





Programa de especialización

DATA DRIVEN DECISIONS 2026

Nivel Intermedio

SESIÓN 4: Ejercicio práctico de limpieza de datos

TRANSFORMAR AHORA,
trascender siempre

Sesión 4:
Ejercicio práctico de limpieza de
datos e introducción a Power
Query y algunas de sus
operaciones

Logro de aprendizaje

Al finalizar la sesión, lograrás:

Realizar una limpieza de datos básica en Power Query de Power BI

Contenido:

1. Actividad individual limpieza de datos Power Query
2. Estructura tabular y granularidad
3. Identificadores y llave compuestas
4. Normalización básica
5. Append, Merge y tipos de unión
6. Power Query avanzado

1.Actividad individual

Limpieza de datos en Power Query

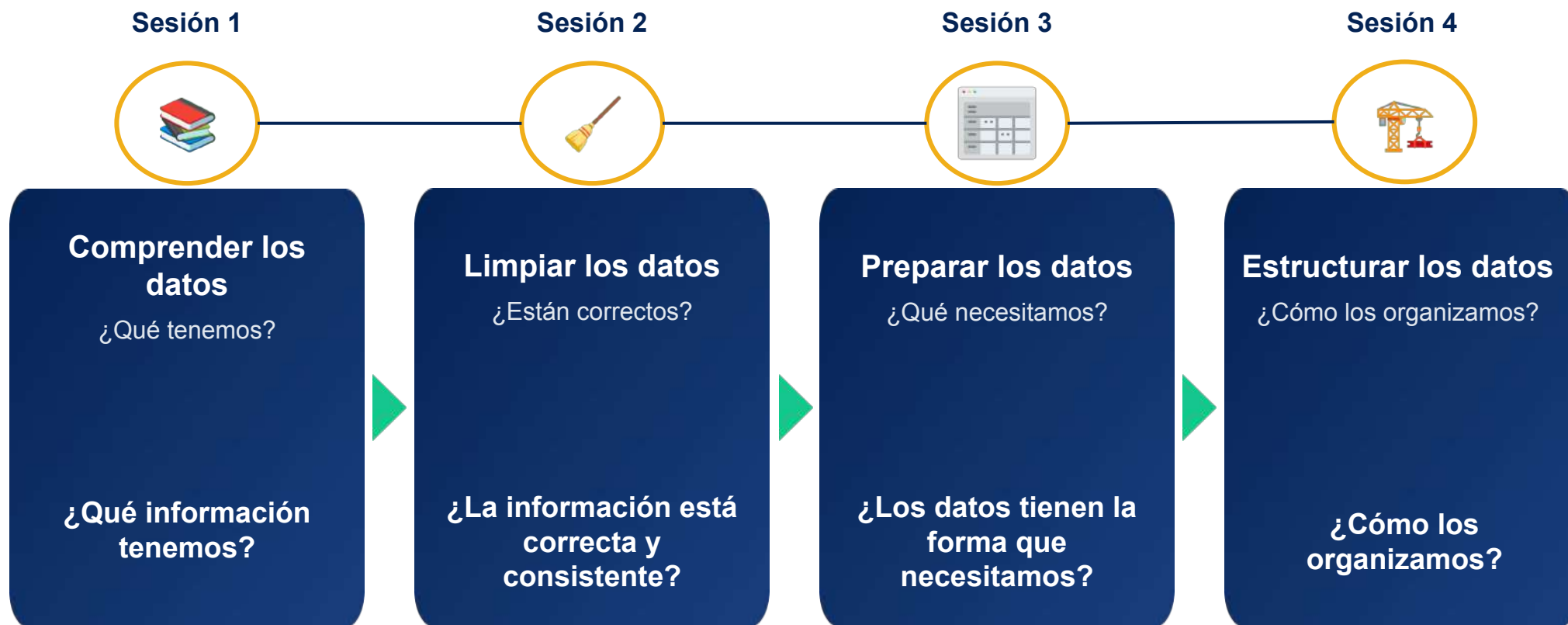
Caso práctico sencillo

Vamos a aplicar el proceso de limpieza en Power Query del excel de la clase pasada



2.Estructura tabular y granularidad

¿Dónde estamos en el proceso?



El reto real al que vamos a llegar

La información rara vez vive en un solo archivo.



El desafío no es “juntar Excel”. Es decidir cómo deben relacionarse para conservar el significado de los datos.

Antes de unir: ¿qué representa una fila?

