



ExecuTrain

Impulsamos tu talento tecnológico



MICROSOFT

RED HAT

VIRTUALIZACIÓN

CIBERSEGURIDAD

DESARROLLO

OFFICE

BIG DATA

BLOCK CHAIN

BASES DE DATOS

GESTIÓN DE
SERVICIOS IT

CLOUD
COMPUTING

METODOLOGÍAS
EN PROYECTOS

SISTEMAS
OPERATIVOS

Y MÁS...



www.executrain.com.mx



¿Por qué ExecuTrain?

ExecuTrain es un proveedor de entrenamiento corporativo a nivel internacional y líder mundial en la capacitación empresarial. Contamos con más de 30 años de Experiencia y con más de 75 mil personas capacitadas a nivel Nacional.

Te guiamos en la definición de tus requerimientos de capacitación, en las diferentes etapas:

- ✓ Detección de necesidades, evaluación de conocimientos, plan de capacitación y seguimiento posterior para elegir el plan de capacitación como tú lo necesitas.
- ✓ El **más amplio catálogo de cursos**, desde un nivel básico hasta los niveles de conocimientos más especializados.
- ✓ En ExecuTrain el material y la **metodología están diseñados por expertos en aprendizaje humano**. Lo que te garantiza un mejor conocimiento en menor tiempo.
- ✓ Tú puedes confiar y estar seguro del aprendizaje porque nuestro **staff de instructores es de primer nivel**, algunos de los cuales son consultores en reconocidas empresas.
- ✓ No pierdas tu tiempo, los cursos están diseñados para un aprendizaje práctico.

Nuestro compromiso es que tú aprendas, si no quedas satisfecho con los resultados del programa, podrás volver a tomar los cursos hasta tu entera satisfacción o la devolución de tu dinero.

Modalidad de Servicio



Cursos en Fecha Calendario

Súmate a nuestros grupos en fechas públicas.



Cursos Privados

On site, en nuestras instalaciones o en línea con instructor en vivo.



Autoestudio con soporte de instructor

Cursos en modalidad autoestudio, con acceso 24/7 a la plataforma de estudio, con soporte de instructor y foros de ayuda

DP-750 / Implement data engineering solutions using Azure Databricks

Domina la Ingeniería de Datos de extremo a extremo con Azure Databricks y Unity Catalog. Este curso pasa de la configuración básica a la implementación de producción, que abarca la configuración del entorno y la gobernanza de nivel empresarial. Aprenda a crear canalizaciones de ingesta sólidas, implementar la seguridad con el catálogo de Unity e implementar cargas de trabajo optimizadas. Al final, tendrá las aptitudes prácticas para implementar, proteger y mantener soluciones escalables de lakehouse que cumplan los estrictos requisitos empresariales.

Perfil del Público

El público objetivo es ingenieros de datos que tienen conocimientos fundamentales de los conceptos de análisis de datos, un conocimiento básico del almacenamiento en la nube y conocimientos sobre los principios de la organización de datos. Deben estar cómodos trabajando con SQL y tener experiencia en el uso de Python, incluidos cuadernos, para tareas de ingeniería de datos. Se espera que los alumnos conozcan bien las áreas de trabajo de Azure Databricks y el catálogo de Unity, junto con la familiaridad con los patrones de acceso a datos y los conceptos básicos de ingeniería de datos y almacenamiento de datos. Además, deben tener conocimientos básicos sobre Azure seguridad, incluidos los Microsoft Entra ID, y estar familiarizados con los aspectos básicos del control de versiones de Git.

Requisitos Previos

- Conocimientos fundamentales de los conceptos de análisis de datos
- Conocimientos básicos de los conceptos de almacenamiento en la nube
- Conocimientos sobre los principios de la organización de datos y SQL

Módulos

Explorar Azure Databricks

Azure Databricks es un servicio en la nube que proporciona una plataforma escalable para el análisis de datos mediante Apache Spark.

