

# BIG DATA





# NUESTRO PROGRAMA

La velocidad de cambio de los modelos de Negocio y la exigencia competitiva hace imperativa la necesidad de actualizar nuestras habilidades en la gestión de grandes cantidades de datos para los Profesionales de Gestión, Análisis de datos, ingeniería y arquitectura de sistemas de cómputo entre otros especialistas que buscan adaptarse para liderar el mercado actual.

En este curso aprenderás a:

- Diseñar modernas Arquitecturas de Datos y comprender el uso de componentes importantes en una Arquitectura de datos como Data Lakes.
- Brindar una visión 360° del rol de la estrategia de datos en la Transformación Digital como habilitador de contemplando incluye arquitecturas y componentes tecnológicos, así como roles y procesos orquestados que permiten resolver las necesidades de negocios.
- Comprender el uso de Herramientas de Big Data en Cloud y cómo ellas encajan en nuevas y modernas arquitecturas de datos, permitiendo explotar sus funcionalidades aprovechando la riqueza de datos semi estructurados o no estructurados para mejorar el contexto y la base de análisis.
- Conocer el ecosistema de Big Data, con Apache Spark como herramienta, aprender a utilizarlo para el procesamiento a gran escala para preparación y análisis de datos, comprender sus características, arquitectura y potencial en un nuevo ecosistema.





# INFORMACIÓN GENERAL



**Inicio:**  
06 de marzo



**Fin:**  
20 de abril



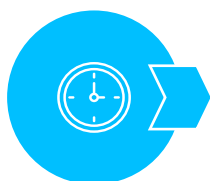
**Modalidad:**  
Live Online



**Duración:**  
6 semanas  
(36 horas lectivas)



**Inversión:**  
S/ 3,850.00



**Horarios\*:**  
Martes y Jueves  
de 19:00 a 22:00 hrs

\*Revisar [cronograma](#).





# PERFIL DEL INTERESADO



Analistas de datos, TI, líderes y especialistas que trabajen con Business Intelligence y/o participen en proyectos de Gestión de Datos.



Ingenieros de Datos, desarrolladores, Arquitectos de Sistemas y Administradores de Bases de Datos, que desean y/o requieran ingresar al mundo del Big Data.



# PERFIL DEL EGRESADO

Al culminar el curso tendrás la oportunidad de trabajar con herramientas modernas, y serás capaz de:



Diseñar modernas Arquitecturas de Datos y comprender el uso de componentes importantes en una Arquitectura de datos como Data Lakes.



Comprender arquitecturas Cloud para diseños de Análisis de datos, así como comprender y ser capaz de utilizar frameworks de procesamiento de datos Core como Apache Hadoop y Apache Spark 3.0 para preparación y Análisis de Datos.



Utilizar y habilitar el uso de lenguajes de consulta como SQL o Python en plataformas de software Libre de manera ágil para el análisis de Datos.



# CRONOGRAMA<sup>1</sup>

Módulo	Sesión	Fecha	Tema	Objetivo
<b>Bienvenida:</b> Sesión de inmersión (Obligatoria)	0	Lunes, 06 de marzo	Apertura, presentación y reglas del curso.  Introducción a las plataformas.	<ul style="list-style-type: none"> <li>➤ Bienvenida e introducción a las herramientas y requisitos necesarios para que el participante pueda llevar el curso adecuadamente.</li> <li>➤ Networking.</li> </ul>
<b>Módulo 1:</b> Foundations de una estrategia Data & Analytics (Obligatoria)	1	Martes, 07 de marzo	Comprender el rol de los datos en la organización, en una visión empresarial Data Driven con la perspectiva de Transformación Digital.	<ul style="list-style-type: none"> <li>➤ Ecosistema del Big Data y presentar la Estructura del curso.</li> <li>➤ Enfoques: Proyectos e Iniciativas D&amp;A o Data Driven Enterprise.</li> <li>➤ Componentes de una estrategia Data Driven.</li> <li>➤ Oportunidades del uso de los datos para la Transformación Digital.</li> </ul>

<sup>1</sup> La Universidad de Ingeniería y Tecnología, se reserva el derecho de postergar, reprogramar o cancelar el curso en caso no se cuente con la cantidad mínima necesaria de participantes, así como modificar la programación de alguna sesión agendada en caso de fuerza mayor y de manera excepcional fuera del horario establecido. De igual manera, se reserva el derecho de reemplazar al expositor que por algún motivo de fuerza mayor no pueda asistir a las sesiones programadas, por un expositor de similar calidad.