Una fila no es “un registro” en abstracto. Representa una unidad concreta del proceso.

Si no sabemos qué representa la fila, tampoco podemos saber qué la identifica.

Granularidad

Es el nivel de detalle con el que está construida una tabla.

Una fila = un estudiante

La tabla tiene una fila por persona. Si el estudiante aparece dos veces, debemos investigar por qué.

Una fila = una matrícula

Un estudiante puede aparecer varias veces porque puede matricularse en varios cursos o periodos.

Una fila = una transacción

La misma persona puede tener decenas de filas y cada una seguir siendo válida.

Una fila puede representar...

Persona

ID estudiante

Matrícula

Estudiante + curso + periodo

Solicitud

N.º de solicitud

Pago

N.º de operación

Documento

Código documental

Actividad

Persona + fecha + actividad

Repetido no siempre significa duplicado

DNI	Curso	Periodo
12345678	Excel	2026-1
12345678	IA Generativa	2026-1
12345678	Excel	2026-2

El DNI se repite tres veces, pero las filas representan matrículas diferentes.

La pregunta clave

**¿Qué columna —o conjunto de columnas— hace
única cada fila?**

Ese será el punto de partida para relacionar la tabla con otras fuentes.

3. Identificadores y llave compuestas

Llave simple

Una sola columna identifica de forma única cada fila.

ID_ESTUDIANTE	Nombre	Facultad
E001	Ana Torres	Gestión
E002	Luis Vega	Derecho
E003	María León	Ingeniería

Clave: ID_ESTUDIANTE

Llave compuesta

Varias columnas, juntas, identifican de forma única una fila.

ID_ESTUDIANTE	ID_CURSO	PERIODO	Nota
E001	C101	2026-1	17
E001	C102	2026-1	15
E001	C101	2026-2	18

ID_ESTUDIANTE + ID_CURSO + PERIODO

La cédula puede identificar a la persona...

...pero no necesariamente a la fila que estás analizando.

Tabla de estudiantes

Una fila = una persona
DNI puede ser suficiente

Tabla de matrículas

Una fila = una matrícula
DNI se repite legítimamente

¿Cómo probar si una llave es realmente única?

Cuenta filas

¿Cuántas filas tiene la tabla?

Cuenta combinaciones únicas

¿Cuántas combinaciones distintas tiene la llave propuesta?

Compara

Si ambos conteos son iguales, la llave es candidata a ser única.

Valida con negocio

La unicidad técnica debe tener sentido en el proceso real.

4.Normalización y estructura tabular

Normalización — en términos prácticos

Organizar la información en tablas que representan entidades o eventos diferentes y relacionarlas mediante identificadores.

No necesitamos memorizar formas normales: necesitamos evitar repetición innecesaria y relaciones ambiguas.

¿Por qué evitar una “tabla gigante”?

Repetición

Datos como nombre de facultad, curso o estudiante se repiten una y otra vez.

Inconsistencia

Un mismo valor puede quedar escrito de formas diferentes en distintas filas.

