



Aprender
Online UTN

Curso de SQL para Ciencia de Datos desde cero

Domina el lenguaje universal de los datos y conviértete en el profesional que transforma tablas en decisiones estratégicas

TEMARIO OFICIAL DEL CURSO

SQL desde Cero para Ciencia de Datos: del Primer SELECT al Análisis Profesional

Domina el lenguaje universal de los datos y conviértete en la persona que transforma tablas en decisiones.

Ficha técnica

Duración total	30 horas estimadas de estudio (8 unidades de 2 a 4 horas cada una, más ejercicios, actividades interactivas, cuestionarios y examen final)
Cantidad de unidades	8 unidades didácticas + Examen Final
Nivel	Inicial a intermedio-avanzado. Comienza desde cero absoluto y llega hasta funciones de ventana, CTE y análisis de datos profesional.
Modalidad	On-demand, 100% autónoma, sin entregables. Evaluación automática mediante cuestionarios con retroalimentación.
Requisitos previos	Ninguno de tipo técnico. Solo se requiere manejo básico de una computadora (abrir programas, descargar y guardar archivos) y una computadora con Windows, macOS o Linux con 2 GB de espacio libre. Todo el software utilizado es gratuito y su instalación se explica paso a paso.

Herramientas	PostgreSQL 18 (última versión estable), con la herramienta gráfica pgAdmin 4 y la consola psql. En la unidad final se muestra la integración con Python (psycopg2) y pandas.
Fuentes	Estándar ISO/IEC 9075 (SQL), documentación oficial de PostgreSQL (postgresql.org) y pgAdmin (pgadmin.org), documentación oficial de Python y pandas.
Proyecto final	No aplica (curso sin entregables). El cierre integrador es la Unidad 08, donde el alumno realiza un análisis completo de la base NovaMarket con técnicas profesionales, y el Examen Final de 20 preguntas que evalúa todo el curso.

Metodología

El curso utiliza una metodología de **aprendizaje autónomo guiado** con un hilo conductor único: la base de datos **NovaMarket**, una tienda online ficticia de tecnología y hogar inteligente. Todos los ejemplos, ejercicios y evaluaciones del curso trabajan sobre estos mismos datos, que crecen en complejidad unidad tras unidad. De este modo, cada concepto nuevo se ancla en un contexto ya conocido.

Cada unidad sigue el ciclo **leer** → **ver** → **practicar** → **autoevaluarse**:

Momento	Recurso	Qué hace el alumno
1. Leer	Contenido y archivos	Estudia la clase completa ejecutando cada consulta en su computadora. Teoría, analogías, código comentado línea por línea, errores comunes y buenas prácticas.
2. Ver	videos	Refuerza los conceptos clave con videos cortos (5 a 15 minutos) que muestran las consultas en vivo.
3. Practicar	Ejercicios	Resuelve ejercicios graduados de menor a mayor dificultad y luego compara con las soluciones explicadas paso a paso.

Momento	Recurso	Qué hace el alumno
4. Autoevaluarse	Actividad Interactiva + Cuestionario	Fija conocimientos con una actividad interactiva y verifica su comprensión con un cuestionario con retroalimentación inmediata.

Qué aprenderá el alumno

- Comprender qué es una base de datos relacional, qué es SQL y por qué es el lenguaje central del mundo de los datos.
- Instalar y utilizar PostgreSQL, el motor de base de datos abierto más importante de la industria, con su herramienta gráfica pgAdmin.
- Escribir consultas `SELECT` con filtros, ordenamientos, alias y límites.
- Dominar los tipos de datos, las funciones de texto, numéricas y de fecha, las expresiones `CASE` y el tratamiento correcto de `NULL`.
- Resumir información con funciones de agregación, `GROUP BY` y `HAVING`.
- Combinar múltiples tablas con todos los tipos de `JOIN` y operaciones de conjuntos.
- Construir consultas complejas y legibles con subconsultas, CTE (`WITH`) y vistas.
- Diseñar, crear y mantener bases de datos: DDL, DML, claves, restricciones, integridad referencial y transacciones.
- Aplicar SQL a la ciencia de datos: funciones de ventana, estadística descriptiva, limpieza y perfilado de datos, e integración con Python y pandas.

Competencias adquiridas

Competencia	Descripción
Consulta de datos	Formular en SQL cualquier pregunta de negocio sobre una base de datos relacional, desde listados simples hasta análisis multicapa.
Modelado relacional	Diseñar tablas correctas con claves primarias y foráneas, restricciones e integridad referencial, aplicando los principios de la normalización.

Competencia	Descripción
Análisis de datos	Producir métricas, rankings, tendencias temporales y segmentaciones (como el análisis RFM) usando agregaciones y funciones de ventana.
Calidad de datos	Detectar y corregir duplicados, valores faltantes, inconsistencias y valores atípicos: el trabajo diario de la ciencia de datos.
Integración con el ecosistema de datos	Conectar una base SQL con Python y pandas para continuar el análisis, graficar o modelar.
Autonomía profesional	Leer documentación oficial, escribir SQL legible y mantenible, y aplicar buenas prácticas de la industria.

