

25 FEBRERO



PROCESAMIENTO DE BIG DATA CON PYSPARK Y CLOUD PLATFORMS

MODALIDAD VIRTUAL

Docente: Lic. Manuel Illanes

Avalado por:





Procesamiento de Big Data con PySpark y Cloud Platforms

Objetivo Principal

Desarrollar competencias prácticas para el procesamiento, transformación y análisis de grandes volúmenes de datos utilizando PySpark, integrándolo con plataformas cloud para la resolución de escenarios empresariales reales.

Objetivos Específicos

- Comprender los fundamentos del ecosistema Big Data y la arquitectura distribuida de Apache Spark.
- Manipular y transformar datos estructurados mediante el uso eficiente de la interfaz PySpark.
- Aplicar operaciones distribuidas optimizadas para el procesamiento de grandes volúmenes de información.
- Integrar flujos de trabajo de PySpark con entornos cloud y arquitecturas de data lake.
- Resolver casos de uso reales aplicados a entornos industriales y estratégicos.

Contenido

Unidad 1: Introducción a Big Data y Arquitecturas Distribuidas

- Concepto de Big Data: volumen, velocidad, variedad, veracidad y valor.
- Diferencias entre procesamiento tradicional y procesamiento distribuido.
- Casos de uso de Big Data en entornos empresariales e industriales.
- Introducción a arquitecturas modernas: Data Warehouse vs. Data Lake.
- Rol de Spark dentro del ecosistema Big Data.

Unidad 2: Fundamentos de Apache Spark y PySpark

- ¿Qué es Apache Spark y por qué se utiliza?
- Arquitectura de Spark: driver, executors y cluster manager.
- PySpark como interfaz de Spark para Python.
- Modos de ejecución: local vs. cluster.
- Conceptos clave: paralelismo, particiones y tolerancia a fallos.

Unidad 3: Estructuras de Datos en PySpark

- Conceptos y casos de uso de RDDs y DataFrames.
- Comparación técnica y creación desde diversas fuentes.
- Inspección y exploración inicial de datos masivos.

Unidad 4: Transformaciones y Acciones en PySpark

- Transformaciones comunes: select, filter, withColumn y groupBy.
- Acciones principales: count, show, collect y take.
- Operaciones de agregación y manejo de grandes volúmenes de datos.
- Optimización básica de consultas en Spark.
- Buenas prácticas para escribir código PySpark eficiente.

Unidad 5: Integración con Almacenamiento y Plataformas Cloud

- Gestión de Data Lakes y almacenamiento distribuido.
- Lectura y escritura optimizada en formatos CSV y Parquet.
- Integración de PySpark con servicios de almacenamiento en la nube.
- Fundamentos de entornos Cloud, escalabilidad y seguridad de datos.

Unidad 6: Casos Prácticos Aplicados a Entornos Empresariales

- Procesamiento y limpieza de datos transaccionales a gran escala.
- Transformación de datos crudos (Data Wrangling) en tablas analíticas.
- Construcción de pipelines de datos eficientes con PySpark.
- Resolución de escenarios reales y recomendaciones de implementación.

Dirigido a

Profesionales de TI, analistas, científicos e ingenieros de datos, así como profesionales de áreas industriales y desarrolladores interesados en especializarse en el procesamiento de Big Data y arquitecturas distribuidas.

Duración

Febrero

Lu	Ma	Mi	Ju	Vi	Sá	Do
23	24	25	26	27	28	

Marzo

Lu	Ma	Mi	Ju	Vi	Sá	Do
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

Horario



Lunes, miércoles, jueves
y viernes

19:30 a 21:30 h.