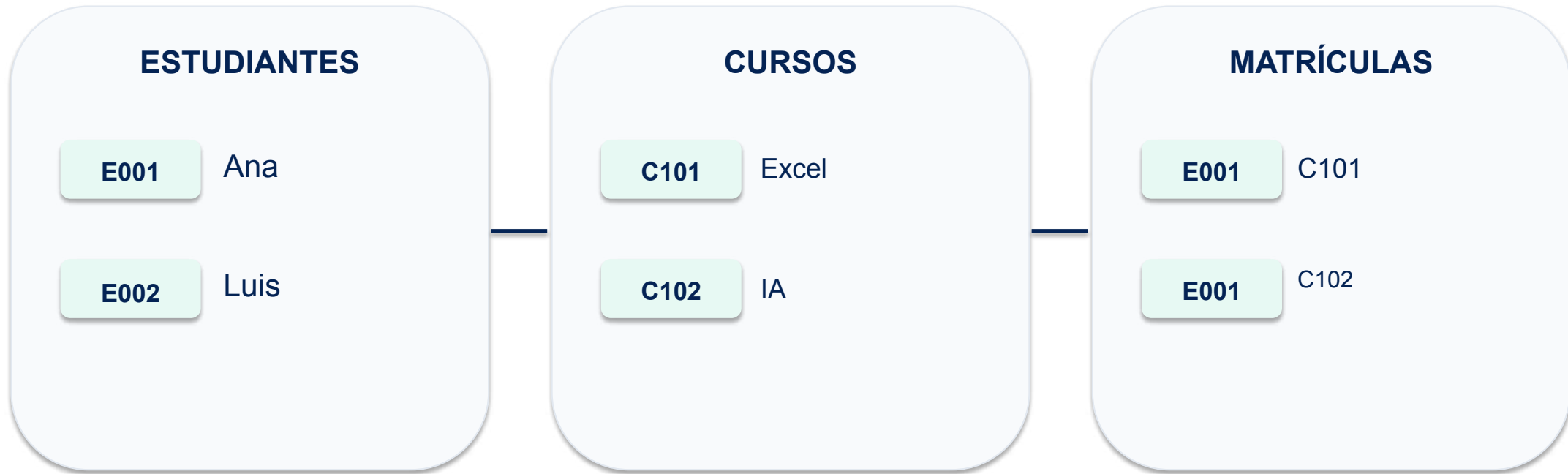
Mantenimiento

Cada cambio exige actualizar múltiples registros y aumenta el riesgo de error.

Ejemplo: una tabla única de matrículas

ID Est.	Nombre	Facultad	Curso	Docente	Periodo
E001	Ana Torres	Gestión	Excel	P. Ruiz	2026-1
E001	Ana Torres	Gestión	IA	L. Soto	2026-1
E002	Luis Vega	Derecho	Excel	P. Ruiz	2026-1
E003	María León	Ingeniería	IA	L. Soto	2026-1

Separar por propósito



Cada tabla conserva su propio nivel de detalle y se conecta mediante llave .

Las tablas se relacionan mediante identificadores



La calidad de una unión depende de la llave usada represente correctamente la relación entre las tablas.

Regla práctica para una buena estructura tabular

Cada fila representa una sola unidad

Evita mezclar niveles de detalle dentro de la misma tabla.

Cada columna representa una variable

No guardes varias variables dentro de una misma celda.

Cada tabla tiene un propósito

Separa maestros, eventos y catálogos cuando corresponda.

Las relaciones usan llave /llaves estables

Evita unir por nombres libres si existe un código o identificador confiable.

5.Append, Merge y tipos de unión

Dos preguntas distintas

¿Quiero más filas?

ANEXAR / APPEND

Mismas columnas + nuevos registros

¿Quiero más columnas?

COMBINAR / MERGE

Información diferente + una llave en común

Append

Agrega filas de tablas que comparten una estructura compatible.

ENERO

ID	Monto
E01	120
E02	90



FEBRERO

ID	Monto
E03	110
E04	70



RESULTADO

ID	Monto
E01	120
E02	90
E03	110
E04	70

¿Cuándo usar Append?

Archivos mensuales

Enero, febrero, marzo... con las mismas columnas.

Sedes o unidades

Cada unidad entrega el mismo formato y queremos consolidarlo.

Históricos

Agregar nuevos periodos a una tabla acumulada.

Exportaciones repetidas

Unir archivos generados por el mismo sistema en momentos distintos.

Merge

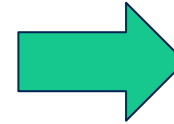
Incorpora columnas de otra tabla a partir de una llave común.

Matrículas

ID	Curso
E01	Excel
E02	IA

Estudiantes

ID	Nombre
E01	Ana
E02	Luis



Columnas incorporadas

ID	Curso	Nombre
E01	Excel	Ana
E02	IA	Luis

La tabla principal conserva sus filas y se enriquece con información de otra fuente.

Ejemplo: Matrículas + Estudiantes

ID_ESTUDIANTE	Curso
E001	Excel
E002	IA
E003	Power BI

ID_ESTUDIANTE	Nombre
E001	Ana
E002	Luis
E003	María

Se relacionan por ID_ESTUDIANTE

Regla mental

¿Quiero más filas? Append

¿Quiero más columnas? Merge

La pregunta parece simple, pero evita una gran cantidad de errores en Power Query.

Antes de hacer Merge

Define exactamente por cuál columna —o columnas— deben coincidir las tablas.

Unir por una columna “parecida” no es lo mismo que unir por una llave válida.

Cardinalidad

Describe cuántas coincidencias puede tener una llave entre dos tablas.

1 : 1

Una llave aparece una vez en A y una vez en B.

1 : N

Una llave aparece una vez en A y varias veces en B.

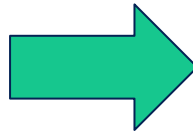
N : N

La misma llave aparece varias veces en ambos lados.

La cardinalidad determina el riesgo de multiplicar filas al combinar.

