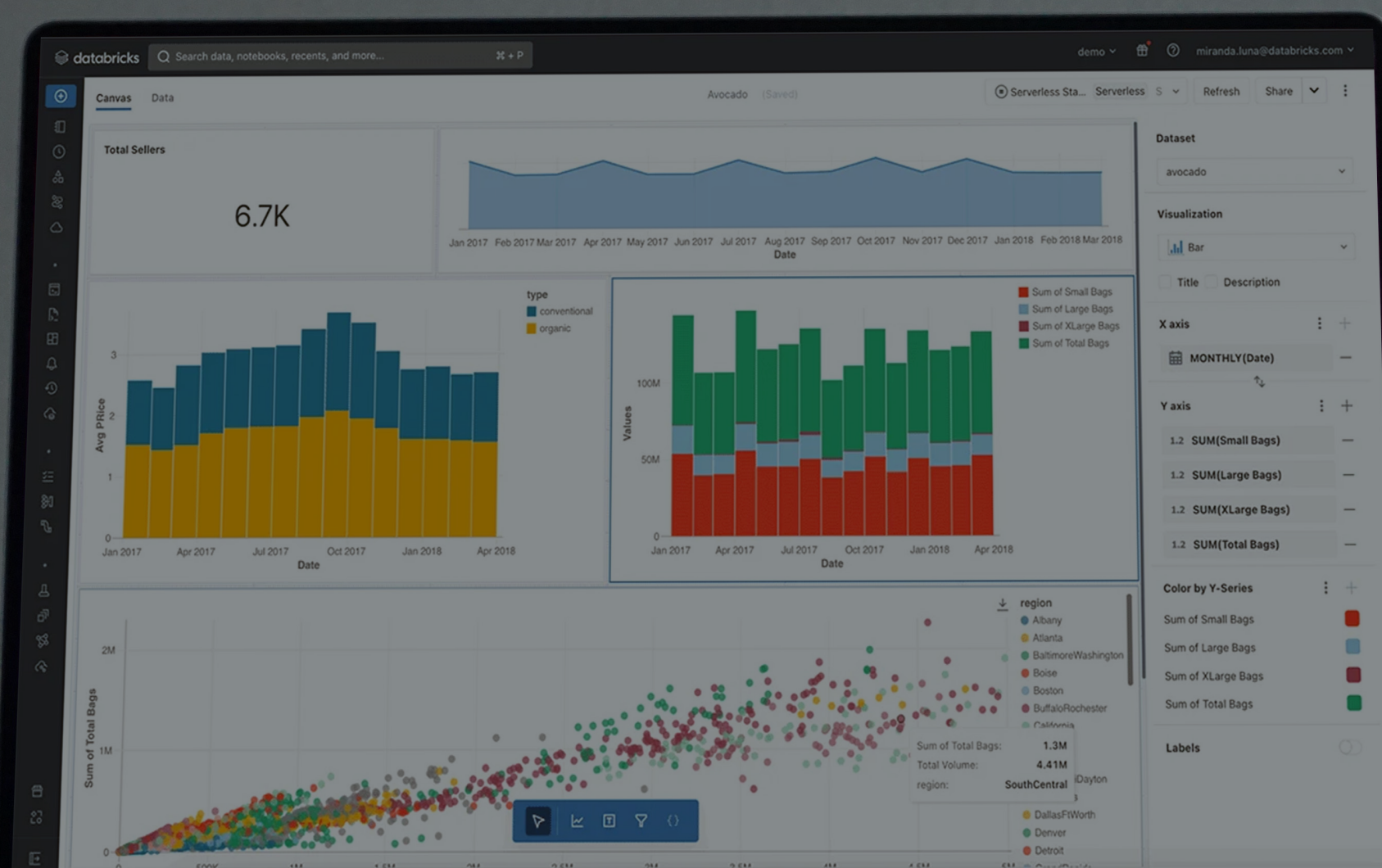




Data Engineer with Databricks

Curso de Especialización





Presentación

Databricks se ha consolidado como una herramienta esencial para integrar, procesar y gobernar grandes volúmenes de información en entornos empresariales y en la nube.

A lo largo del curso y a través de un enfoque práctico y basado en proyectos, los participantes aprenderán a diseñar arquitecturas Lakehouse modernas, construir pipelines de datos escalables, implementar procesos de streaming, gestionar la gobernanza de datos y aplicar buenas prácticas de automatización, DevOps y MLOps dentro del ecosistema Databricks.