8 horas reloj

Modalidad

• **Virtual** para toda Bolivia y LATAM

Alcance

Este programa proporciona las herramientas fundamentales para dominar el procesamiento de datos a gran escala utilizando PySpark. El contenido abarca desde la arquitectura distribuida de Apache Spark hasta la implementación de transformaciones complejas y la integración con entornos cloud y data lakes. Al finalizar, el participante será capaz de construir pipelines de datos eficientes, optimizar consultas sobre volúmenes masivos de información y resolver desafíos técnicos reales mediante el uso de infraestructuras modernas de Big Data.

Inversión

PRECIO
NORMAL

680 Bs.

PREVENTA
(HASTA EL 18 DE FEBRERO)

630 Bs.

PRECIO CORPORATIVO
(3 O MÁS PERSONAS)

580 Bs.

- Descuento del 20% para afiliados de la S.I.B. Santa Cruz



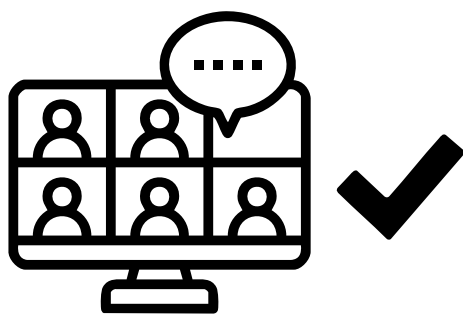


Lic. Manuel Illanes

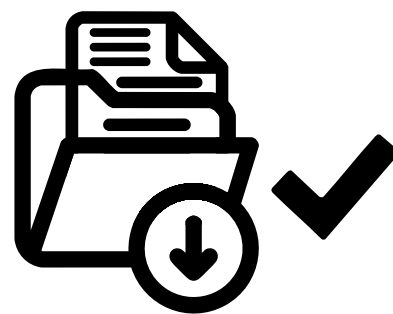
Profesional con sólida formación académica y especialización en la gestión estratégica de información. Su enfoque se centra en la optimización de procesos organizacionales a través de herramientas tecnológicas avanzadas. Docente con experiencia en:

- Docencia universitaria en instituciones de educación superior pública y privada.
- Especialista en capacitación técnica para entornos empresariales y programas de diplomado.
- Experto en análisis de datos aplicado a la gestión y toma de decisiones organizacionales.
- Diseño e impartición de cursos especializados en productividad y entornos digitales.
- Consultoría interna para la mejora de flujos de trabajo y gestión del conocimiento.

¿Qué Incluye tu Inscripción?



Clases virtuales en vivo por GOOGLE MEET junto al docente y tus compañeros, una oportunidad de networking.



Material digital de apoyo disponible de manera indefinida en la nube, desde donde podrás ver y/o descargar tus clases grabadas y documentos.
Factura.



Certificado con valor curricular con **11 horas académicas**, avalado por Wibel, marca oficial registrada de GRUPO BELCAS con matrícula de comercio e identificación tributaria: 375983023 y el Colegio de Ingenieros Industriales de Santa Cruz.

¿Cómo Inscribirte?

1 Ingresa a :

wibel.net 

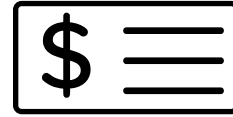
Inscripciones habilitadas 24/7



Visítanos a la Calle Jaurú #2410 , Entre 2 y 3 Anillo, Entre Av Paraguá y Av. Guapay, Santa Cruz de la Sierra, Bolivia.

HORARIO DE ATENCIÓN:
De 08:30 a 12:30 y de 14:30 a 18:30 de lunes a viernes y los sábados de 08:30 a 12:30.

2



Comparte tu comprobante digital de pago donde indique el monto, fecha y hora del pago efectuado; a uno de los siguientes contactos de WhatsApp:



+591 69956117

+591 78528175

Métodos de Pago:



Tigo Money: 76070714
Titular: BELCAS S.R.L.



Nº Cuenta: 1041416494
Cuenta Corriente
Nit: 375983023

QR:



Titular: de BELCAS S.R.L.
Nº Cuenta: 1041416494 / Cuenta
Corriente NIT: 375983023