Relación 1 a 1

ID	Correo
E01	a@pucp.edu.pe
E02	b@pucp.edu.pe

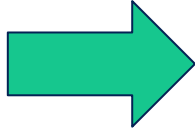


ID	Facultad
E01	Gestión
E02	Derecho

Cada ID tiene una sola coincidencia en ambos archivos.

Relación 1 a muchos

ID	Nombre
E01	Ana
E02	Luis



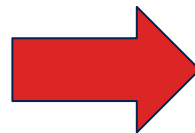
ID	Curso
E01	Excel
E01	IA
E02	Power BI

Un estudiante aparece una vez en el maestro, pero puede tener varias matrículas.

Relación muchos a muchos

Es la que exige mayor cuidado al hacer Merge.

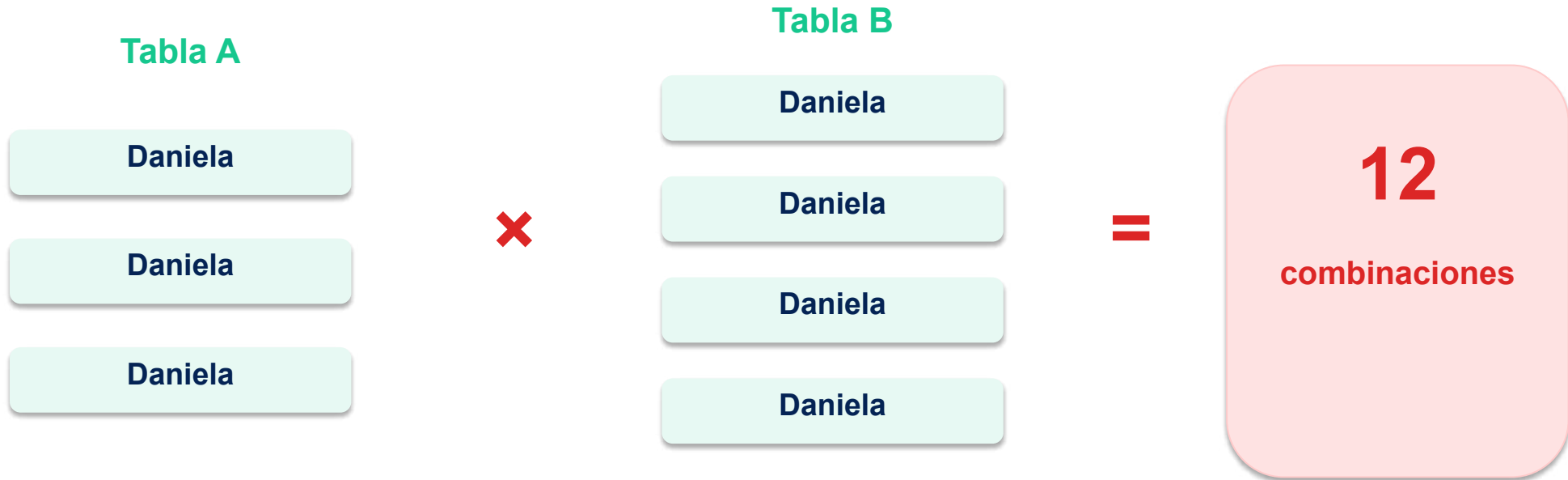
Nombre	Periodo
Ana	2026-1
Ana	2026-2
Luis	2026-1



Nombre	Actividad
Ana	Taller
Ana	Mentoría
Luis	Taller

Si unimos solo por Nombre, cada “Ana” del lado izquierdo puede combinarse con cada “Ana” del lado derecho.

