento de la semana. Recorré sus cinco etapas: ¿dónde se capta, cómo viaja, dónde se guarda, quién lo procesa y cómo te llega?

CAPÍTULO 8

Gobernanza: el HACCP del dato

En una planta pesquera nadie discute el HACCP. Es el sistema que define qué puntos del proceso se controlan, quién se hace responsable, cómo se registra y qué se hace si algo sale de rango. Gracias a eso un lote se puede seguir hacia atrás y hacia adelante, y se puede demostrar que se hizo bien.

Con los datos hace falta lo mismo. A ese sistema de reglas se lo llama **gobernanza de datos**. Pensala como un **HACCP del dato**: cinco controles que protegen la calidad de la información del mismo modo que el HACCP protege la del producto.

Control	Qué pregunta responde	En la pesca	Dónde se aplica
Propiedad	¿Quién es responsable de este dato?	El patrón responde por el parte; el jefe de turno, por la planilla de línea	Captación
Calidad	¿Está completo, bien escrito y es creíble?	Controlar el calibre de la balanza antes de pesar	Clasificado
Estándares	¿Todos lo anotan igual?	Una misma unidad (kilos), una misma fecha, un mismo nombre de especie	Línea de proceso, envasado
Trazabilidad	¿De dónde viene y por dónde pasó?	El lote con su etiqueta, de la captura al despacho	Etiquetado
Acceso	¿Quién puede verlo, cambiarlo o borrarlo?	Sólo calidad cierra el control de recepción; la llave de la cámara	Cámara (warehouse)

La columna “Dónde se aplica” es la propuesta del curso para ubicar cada control en un punto del recorrido. No quiere decir que sólo se aplique ahí: la propiedad, por ejemplo, también importa en cada estación siguiente. Es una guía para empezar.

La trazabilidad, un viejo conocido

La trazabilidad merece una mención aparte porque en la pesca ya es una exigencia. Los programas de certificación, la lucha contra la pesca ilegal (la sigla IUU) y los requisitos de los mercados que compran nuestro producto piden poder demostrar el origen de cada lote. Esa demostración sólo es posible si los datos también se guardaron con etiqueta.

Una escala para ir ganando gobernanza

La gobernanza no se instala de un día para el otro. Una planta no pasó de cero al HACCP: primero ordenó las buenas prácticas, después sumó los controles y recién entonces buscó una certificación. Con los datos pasa lo mismo, y conviene avanzar por niveles, ganando de a un control por vez.

Nivel	Cómo está el dato	Qué se hace	Control que se gana	Señal de que llegaste
0. Cuaderno	En papel, en el celular o en la cabeza del jefe de turno	Nada formal	Ninguno	—
1. Ordenado	En planillas, y todos anotan lo mismo y de la misma forma	Elegir un dato prioritario, darle un dueño y fijar nombre, unidad y formato	Propiedad y estándares	Si preguntás quién carga ese dato, responden sin dudar
2. Etiquetado	El lote une captura, planta y despacho	El lote pasa a ser la llave común, y se registra quién cargó cada dato y cuándo	Trazabilidad	Se puede seguir un lote de punta a punta en minutos, no en días
3. Controlado	El dato se revisa solo	Reglas automáticas (rangos, faltantes, parte contra balanza), alertas y permisos por rol	Calidad y acceso	Un dato mal cargado salta antes de llegar al tablero

Nivel	Cómo está el dato	Qué se hace	Control que se gana	Señal de que llegaste
4. Integrado	Un solo lugar y un solo lenguaje	Pizarrón único, lake y warehouse, y gobernanza como rutina: responsable, revisión periódica y auditoría	Todos, sostenidos en el tiempo	Se puede mostrar a una auditoría o a un cliente sin armar nada especial



Escala de madurez de datos en la cadena pesquera, del cuaderno al tablero integrado. Ilustración.

La regla para avanzar es no subir de nivel con todos los datos a la vez, sino con **uno solo**: el peso de descarga, la temperatura de cámara, el parte de pesca. Un dato bien gobernado de punta a punta vale más que diez a medias.

Por eso la pregunta de Fishy del capítulo 1 se resuelve de a poco. Para que pueda explicar un rendimiento hacen falta la temperatura de cámara y el control de calidad de ese lote, con dueño, formato común y lote como llave. Es decir, llegar al nivel 2 con esos dos datos.

