Reactor
COURSES

BROCHURE DEL PROGRAMA



ARQUITECTURA E INGENIERÍA · INTEGRACIÓN Y DATOS

Data Platform Architecture

Del dato crudo al consumo confiable

Diseña y construye la plataforma de datos de una empresa: ingesta batch y en tiempo real, almacenamiento analítico, modelado, calidad verificable, gobierno y costo por consulta bajo control.

Lo que vas a lograr Aprendes a construir la plataforma donde una empresa reúne, limpia y analiza sus datos, con calidad comprobable y costo controlado.

- Lakehouse
- CDC y streaming
- Modelado analítico
- Contratos de datos
- Linaje y gobierno
- Costo por consulta

horas en vivo	sesiones	semanas	plataforma de datos completa
90	30	15	1

INICIO
Por confirmar

PRECIO
US$420 pago único · USD

MODALIDAD
Clases en vivo por Zoom · Avanzado

reactorcourses.com/programas/data-platform-architecture

EL PROGRAMA

Mover datos es fácil. Que alguien confíe en el número que sale al final es el trabajo.

Diseña y construye la plataforma de datos de una empresa: ingesta batch y en tiempo real, almacenamiento analítico, modelado, calidad verificable, gobierno y costo por consulta bajo control. Un programa de **90 horas** para ingenieros que ya mueven datos y están cansados de que el tablero de la mañana no cuadre con el sistema que lo alimenta.

Duración	90 horas · 30 sesiones	Calendario	15 semanas · 2 sesiones por semana de 3 h
Modalidad	Clases en vivo por Zoom	Grabaciones	Disponibles por 12 meses
Credenciales	3 certificados con insignia de Reactor Courses	Nivel	Avanzado
Inicio	Por confirmar	Horario	Martes y jueves, 19:00–22:00 (GMT-5)
Precio	US$420 · pago único		

Casi toda plataforma de datos empieza igual: unos cuantos procesos que copian tablas de noche. Y casi todas terminan igual: nadie sabe de dónde salió el número, quién lo rompió ni a qué hora estuvo bien. La diferencia no está en la herramienta, está en el diseño.

01

Construir el flujo

Ingesta, almacenamiento, modelado y transformación

El camino completo del dato: capturarlo del sistema origen sin castigarlo, guardarlo en un formato que admita evolución y viajes en el tiempo, modelarlo para que responda preguntas de negocio y transformarlo con dependencias explícitas y ejecución reproducible.

- Ingesta por lotes, incremental y captura de cambios del origen
- Datos en movimiento: registro de eventos, ventanas y llegadas tardías
- Almacenamiento analítico con formato de tabla abierto y evolución de esquema
- Modelado para consumo y transformación con dependencias y pruebas

02

Hacerlo confiable

Calidad, contratos, gobierno y costo

Lo que separa una plataforma de un montón de procesos: contratos entre quien produce y quien consume, calidad medida y no supuesta, linaje que responde de dónde salió cada número, control de acceso al dato sensible y una factura que alguien entiende y puede bajar.

- Contratos de datos, versionado y ruptura anunciada
- Calidad verificable, acuerdos de frescura y detección de anomalías
- Catálogo, linaje, propiedad del dato y clasificación de sensibilidad
- Costo por consulta y por proceso, con dueño identificable

> Un dato sin dueño, sin contrato y sin linaje no es un activo: es un rumor con formato de tabla.

PARA QUIÉN ES

¿Es el programa indicado para ti?

Está pensado para ingenieros que escriben SQL y Python y quieren construir la plataforma de datos de una empresa, de punta a punta.

Es para ti si

- ✓ Ingenieros de datos que ya mantienen procesos en producción y quieren dejar de apagar incendios.
- ✓ Ingenieros de backend que heredaron la plataforma de datos porque no había nadie más.
- ✓ Analistas técnicos y desarrolladores de BI con SQL sólido que quieren pasar al lado de la ingeniería.
- ✓ Arquitectos que deben decidir la plataforma de datos de su empresa y quieren decidir con criterio.
- ✓ Ingenieros de plataforma que dan servicio a equipos de datos y necesitan entender su problema.

Te conviene otro programa

- – Quien no escriba SQL con soltura. Aquí se asume desde la primera sesión y no hay nivelación.
- – Quien busque un curso de ciencia de datos, estadística o modelos predictivos. No es este programa.
- – Quien busque un curso de una herramienta específica ni preparación para un examen de certificación.
- – Quien quiera aprender a hacer tableros. Aquí se construye la plataforma que alimenta los tableros.
- – Quien no pueda dedicar entre 10 y 13 horas por semana durante 15 semanas.
- – Quien no pueda crear una cuenta de nube propia para los dos módulos que despliegan fuera de su máquina.

EL RESULTADO

Terminas con una plataforma que se levanta con un comando y con la evidencia de que sus datos son confiables.

Todo lo que construyes vive en un repositorio que puedes mostrar. Es el material con el que se gana una entrevista de ingeniería de datos o se sustenta la reconstrucción de la plataforma de tu empresa.

01 **Ingesta por lotes** Carga incremental con marcas de agua, reintentos y reproceso de un rango histórico.