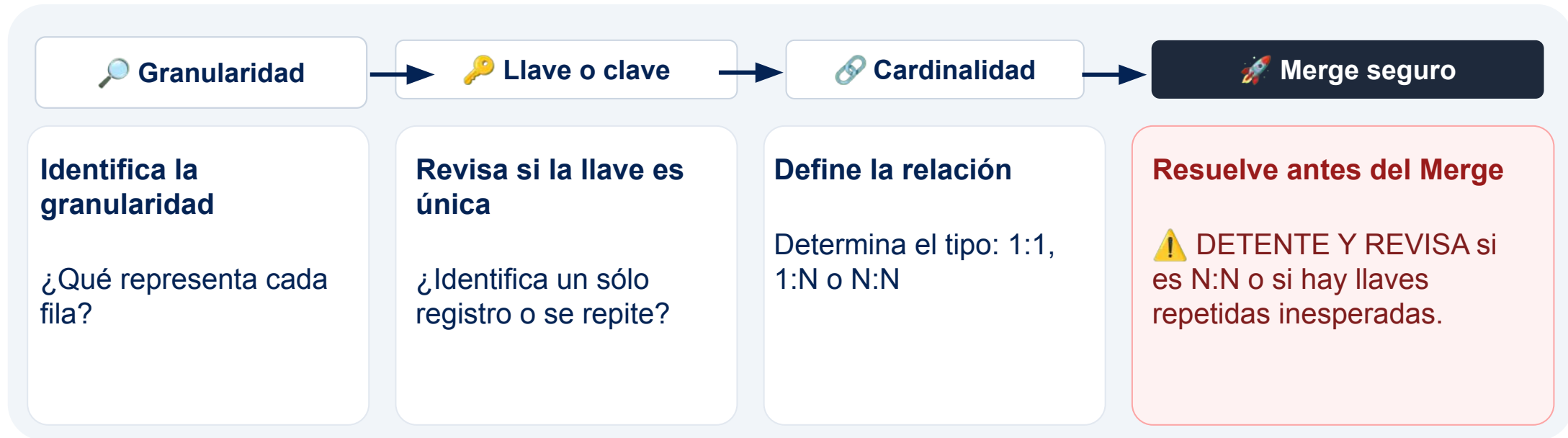
Un Merge puede ejecutarse... y estar mal



Power Query hace exactamente lo que le pedimos; no sabe si la relación tiene sentido para el negocio.

La buena práctica antes de hacer un Merge

Antes de hacer un Merge, verifica la cardinalidad de tus llaves



⚠ Consejo clave
Una **llave repetida en ambas tablas puede multiplicar registros de forma exponencial** al hacer un Merge, distorsionando la granularidad de los datos. No evites N:N ciegamente; comprende la lógica de la multiplicidad antes de proceder.

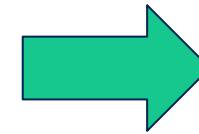
5. Tipos de unión

Left Outer

Conserva todas las filas de la tabla principal (izquierda) y trae coincidencias de la segunda (derecha)

A: ID	Solicitud
1	S01
2	S02
3	S03

B: ID	Nombre
1	Ana
2	Luis



Resultado

1 ✓
2 ✓
3 NULL

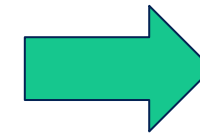
Útil cuando la tabla izquierda define el universo que quieres conservar.

Inner

Conserva únicamente los registros que existen en ambas tablas.

A: ID	Dato
1	A
2	B
3	C

B: ID	Dato
1	X
2	Y
4	Z



Resultado

1

2

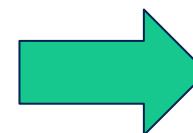
Resultado: solo las llave 1 y 2, porque son las coincidencias comunes.

Full Outer

Conserva todos los registros de ambos archivos, coincidan o no.

A: ID	Dato
1	A
2	B
3	C

B: ID	Dato
1	X
2	Y
4	Z



Resultado

1
2
3
4

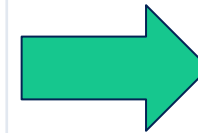
Útil para conciliaciones: permite ver coincidencias y también faltantes de ambos lados.

Left Anti

Devuelve los registros de la tabla principal que NO encontraron coincidencia.

A: ID	Solicitud
1	S01
2	S02
3	S03

B: ID	Nombre
1	Ana
2	Luis



Resultado

3

Ideal para detectar pendientes, faltantes, códigos sin maestro o registros no conciliados.

¿Qué tipo de unión elegir?

Necesidad	Tipo de unión
Conservar todo mi archivo principal	Left Outer

¿Qué tipo de unión elegir?

Necesidad	Tipo de unión
Conservar todo mi archivo principal	Left Outer
Quedarme sólo con coincidencias	Inner

¿Qué tipo de unión elegir?

Necesidad	Tipo de unión
Conservar todo mi archivo principal	Left Outer
Quedarme solo con coincidencias	Inner
Ver todo, coincida o no	Full Outer

¿Qué tipo de unión elegir?

Necesidad	Tipo de unión
Conservar todo mi archivo principal	Left Outer
Quedarme solo con coincidencias	Inner
Ver todo, coincida o no	Full Outer
Encontrar lo que falta en la segunda tabla	Left Anti

El tipo de unión depende de la pregunta del proceso, no de una preferencia técnica.