Como punto de partida, es razonable suponer que muchas empresas del sector están entre el nivel 0 y el 1. Lo que suele empujar al nivel 2 viene de afuera: las exigencias de trazabilidad de los organismos de control, los mercados y las certificaciones. Es una hipótesis para contrastar con tu empresa, no una medición.

IDEA CLAVE

El HACCP controla el producto. La gobernanza controla el dato. Y sin un dato confiable no hay trazabilidad, ni tablero, ni asistente de inteligencia artificial que se pueda creer.

PARÁ Y PENSÁ

Autodiagnóstico. Ubicá hoy a tu empresa en la escala, de 0 a 4, y fundamentá por qué: ¿qué señal de ese nivel ya cumplís y cuál te falta para el siguiente? Después elegí un solo dato y planteate qué harías para subirlo un nivel.

PARÁ Y PENSÁ

Elegí un dato de tu puesto y respondé los cinco controles: ¿quién es el dueño? ¿Cómo sabés que está bien? ¿Todos lo anotan igual? ¿Se puede rastrear de dónde salió? ¿Quién puede cambiarlo?

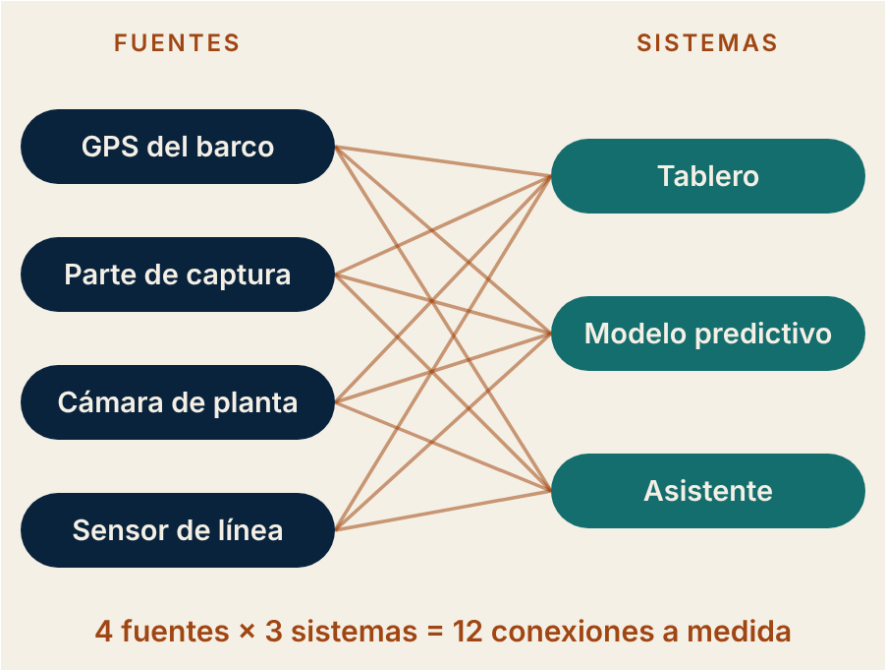
CAPÍTULO 9

De la maraña al pizarrón único

Hasta acá hablamos del dato histórico, el que ya pasó y descansa en la cámara. Pero hay otra cara: el dato que se está generando ahora mismo y que alguien necesita ver en este momento. Para eso, el problema pasa por cómo se comunican las fuentes con los sistemas que las usan.

La maraña de conexiones

Imaginá una empresa con cuatro fuentes de datos, por ejemplo GPS, sensores de bodega, balanza y calidad. Y tres sistemas que las necesitan: un tablero, un sistema de gestión y un asistente como Fishy. Si cada sistema se conecta por su cuenta con cada fuente, el resultado es **4 × 3 = 12 conexiones**. Si se suma una fuente nueva, se suman tres conexiones más. Si cambia una, hay que revisar las tres.