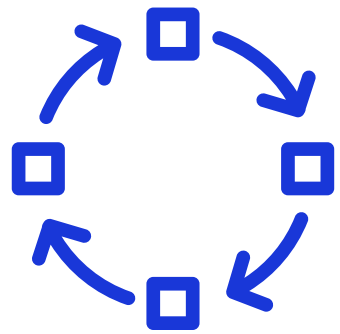
Conocimientos previos

- *Conocimientos básicos de SQL y Python.*
- *Familiaridad con conceptos generales de bases de datos y procesamiento de datos.*

Objetivo:

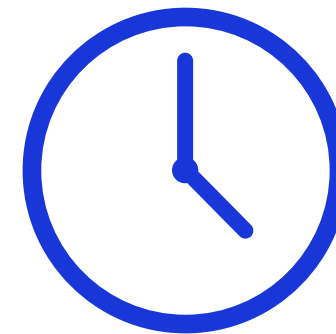
- Capacitar a los participantes en el diseño, implementación y optimización de soluciones de ingeniería de datos sobre Databricks, permitiéndoles construir pipelines modernos, administrar plataformas Lakehouse y aplicar prácticas de procesamiento, gobierno y analítica de datos en entornos empresariales.

Metodología



Sesiones sincrónicas basadas en casuísticas (*casos reales*).

Duración



El curso tiene una duración de **30 horas cronológicas**.

Beneficios



Certificación con **validez internacional**.



Instructores y Consultores especialistas en Data y Analytics, reconocidos a nivel latinoamericano.



Acompañamiento, asesoramiento y desarrollo a los alumnos.



Cuenta educativa de Microsoft Learning Partner (acceso a Office 365 online).



Acceso a la comunidad y Bolsa de Trabajo EXCLUSIVA para nuestros estudiantes.



Temario

Tema 1

Ecosistema Databricks y Arquitectura Lakehouse

- **Arquitectura Lakehouse:** por qué supera al Data Warehouse y al Data Lake por separado.
- **Databricks Platform:** workspace, clusters, Unity Catalog, Repos y estructura de proyectos.
- **Delta Lake 3.x:** fundamentos, ACID transactions, transaction log y checkpoint.
- **Unity Catalog:** gobernanza unificada multi-cloud, metastore, catálogos y schemas.
- **Lab:** Configuración del entorno, Azure CLI, SDK y primer pipeline con ADLS + Synapse Serverless.

Tema 2

Apache Spark Avanzado en Databricks

- **Arquitectura interna de Spark:** DAG, stages, shuffles y optimización del plan de ejecución.
- **Photon Engine:** el motor vectorizado de Databricks y cuándo activa su máximo rendimiento.
- **Spark UI y Databricks Query Profiler:** diagnóstico y tuning de jobs lentos.
- **Optimizaciones avanzadas:** broadcast joins, AQE (Adaptive Query Execution), salting y skew.
- **Lab:** Diagnóstico y optimización de un pipeline Spark con Photon + Query Profiler.

Tema 3

Delta Lake Avanzado y Open Table Formats

- **Delta Lake:** MERGE, UPSERT, DELETE y manejo de CDC (Change Data Capture).
- Time Travel, RESTORE y auditoría con el transaction log de Delta.
- **Liquid Clustering:** reemplazando particionado tradicional con clustering dinámico.
- **Comparativa práctica:** Delta Lake vs Apache Iceberg vs Apache Hudi en 2025.
- **Lab:** Pipeline CDC con MERGE + Time Travel + migración de tabla a Liquid Clustering.



Tema 4

Ingesta de Datos: Auto Loader y Streaming

- **Auto Loader (cloudFiles):** ingesta incremental desde GCS, ADLS y S3 con schema inference.
- **Structured Streaming en Databricks:** triggers, checkpoints y exactly-once semantics.
- **Ingesta desde Kafka / Confluent Cloud** con Spark Structured Streaming.
- **Foreachbatch pattern y Stateful Streaming** para casos avanzados de tiempo real.
- **Lab:** Pipeline de ingesta streaming con Auto Loader + Kafka + Structured Streaming a Delta

Tema 5

Delta Live Tables (DLT): Pipelines Declarativos

- **Delta Live Tables:** arquitectura, tablas live, tablas streaming y expectations (DQ integrada).
- **DLT Expectations:** @dlt.expect, @dlt.expect_or_drop y @dlt.expect_or_fail para calidad.
- **Patrón Medallion con DLT:** Bronze → Silver → Gold de forma declarativa y mantenible.
- DLT Enhanced Autoscaling, pipeline graphs y monitoreo en tiempo real.
- **Lab:** Pipeline Medallion completo con DLT + Expectations + monitoreo de calidad de datos.