02 **Captura de cambios** Replicación desde el registro de la base origen sin consultarla ni bloquearla.

03 **Ingesta en tiempo real** Registro de eventos particionado, consumo con ventanas y manejo de llegadas tardías.

04 **Almacén analítico** Formato de tabla abierto con evolución de esquema, viaje en el tiempo y compactación.

05 **Capas del modelo** Bruto, refinado y de consumo, con propiedad, convenciones y criterios de promoción.

06 **Modelo dimensional** Hechos, dimensiones y manejo de cambios históricos con su justificación.

07 **Transformación con pruebas** Dependencias explícitas, ejecución incremental y pruebas que corren en cada cambio.

08 **Orquestación** Flujos programados y por evento, con reintentos, reproceso y alertas accionables.

09 **Contratos de datos** Esquema, semántica, frescura y compatibilidad, verificados automáticamente.

10 **Calidad y frescura** Reglas de calidad ejecutables, acuerdos de frescura y detección de anomalías.

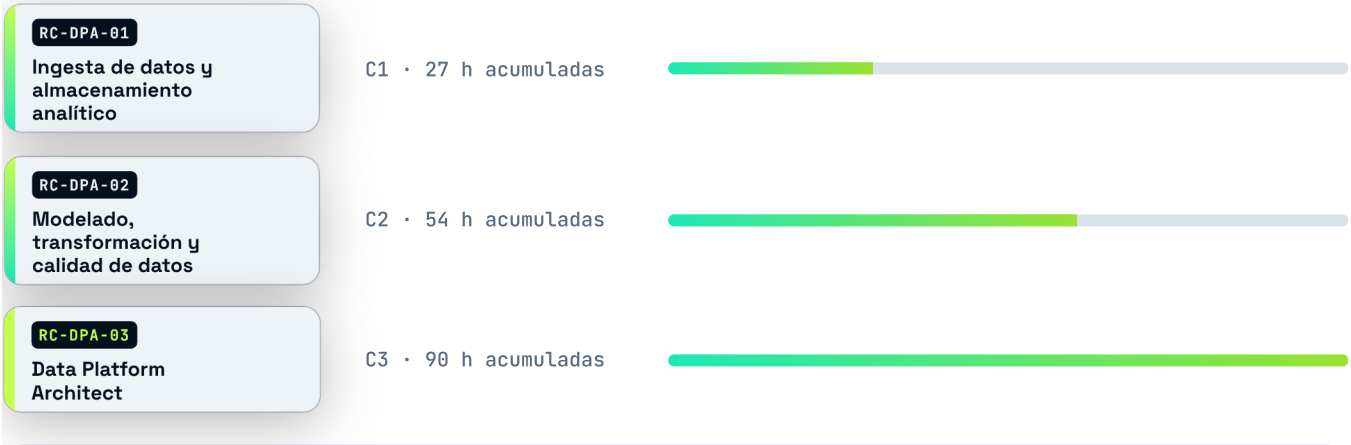
11 **Catálogo y linaje** De dónde salió cada columna, quién la usa y qué se rompe si cambia.

12 **Costo y defensa** Costo por consulta y por proceso con plan de optimización, sustentado ante el grupo.

RUTA DE CERTIFICACIÓN

3 credenciales progresivas, no una al final.

Cada etapa cierra un bloque completo de competencias. Si algo te obliga a pausar en la semana 9, ya tienes dos credenciales defendibles en tu CV.



CERTIFICACIÓN INTEGRAL

Data Platform Architecture — Del dato crudo al consumo confiable (90 horas)

Requiere entregar la plataforma completa, levantarla desde cero y demostrar la calidad y el linaje de sus datos. Los certificados son emitidos por Reactor Courses y acreditan las horas y competencias del programa.

Cada certificado de Reactor Courses viene con su insignia, aparte de cualquier certificación oficial.

Cualquiera puede comprobar un certificado de Reactor Courses con su código en reactorcourses.com/verificar.

TEMARIO

Las 30 sesiones, abiertas y sin letra chica.

90 horas repartidas en 10 módulos. Así sabes qué verás, qué practicarás y qué entregarás en cada una de las 15 semanas.

MÓDULO 1 LA PLATAFORMA DE DATOS COMO PRODUCTO

9 h · 3 sesiones

01 Qué plataforma necesita esta empresa y cuál no

consumidores de datos y lo que cada uno necesita de verdad · almacén, lago y arquitectura de lago con formato de tabla · arquitectura centralizada frente a dominios con propiedad distribuida · carga transaccional y carga analítica: por qué se separan · señales de que la plataforma actual ya no da · capacidades mínimas de una plataforma viable · métricas de la plataforma: frescura, confianza y costo · el proyecto sobredimensionado que nadie usa

LABORATORIO: Entorno local levantado y mapa de consumidores y necesidades del caso del programa

02 Capas, propiedad y convenciones

capa bruta, refinada y de consumo · qué se permite y qué se prohíbe en cada capa · criterios de promoción entre capas · propiedad del dato y responsable identificable · convenciones de nombres, esquemas y particiones · zona de aterrizaje inmutable y su valor en una auditoría · reprocesar desde el origen sin depender del sistema fuente · documentación que se genera sola