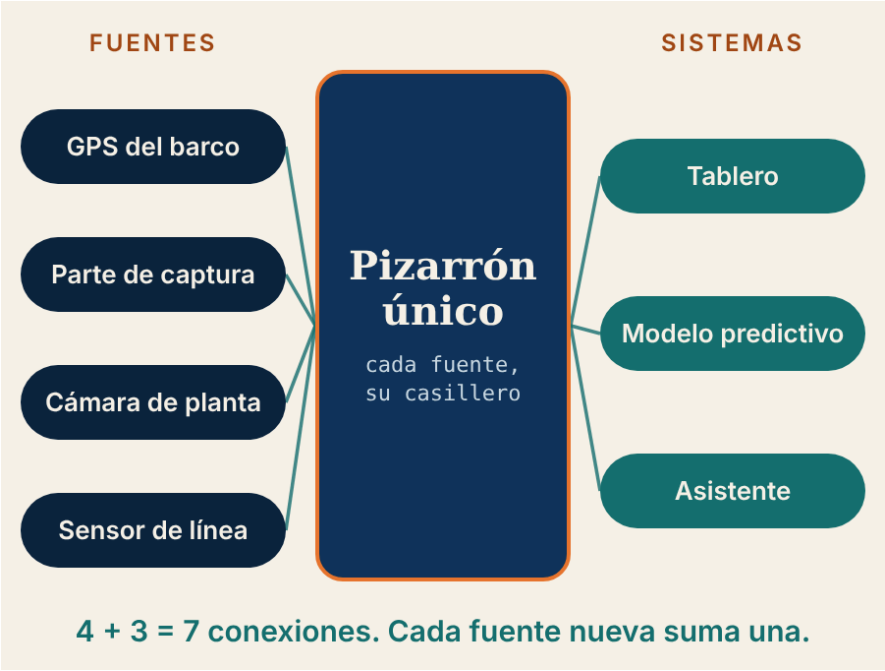


Cuatro fuentes por tres sistemas: doce conexiones a mantener.

El pizarrón único

La alternativa es la del pizarrón de la planta. Cada puesto anota su dato en el mismo pizarrón, y cualquiera que necesite saber algo va a mirar ahí. Nadie va puesto por puesto a preguntar.

En datos se llama **Unified Namespace** (UNS), algo así como “espacio de nombres unificado”. Cada fuente publica su dato en un lugar fijo del pizarrón, y cada sistema lee del lugar que le interesa. El resultado es **4 + 3 = 7 conexiones**, y agregar una fuente nueva suma una sola.