Tema 6

Databricks Workflows y Orquestación

- **Databricks Workflows:** jobs, tasks, dependencias y multi-task pipelines.
- Orquestación con parámetros dinámicos, task values y repair & re-run.
- Integración de Workflows con dbt Core para transformaciones SQL en Databricks.
- **Airflow + Databricks:** DatabricksRunNow Operator y DatabricksSubmitRunOperator
- **Lab:** Orquestación end-to-end con Databricks Workflows + dbt + alertas por email/Slack



Tema 7

dbt en Databricks: Transformaciones SQL a Escala

- **dbt Core vs dbt Cloud con Databricks:** setup, profiles y conexión con Unity Catalog.
- **Modelos dbt:** materialización como table, view, incremental y dynamic tables en Databricks
- **dbt incremental models con Delta:** merge strategy, unique_key y configuración avanzada.
- **Tests en dbt:** built-in tests, dbt-expectations y Great Expectations integrado.
- **Lab:** Proyecto dbt completo sobre Databricks: staging → marts + tests + documentación

Tema 8

Databricks SQL y Analítica Self-Service

- **Databricks SQL Warehouses:** serverless vs classic, T-shirt sizing y concurrency scaling.
- **Genie:** analítica conversacional con IA generativa sobre datos en Databricks (2024).
- **AI/BI Dashboards nativos de Databricks:** visualizaciones, filtros y publicación.
- **Query Federation con Databricks:** consultas sobre PostgreSQL, MySQL y Snowflake sin mover datos.
- **Lab:** Dashboard analítico con Databricks SQL + Genie Spaces + Query Federation.

Tema 9

MLOps y Feature Engineering para Data Engineers

- **MLflow en Databricks:** tracking, model registry, Model Serving y A/B testing.
- **Databricks Feature Store / Feature Engineering:** crear, almacenar y servir features.
- **AutoML en Databricks:** generación automática de modelos baseline con código exportable.
- **Mosaic AI (DBRX):** integración de LLMs en pipelines de datos con Databricks.
- **Lab:** Pipeline MLOps completo: Feature Store → AutoML → MLflow Registry → Model Serving.



Tema 10

laC, DevOps, Seguridad y Proyecto Final

- **Terraform + Databricks**
Provider: infraestructura como código para workspaces y clusters.
- **CI/CD para Databricks:**
Databricks Asset Bundles (DAB) + GitHub Actions en producción.
- **Seguridad avanzada:** Unity Catalog row/column filters, data masking y audit logs.
- **Optimización de costos:**
cluster policies, spot instances, autoscaling y cost tagging.
- **Proyecto Final:**
Diseño y presentación de arquitectura Lakehouse end-to-end en Databricks





Instructor



Tony Trujillo
Data Architect and IA Architect

Arquitecto de datos e instructor especializado en Big Data, Cloud Computing e Ingeniería de Datos. Cuenta con formación en Ingeniería de Sistemas, Administración y Dirección de Empresas, un MBA y un Diplomado en Gerencia de Proyectos.

Con experiencia en el diseño de arquitecturas de datos, plataformas cloud (AWS, Azure y Google Cloud), procesos ETL, Big Data, Machine Learning y Deep Learning. Ha participado en proyectos para los sectores de retail, banca, finanzas y telecomunicaciones, destacando por el desarrollo de soluciones de datos que optimizan procesos y apoyan la toma de decisiones.

Certificación

Al finalizar el curso, el alumno obtendrá doble certificación a nombre de la institución.





Empresas líderes del mercado que confían en nosotros



Los Quenuales



GLENCORE





Métodos de pago



BCP - Banco de Crédito
Nombre: Levo Learning SAC
RUC: 20606130661
Cuenta Corriente: 19171497185076
CCI: 00219117149718507658



BBVA - Banco Continental
Nombre: Levo Learning SAC
RUC: 20606130661
Cuenta Corriente: 00107330200144657
CCI: 01173300020014465730



LEVO LEARNING CENTER SAC
@levolearning



Información Adicional

Importante

Las fechas previstas para el inicio del curso ***están sujetas a cambios basados en la cantidad de personas registradas.***

Nuestra prioridad es asegurar que todos los inscritos puedan participar y aprovechar al máximo esta experiencia educativa. Se notificará con antelación cualquier ajuste necesario.

Sé disruptivo en el mundo tecnológico

LLC

Levo Learning Center

Educación Continua
Av. Pershing 890
Lima - Perú

www.levolearning.edu.pe
Cel: +51 986 619 728
informes@levolearning.edu.pe

Avalados por:



Microsoft
Learning Partner



Microsoft
Partner Network