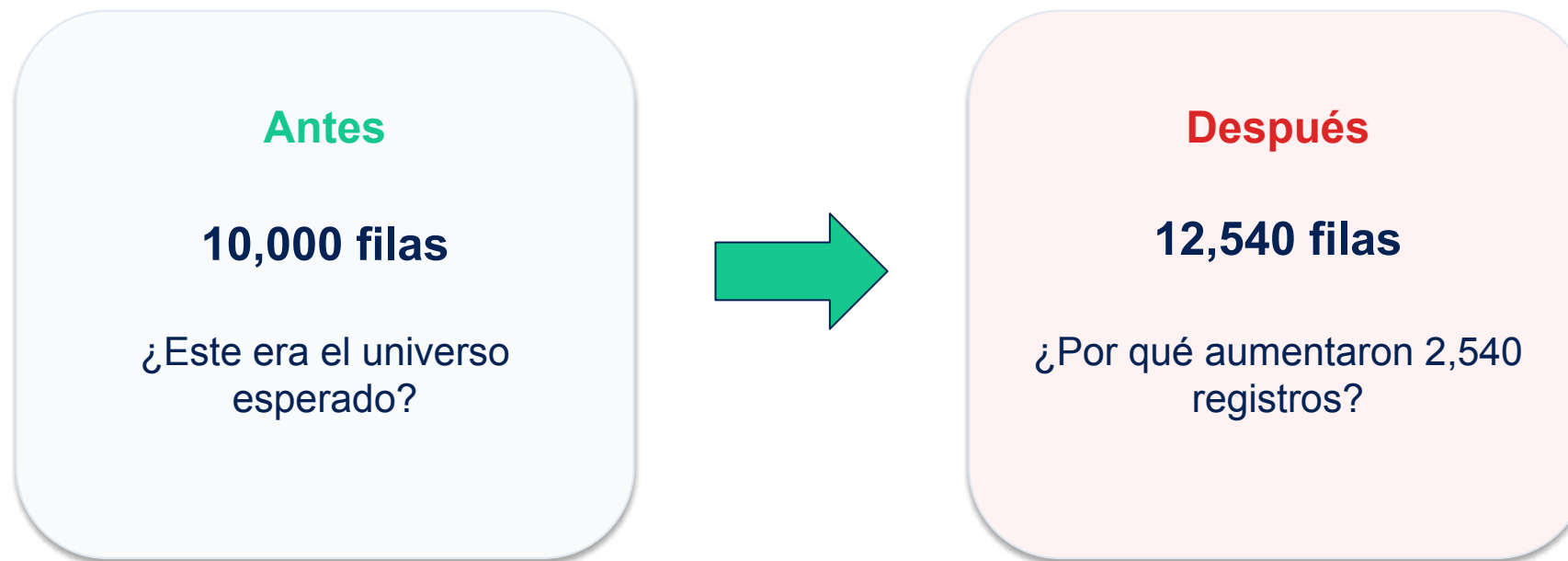
Un registro sin coincidencia también es información

**NULL después de un Merge no siempre significa error.
Puede revelar un pendiente, una ausencia o una
diferencia entre fuentes.**

Antes de eliminarlo, interpreta qué significa para el proceso.

Validar el resultado

Compara el número de filas



Un aumento inesperado suele ser una pista de llave repetidas o una relación muchos a muchos.

Revisa los NULL generados por el Merge

ID	Nombre	Facultad
E001	Ana	Gestión
E002	Luis	Derecho
E003	NULL	NULL

Pregúntate: ¿E003 no existe en el maestro o usamos la llave equivocada?

Buscar duplicados inesperados

ID_ESTUDIANTE	Curso	Periodo
E001	Excel	2026-1
E001	Excel	2026-1
E002	IA	2026-1

Si el nivel de detalle esperado era matrícula única, estas dos filas deben investigarse.

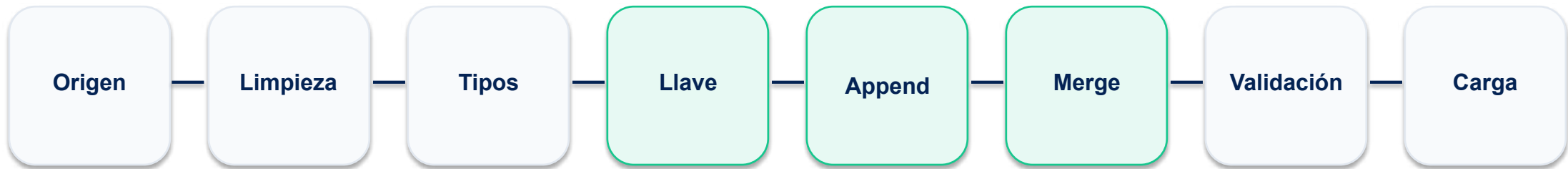
Verifica el nivel de detalle final

Después de combinar, vuelve a responder: ¿qué representa una fila ahora?