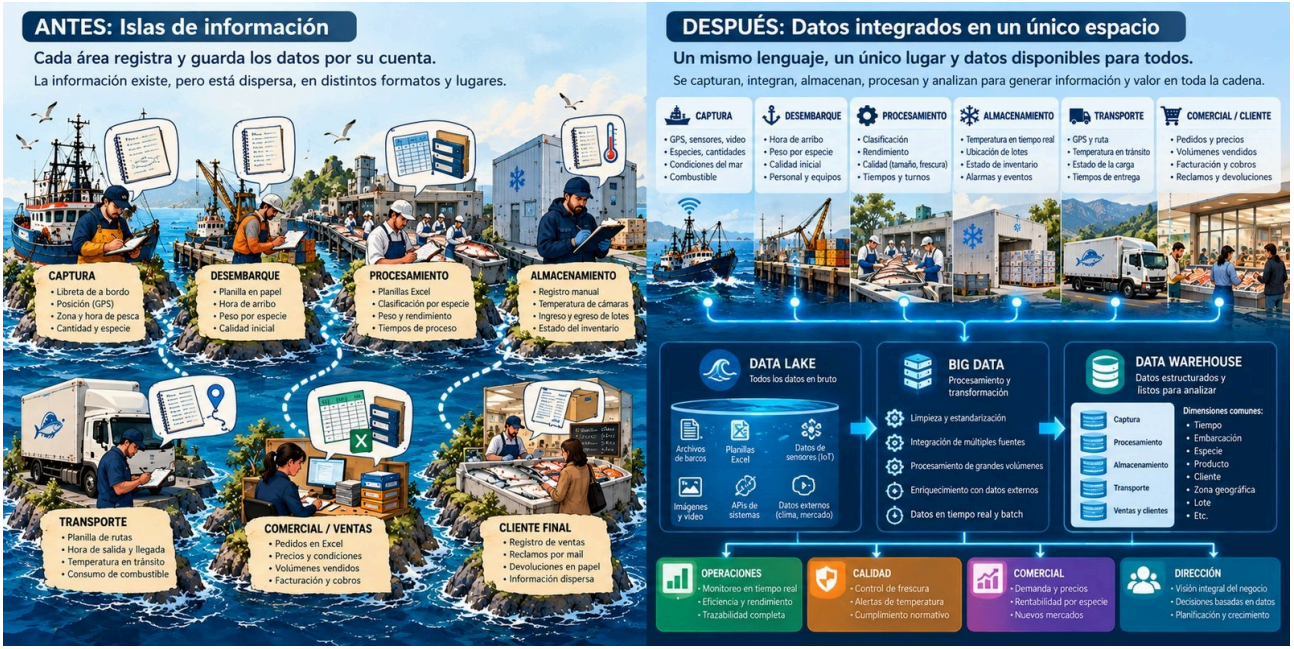


Con el pizarrón único, las conexiones son siete.

Cada lugar del pizarrón tiene una dirección con nombre, ordenada de lo general a lo particular, como la dirección de un puesto en la planta:

```
empresa / flota / barco1 / gps / posición
empresa / flota / barco1 / bodega / temperatura
empresa / planta / recepción / balanza / peso
```

Con estas direcciones, cualquiera sabe dónde buscar y de dónde viene lo que encuentra.



Antes, islas de información. Después, datos integrados en un único espacio. Ilustración.

Dato de ahora, dato de antes

	Pizarrón único (UNS)	**Lake y warehouse**
Qué muestra	El dato de ahora	El dato histórico
Pregunta típica	¿Cuánta temperatura hay ahora en la cámara 2?	¿Cómo viene la temperatura de la cámara 2 en el último mes?
En la pesca	El pizarrón de la planta	Los registros archivados

PARA QUIEN QUIERA PROFUNDIZAR

Para que las fuentes publiquen en el pizarrón suele usarse un protocolo llamado MQTT, que funciona por “publicar y suscribirse”: quien genera el dato lo publica, y quien lo necesita se suscribe. No hace falta saberlo para seguir el curso.

IDEA CLAVE

Una maraña de conexiones crece rápido y se rompe fácil. Un pizarrón único crece despacio y se entiende. Cuando una empresa empieza a sumar sensores, esa diferencia pesa.

PARÁ Y PENSÁ

Contá cuántas fuentes y cuántos sistemas hay en tu empresa. ¿Cuántas conexiones habría en la maraña? ¿Y con un pizarrón único?

CAPÍTULO 10

El dato del futuro

Hasta ahora el dato tenía dos tiempos: el de **ahora**, que vive en el pizarrón, y el de **antes**, que descansa en el lake y en el warehouse. Hay un tercero que cada vez pesa más: el dato de **lo que todavía no pasó**.

Es el que produce un modelo de inteligencia artificial cuando mira los datos de antes y los de ahora y anticipa lo que viene: la temperatura que va a tener la cámara 2 dentro de dos horas, el rendimiento esperado de un lote, la captura probable de la semana. No lo midió ningún sensor ni lo anotó ninguna persona. Lo **generó un modelo**.

La comparación con el pronóstico del tiempo

Todos entendemos la diferencia entre el parte de ayer, la temperatura que marca el termómetro ahora y el pronóstico de mañana. Los tres son información, pero no valen lo mismo. El pronóstico es útil porque permite decidir antes de que pase, y es riesgoso si se lo confunde con un hecho. Con los datos de la empresa pasa exactamente lo mismo.

	Dato de antes	Dato de ahora	Dato del futuro
Qué es	Lo que pasó y quedó registrado	Lo que está pasando	Lo que un modelo estima que va a pasar
De dónde sale	Lake y warehouse	Pizarrón único (UNS)	Un modelo de IA alimentado con los dos anteriores
Pregunta típica	¿Cómo vino la cámara 2 el último mes?	¿Cuánto marca la cámara 2 ahora?	¿Qué temperatura va a tener la cámara 2 esta noche?
Es un hecho	Sí	Sí	No: es una estimación
Sirve para	Entender y explicar	Controlar y reaccionar	Anticipar y decidir antes

Qué tipos de dato del futuro existen

- **Predicciones:** el valor que se espera más adelante. Por ejemplo, la temperatura de una cámara en las próximas horas o la cantidad de cajones que va a llegar mañana.
- **Alertas anticipadas:** una probabilidad de que algo ocurra. Por ejemplo, la chance de que un lote quede fuera de especificación si la cámara sigue subiendo.
- **Datos completados:** un valor que faltaba y el modelo estima a partir de los demás, como la lectura de un sensor que estuvo caído un rato.
- **Datos sintéticos:** datos inventados por un modelo que se parecen a los reales, y que sirven para practicar o entrenar sin exponer información sensible.