LABORATORIO: Estructura de capas y convenciones implementadas en el almacenamiento de objetos local

03 Repositorio, entornos y flujo de trabajo

el repositorio de la plataforma y su organización · entornos de desarrollo y producción con datos distintos · control de versiones para transformaciones y esquemas · revisión de cambios que afectan datos publicados · integración continua sobre una muestra representativa · gestión de credenciales de los sistemas origen · reproducibilidad del entorno completo · criterio de listo para publicar

LABORATORIO: Repositorio inicializado con entornos separados y una tubería de integración continua funcionando

MÓDULO 2 INGESTA POR LOTES Y CAPTURA DE CAMBIOS

9 h · 3 sesiones

04 Ingesta por lotes e incremental

carga completa frente a carga incremental · marca de agua y detección de registros nuevos · registros modificados y borrados en el origen · idempotencia de la carga y reejecución segura · reproceso de un rango histórico sin duplicar · límite de velocidad para no castigar al sistema origen · ventana de ingesta y acuerdo con el equipo dueño · archivos que llegan tarde, incompletos o corruptos

LABORATORIO: Carga incremental con marca de agua, reejecutable y con reproceso de un rango histórico

05 Captura de cambios desde el registro de la base

por qué consultar la base origen no escala · lectura del registro de transacciones del motor · eventos de inserción, modificación y borrado · captura inicial y transición al flujo continuo · orden de eventos y su clave de partición · cambios de esquema en el origen y su propagación · retraso de replicación y su monitoreo · el impacto real sobre la base de producción

LABORATORIO: Captura de cambios desde una base transaccional hasta la capa bruta, con captura inicial y flujo continuo

06 Ingesta desde APIs y archivos

paginación, límites de tasa y reintentos · autenticación y rotación de credenciales de origen · esquemas semiestructurados y su aplanado · archivos planos, codificaciones y separadores traicioneros · detección de cambios de esquema no anunciados · conservar la carga original tal como llegó · registro de ingesta y trazabilidad del origen · cuándo conviene comprar un conector en lugar de escribirlo

LABORATORIO: Conector propio a una API paginada con límite de tasa, reintentos y registro de ingesta — compuerta Ingestion Ready

MÓDULO 3 DATOS EN MOVIMIENTO

9 h · 3 sesiones

07 Registro de eventos y semántica de entrega

registro particionado, orden y retención · clave de partición y su efecto en el orden · grupos de consumidores y reparto de particiones · compensación de desfase y reproceso desde el inicio · entrega al menos una vez y consumidores idempotentes · esquema del evento y registro de esquemas · evolución compatible del evento · el evento como contrato entre equipos

LABORATORIO: Productor y consumidor con reproceso completo desde el inicio del registro y sin duplicar efectos

08 Procesamiento con ventanas y tiempo de evento

tiempo de evento frente a tiempo de proceso · ventanas fijas, deslizantes y de sesión · marcas de agua y datos que llegan tarde · resultados parciales y su corrección posterior · agregaciones con estado y su almacenamiento · reinicio de un proceso con estado · exactitud frente a latencia: la decisión explícita · cuándo el tiempo real no aporta nada

LABORATORIO: Agregación por ventanas de tiempo de evento con manejo explícito de registros tardíos

09 Unificar lo continuo y lo periódico

una sola definición de la métrica para dos caminos · arquitecturas de doble ruta y su costo de mantenimiento · convergencia entre el resultado continuo y el recalculado · reconciliación periódica y detección de divergencia · corrección de un histórico ya publicado · señales de que el tiempo real es innecesario · escoger la latencia por caso de uso, no por moda · documentar la garantía de frescura de cada tabla

LABORATORIO: Métrica de negocio calculada por ambas rutas, con reconciliación automática y alerta de divergencia

MÓDULO 4 ALMACENAMIENTO ANALÍTICO

9 h · 3 sesiones

10 Formatos columnares y organización física

formato columnar, compresión y codificaciones · estadísticas de bloque y descarte de lectura · tamaño de archivo y el problema de los archivos pequeños · particionado físico y su elección · ordenamiento y agrupamiento por columnas de filtro · impacto de la organización en el costo por consulta · leer solo lo necesario: cómo se logra · medir el efecto de cada decisión física

LABORATORIO: Comparar cuatro organizaciones físicas de la misma tabla midiendo datos leídos y tiempo

11 Formato de tabla abierto

qué agrega un formato de tabla sobre archivos sueltos · transacciones sobre almacenamiento de objetos · evolución de esquema sin reescribir el histórico · viaje en el tiempo y consulta de una versión anterior · actualización y borrado a nivel de fila · compactación y limpieza de versiones antiguas · lectores y escritores concurrentes · portabilidad entre motores de consulta

LABORATORIO: Tabla con evolución de esquema, borrado por fila, viaje en el tiempo y compactación programada

12 Motores de consulta y rendimiento

motor embebido, distribuido y gestionado · leer un plan de ejecución analítico · estrategias de unión y su costo · estadísticas, cardinalidad y decisiones del optimizador · consulta que escanea todo por un filtro mal escrito · caché de resultados y su validez · concurrencia y aislamiento entre cargas · elegir el motor según el patrón de consulta