Programa detallado por unidad

Unidad 01 — Fundamentos de bases de datos y primer contacto con SQL

Tiempo estimado: 3 horas

Temas y subtemas

- El mundo de los datos: qué es un dato, qué es información, por qué las planillas de cálculo se quedan cortas.
- Bases de datos relacionales: tablas, filas, columnas; el modelo relacional; sistemas de gestión (SGBD) y panorama actual (PostgreSQL, MySQL, SQL Server, Oracle).
- Qué es SQL: historia, el estándar ISO/IEC 9075, dialectos, y las sublenguas del lenguaje (DQL, DDL, DML, DCL, TCL).
- Instalación paso a paso de PostgreSQL 18 y pgAdmin 4 en Windows, macOS y Linux.
- La base NovaMarket: presentación del caso, su modelo de datos y carga del script inicial.
- Primeras consultas: `SELECT * FROM tabla`, selección de columnas, comentarios y reglas de escritura del lenguaje.

Resultados esperados

El alumno explica con sus palabras qué es una base de datos relacional y qué es SQL, tiene su entorno de trabajo instalado y funcionando, carga la base NovaMarket y ejecuta sus primeras consultas de selección de columnas.

Unidad 02 — SELECT: consultar, filtrar y ordenar datos

Tiempo estimado: 3,5 horas

Temas y subtemas

- Anatomía completa de una consulta: orden de escritura vs. orden lógico de evaluación.
- Filtrado con `WHERE` : operadores de comparación, `AND / OR / NOT` , precedencia y paréntesis.
- Operadores especiales: `BETWEEN` , `IN` , `LIKE` y comodines, `IS NULL` .
- Alias con `AS` , eliminación de duplicados con `DISTINCT` .
- Ordenamiento con `ORDER BY` (múltiples criterios, ASC/DESC, NULL al ordenar) y paginación con `LIMIT / OFFSET` .

Resultados esperados

El alumno responde preguntas de negocio reales filtrando y ordenando datos con precisión, comprende cómo el motor evalúa cada cláusula y evita los errores clásicos con NULL y con la lógica booleana.

Unidad 03 — Tipos de datos, funciones y expresiones

Tiempo estimado: 3,5 horas

Temas y subtemas

- Tipos de datos de PostgreSQL: números enteros y decimales (INTEGER, NUMERIC), texto (TEXT, VARCHAR), fechas (DATE, TIMESTAMP) y booleanos (BOOLEAN); su relación con el estándar SQL.
- Funciones de texto: `UPPER`, `LOWER`, `LENGTH`, `SUBSTR`, `TRIM`, `REPLACE`, concatenación.
- Funciones numéricas: `ROUND`, `ABS`, aritmética y el peligro de la división entera.
- Fechas y horas: formatos ISO 8601, `CURRENT_DATE`, `AGE()`, `EXTRACT()`, `TO_CHAR()`, cálculos de edad y antigüedad.
- Lógica condicional con `CASE`; funciones `COALESCE` y `NULLIF`; la lógica de tres valores y el tratamiento profesional de NULL.
- Conversión de tipos con `CAST`.

Resultados esperados

El alumno transforma y limpia valores de cualquier tipo, construye columnas calculadas y categorías con CASE, maneja NULL sin sorpresas y entiende las diferencias de tipos entre motores.

Unidad 04 — Agregación y agrupamiento: el corazón del análisis

Tiempo estimado: 4 horas

Temas y subtemas

- Funciones de agregación: `COUNT`, `SUM`, `AVG`, `MIN`, `MAX` ; `COUNT(*)` vs. `COUNT(columna)` vs. `COUNT(DISTINCT)`.
- `GROUP BY` : qué significa agrupar, agrupación por una y varias columnas, la regla de oro de las columnas del `SELECT`.
- `HAVING` vs. `WHERE`: filtrar grupos vs. filtrar filas.
- El orden lógico completo de evaluación de una consulta.
- KPIs de negocio reales: ticket promedio, ventas por mes, productos más vendidos, tasa de reseñas.
- Agregación y `NULL`; errores clásicos del agrupamiento.

Resultados esperados

El alumno calcula indicadores de negocio a partir de datos crudos, elige correctamente entre `WHERE` y `HAVING` y explica el orden de evaluación de una consulta con agregación.