- Introducción
- Comienza con Azure Databricks
- Identificación de las cargas de trabajo de Azure Databricks
- Descripción de los conceptos clave
- Gobierno de datos mediante Unity Catalog y Microsoft Purview

- Ejercicio: Explorar Azure Databricks
- Valoración del módulo
- Resumen

Descripción de la arquitectura de Azure Databricks

La arquitectura de Azure Databricks separa el control y los planos de proceso al organizar los recursos a través de una estructura jerárquica. En este módulo se explora cómo funciona la jerarquía de cuentas, las diferencias entre los planos de proceso clásico y sin servidor, así como las distintas opciones de almacenamiento disponibles, incluido el

almacenamiento predeterminado, el almacenamiento externo y el almacenamiento administrado del catálogo de Unity para organizar y gobernar los datos.

- Introducción
- Descripción de la arquitectura de Azure Databricks
- Comprender el almacenamiento administrado del Unity Catalog
- Comprender el almacenamiento externo
- Descripción del almacenamiento predeterminado (proceso sin servidor)
- Evaluación del módulo
- Resumen

Descripción de las integraciones de Azure Databricks

Azure Databricks se integra con varios servicios de Microsoft para proporcionar funcionalidades de ingeniería, análisis e inteligencia artificial de un extremo a otro. En este módulo se explora cómo funciona Azure Databricks con Microsoft Fabric, Power BI, Visual Studio Code, Power Platform, Copilot Studio, Microsoft Purview y Microsoft Foundry para habilitar soluciones completas que combinan funcionalidades de data lakehouse con inteligencia empresarial, desarrollo de aplicaciones e inteligencia artificial conversacional.

- Introducción
- Descripción de la integración con Microsoft Fabric
- Descripción de la integración con Power BI
- Descripción de la integración con VS Code
- Descripción de la integración con Power Platform
- Descripción de la integración con Copilot Studio
- Descripción de la integración con Microsoft Purview
- Descripción de la integración con Microsoft Foundry
- Evaluación del módulo
- Resumen

Selección y configuración de recursos de cómputo en Azure Databricks

Azure Databricks proporciona varias opciones de proceso optimizadas para diferentes cargas de trabajo. En este módulo se explora cómo elegir el tipo de proceso adecuado, configurar las opciones de rendimiento, administrar permisos de acceso e instalar bibliotecas. Aprenderá a usar el proceso sin servidor frente al proceso clásico, cómo optimizar los clústeres para el costo y el rendimiento, y los procedimientos recomendados para proteger los recursos de proceso.

- Introducción
- Elige un tipo de computación adecuado
- Configurar rendimiento de cómputo
- Configuración de características de proceso
- Instalación de bibliotecas para cómputo
- Configurar acceso a la computación
- Ejercicio: Selección y configuración de la computación en Azure Databricks
- Valoración del módulo
- Resumen

Creación y organización de objetos en el catálogo de Unity

El espacio de nombres de tres capas de Unity Catalog (catálogos, esquemas y objetos) proporciona una base flexible para organizar los recursos de datos al tiempo que mantiene una gobernanza centralizada. En este módulo se explora cómo crear catálogos para el aislamiento del entorno, organizar esquemas dentro de esos catálogos y crear tablas, vistas y volúmenes para datos estructurados y no estructurados. Aprenderá a implementar catálogos externos para el acceso a bases de datos externas, aplicar convenciones de nomenclatura eficaces y configurar instrucciones de AI/BI Genie para mejorar la detectabilidad de los datos.

- Introducción
- Aplicación de convenciones de nomenclatura
- Creación de un catálogo
- Crear esquema
- Creación de tablas y vistas
- Creación de volúmenes

- Implementar operaciones DDL
- Implementar catálogo extranjero
- Configuración de las instrucciones de AI/BI Genie
- Ejercicio: Creación y organización de objetos en el catálogo de Unity
- Prueba de conocimientos
- Resumen

Proteger objetos de catálogo de Unity

El catálogo de Unity proporciona gobernanza y seguridad centralizadas para los recursos de datos en Azure Databricks. En este módulo se explora cómo proteger objetos de Catálogo de Unity mediante estrategias de control de acceso, permisos específicos, administración de credenciales y mecanismos de autenticación. Aprenderá a implementar la seguridad de tabla y de nivel de esquema, aplicar el filtrado de filas y columnas, acceder de forma segura a secretos desde Azure Key Vault y autenticar el acceso a datos mediante entidades de servicio e identidades administradas.