LABORATORIO: Optimizar tres consultas lentas y documentar la mejora en datos leídos, tiempo y costo estimado

MÓDULO 5 MODELADO PARA CONSUMO

9 h · 3 sesiones

13 Modelado dimensional

hechos, dimensiones y grano de la tabla de hechos · elegir el grano antes que cualquier otra cosa · dimensiones conformadas y métricas comparables · hechos aditivos, semiaditivos y no aditivos · dimensión de fecha y calendarios de negocio · claves sustitutas y su necesidad real · modelado ancho frente a modelado normalizado · el modelo que responde la pregunta que nadie hizo todavía

LABORATORIO: Modelo dimensional del caso con grano declarado y dimensiones conformadas

14 **Historia y cambios en el tiempo**

dimensiones que cambian lentamente y sus estrategias · sobrescribir, versionar o mantener columna previa · vigencia temporal y consultas a una fecha dada · hechos que se corrigen después de publicados · reproceso histórico y su efecto en informes ya emitidos · instantáneas periódicas y su costo · borrado de datos personales en un modelo histórico · explicar a negocio por qué el número de ayer cambió

LABORATORIO: Dimensión con versionado histórico y consulta que reproduce el estado a una fecha pasada

15 **Métricas y capa semántica**

definición única de cada métrica de negocio · métricas que se calculan distinto en cada informe · capa semántica y su mantenimiento · dimensiones permitidas por métrica · agregaciones previas y su invalidación · diccionario de negocio y su dueño · control de acceso a nivel de fila y de columna · publicar una métrica: requisitos mínimos

LABORATORIO: Capa semántica con métricas definidas una sola vez y pruebas de consistencia entre consumos

MÓDULO 6 TRANSFORMACIÓN Y ORQUESTACIÓN — 9 h · 3 sesiones

16 **Transformación declarativa con dependencias**

grafo de dependencias derivado del código · materialización: vista, tabla e incremental · estrategia incremental y su clave de unicidad · reconstrucción completa frente a incremental · modularidad y reutilización sin duplicar lógica · pruebas de datos ejecutadas en cada cambio · documentación y linaje generados del propio código · el modelo que nadie se atreve a tocar

LABORATORIO: Capa de transformación con dependencias explícitas, materializaciones incrementales y pruebas

17 **Orquestación y recuperación**

flujos por horario y por evento · dependencia entre activos de datos, no entre tareas · reintentos, tiempos de espera y ejecución atrasada · reproceso de un rango con control de concurrencia · ejecución parcial y reanudación · detección de un flujo que no corrió · alertas accionables frente a correo ignorado · el flujo que corre dos veces y duplica todo

LABORATORIO: Orquestación completa con reproceso de un rango histórico y recuperación de una ejecución fallida — compuerta Model Ready

18 **Entrega continua de cambios de datos**

probar una transformación antes de publicarla · entorno de desarrollo con datos representativos · detección de cambios incompatibles en el esquema publicado · despliegue de un cambio de modelo sin romper consumidores · reversión de un cambio de datos ya materializado · control de costo de la integración continua · revisión entre pares de un cambio de métrica · registro de cambios visible para los consumidores

LABORATORIO: Tubería que detecta un cambio incompatible y bloquea la publicación antes de romper a un consumidor

MÓDULO 7 CALIDAD, CONTRATOS Y CONFIABILIDAD — 9 h · 3 sesiones

19 **Calidad de datos verificable**

dimensiones de calidad: completitud, unicidad, validez, consistencia · reglas ejecutables en lugar de acuerdos verbales · prueba en el origen, en la transformación y en la salida · umbral de tolerancia y qué hacer al superarlo · bloquear la publicación frente a publicar con aviso · cuarentena de registros que no cumplen · métricas de calidad y su evolución en el tiempo · la regla de calidad que nadie mira

LABORATORIO: Batería de reglas de calidad que bloquea la publicación y deja los registros rechazados en cuarentena

20 **Contratos de datos**

qué contiene un contrato: esquema, semántica, frescura y volumen · productor y consumidor: quién se compromete a qué · verificación automática del contrato en cada carga · cambio compatible e incompatible · depreciación anunciada y período de gracia · versionado de la tabla publicada · consecuencia de romper un contrato · negociar un contrato con un equipo que no quiere firmarlo

LABORATORIO: Contrato de datos publicado y verificado automáticamente, con simulación de una ruptura

21 **Frescura, anomalías y confiabilidad**

acuerdo de frescura por tabla y su medición · detección de tabla que dejó de actualizarse · anomalías de volumen, de distribución y de nulos · línea base histórica y estacionalidad · falsa alarma y fatiga de alertas · incidente de datos: clasificación y comunicación · revisión posterior de un número incorrecto publicado · confianza como métrica del equipo de datos

LABORATORIO: Monitoreo de frescura y detección de anomalías sobre las tablas publicadas, con incidente simulado