Unidad 05 — JOIN: combinar tablas y modelar relaciones

Tiempo estimado: 4 horas

Temas y subtemas

- Por qué los datos viven en varias tablas: claves primarias y foráneas, relaciones 1-a-muchos y muchos-a-muchos.
- `INNER JOIN` paso a paso; alias de tabla; unión de 3, 4 y 5 tablas.
- `LEFT JOIN` (y `RIGHT/FULL`): encontrar lo que falta; el patrón anti-join con `IS NULL`.
- Self join, `CROSS JOIN` y productos cartesianos accidentales.
- Operaciones de conjuntos: `UNION`, `UNION ALL`, `INTERSECT`, `EXCEPT` .
- `JOIN` + `GROUP BY`: los informes multi-tabla que se usan en la industria.

Resultados esperados

El alumno combina cualquier número de tablas eligiendo el `JOIN` correcto, detecta filas sin correspondencia y produce informes completos que cruzan clientes, pedidos, productos y reseñas.

Unidad 06 — Subconsultas, CTE y vistas

Tiempo estimado: 4 horas

Temas y subtemas

- Subconsultas escalares, de columna y de tabla; dónde pueden aparecer.
- Subconsultas con `IN / NOT IN` (y su trampa con NULL), `EXISTS / NOT EXISTS`.
- Subconsultas correlacionadas: cómo funcionan y cuándo evitarlas.
- CTE con `WITH`: legibilidad profesional, CTE encadenadas, introducción a las CTE recursivas.
- Vistas: `CREATE VIEW`, cuándo usarlas y sus límites.

Resultados esperados

El alumno descompone problemas complejos en pasos legibles con CTE, elige entre subconsulta, JOIN y EXISTS según el caso, y encapsula consultas frecuentes en vistas.

Unidad 07 — DDL y DML: crear, poblar y mantener bases de datos

Tiempo estimado: 4 horas

Temas y subtemas

- `CREATE TABLE` completo: tipos, `PRIMARY KEY`, `FOREIGN KEY`, `NOT NULL`, `UNIQUE`, `CHECK`, `DEFAULT`.
- Integridad referencial: qué impide una clave foránea; acciones `ON DELETE / ON UPDATE`.
- Normalización explicada sin dolor: 1FN, 2FN, 3FN con ejemplos de NovaMarket.
- `INSERT`, `UPDATE`, `DELETE` seguros: el patrón SELECT-antes-de-modificar.
- Transacciones: `BEGIN`, `COMMIT`, `ROLLBACK`, propiedades ACID.
- `ALTER TABLE`, `DROP`, e índices: `CREATE INDEX` y cuándo acelera (o no) una consulta.

Resultados esperados

El alumno diseña y crea desde cero un esquema íntegro y normalizado, modifica datos sin accidentes usando transacciones y comprende el rol de los índices.

Unidad 08 — SQL para ciencia de datos: ventanas, limpieza y análisis

Tiempo estimado: 4 horas

Temas y subtemas

- El rol de SQL en el flujo de trabajo de ciencia de datos.
- Funciones de ventana: `OVER, PARTITION BY, ORDER BY` ; rankings (`ROW_NUMBER, RANK, DENSE_RANK, NTILE`), desplazamientos (`LAG, LEAD`), acumulados y promedios móviles.
- Estadística descriptiva en SQL: media, mediana, desviación estándar, percentiles, detección de outliers.
- Perfilado y limpieza de datos: duplicados, faltantes, inconsistencias, estandarización.
- Análisis RFM (Recencia, Frecuencia, Monto) completo de los clientes de NovaMarket.
- SQL + Python: leer consultas de PostgreSQL con `psycopg` y pandas; dónde termina SQL y dónde empieza Python.

Resultados esperados

El alumno realiza un análisis de datos completo y profesional: perfila y limpia los datos, construye métricas avanzadas con funciones de ventana, segmenta clientes con RFM y conecta los resultados con Python/pandas.

Evaluación y aprobación

- **Cuestionario por unidad:** 5 preguntas con retroalimentación inmediata (formato GIFT), que evalúan comprensión y no memoria.
- **Actividad interactiva por unidad (H5P):** fija los conocimientos clave de forma visual y autoevaluable.
- **Examen Final:** 20 preguntas de dificultad intermedia-alta que integran todo el curso.
- **Criterio sugerido de aprobación:** promedio $\geq 60\%$ en cuestionarios de unidad y $\geq 60\%$ en el Examen Final. Intentos ilimitados sugeridos: el objetivo es aprender.
- **Sin entregables:** no hay proyectos ni tareas para corregir manualmente.

SQL desde Cero para Ciencia de Datos: del Primer SELECT al Análisis Profesional

8 unidades · 30 horas · Nivel inicial a intermedio-avanzado · Moodle 5.2+ · Basado en ISO/IEC 9075 y PostgreSQL 18



Aprender
Online UTN



UNIVERSIDAD
TECNOLÓGICA
NACIONAL
FACULTAD REGIONAL
RECONQUISTA

aprenderonline.frrq.utn.edu.ar



Aprender Online se reserva el derecho de modificar este programa, o cualquiera de los componentes del curso, cuando lo considere necesario a fin de generar mejoras o actualizaciones.