Módulo	Sesión	Fecha	Tema	Objetivo
<b>Módulo 2:</b> Introducción a arquitectura de datos (Obligatoria)	2	Jueves, 09 de marzo	Comprender las arquitecturas de datos, su importancia y rol dentro de la transformación digital, además introduciremos a algunos proveedores como cloud computing y sus arquitecturas de referencia.	<ul style="list-style-type: none"><li>➤ Arquitectura Empresarial: Tecnología, procesos y personas.</li><li>➤ Arquitectura de Datos: la importancia del diseño, Componentes, función, configuración y patrones.</li><li>➤ Roles y responsabilidades de los involucrados en la gestión de Datos en Big Data.</li></ul>
<b>Módulo 3:</b> Fundamentos de Apache Hadoop y Spark (Obligatoria)	3	Martes, 14 de marzo	Conocer la arquitectura y diseño del ecosistema Big Data basado en Apache Hadoop y Apache Spark, sus componentes y el uso como plataformas complementario para procesamiento y análisis Big Data.	<ul style="list-style-type: none"><li>➤ Arquitectura y Diseño del ecosistema Big Data basado en Apache Hadoop y Apache Spark.</li><li>➤ Componentes, lenguajes y Frameworks de Procesamiento.</li></ul>
<b>Módulo 4:</b> Tecnologías Big Data Cloud (Obligatoria)	4	Jueves, 16 de marzo	Conocer diferentes herramientas de gestión de datos en cloud clasificados en herramientas de integración de datos, almacenamiento y procesamiento.	<ul style="list-style-type: none"><li>➤ Herramientas de gestión de Datos en Cloud: Integración de Datos, almacenamiento y procesamiento.</li></ul>





Módulo	Sesión	Fecha	Tema	Objetivo
<b>Módulo 5:</b> Iniciando con Apache Spark (Obligatoria)	5	Martes, 21 de marzo	Introducir al uso de Apache Spark, conocer sus estructuras y tipos de datos de forma práctica comprendiendo el uso de las APIs para analizar grandes datos con Apache Spark SQL.	<ul style="list-style-type: none"><li>Introducir al uso de Apache Spark, conocer sus estructuras y tipos de datos de forma práctica comprendiendo el uso de las APIs para analizar grandes datos con Apache Spark SQL.</li><li>Mi Primer Cluster.</li><li>Sintaxis básica de interacción con la información usando Spark.</li></ul>
	6	Jueves, 23 de marzo	Comprender el uso de Spark a través de sus funciones y capacidades prácticas.	<ul style="list-style-type: none"><li>Diseño de programas de procesamiento de datos y patrones en un Ecosistema de datos.</li></ul>
	7	Martes, 28 de marzo		<ul style="list-style-type: none"><li>Comprender el uso de Spark y sus funciones a través de carga, almacenamiento y modificación de datos con diferentes Datasets.</li></ul>
8	Jueves, 30 de marzo			
Semana Santa				





Módulo	Sesión	Fecha	Tema	Objetivo
<b>Módulo 7:</b> Análisis de datos (Obligatoria)	9	Martes, 11 de abril	Comprender el uso de Apache Spark y su relación con la ciencia de datos en gran escala.	▶ Desarrollo de un caso de análisis de datos tradicional utilizando Apache Spark y desarrollar casos de uso con datos reales.
	10	Jueves, 13 de abril		
	11	Martes, 18 de abril		
<b>Módulo 9:</b> Conclusión y Capstone Project (Obligatoria)	12	Jueves, 20 de abril	Determinar qué tecnología o combinación de estas es la mejor para una tarea o reto determinado.	▶ Determinar qué tecnología o combinación de estas es la mejor para una tarea o reto determinado.  ▶ Conclusiones y presentación del Capstone Project.

**N° de Horas síncronas: 36<sup>2</sup>**

<sup>2</sup>**Síncronas:** Sesiones live online con horario establecido.

# EXPERTOS



## LUIS CHÁVEZ

→ **Data Architect**  
at **Telefónica HISPAM**

Digital Expert con especialización en transformación Digital por University of California, Berkeley, Haas School of Business. Especialista en diseño de implementación de arquitectura de Datos, desarrollo de aplicaciones y plataformas Big Data; certificado con Cloudera, Databricks, Microsoft Azure, así como con Especialización en Análisis de Datos por la Universidad de Washington y The Caltech-JPL Summer School on Big Data Analytics. Ex Arquitecto de datos en Centro de Excelencia Big Data del Banco de Crédito del Perú. Actualmente es Data Architect at Telefónica HISPAM.

## DIEGO BENAVIDES

**Data Center of Excellence Leader** ←  
at **Telefónica HISPAM**

Científico y Arquitecto de datos Senior, con experiencia en análisis predictivo y aprendizaje automático trabajando en soluciones de telecomunicaciones, scores de crédito, segmentación de clientes, motores de recomendación, gestión de registros para sistemas de seguridad y proyectos de innovación utilizando capacidades de inteligencia artificial como visión artificial, robótica, IoT e interacción humana. Master en Sistemas Mecatrónicos por la Universidad de Brasilia. Ex Senior Data Architect at Belcorp, actualmente es Data Center of Excellence Leader at Telefónica HISPAM.