MÓDULO 8 GOBIERNO, PRIVACIDAD Y SEGURIDAD DEL DATO — 9 h · 3 sesiones

22 **Catálogo y linaje**

catálogo técnico y glosario de negocio · linaje a nivel de tabla y a nivel de columna · análisis de impacto antes de un cambio · descubrimiento: encontrar el dato sin preguntar · metadatos que se recolectan solos · propiedad, criticidad y clasificación de cada activo · activos huérfanos y su retiro · el catálogo que se llena una vez y se abandona

LABORATORIO: Catálogo poblado automáticamente con linaje a nivel de columna y análisis de impacto de un cambio

23 **Privacidad y datos personales**

identificación y clasificación de datos personales · minimización: no ingerir lo que no se necesita · seudonimización, enmascaramiento y cifrado · control de acceso a nivel de fila y de columna · propósito de uso y su registro · derecho de supresión sobre un histórico inmutable · retención y borrado automático · auditoría de accesos al dato sensible

LABORATORIO: Enmascaramiento y control de acceso por rol, con procedimiento de supresión sobre el histórico

24 **Gobierno operativo**

quién aprueba una tabla nueva en la capa de consumo · estándares aplicados automáticamente y no por revisión manual · excepciones con dueño y vencimiento · gestión de acceso: solicitud, aprobación y revisión periódica · compartir datos con terceros de forma controlada · evidencia automática para una auditoría · gobierno que habilita frente a gobierno que frena · costo del gobierno mal diseñado

LABORATORIO: Estándares aplicados en la tubería, con excepciones registradas y evidencia de cumplimiento — compuerta Trust Ready

MÓDULO 9 SERVICIO A CONSUMIDORES Y COSTO — 9 h · 3 sesiones

25 **Servir a análisis, aplicaciones y modelos**

consumo desde herramientas de visualización · extracción reproducible para ciencia de datos · conjunto de entrenamiento con corte temporal correcto · fuga de información desde el futuro · servir datos a una aplicación con baja latencia · invertir el flujo: devolver datos al sistema operacional · compartir sin copiar · acuerdo de servicio por tipo de consumidor

LABORATORIO: Tres consumos servidos desde el mismo modelo: visualización, extracción reproducible y consulta de baja latencia

26 **Costo de la plataforma de datos**

dónde se va el dinero: almacenamiento, cómputo y movimiento · costo por consulta y por proceso · atribución de costo por equipo y por activo · consulta cara ejecutada cada cinco minutos · retención por capa y su costo acumulado · compresión, particionado y ahorro medible · cómputo bajo demanda frente a capacidad reservada · mostrar el costo a quien puede reducirlo

LABORATORIO: Informe de costo por activo y por consumidor, con tres optimizaciones aplicadas y medidas

27 **Operación y evolución de la plataforma**

guardia del equipo de datos y manual de incidente · prioridad de recuperación por criticidad del activo · recuperación ante un reproceso equivocado · copias, retención y prueba de restauración · migración de motor o de formato sin detener el servicio · deuda de modelo y su vencimiento · comunicación con consumidores durante una degradación · informe de salud de la plataforma

LABORATORIO: Simulacro: reproceso equivocado publicado, detección, comunicación y restauración verificada

MÓDULO 10 PLATAFORMA FINAL Y DEFENSA

9 h · 3 sesiones

28 Consolidación de la plataforma

integración de ingesta, almacenamiento, modelo y gobierno · levantar la plataforma completa desde cero · verificación contra las capacidades comprometidas · coherencia entre contratos, pruebas y linaje · limpieza de activos huérfanos y accesos sobrantes · costo final y proyección a doce meses · empaquetado del repositorio para mostrarlo · qué se entrega y en qué formato

LABORATORIO: Plataforma consolidada, levantada desde cero y verificada contra sus contratos y reglas de calidad

29 Revisión cruzada y auditoría de confianza

auditar la plataforma de otro equipo · rastrear un número del tablero hasta el sistema origen · buscar la métrica definida dos veces · detectar la tabla sin dueño, sin prueba y sin contrato · verificar el manejo de datos personales · hallazgos priorizados por riesgo de negocio · corrección y verificación con la prueba que falló · preparación de la defensa

LABORATORIO: Auditoría cruzada: rastrear tres números hasta su origen en la plataforma de otro equipo y documentar hallazgos

30 Defensa de la plataforma de datos

¿de dónde salió este número y quién responde por él? · ¿qué pasa si el sistema origen cambia una columna mañana? · ¿cómo te enteras de que una tabla dejó de actualizarse? · ¿qué se rompe si borro esta columna? · ¿dónde vive el dato personal y quién puede verlo? · ¿cuánto cuesta esta consulta y quién la paga? · ¿cómo reprocesas seis meses sin duplicar nada? · ¿qué eliminarías de tu plataforma si tuvieras que abaratarla a la mitad?

LABORATORIO: Defensa individual: rastreo en vivo de un número hasta el origen y sustentación de las decisiones

HERRAMIENTAS

Todo lo que vas a tocar con las manos.