Si la respuesta cambió sin que lo esperáramos, la transformación modificó la granularidad.

6.Power Query como flujo reproducible

Power Query = secuencia de transformaciones



Cada paso queda registrado y puede ejecutarse de nuevo cuando llegan archivos actualizados.

Transformaciones frecuentes en Power Query

Cambiar tipos

Texto, número,
fecha

Limpiar texto

Trim, mayúsculas,
reemplazos

Crear llave

Combinar columnas

Append

Consolidar periodos

Merge

Enriquecer tablas

Expandir

Seleccionar
columnas del Merge

Agrupar

Resumir por
categorías

Pivot / Unpivot

Cambiar estructura
de columnas y filas

La ventaja: reproducibilidad

Hoy

Limpias y combinas los archivos una vez.

Mañana

Reemplazas o agregas los archivos actualizados.

Actualizar

Power Query repite los pasos y reconstruye el resultado.

La transformación deja de depender de recordar qué hicimos manualmente.



7. Gemini como copiloto de Power Query

Prompt limitado

“Tengo varios Excel. Dime cómo unirlos en Power Query.”

Gemini tendrá que asumir la estructura, las llaves , el tipo de unión y qué resultado esperas.

Anatomía de un buen prompt para Power Query

Granularidad

Qué representa una fila en cada archivo.

Validación

Qué debe comprobar Gemini antes y después del procedimiento.

Problemas conocidos

Duplicados, formatos, espacios, códigos inconsistentes.

llave

Qué columnas identifican las filas y relacionan las tablas.

Objetivo

Qué tabla quieres construir al final.

Tipo de unión

Qué información debe conservarse.



Prompt modelo

Describe el problema antes de pedir el paso a paso.

Tengo tres archivos de Excel.

- Matrículas: una fila representa la matrícula de un estudiante en un curso y periodo. La clave es ID_ESTUDIANTE + ID_CURSO + PERIODO.
- Estudiantes: una fila representa un estudiante. La clave es ID_ESTUDIANTE.
- Cursos: una fila representa un curso. La clave es ID_CURSO.

Necesito conservar todas las matrículas e incorporar el nombre del estudiante y del curso. Explícame paso a paso cómo hacerlo en Power Query. Antes de darme el procedimiento, verifica si las relaciones pueden generar duplicados y dime qué debería revisar al final.

Gemini puede explicar; el criterio sigue siendo humano

La IA puede sugerir pasos, código M y validaciones. Pero no conoce por sí sola qué representa cada fila ni qué relación es correcta para tu proceso.

Primero define la lógica de datos; después usa la IA para acelerar la ejecución.

Caso práctico: avanzado (próxima clase)

Solicitudes_enero.xlsx

Misma estructura que febrero

Solicitudes_febrero.xlsx

Debe anexarse a enero

Personas.xlsx

Maestro de personas

Unidades.xlsx

Maestro de unidades

Los archivos incluyen errores de formato, llaves repetidas y algunos registros sin correspondencia.



Paso 1. Diagnóstico

- ¿Qué representa una fila en cada archivo?
- ¿Cuál es el identificador?
- ¿Existe alguna llave compuesta?
- ¿Qué problemas deben corregirse antes de unir?



Paso 2. Diseña la estrategia de unión

Append

¿Qué archivos agregan nuevas filas?

Merge

¿Qué archivos incorporan nuevas columnas?

Join

¿Qué registros deben conservarse cuando no hay coincidencia?