Los dos primeros miran hacia adelante. Los dos últimos no son del futuro, pero tienen lo mismo en común: no los midió nadie. Por eso se manejan con el mismo cuidado.

> **EJEMPLO**
> Un modelo mira el historial de temperatura de la cámara 2 y la carga que tiene hoy, y estima que a las 22 hs va a estar por encima de lo permitido. Con ese aviso, el encargado puede mover producto o revisar el equipo **antes** de que el lote se afecte. Sin el modelo, se enteraría después, cuando el dato de ahora ya mostró el problema.

Dónde vive el dato del futuro

El dato del futuro tiene que recorrer el mismo camino que cualquier otro: captarse, viajar, almacenarse y llegar a quien decide. Pero tiene una particularidad: **no se mezcla con el dato medido**. Se guarda aparte y con su propia dirección, para que nadie confunda lo que pasó con lo que se espera que pase.

```
empresa / planta / camara2 / temperatura
empresa / planta / camara2 / prediccion / temperatura
```

La primera dirección guarda lo que mide el sensor. La segunda, lo que estima el modelo.

Cuando pase el tiempo, el dato del futuro se convierte en dato de antes, y ahí aparece lo mejor que tiene: se puede **comparar lo que el modelo**

dijo con lo que realmente ocurrió. Esa diferencia es el error del modelo, y es lo que permite saber si se puede confiar en él.

El HACCP del dato, también para el futuro

Un dato generado por un modelo necesita más control que uno medido, no menos. Los cinco controles de la gobernanza se aplican igual, con un agregado: hay que poder decir **quién lo generó**.

Control	Qué pide para el dato del futuro
Propiedad	Quién es responsable del modelo y de lo que dice
Calidad	Cuánto se equivocó el modelo en el pasado, medido contra lo que realmente pasó
Estándares	Unidad, formato y horizonte claros: ¿es para dentro de una hora o de una semana?
Trazabilidad	Qué modelo, qué versión, con qué datos y cuándo generó cada valor
Acceso	Quién puede verlo y quién puede decidir en base a él

ETIQUETA MÍNIMA DEL DATO DEL FUTURO

Igual que el cajón en la bodega, todo dato generado por un modelo debería llevar su etiqueta: **que es estimado, qué modelo y versión lo generó, cuándo, para qué momento, y con qué margen de confianza**. Sin esa etiqueta, una predicción se confunde con un dato medido.

Los riesgos

- **Confundir estimación con hecho.** Una predicción puesta en un tablero sin aclarar que lo es termina tomándose por una medición.
- **Que el modelo se equivoque sin que nadie lo note.** Los modelos envejecen: lo que aprendieron del año pasado puede no valer cuando cambian la flota, la temporada o el equipo. Por eso se compara siempre con lo que ocurrió.
- **Alimentar un modelo con datos de otro modelo.** Si las estimaciones se vuelven a guardar como si fueran datos reales, los errores se acumulan sin que se vean.
- **Decidir sin saber quién responde.** Si un modelo recomienda y nadie es dueño de esa recomendación, nadie la revisa.

Por qué esto es el final del viaje

El recorrido del dato empezó con un lance y terminó en un tablero. El dato del futuro cierra el círculo: lo que hoy se mide, se guarda y se ordena es la materia prima para anticipar lo que viene. Y un modelo es tan bueno como los datos con los que aprende. Por eso todo lo que vimos antes, la bodega etiquetada, el ETL, la gobernanza, no es burocracia: es lo que permite que el día de mañana el dato del futuro sea confiable.

> **IDEA CLAVE**
>
> El dato del futuro no se mide: se genera. Por eso hay que guardarlo aparte, etiquetarlo como estimado, decir quién lo generó y compararlo siempre con lo que realmente pasó.

PARÁ Y PENSÁ

Elegí una decisión de tu empresa que hoy se toma cuando el problema ya pasó, por ejemplo una cámara que se pasó de temperatura o un lote que salió con otro calibre. ¿Qué dato del futuro te habría permitido decidir antes? ¿Con qué datos de antes y de ahora habría que alimentar al modelo?