Lenguajes

- SQL avanzado
- Python
- Funciones de ventana
- Notación de configuración declarativa

Almacenamiento

- Almacenamiento de objetos
- Parquet
- Formato de tabla abierto
- Particionado y compactación
- Viaje en el tiempo

Procesamiento

- Spark
- DuckDB
- Motor de consulta distribuido
- Procesamiento incremental

Movimiento del dato

- Captura de cambios
- Registro de eventos particionado
- Conectores de ingesta
- Reproceso histórico

Transformación

- Transformación declarativa con dependencias
- Pruebas de datos
- Materializaciones incrementales
- Documentación generada

Orquestación

- Flujos programados y por evento
- Reintentos y reproceso
- Dependencias entre activos de datos
- Alertas de proceso

Calidad y gobierno

- Reglas de calidad ejecutables
- Contratos de datos
- Linaje abierto
- Catálogo y clasificación
- Enmascaramiento

Operación y costo

- Métricas de frescura
- Detección de anomalías
- Costo por consulta
- Etiquetado y asignación
- Git y CI

Los logos son marcas de sus respectivos dueños. Se muestran solo para identificar las herramientas que se usan en el programa y no implican aval, patrocinio ni afiliación.

CÓMO FUNCIONA

Clases en vivo, un conjunto de datos sucio de verdad y una plataforma que crece cada semana.

01 **Clases en vivo por Zoom** 30 sesiones de 3 horas, 2 veces por semana. Cámara y micro abiertos: se construye y se depura en conjunto.

02 **Todo queda grabado** Si te pierdes una sesión o vives en otro huso horario, la grabación queda disponible. Acceso por 12 meses.

03 **25% concepto / 75% laboratorio** Casi todo el tiempo escribes SQL, Python y definiciones de transformación sobre datos que vienen rotos a propósito.

04 **Un caso que dura 15 semanas** Los mismos sistemas origen desde la sesión 01: cada módulo agrega una capa y el acumulado es tu plataforma.

05 **Los datos llegan sucios** Duplicados, nulos inesperados, esquemas que cambian sin avisar y registros que llegan tarde. Como en producción.

06 **Tres compuertas de calidad** Ingestion Ready, Model Ready y Trust Ready. No se aprueba una tabla sin dueño, sin prueba y sin contrato.

Horarios · Primer grupo

Martes	19:00 - 22:00
Jueves	19:00 - 22:00

Horario referido a GMT-5 (Perú, Colombia, Ecuador y Panamá; en Ciudad de México es una hora menos).

Los días y el horario son la propuesta para este grupo y se confirman junto con la fecha de inicio, antes de abrir la matrícula. El calendario definitivo se entrega al confirmar la matrícula.

REQUISITOS

Lo que necesitas traer a la sesión 01.

No hay nivelación previa. La sesión 01 asume SQL sólido y capacidad de escribir un script de Python que no sea trivial.

SQL

- ✓ Uniones, agregaciones y subconsultas
- ✓ Funciones de ventana
- ✓ Leer un plan de ejecución
- ✓ Entender qué hace lenta una consulta

Python

- ✓ Scripts con funciones y módulos
- ✓ Manejo de archivos y formatos
- ✓ Consumo de APIs
- ✓ Entornos virtuales y dependencias

Datos

- ✓ Modelo relacional y normalización
- ✓ Tipos, nulos y claves
- ✓ Diferencia entre carga transaccional y analítica
- ✓ Nociones de formatos columnares

Sistemas

- ✓ Línea de comandos y Linux básico
- ✓ Contenedores y Docker Compose
- ✓ Variables de entorno y configuración
- ✓ Lectura de logs

Ingeniería

- ✓ Git con ramas y pull requests
- ✓ Nociones de CI
- ✓ Pruebas automatizadas a nivel conceptual
- ✓ Revisión de código entre pares

Entorno

- ✓ Equipo con 16 GB de RAM recomendado
- ✓ Docker instalado y funcionando
- ✓ Cuenta propia de nube para dos módulos
- ✓ Editor con soporte de SQL y Python

Costos de herramientas que asumes tú `US$40 a 90 aprox.`

Estos costos NO están incluidos en el precio del programa. Se declaran por transparencia: cada alumno administra su propia cuenta y su propio presupuesto.

- La mayor parte del laboratorio corre local con contenedores: almacenamiento de objetos, motor de consulta y registro de eventos
- Dos módulos despliegan en una nube pública para trabajar con un almacén analítico gestionado y su modelo de costos
- Se trabaja sobre los niveles gratuitos (free tier) y créditos gratuitos cuando aplican, con presupuestos y alertas configurados desde la primera sesión

PRECIO

90 horas para que el número del tablero por fin se pueda defender.

US$420

Pago único · USD · Incluye impuestos

Todavía no hay fecha de inicio. Escríbenos y te avisamos cuando la confirmemos: no cobramos nada antes de abrir la matrícula.

