

Trabajo Práctico de Bases de Datos

Grupo 5: LOS_LINDOS

Curso: K3522

Migración de datos, generación de modelo
transaccional y modelo de inteligencia de negocios
(Business Intelligence)

Integrantes:

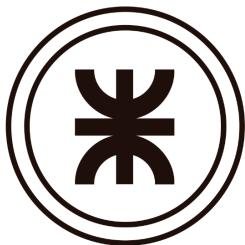
Agostini Dohmen, Yanel - 2079112

Nardelli Lorenzo, Federico Alejandro Ariel - 2099111

Vultaggio, Martín Salvador - 2092347

Vultaggio, Pablo Nicolás - 2095488

Fecha de entrega: 21/11/2025



UTn.BA

UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL BUENOS AIRES

Índice

Procedimientos de tablas nivel 1, 2 y 3	2
Procedimientos de tablas nivel 4, 5 y 6	3
Detalle sobre dimensiones, consideración sobre tiempo	4
Explicación en general, tabla sobre tablas de hechos	5
Tablas sobre tablas de hechos, explicación de vistas	6
Explicación de vistas, consideraciones importantes	7
Consideraciones importantes	8

Estrategia para el modelo transaccional

Procedimientos de Tablas de Nivel 1

- Tabla Estado: Lo extrae de Inscripción_Estado. Extrae y normaliza los nombres de los estados de inscripción
- Tabla Dia_Semana: Extrae de la columna Curso_Dia y normaliza los días de la semana en los que se dictan los cursos
- Tabla Categoría: Normaliza las categorías únicas de los cursos desde la columna Curso_Categoría
- Tabla Turno: Normaliza los tipos de turnos disponibles desde la columna Curso_Turno
- Tabla Institucion: Normaliza los datos únicos de las instituciones, provenientes de las columnas Institucion_Cuit, Institucion_Nombre y Institucion_RazonSocial
- Tabla Provincia: Lo extrae de Sede_Provincia, Profesor_Provincia y Alumno_Provincia mediante un UNION. Consolida y normaliza la lista unica de todas las provincias mencionadas en las diferentes entidades
- Tabla Direccion: Lo extrae de Sede_Direccion, Profesor_Direccion y Alumno_Direccion mediante un UNION. Consolida y normaliza la lista unica de todas las direcciones mencionadas en las diferentes entidades
- Tabla Localidad: Lo extrae de Sede_Localidad, Profesor_Localidad y Alumno_Localidad mediante un UNION. Consolida y normaliza la lista unica de todas las localidades mencionadas en las diferentes entidades
- Tabla Medio_Pago: Lo extrae de Pago_MedioPago, normalizando los tipos unicos de medios de pago
- Tabla Pregunta: Lo extrae de Encuesta_Pregunta1 a Encuesta_Pregunta4 mediante UNION, consolidando y normalizando los enunciados de las preguntas de la encuesta

Procedimiento de Tablas de Nivel 2

- Tabla Sede: Las sedes son las ubicaciones físicas donde se dictan los cursos. Tiene relaciones con Localidad, Direccion y Provincia para normalizar las direcciones y ubicaciones.
 - Tabla Alumno: Almacena la información de los alumnos, relacionándolos con las direcciones, localidades y provincias correspondientes. También incluye detalles personales como nombre, apellido, DNI y teléfono para su identificación.
 - Tabla Profesor: Almacena la información de profesores y los asocia a sus respectivas direcciones, localidades y provincias. Los profesores tienen un perfil detallado (nombre, apellido, fecha de nacimiento, etc.).
- Tabla Curso: Relaciona a los cursos con una sede, un profesor, una categoría, un turno, y define las fechas de inicio y fin, así como el precio.
- Tabla Inscripción_Curso: Registra las inscripciones de los alumnos en los cursos, relacionando cada inscripción con un alumno y un curso, incluyendo la fecha de inscripción.
 - Tabla Estado_de_Inscripcion: Registra el estado de cada inscripción ("Pendiente", "Aprobado"), relacionado con cada inscripción en los cursos.