CAPÍTULO 11

El juego: armá el viaje de un dato

Para que todo esto se te quede, vamos a hacer una dinámica. Se juega en grupos de tres o cuatro y lleva unos veinte minutos.

Los materiales

Cada grupo recibe un conjunto de **fichas**, cada una con una fuente de datos del capítulo 3. Y un tablero con las cinco etapas del capítulo 7 dibujadas como casilleros: Captación, Transporte, Almacenamiento, Procesamiento y Salida.

Cómo se juega

1. Elijan una ficha, por ejemplo *parte de pesca manuscrito*.
2. Ubíquenla en el primer casillero, Captación, y completen: ¿quién lo genera? ¿En qué formato?
3. Avancen por los casilleros y respondan en cada uno qué le pasa al dato. ¿Cómo viaja? ¿Dónde se guarda? ¿Quién lo clasifica y lo etiqueta? ¿Quién lo usa al final?
4. Cuando terminen, completen la **ficha de gobernanza**: propiedad, calidad, estándares, trazabilidad y acceso.

Un ejemplo resuelto

Tomamos el parte de pesca manuscrito del patrón:

Etapa	Qué pasa con el dato
Captación	El patrón anota a mano lances, horas y kilos estimados por especie. Texto libre en papel.
Transporte	Al llegar a puerto, el parte se entrega en la oficina o se fotografía y se manda por mensaje.
Almacenamiento	La foto o el escaneo va al data lake, sin modificar. Se conserva el original.
Procesamiento	Un modelo de lenguaje, como los que vimos en la Unidad 3, lee el parte y lo convierte en filas de una tabla. Se clasifica, se chequea contra la balanza y se etiqueta con el lote.
Salida	Los kilos del parte aparecen en el tablero y Fishy puede usarlos para responder.

Y la ficha de gobernanza de ese mismo dato:

Control	Respuesta
Propiedad	El patrón del barco
Calidad	Se compara contra la balanza de descarga; las diferencias grandes se revisan
Estándares	Misma lista de especies y mismas unidades que el resto de la empresa
Trazabilidad	Cada fila queda asociada al lance y al lote
Acceso	Edita el responsable de captura; el resto sólo lee

IDEA CLAVE

Pasar de un parte escrito a mano a una fila de tabla es lo que hace posible que el dato viaje. Es una tarea en la que los modelos de lenguaje ayudan mucho, siempre que alguien revise el resultado.

PARÁ Y PENSÁ

Hacé el ejercicio con una ficha propia. Elegí la fuente que más te complica y recorré las cinco etapas. Lo que completes acá es la base del mini-entregable.

CAPÍTULO 12

Volvemos a Fishy

Al principio le preguntamos a Fishy por qué había bajado el rendimiento de la merluza, y nos contestó a medias, porque le faltaban la temperatura de cámara y el control de calidad de esa jornada.

Ahora sabemos por qué. Esos dos datos estaban en un cajón y en una carpeta. Ya conocemos lo que hace falta para que Fishy los vea:

- que se **capten** con un responsable y un formato claros,
- que **viajen** hasta un lugar común,
- que se **almacenen** primero en la bodega y después, ya preparados, en la cámara,
- que se **procesen** con clasificado, envasado y etiquetado, y
- que lleguen a la **salida** con una gobernanza que permita confiar en ellos.

Cuando eso está, la misma pregunta puede tener una respuesta entera: el rendimiento bajó, y bajó porque la cámara de ese lote estuvo por encima de la temperatura esperada durante varias horas.



El panorama completo: seis etapas, un espacio unificado de nombres y, debajo, lake, procesamiento y warehouse. Ilustración.

IDEA CLAVE

Un asistente de inteligencia artificial no es más inteligente que los datos que alcanza a ver. Armar el viaje del dato es lo que lo vuelve útil.

El mini-entregable

Para cerrar la unidad, completá esta ficha sobre **un dato de tu empresa**. Una carilla alcanza.