Incluye

- ✓ 30 sesiones en vivo por Zoom (90 horas lectivas)
- ✓ Grabación de todas las clases con acceso por 12 meses
- ✓ Sistemas origen simulados, conjuntos de datos sucios y repositorio de referencia
- ✓ Revisión de entregables en las 3 compuertas de calidad
- ✓ Revisión de código y de modelo entre pares en cada módulo
- ✓ Plataforma final con acompañamiento y defensa técnica
- ✓ Los 3 certificados progresivos + la certificación integral
- ✓ Canal privado del grupo para dudas entre sesiones
- ✓ Actualizaciones del material por 12 meses

No incluye

- – Tu cuenta de nube y su consumo en los dos módulos que despliegan fuera de tu máquina (aprox. US$40 a 90)
- – Exámenes ni certificaciones de terceros: este programa no prepara para ninguno
- – Licencias comerciales: el laboratorio corre con software de código abierto y niveles gratuitos

Métodos de pago

Tarjeta de crédito · Tarjeta de débito internacional · PayPal · Apple Pay · Google Pay

Garantía de 14 días. Si dentro de los primeros 14 días del programa decides que no es para ti, te devolvemos el 100 % de lo pagado. No pedimos justificación.

Paddle actúa como comerciante registrado (merchant of record): emite el comprobante y gestiona los impuestos de cada país.

PREGUNTAS FRECUENTES

Preguntas frecuentes sobre el programa.

¿Las clases son en vivo o grabadas?

En vivo por Zoom: 30 sesiones de 3 horas, 2 veces por semana. Todas quedan grabadas y disponibles por 12 meses, así que puedes recuperar una clase o repasar una transformación con calma.

¿Cuándo empiezan las clases?

La fecha de inicio y el calendario definitivo todavía no están confirmados: se publican en esta misma ficha antes de que se abra la matrícula. Los días propuestos son Martes y jueves en horario de GMT-5 (Perú, Colombia, Ecuador y Panamá; en Ciudad de México es una hora menos). Si quieres que te avisemos cuando se confirmen, escribe a soporte@reactorcourses.com.

¿Es un curso de ciencia de datos o de machine learning?

No. Es un curso de ingeniería de plataforma de datos. Se construye la infraestructura y el modelo de datos que un equipo de ciencia de datos consume, y hay una sesión dedicada a servirles datos de forma reproducible, pero no se entrenan modelos ni se hace estadística.

¿Sobre qué herramientas se trabaja?

El laboratorio corre con software de código abierto y estándares abiertos: almacenamiento de objetos, formato de tabla abierto, motores de consulta, transformación declarativa y orquestación. La razón es que los conceptos se transfieren a cualquier plataforma comercial, mientras que aprender una herramienta cerrada solo sirve mientras tu empresa la pague.

¿Cuánto voy a gastar en la nube?

Aproximadamente US$40 a 90 en todo el programa. La mayor parte del laboratorio corre local con contenedores; solo dos módulos despliegan en una nube pública para trabajar con un almacén analítico gestionado y su modelo de costos real, que es justamente lo que no se puede simular.

¿Este programa prepara para alguna certificación?

No. No prepara ni acredita para ninguna certificación de Microsoft, AWS, Google, Databricks, Snowflake ni ninguna otra organización, y Reactor Courses no tiene relación ni aval de ellas.

¿Cuánto SQL hace falta realmente?

Bastante. Debes poder escribir una consulta con varias uniones y funciones de ventana sin buscar la sintaxis, y leer un plan de ejecución para entender por qué algo tarda. Si tu SQL es de nivel select con where, este programa te va a costar mucho.

¿Puedo usar los datos de mi empresa?

No en clase, y por una razón simple: son datos de terceros y casi siempre hay información personal de por medio. El programa trae sistemas origen simulados con la suciedad suficiente para que el ejercicio sea real. Lo que aprendas se aplica después, en tu entorno y con tus controles.

¿Cuánto tiempo debo dedicarle por semana?

Seis horas de clase más entre cuatro y siete horas de laboratorio. Considera entre 10 y 13 horas semanales durante 15 semanas. Los procesos de datos tardan en correr y conviene dejar margen.

¿Qué métodos de pago aceptan?

El pago lo procesa Paddle y acepta Tarjeta de crédito, Tarjeta de débito internacional, PayPal, Apple Pay, Google Pay. El cobro es en dólares. Importante: no procesamos Yape, Plin, OXXO, Mercado Pago ni transferencias locales. Si tu banco rechaza compras internacionales, habilítalas antes de pagar o usa PayPal.

¿Puedo pagar en cuotas?

Por ahora el programa se cobra en un pago único de US$420. Algunos bancos permiten diferir la compra en cuotas desde su propia app o banca en línea; eso depende de tu emisor, no de nosotros.

¿Emiten factura o comprobante?

Sí. Paddle actúa como comerciante registrado (merchant of record): emite el comprobante y gestiona los impuestos de cada país. Lo recibirás por correo automáticamente al completar el pago. Si necesitas datos fiscales de empresa, cárgalos en el campo correspondiente de la ventana de pago.

¿Cuál es la política de reembolso?