Procedimientos de Tabla de Nivel 3

- Tabla Factura: Asocia las facturas con los alumnos. Contiene la información del número de factura, fechas de emisión y vencimiento, y el importe total. Relacionado con la tabla Alumno para saber quién hizo el pago.

Procedimientos de Tablas de Nivel 4

- Tabla Curso_x_Alumno: Tabla de relación que asocia a alumnos con cursos a los que están inscritos, permitiendo relación de muchos a muchos entre alumnos y cursos
- Tabla Curso_x_Dia: Relaciona los cursos con los días de la semana que se dictan, permite asociar múltiples cursos a múltiples días.
- Tabla Trabajo_Practico: Registra las calificaciones de los trabajos prácticos de los alumnos , relacionando un alumno con un curso y su nota
- Tabla Final: Registra los exámenes finales de los cursos, con su fecha, hora y descripción del examen final
- Tabla Encuesta: Asocia las encuestas a cursos específicos para juntar la retroalimentación de los alumnos, guardando la fecha y observaciones
- Tabla Detalle_factura: Guarda los detalles de cada factura (el importe de cada curso por ejemplo)
- Tabla Pago: Registra los pagos hechos para las facturas, relacionando las facturas con los medios de pago

Procedimientos de Tablas de Nivel 5

- Tabla Parcial: Registra los exámenes parciales de los cursos, especificando el nombre y la descripción del módulo.
- Tabla Inscripcion_de_final: Registra las inscripciones de los alumnos en los exámenes finales, lo que permite saber si los alumnos están inscritos en cierto final
- Tabla Respuesta: Vincula las respuestas dadas por los alumnos con las pregunta y encuestas correspondientes, permite evaluar los resultados de las encuestas

Procedimientos de Tablas de Nivel 6

- Tabla Parcial_de_Alumno: Guarda los resultados de cada alumno en cada parcial (su nota, la instancia y la presencia),, relacionado directamente con entidades Alumno y Parcial
- Tabla Evaluación_de_final: Registra los resultados de los alumnos en los finales, guarda también el profesor que evalúa, la nota y asistencia

Estrategia para el modelo de inteligencia de negocio

Se implementó un modelo estrella: tablas de hechos centrales rodeadas de tablas de dimensiones que la contextualizan. Esto permite análisis rápidos, intuitivos y altamente desnormalizados, ideal para consultas analíticas (OLAP) en Power BI.

Dimensión	Qué representa
BI_DIMENSION_Tiempo	Cada fila representa un período en el sistema, un período está compuesto por un año, semestre y mes.
BI_DIMENSION_Sede	Cada sede física de la institución (nombre, localidad, provincia, institución, etc.)
BI_DIMENSION_Turno	Los distintos turnos posibles (Mañana, Tarde o Noche)
BI_DIMENSION_Categoría_Curso	Las categorías o áreas temáticas de los cursos.
BI_DIMENSION_Rango_Etario_Alumno	Grupos etarios de alumnos: <25, 25-35, 35-50, >50
BI_DIMENSION_Rango_Etario_Profesor	Grupos etarios de profesores: 25-35, 35-50, >50
BI_DIMENSION_Bloque_Satisfacción	Bloques de calificación de encuestas: Satisfechos (7-10), Neutrales (5-6), Insatisfechos (0-4)
BI_DIMENSION_Medio_Pago	Formas de pago disponibles

Para los valores de las dimensiones de tiempo se tomaron los valores:

Años: Entre 2019 y 2025

Cuatrimestre: 1 y 2

Mes: Entre 1 y 12

Notamos que en la consigna se usó de manera indiscriminada tanto la palabra “Cuatrimestre” como “Semestre”, por lo que es importante saber que en nuestra implementación consideramos que ambas se tratan de lo mismo, considerando como al primer “Semestre” o “Cuarto trimestre” a los primeros 6 meses del año y el resto, al segundo.

Un aspecto de nuestra resolución importante que queremos mencionar es que, como consideramos que las distintas vistas requieren información agrupada de distintas maneras, todas requieren agrupación distinta para conseguir información. Por lo que no tendría sentido un **HIPERCUBO** que contenga toda la información agrupada de maneras que no tengan sentido. Por lo que nos pareció más eficiente generar tablas de hechos distintas (**MULTICUBO**) para cada parte de información que requiera distinta agrupación y además se relacionen con alguna entidad específica del modelo transaccional.