Paso 3. Pídele a Gemini el procedimiento

Entrégale el diagnóstico

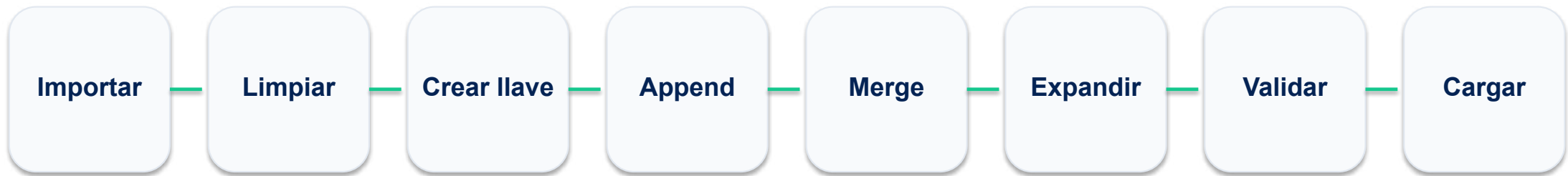
Granularidad, llaves, problemas conocidos y resultado esperado.

Pide el paso a paso

Solicita acciones específicas en Power Query y las validaciones que debes realizar.

No le pidas solo “unir archivos”: entrégale la lógica que ya definiste.

Paso 4. Ejecuta en Power Query



Si un paso no coincide con tu diagnóstico, cuestiona la recomendación de Gemini y corrígela.

Paso 5. Valida

Filas

¿El número de filas cambió como esperaba?

Nulos

¿Qué registros no encontraron coincidencia?

Duplicados

¿Aparecieron combinaciones inesperadas?

Granularidad

¿Qué representa una fila del resultado final?



Checklist final



¿Sé qué representa una fila en cada tabla?



¿Definí una llave válida?



¿Diferencié Append de Merge?



¿Elegí el tipo de unión correcto?



¿Validé filas, nulos y duplicados?



¿El flujo quedó reproducible en Power Query?

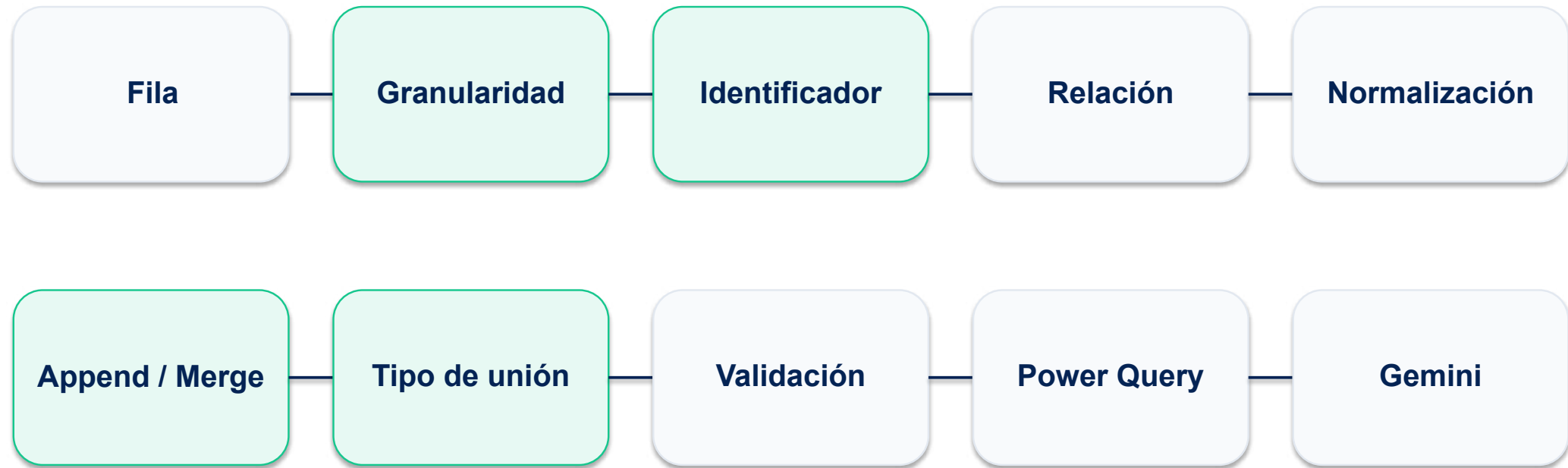


¿Gemini recibió suficiente contexto?



¿Revisé críticamente su propuesta?

Resumen de la sesión



Primero entendemos la lógica. Después usamos las herramientas para ejecutarla.

Evaluación de la Sesión



<https://docs.google.com/forms/d/e/1FAIpQLSefDWVWtRfqMuFD1mJMU3xsePBLH3zG-sga1nLQpqBigGBwyA/viewform?usp=dialog>

Fin Sesión 4



La *transformación digital*
es parte de nuestro

Plan
Estratégico
Institucional

23
27



Conoce más ↗