\* La Universidad de Ingeniería y Tecnología, se reserva el derecho de reemplazar al expositor que por algún motivo de fuerza mayor no pueda asistir a las sesiones programadas, por un expositor de similar calidad.



# SOBRE EL CURSO

## METODOLOGÍA

Durante el Curso Especializado en Big Data Cloud and Insights, se enseñará a aplicar el análisis sobre datos estructurados, así como datos semiestructurados y no estructurados complejos, permitiendo complementar sus habilidades de análisis utilizando herramientas de Big Data como Apache Spark, y; componente de procesamiento y análisis de macro datos en nubes públicas (cloud computing), a través de código SQL, permitiendo comprender herramientas aplicables en cualquier entorno empresarial On-Premise o Cloud como Snowflake, Cloudera, Oracle, Google Cloud Platform, Azure o AWS. Además, utilizaremos datasets públicos de repositorios conocidos como UCI Machine Learning Repository, Kaggle y Google Dataset Search, que nos permitirán poner en práctica el conocimiento de estas tecnologías tomando en cuenta casos reales de industria.

Requisitos técnicos del equipo a utilizar:

- Una computadora con al menos 8 GB de RAM es altamente deseada.
- Internet Fluido.
- Navegador Chrome.
- Un sistema operativo de 64 bits (Windows o Linux).



# SOBRE EL CURSO

## METODOLOGÍA

Requisitos de conocimiento previo:

Para un adecuado desarrollo dentro del curso, es necesario que el participante posea conocimientos básicos de SQL y Python.

Compartimos dos cursos gratuitos para su nivelación, en caso sea necesario:

- [Introducción a SQL con Oracle 11g.](#)
- [Introduction to SQL - DATACAMP.](#)
- No se requiere conocimiento previo de Apache Hadoop o Apache Spark.
- Conocimiento básico con la línea de comando de Linux (Deseable).
- Habilidades informáticas básicas.



# **SOBRE EL CURSO**

## **METODOLOGÍA**

Requisitos de conocimiento previo:

El presente curso otorga los conocimientos necesarios en un 70% para la preparación del examen de Certificación Apache Spark 3 - Databricks Certified Associate Developer, que es un aval otorgado por la compañía Databricks, de forma independiente, el cual constata las competencias adquiridas en el uso de Apache Spark 3.0.

El curso está desarrollado en 36 horas lectivas en total de sesiones síncronas y tiene previsto iniciar en el mes marzo y culminar en el mes de abril. El horario regular de las sesiones es martes y jueves de 19:00 a 22:00 hrs., teniendo la sesión de inmersión en día distinto de acuerdo al cronograma detallado.



# EVALUACIÓN

La evaluación es vigesimal siendo la nota mínima aprobatoria 13.00.

El curso combina la teoría con la práctica, por lo que se requiere del compromiso del participante en las diferentes actividades preparadas a lo largo de las clases.

Es importante y necesaria la asistencia a las sesiones programadas para poder seguir la ilación del desarrollo del curso. De presentar algún inconveniente, puede reportarlo con el Asistente Virtual. De igual manera, se sugiere revisar la grabación de la sesión que será subida a la plataforma tecnológica. Si, en caso, el participante no puede asistir a alguna de las sesiones programadas, entonces deberá ponerse en contacto con el grupo que se le asignó.

**Inasistencia:** Se considera como inasistencia las faltas justificadas e injustificadas. El participante debe revisar la grabación, realizar las tareas y practicar el desarrollo de la clase antes de la siguiente clase síncrona. Las sesiones de inmersión y de conclusión y Capstone Project son de carácter obligatorio. Las sesiones de práctica se desarrollarán a lo largo del curso en las sesiones correspondientes a los módulos 3, 5, 6 y 7, por tal motivo estas clases son cruciales y se recomienda evitar la inasistencia.

## Evaluación

Participación en clase (Aporte*)	<b>30 %</b>
Casos de estudio	<b>30 %</b>
Capstone Project	<b>40 %</b>
Total	<b>100 %</b>

\* Se considera como aporte, aquello que busca contribuir o complementar los conocimientos que se van adquiriendo a lo largo de las sesiones de clases.





# EVALUACIÓN

**Caso de estudio:** El participante seleccionará una empresa Data Driven y analizará sus características de uso de análisis de datos para la toma de decisiones y discutirá algunos aspectos asociados a la estrategia la generación de valor, así como aspectos operativos importantes que ayudarán a la comprensión de gestión de estrategias Data Driven.