Nuestra premisa fue intentar maximizar la cantidad de dimensiones usadas por cada tabla de hechos (siempre que una dimensión tenga sentido para una entidad determinada), esto permite que la agrupación de la tabla de hechos sea mucho mayor, y que, el universo de vistas que se puedan generar sea mucho mayor, no solo las que se pidieron en la consigna, si no que podamos generar vistas que agrupen un montón de combinaciones de dimensiones posibles.

Esta decisión permite a cada vista agrupar en las dimensiones que les interese, ignorando las que no necesite, la posibilidad de elección es mucho mejor, que si se generaran tablas de hechos con pocas dimensiones. Además, elegir muchas dimensiones produce mucha mayor calidad en la información a la hora de analizar los hechos, y se pueden atribuir las distintas piezas de información a distintas causas, aún si quizás pudiese parecer que una dimensión determinada no haga diferencia en el agrupamiento de la información.

Por ejemplo, pudiese parecer que no tienen nada que ver la categoría de un curso con el porcentaje de desvío de pagos, pero quizás, tras un profundo análisis, descubrimos que esa categoría de curso es la más atendida por estudiantes que ya trabajan y por eso suelen pagar a tiempo, en cambio quizás en otras categorías de cursos, los pagos suelen hacerse a destiempo.

Tabla de hechos	Granularidad	Métricas clave principales	Vistas asociadas
Inscripciones	Tiempo Sede Rango etario alumno Rango etario profesor Turno Categoría de curso	Inscripciones totales, aceptadas y rechazadas	1 y 2
Cursadas	Tiempo Sede Rango etario alumno Rango etario profesor Turno Categoría de curso	Cantidad de aprobados, desaprobados, cantidad total de cursadas y tiempo promedio de finalización de curso en día	3 y 4
Finales	Tiempo Sede Rango etario alumno Rango etario profesor Turno	Promedio de notas, cantidad de ausentes y la cantidad de inscriptos	5 y 6

	Categoría de curso		
Pagos	Tiempo Sede Rango etario alumno Rango etario profesor Turno Categoría de curso Medio de pago	Cantidad de pagos total, pagos en y fuera de término, ingresos totales e ingresos en y fuera de término	7 y 9
Facturas	Tiempo Sede Rango etario alumno Rango etario profesor Turno Categoría de curso	Cantidad total de facturas, facturas pagadas, facturas sin pagar, facturación esperada, importes adeudados y facturados	8
Encuestas	Tiempo Sede Rango etario profesor Turno Categoría de curso Bloque de satisfacción	Cantidad de respuestas que cumplen el determinado nivel de satisfacción	10

Las tablas de hechos no poseen la información final de las vistas, pero facilitan el acceso a estas a información 100% desnormalizada para que las vistas, mediante un leve agrupamiento y cálculo procesamiento de datos, genere la información que interesa.

1. VISTA_TOP3_CATEGORIA_TURNO_INSCRIPCIONES

Agrupa la tabla de hechos BI_FACT_INSCRIPCIONES por año, sede, categoría y turno, suma la cantidad de inscripciones y aplica ROW_NUMBER() particionado por año y sede para quedarse con las 3 combinaciones categoría y turno con más inscriptos.

2. VISTA_TASA_RECHAZO

Sobre BI_FACT_INSCRIPCIONES agrupa por mes y sede, suma inscripciones totales, rechazadas y aceptadas, y calcula los porcentajes de rechazo y confirmación sobre el total de inscripciones del mes.

3. VISTA_PORCENTAJE_APROBACION_Y_DESAPROBACION

Usa BI_FACT_CURSADAS, agrupa por año y sede, suma cursadas totales, aprobadas y desaprobadas, y calcula el % de aprobación y desaprobación de cursadas por sede y año.

4. VISTA_PROMEDIO_FINALIZACION

Parte de BI_FACT_CURSADAS, agrupa por año y categoría de curso, y promedia los días promedio de finalización ya calculados en la tabla de hechos (AVG sobre promedio_finalizacion_dias).

