

DATA ENGINEERING ON GCP

Duración: 4 Días **Código del Curso: IFCT171PO**

Temario:

PROGRAMAS DE CAPACITACIÓN Y TRANSFORMACIÓN DIGITAL DE ÁMBITO ESTATAL, DIRIGIDOS PRIORITARIAMENTE A LAS PERSONAS OCUPADAS

Esta formación está subvencionada al 100% por el Servicio de Empleo Público Estatal (SEPE).

El programa de formación gratuita de SEPE es la herramienta perfecta para actualizarte, adquirir nuevas habilidades laborales, y adaptarte a los requerimientos actuales del mercado de trabajo.

Tienes disponibles cursos gratuitos en áreas como la programación, la seguridad informática, la nube, gestión de servicios IT, entre otros.

QUÉ INCLUYE?

El programa incluye sesiones de formación en diferentes tecnologías, la asistencia es totalmente gratuita, y el programa está dirigido prioritariamente a profesionales de diversos sectores de actividad.

Las formaciones están dirigidas por un instructor experto en áreas como la programación, la seguridad informática, la nube, gestión de servicios IT, entre otras.

QUIÉNES PUEDEN ACCEDER A LA FORMACIÓN?

- Este programa es compatible para las empresas con el uso de otro tipo de fondos de ayuda para la formación, estatales o autonómicos (Fundae, subvenciones directas, etc)
- La formación es gratuita tanto para la empresa como para el empleado, sin entrar en conflicto con el programa de bonificación de Fundae.
- Este programa es de ámbito estatal y está destinado principalmente a profesionales en activo que desempeñen su labor en los siguientes sectores de actividad: Economía Digital

OBJETIVO GENERAL

Diseñar y construir sistemas de procesamiento de datos en Google Cloud Platform, crear canales de datos de extremo a extremo, analizar datos y llevar a cabo el aprendizaje automático, tanto con datos estructurados, como no estructurados y de transmisión.

Dirigido a:

Desarrolladores responsables de la administración de las transformaciones big data, incluyendo: Extracción, carga, transformación, limpieza y validación de datos. Diseño de pipelines y arquitecturas para procesamiento de datos. Creación y administración de machine learning y modelos estadísticos. Realizar consultas a dataset, visualización de resultados de consultar y creación de reportes.

Objetivos:

- Diseñar y construir sistemas de procesamiento de datos en Google Cloud Platform.
- Procesar datos batch y transferencia continua implementando canales de autoescalado de datos en Cloud Dataflow.
- Conseguir información de negocio de datasets enormes utilizando Google BigQuery.
- Evaluar y predecir utilizando modelos machine learning utilizando Tensorflow y Cloud ML.
- Hacer uso de datos no estructurado utilizando Spark y APIs ML en Cloud Dataproc.
- Habilitar información instantánea de la transferencia continua de datos.
- Habilitar información instantánea de la transferencia continua de datos.
- Proporcionar información, consejos y sugerencias de cara a la realización del examen de certificación asociado al presente curso.
- Revisar en profundidad varios ejemplos de estudio de casos.
- Identificar aquellas áreas de estudio en las deban mejorar.

Prerequisitos:

- Estar registrado como empleado activo en la Seguridad Social, ya sea como trabajador por cuenta propia o ajena, o como solicitante de empleo en el Servicio Público de Empleo Estatal (SEPE).

- Para cada curso al que el alumno quiera asistir, recopilar y devolver al centro de formación la documentación que se le solicitará.
 - Atender al menos a un 75% de la extensión del curso para poder recibir el diploma de asistencia correspondiente.
 - Contar con las habilidades necesarias para utilizar la plataforma virtual WEBEX, que se emplea en la formación.
-

Contenido:

1. DESCRIPCIÓN GENERAL DE GOOGLE CLOUD DATAPROC..

1.1. Creando y administrando clústeres.

1.2. Aprovechando los tipos de máquina personalizados y los nodos de trabajadores preferentes.

1.3. Escalar y eliminar clústeres.

1.4. Laboratorio: Creación de clústeres Hadoop con Google Cloud Dataproc

2. EJECUCIÓN DE TRABAJOS DE DATAPROC.

2.1. Ejecución de trabajos Pig y Hive.

2.2. Separación de almacenamiento y cálculo.

2.3. Laboratorio: ejecutar Hadoop y Spark Jobs con Dataproc.

2.4. Laboratorio: enviar y monitorear trabajos.

3. INTEGRANDO DATAPROC CON GOOGLE CLOUD PLATFORM.

3.1. Personalice el clúster con acciones de inicialización.

3.2. Soporte BigQuery.

3.3. Laboratorio: Aprovechamiento de Google Cloud Platform Services.

4. DAR SENTIDO A LOS DATOS NO ESTRUCTURADOS CON LAS API.

4.1. API de Aprendizaje automático de Google.

4.2. Casos de uso de ML comunes.

4.3. Invocando ML API.

5.10. Rendimiento y fijación de precios.

6. Canalizaciones de datos de escalado automático sin servidor con Dataflow.

6.1. El modelo de programación Beam.

6.2. Canalizaciones de datos en Beam Python.

6.3. Canalizaciones de datos en Beam Java.

6.4. Laboratorio: escribir una canalización de flujo de datos.

6.5. Procesamiento escalable de Big Data usando Beam.

6.6. Laboratorio: MapReduce en Dataflow.

6.7. Incorporando datos adicionales.

6.8. Laboratorio: entradas laterales.

6.9. Manejo secuencial de datos.

6.10. Arquitectura de referencia GCP.

6.11. Aprendizaje automático sin servidor con TensorFlow en Google Cloud Platform.

7. PRIMEROS PASOS CON EL APRENDIZAJE AUTOMÁTICO.

7.1. ¿Qué es el aprendizaje automático (ML)?

7.2. ML efectivo: conceptos y tipos.

7.3. Conjuntos de datos ML: generalización.

7.4. Laboratorio: explore y cree conjuntos de datos ML.

8. CREACIÓN DE MODELOS ML CON TENSORFLOW.

8.12. Sesión 10: Ingeniería de funciones.

8.13. Creando buenas características.

8.14. Transformando las entradas.

8.15. Características sintéticas.

8.16. Pre-procesamiento con Cloud ML.

8.17. Laboratorio: ingeniería de funciones.

8.18. Creación de sistemas de transmisión flexibles en Google Cloud Platform.

9. ARQUITECTURA DE LAS TUBERÍAS DE ANÁLISIS DE TRANSMISIÓN.

9.1. Procesamiento de datos de flujo: desafíos.

9.2. Manejo de volúmenes de datos variables.

9.3. Tratar con datos desordenados / tardíos.

9.4. Laboratorio: Diseño de canalización de transmisión.

10. INGESTIÓN DE VOLÚMENES VARIABLES.

10.1. ¿Qué es Cloud Pub / Sub?

10.2. Cómo funciona: temas y suscripciones.

10.3. Laboratorio: Simulador.

11. IMPLEMENTACIÓN DE TUBERÍAS DE TRANSMISIÓN.

11.1. Desafíos en el procesamiento de flujo.

11.2. Manejar datos tardíos: marcas de agua, desencadenantes y acumulación.

4.4. Laboratorio: Agregar capacidades de aprendizaje automático al análisis de Big Data.	8.1. Comenzando con TensorFlow.	11.3. Laboratorio: secuencia de procesamiento de datos de flujo para datos de tráfico en vivo.
4.5. Análisis de datos sin servidor con Google BigQuery y Cloud Dataflo.	8.2. Laboratorio: utilizando Tensor Flow Learn.	12. ANÁLISIS Y CUADROS DE MANDO DE STREAMING.
5. ANÁLISIS DE DATOS SIN SERVIDOR CON BIGQUERY.	8.3. TensorFlow gráficos y bucles + laboratorio.	12.1. Streaming analytics: de datos a decisiones.
5.1. Qué es BigQuery?	8.4. Laboratorio: uso de TensorFlow de bajo nivel + detención temprana.	12.2. Consulta de datos de transmisión con BigQuery.
5.2. Consultas y funciones.	8.5. Monitorización de entrenamiento ML.	12.3. ¿Qué es Google Data Studio?
5.3. Laboratorio: escribir consultas en BigQuery.	8.6. Laboratorio: Gráficos y gráficos del entrenamiento TensorFlow.	12.4. Laboratorio: construya un tablero en tiempo real para visualizar los datos procesados.
5.4. Cargando datos en BigQuery.	8.7. Sesión 9: Modelos ML de escalabilidad con CloudML.	13. ALTO RENDIMIENTO Y BAJA LATENCIA CON BIGTABLE.
5.5. Exportar datos de BigQuery.	8.8. ¿Por qué Cloud ML?	13.1. ¿Qué es Cloud Spanner?
5.6. Laboratorio: carga y exportación de datos.	8.9. Empaquetado de un modelo TensorFlow.	13.2. Diseñando el esquema de Bigtable.
5.7. Campos anidados y repetidos.	8.10. Entrenamiento de extremo a extremo.	13.3. Ingerir en Bigtable.
5.8. Consultando múltiples tablas.	8.11. Laboratorio: ejecute un modelo ML localmente y en la nube.	13.4. Laboratorio: transmisión a Bigtable
5.9. Laboratorio: consultas complejas.		

Información Adicional:

+ Información e inscripciones: es.comercial@skillssoft.com

Más información:

Para más información o para reservar tu plaza llámanos al (34) 91 425 06 60

info.cursos@globalknowledge.es

www.globalknowledge.com/es-es/

Global Knowledge Network Spain, C/ Retama 7, 6ª planta, 28045 Madrid