- Introducción
- Descripción del ciclo de vida de las consultas
- Implementación de estrategias de control de acceso
- Entender el control de acceso granular
- Implementar filtrado de filas y enmascaramiento de columnas
- Acceso a secretos de Azure Key Vault
- Autenticación del acceso a datos con entidades de servicio
- Autenticación del acceso a recursos con identidades administradas
- Ejercicio: Protección de objetos de catálogo de Unity
- Valoración del módulo
- Resumen

Controlar objetos de catálogo de Unity

En este módulo se tratan los procedimientos de gobernanza esenciales en el Catálogo de Unity, lo que le permite proteger, supervisar y administrar el patrimonio de datos de forma eficaz. Aprenderá a implementar un control de acceso específico, realizar un seguimiento del linaje de datos, configurar

registros de auditoría y compartir datos de forma segura.

- Introducción
- Creación y conservación de definiciones de tabla
- Configuración de ABAC con etiquetas y directivas
- Aplicación de directivas de retención de datos
- Configuración y administración del linaje de datos
- Configuración del registro de auditoría
- Diseñar una estrategia segura de Delta Sharing
- Ejercicio: Control de objetos de catálogo de Unity
- Evaluación del módulo
- Resumen

Diseño e implementación del modelado de datos con Azure Databricks

El modelado de datos efectivo forma la base de una plataforma de datos eficaz y fácil de mantener. En este módulo se explora cómo diseñar lógica de ingesta, seleccionar las herramientas adecuadas y los formatos de tabla, implementar esquemas de creación de particiones, administrar dimensiones que cambian lentamente, elegir la granularidad de datos adecuada y optimizar el rendimiento de las tablas mediante estrategias de agrupación en clústeres en Azure Databricks con Unity Catalog.

- Introducción
- Diseño de la lógica de ingesta y la configuración del origen de datos
- Elección de una herramienta de ingesta de datos
- Elección de un formato de tabla de datos
- Diseñar e implementar un esquema de partición de datos
- Elegir un tipo de dimensión de variación lenta (SCD)
- Implementar una dimensión de cambio lento (SCD) tipo 2
- Diseñar e implementar una tabla temporal (historial) para registrar los cambios a lo largo del tiempo

- Elección de la granularidad en una columna o tabla en función de los requisitos
- Elección de tablas administradas frente a no administradas
- Diseño e implementación de una estrategia de agrupación en clústeres
- Ejercicio: Diseño e implementación del modelado de datos con Azure Databricks
- Prueba de conocimientos
- Resumen

Ingesta de datos en el catálogo de Unity

La ingesta de datos es una funcionalidad fundamental para cualquier plataforma de datos. En este módulo se explora el conjunto completo de técnicas disponibles en Azure Databricks para cargar datos en tablas de Catálogo de Unity. Aprenderá a usar conectores administrados con Lakeflow Connect, escribir código de ingesta personalizado en cuadernos, aplicar comandos SQL para la carga de archivos por lotes, procesar fuentes de captura de datos modificados, configurar la ingesta de streaming desde buses de mensajes, configurar el cargador automático para la detección automática de archivos y orquestar flujos de trabajo de ingesta con canalizaciones declarativas de Spark de Lakeflow.

- Introducción
- Ingesta de datos con Lakeflow Connect
- Ingerir datos con notebooks
- Ingesta de datos con métodos SQL
- Ingesta de datos con fuente CDC
- Ingesta de datos con Spark Structured Streaming
- Ingesta de datos con cargador automático
- Ingesta de datos con canalizaciones declarativas de Lakeflow Spark
- Ejercicio: Ingesta de datos en el catálogo de Unity
- Valoración del módulo
- Resumen

Limpieza, transformación y carga de datos en el catálogo de Unity

La ingeniería de datos requiere transformar los datos sin procesar en formatos limpios y bien estructurados listos para el análisis. En este módulo se exploran

técnicas para el perfilado de la calidad de los datos, la selección de tipos de columna adecuados, la resolución de duplicados y valores NULL, la aplicación de transformaciones de filtrado y agregación, la combinación de conjuntos de datos con uniones y operadores de conjuntos, el volver a dar forma a los datos mediante pivotación y desnormalización, y la carga de datos transformados mediante estrategias de apéndice, sobrescritura y combinación.