Campo	**Tu respuesta**
Dato elegido	¿Cuál es? ¿En qué paso del recorrido nace?
Tipo	Tabla, texto libre o imagen/audio/video
Dónde queda hoy	Bolsillo, cajón, carpeta o sistema
Captación	¿Quién lo genera y cómo?
Transporte	¿Cómo llega a otro lugar?
Almacenamiento	¿Qué debería guardarse en la bodega? ¿Y en la cámara?
Procesamiento	¿Qué clasificado, procesamiento, envasado y etiquetado necesita?
Salida	¿Quién lo usaría y para decidir qué?
Gobernanza	Propiedad, calidad, estándares, trazabilidad y acceso
Pregunta para Fishy	Una pregunta que hoy no se puede responder y con ese dato sí

Lo que viene

Con el viaje del dato armado, el curso sigue con la Unidad 6, a cargo de Ariel Giamportone: **modelos predictivos**. Ahí vamos a usar los datos que ya están limpios, ordenados y etiquetados para anticipar lo que todavía no pasó. Es la razón por la que dedicamos tanto tiempo a esta unidad: un modelo predictivo sólo es tan bueno como el dato con el que se entrena.

Buen viaje.

Glosario

Todas las palabras del libro, dichas en idioma pesquero.

Fuente de datos

En el recorrido: Captación

El lugar donde nace un dato: el GPS, la balanza, el parte del patrón.

Silo de datos

En el recorrido: Captación

El dato que se queda en el bolsillo del capitán, en el cajón o en la carpeta, y no llega a nadie más.

Data lake

En el recorrido: Bodega del barco

La bodega del dato: guarda todo tal como llegó, sin ordenar pero con su etiqueta de origen.

Extracción (E)

En el recorrido: Descarga

Bajar los datos de la bodega para empezar a trabajarlos.

Staging

En el recorrido: Recepción en planta

La zona de paso donde los datos extraídos esperan, temporalmente, antes de transformarse. Se vacía al terminar la carga.

Transformación (T)

En el recorrido: Clasificado, línea de proceso, envasado y etiquetado

Todo el trabajo de dejar el dato limpio, ordenado, cruzado y etiquetado.

ELT

En el recorrido: Estiba primero, procesa después

La variante del ETL que carga los datos antes de transformarlos, y los transforma dentro del almacén.

Carga (L)

En el recorrido: Estiba en cámara

Guardar los datos ya preparados en el warehouse.

ETL

En el recorrido: Todo el camino de la bodega a la cámara

Extracción, transformación y carga: el recorrido completo del dato.

Data warehouse

En el recorrido: Cámara frigorífica

El depósito de datos ya preparados, listos para consultar.

Big Data

En el recorrido: Una idea, no un lugar

El nombre que se da a la situación en que los datos son tantos, tan rápidos y tan variados que las herramientas comunes no alcanzan.

Las 5 V

En el recorrido: Volumen, velocidad, variedad, veracidad y valor

Las cinco características que describen el Big Data.

Gobernanza de datos

En el recorrido: HACCP del dato

Reglas de propiedad, calidad, estándares, trazabilidad y acceso.

Trazabilidad

En el recorrido: Etiqueta del lote

Poder rastrear de dónde viene un dato y por dónde pasó.

Unified Namespace (UNS)

En el recorrido: Pizarrón de la planta

Un espacio único, con direcciones claras, donde cada fuente publica su dato y cada sistema lo lee. Es el dato de ahora.

Dato del futuro

En el recorrido: El pronóstico del tiempo

Dato que no se midió sino que lo generó un modelo de IA: una predicción, una alerta anticipada o un dato completado. Se guarda aparte, etiquetado como estimado y se compara con lo que realmente pasó.

Modelo de IA

En el recorrido: El que hace el pronóstico

Programa que aprende de los datos de antes y de ahora para estimar lo que todavía no pasó.

Este material fue desarrollado por **Damian Giacone** (www.damiangiacone.com) para el curso **IA aplicada a la Producción Pesquera** de la **Universidad Tecnológica Nacional, Facultad Regional Chubut**.

© Damian Giacone

Esta obra está bajo licencia **Creative Commons Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0)**.

Podés copiarla, distribuirla y adaptarla en cualquier medio o formato, siempre que reconozcas la autoría de Damian Giacone, indiques si hiciste cambios y **no la uses con fines comerciales**. Esta obra no puede ser comercializada.

creativecommons.org/licenses/by-nc/4.0/deed.es