5. VISTA_PROMEDIO_NOTAS_FINALES

Sobre BI_FACT_FINALES agrupa por semestre, rango etario alumno y categoría, suma inscriptos y ausentes, y promedia la columna promedio_de_notas.

6. VISTA_TASA_AUSENTISMO_FINALES

Con BI_FACT_FINALES agrupa por semestre y sede, suma inscriptos y ausentes al final, y calcula el porcentaje de ausentismo (y su complemento: presentes).

7. VISTA_DESVIO_DE_PAGOS

Usa BI_FACT_PAGOS, agrupa solo por semestre (ignora el resto de dimensiones), suma pagos totales, pagos en término y fuera de término, y calcula los porcentajes correspondientes.

8. VISTA_TASA_DE_MOROSIDAD_MENSUAL

Sobre BI_FACT_FACTURAS agrupa por mes, suma facturación esperada, importe cobrado y adeudado, y calcula la tasa de morosidad.

9. VISTA_TOP3_CATEGORIAS_INGRESOS

Parte de BI_FACT_PAGOS, agrupa por año, sede y categoría, suma ingresos totales y aplica ROW_NUMBER() particionado por año y sede, para devolver las 3 categorías con mayores ingresos.

10. VISTA_INDICE_DE_SATISFACCION

Sobre BI_FACT_ENCUESTA agrupa por año, sede y rango etario profesor, calcula % de satisfechos y % de insatisfechos, aplica la fórmula $((\% \text{ satisfechos} - \% \text{ insatisfechos}) + 100)/2$ y lo devuelve con 2 decimales.

Consideraciones importantes:

- En la ejecución de el script del modelo de inteligencia de negocios, salta 1 warning, esto fue revisado por el grupo y se llegó a la conclusión de que **está bien que esto suceda**, ya que ocurre cuando se intenta averiguar el promedio de notas cuando resulta que ningún estudiante atendió a los finales para esas dimensiones determinadas, entonces el promedio de notas de los finales cuando los alumnos no se anotaron, es NULL.
- Hay casos que las tablas de hechos o las propias vistas dan información que podría complementarse entre sí, por ejemplo, dar la cantidad de finales, la cantidad de aprobados y la cantidad de desaprobados. En estos casos teniendo ya la cantidad total de finales y la cantidad de aprobados, se podría obtener la cantidad de desaprobados restando a la cantidad total de finales la cantidad de aprobados, sin embargo nosotros elegimos no hacerlo de esta manera, y utilizar una subquery de todas formas para obtener este valor restante, ya que de

esta forma, si llegamos a un resultado consistente (efectivamente total = aprobados + desaprobados), se verifica la efectividad y exactitud de las subqueries utilizadas.

- Cuando se habla de “top 3 de categorías y turnos” en la vista 1, nosotros entendemos que habla de las 3 duplas de categorías y turnos con más inscripciones para las dimensiones especificadas, pudiendo ser las tres del mismo turno (variando la categoría) o la misma categoría (variando el turno).
- Tratamos a la información de “facturas” y “pagos” en distintas tablas de hechos ya que queremos usar la dimensión “medio de pago” para los pagos, y esto no tendría sentido para la tabla de facturas, ya que obtiene la cantidad de facturas sin pagar, y como justamente no tienen pago registrado, no tienen medio de pago, y la dimensión querría en NULL, lo cual es imposible. Aprovechamos esto para, averiguar toda la información posible en ambas tablas de hechos, y a pesar de sus similitudes, estas buscan información diferente, por ejemplo, la tabla de pagos obtiene los montos pagados fuera de término, pero pagados al fin, y la de facturas obtiene los montos que directamente ni se pagaron, y se utilizan en las vistas respectivas de forma adecuada.
- Cuando se habla de obtener información “por mes”, nosotros entendemos que se habla de un mes de un año determinado, por lo que también se agrupa por año. Por ejemplo, cuando queremos averiguar las inscripciones, no averiguamos las inscripciones que hayan sido en Mayo (pero que podrían ser de distintos años), si no que averiguamos las inscripciones que ocurrieron en Mayo de 2019, Mayo de 2020, Mayo de 2021, etc. Análogamente hacemos este cuando se pide agrupar por semestre.