**Capstone Project:** Será planteado durante las primeras semanas para que el estudiante pueda ir planteando versiones evolutivas de las soluciones con lo aprendido y pueda ajustar a medida que evoluciona las sesiones de aprendizaje. Principalmente, está asociado al planteamiento de algún proceso de negocio u objetivo estratégico y su propuesta de resolución considerando propuesta base de arquitectura de datos y técnicas de análisis con la recomendación sustentada de alguna tecnología a optar.



# DEDICACIÓN ACADÉMICA

Las horas de dedicación fuera de las sesiones de clase son de mínimo 4 horas a la semana. Esto comprende revisión del material del curso sin horario establecido y tiempo aproximado para reuniones grupales, en caso sea necesario.



# ACOMPañAMIENTO VIRTUAL

El participante del curso contará con el apoyo de un Asistente Académico Virtual a lo largo de las sesiones ante cualquier consulta relacionada al curso.

Para otras preguntas fuera del horario de clases:



**De lunes a viernes de 10:00 a 13:00 y de 17:00 a 20:00 hrs., con un nivel de respuesta aproximado de 48 horas hábiles.**



# MATERIAL DEL PROGRAMA

El participante accederá al material relacionado al contenido del curso; para esto se habilitará el acceso online dentro de nuestra plataforma de aprendizaje. Todo el material estará disponible en formato digital. Puede entregarse en inglés dado que es el idioma que muestra los últimos estudios realizados de manera internacional. El material recomendado o adicional podría tener un costo.

Las clases síncronas serán grabadas (salvo restricciones indicadas en la normativa de Protección de Datos Personales) y se alojará el video a través de nuestras plataformas de aprendizaje que permitan la revisión posterior del participante. Dicha grabación se puede visualizar únicamente en línea, bajo ningún motivo se permite su descarga total o parcial a ningún dispositivo.

El vídeo se facilita únicamente para el participante del curso. El uso del material y videos son exclusivos para la enseñanza del curso en el cual el participante está inscrito. Asimismo, por protección de la propiedad intelectual, la descarga, copia, reproducción y/o compartida del material del curso, de manera parcial o total, está prohibido. La Universidad es titular de todos los derechos de propiedad intelectual referentes al contenido y se reserva las acciones legales que puedan tomarse en caso el participante infrinja esta disposición.



# HERRAMIENTAS DE APRENDIZAJE\*

Harás uso de las siguientes plataformas tecnológicas que permitirán el logro de tu aprendizaje:



**Zoom**

Sistema de videoconferencia multiplataforma que permite la interacción entre participantes y expositores en tiempo real.



**Slack**

Herramienta de comunicación que permite el networking entre los participantes.



**Google Drive**

Permite centralizar de manera ordenada y secuencial los diferentes recursos digitales que son el soporte para lograr los aprendizajes esperados.



**Miro**

Plataforma que permite un trabajo colaborativo de manera virtual.

\*Sujeto a modificaciones por Educación Ejecutiva UTEC sin previo aviso.



# CERTIFICACIÓN



UTEC otorgará un Certificado digital\* a los participantes que aprueben el “Curso Especializado en Big Data” en un plazo máximo de 15 días hábiles posterior a la fecha de cierre\*\*. El documento es firmado por la Dirección de Educación Ejecutiva.

Texto del documento:

Por haber aprobado el Curso Especializado en Big Data, desarrollado en 36 horas lectivas de clase, del (día y mes de inicio) al (día y mes de fin) del (año).

*\*El certificado se envía de manera digital al correo registrado durante el proceso de venta, a través de la cuenta:  
[comunicacioneseducacionejecutiva@utec.edu.pe](mailto:comunicacioneseducacionejecutiva@utec.edu.pe)*

*\*\*Fecha de cierre del Curso: se entiende cuando se ha entregado y evaluado el trabajo final, según sea el caso.*



# Reglamento de Educación Ejecutiva

Puedes revisar el reglamento [AQUÍ](#)

## INFORMES

Supervisora de Ventas



Taina Barboza



ybarboza@utec.edu.pe



Acerca de UTEC

La Universidad de Ingeniería y Tecnología (UTEC) viene desarrollando la ingeniería que se requiere en el mercado actual y para las empresas y sociedad del futuro. Una ingeniería que va de la mano con la investigación y la creación de soluciones tecnológicas de vanguardia, comprometida con las necesidades sociales y la sostenibilidad. UTEC es una propuesta de educación superior dedicada a formar profesionales capaces de llevar su ingenio a la práctica. Además, UTEC es la única opción universitaria en el mercado que cuenta con una propuesta curricular de aprendizaje activo y basado en proyectos, tal como lo hacen los planes de estudios de las mejores escuelas de ingeniería del mundo.

© 2022 UTEC  
All Rights Reserved.