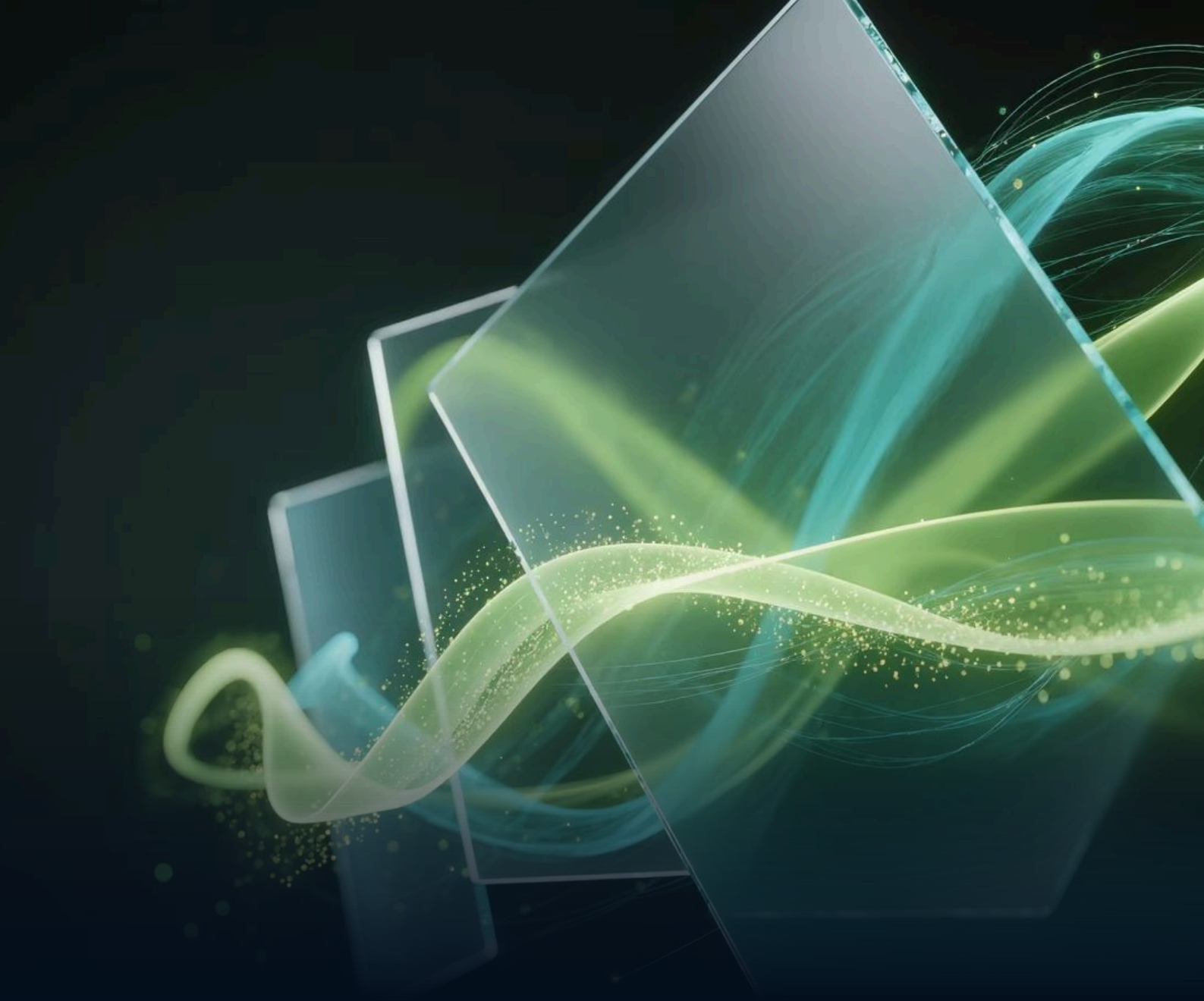
Tienes 14 días desde el inicio del programa para pedir el reembolso completo escribiendo a soporte@reactorcourses.com. Publicamos el temario completo y los requisitos justamente para que decidas informado y nadie llegue a ese punto.

¿Los certificados tienen validez universitaria?

No. Son certificados profesionales emitidos por Reactor Courses que acreditan horas y competencias. No son un grado académico ni una certificación oficial de ninguna organización. Lo que de verdad te va a servir en una entrevista es el repositorio de la plataforma y poder rastrear un número hasta su origen.

¿Este programa se cruza con otros del catálogo?

Cada programa de Reactor Courses es un producto cerrado: trae sus propios fundamentos, su propio proyecto final y no da por visto nada de otro curso. Otros programas de la escuela tocan almacenamiento, colas o costos desde el ángulo de la infraestructura o del sistema; aquí el objeto es el dato mismo: su contrato, su calidad, su linaje y quién responde por él. Si cursas dos programas vas a volver a ver algunos temas, y no hay convalidaciones ni descuentos entre programas: preferimos decirlo antes y no que lo descubras pagando.



Reactor
COURSES

Ingeniería de software en la era de los agentes

¿Te quedó alguna duda?

Escríbenos a info@reactorcourses.com. La ficha al día, con el temario interactivo y la inscripción, está en reactorcourses.com/programas/data-platform-architecture.

`reactorcourses.com/programas/data-platform-architecture`

Información vigente al 23 de noviembre de 2026. El precio, las fechas y el temario pueden cambiar: la versión al día es la de la web.

Reactor Courses no está afiliado a Microsoft, Amazon Web Services, OpenAI, Anthropic ni Google. Los nombres de productos se mencionan únicamente con fines descriptivos del contenido académico.