- Introducción
- Generación de perfiles de datos
- Elegir tipos de datos de columna
- Resolución de duplicados y valores NULL
- Transformación de datos con filtros y agregaciones
- Transformación de datos con uniones y operadores de conjuntos
- Transformar datos con desnormalización y pivotes
- Carga de datos con combinación, inserción y anexión
- Ejercicio: Limpieza, transformación y carga de datos en el catálogo de Unity
- Valoración del módulo
- Resumen

Implementación y administración de restricciones de calidad de datos con Azure Databricks

En este módulo se exploran las estrategias para mantener una alta calidad de datos en Azure Databricks. Aprenderá a implementar comprobaciones de validación, aplicar esquemas, administrar el desfase de esquema y usar expectativas de canalización para garantizar la integridad de los datos en todas las canalizaciones de datos.

- Introducción
- Implementación de comprobaciones de validación
- Implementación de comprobaciones de tipos de datos
- Detectar y gestionar la desviación del esquema
- Administrar calidad de datos con expectativas de canalización

- Ejercicio: Implementación y administración de restricciones de calidad de datos con Azure Databricks
- Valoración del módulo
- Resumen

Diseño e implementación de canalizaciones de datos con Azure Databricks

Aprenda a diseñar e implementar canalizaciones de datos sólidas en Azure Databricks mediante cuadernos y canalizaciones declarativas de Spark de Lakeflow, que incluyen orquestación, control de errores y lógica de tareas.

- Introducción
- Orden de diseño de las operaciones de una canalización
- Elegir Notebook o Lakeflow Pipelines
- Lógica de trabajo de Design Lakeflow
- Diseño del control de errores en canalizaciones y trabajos
- Crear una canalización con cuadernos
- Crear una canalización con las canalizaciones declarativas de Lakeflow Spark
- Ejercicio: Diseño e implementación de canalizaciones de datos con Azure Databricks
- Valoración del módulo
- Resumen

Implementación de trabajos de Lakeflow con Azure Databricks

Este módulo le guía por el proceso de implementación de trabajos de Lakeflow en Azure Databricks. Aprenderá a crear trabajos, configurar desencadenadores y programaciones, configurar alertas y administrar reinicios automáticos para garantizar la ejecución confiable de la canalización de datos.

- Introducción
- Crear configuración e instalación del trabajo
- Configuración de desencadenadores de trabajo
- Programar un trabajo
- Configuración de alertas de trabajo
- Configurar reinicios automáticos

- Ejercicio: Implementación de trabajos de Lakeflow con Azure Databricks
- Valoración del módulo
- Resumen

Implementación de procesos de ciclo de vida de desarrollo en Azure Databricks

Azure Databricks se integra con los procedimientos de desarrollo establecidos a través de carpetas de Git para el control de versiones y Databricks Asset Bundles para implementaciones de infraestructura como código. En este módulo se exploran las mejores prácticas de control de versiones con Git, los flujos de trabajo de bifurcación y solicitudes de incorporación de cambios, las estrategias de prueba exhaustivas y la implementación de paquetes mediante la CLI en distintos entornos.

- Introducción
- Aplicación de procedimientos recomendados de control de versiones de Git
- Gestionar ramas y solicitudes de extracción
- Implementación de la estrategia de pruebas
- Configuración y empaquetado de DAB
- Desplegar paquete con la CLI de Databricks
- Ejercicio: Implementación de procesos de ciclo de vida de desarrollo en Azure Databricks
- Valoración del módulo
- Resumen

Supervisión, solución de problemas y optimización de cargas de trabajo en Azure Databricks

La supervisión y la optimización son esenciales para ejecutar cargas de trabajo de datos confiables y rentables en Azure Databricks. En este módulo se exploran las métricas de consumo de clúster, la solución de problemas de trabajos de Lakeflow, el diagnóstico de trabajos de Spark, la optimización del rendimiento para el almacenamiento en caché, el sesgo, el desbordamiento y los problemas aleatorios, y el streaming de registros a Azure Log Analytics.

- Introducción
- Supervisión y administración del consumo de clústeres
- Solución de problemas y reparación de trabajos de Lakeflow

- Solucionar problemas de trabajos y cuadernos de Spark
- Investigar el almacenamiento en caché, la distorsión, el derrame y la mezcla.
- Implementación del streaming de registros con Azure Log Analytics
- Ejercicio: Supervisión, solución de problemas y optimización de cargas de trabajo en Azure Databricks
- Valoración del módulo
- Resumen