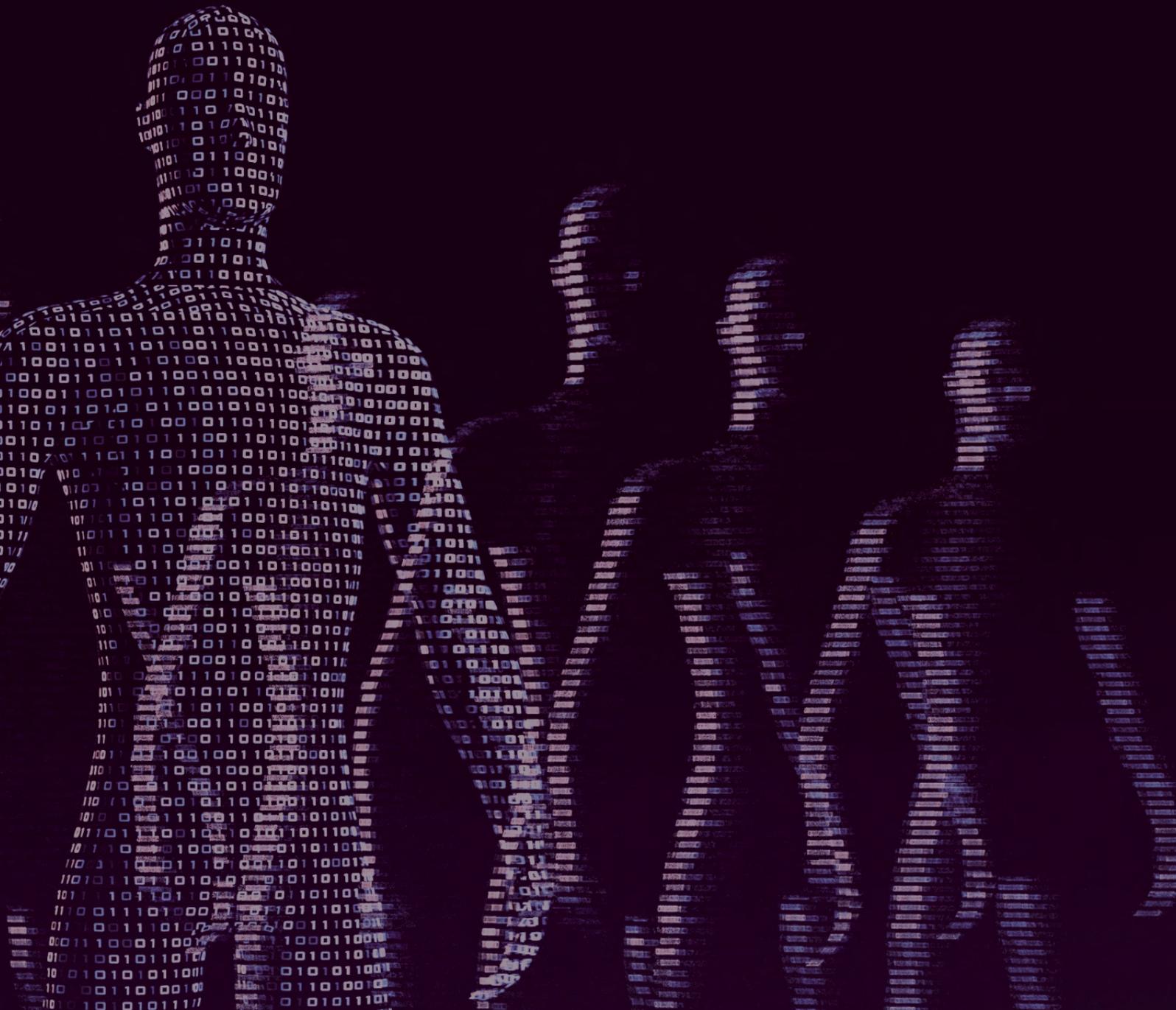


1ª Edición Becas PUE Talent

Curso oficial Administrador Training: CDP Private Cloud Base



PUE Talent es una iniciativa diseñada para el reclutamiento de talento que quiera incorporarse a la empresa de referencia internacional en Big Data, reconocida recientemente como Cloudera Platinum Partner. Enmarcada en PUE Talent, te ofrecemos esta beca de formación y certificación oficial valorada en 2.000€ con la que podrás realizar de forma gratuita el curso y certificación Administrator Training: CDP Private Cloud Base.

Las plazas son limitadas y se asignarán tras un proceso de selección en donde se tendrán en cuenta los requisitos de acceso que encontrarás en la descripción de las becas.

REQUISITOS

1. Experiencia en Administración de Sistemas UNIX.
2. Conocimiento de lenguajes de scripting para automatización de tareas: Bash imprescindible.
3. Nivel de inglés (mínimo B2).

ACERCA DE ESTE CURSO

Con el curso de formación para administradores de Cloudera CDP Private Cloud Base obtendrás una comprensión integral de todos los pasos necesarios para operar y mantener clústeres locales utilizando Cloudera Manager, desde la instalación y configuración hasta el equilibrio de carga y el ajuste. Te prepararás para los desafíos a los que se enfrentan habitualmente los administradores que ejecutan CDP Private Cloud Base.

OBJETIVOS DEL CURSO

1. Conocer en profundidad la topología de un clúster de Cloudera típico y el papel que juegan los componentes principales en el clúster
2. Instalar Cloudera Manager y CDP
3. Utilizar Cloudera Manager para crear, configurar, implementar y monitorear un clúster
4. Descubrir qué herramientas proporciona Cloudera para ingerir datos de fuentes externas en un clúster
5. Configurar los componentes del clúster para un rendimiento óptimo
6. Averiguar qué tareas rutinarias son necesarias para mantener un clúster, incluida la actualización a una nueva versión de CDP
7. Investigar acerca de la detección, resolución de problemas y reparación de problemas
8. Comprender las funciones de seguridad clave de Cloudera

CONTENIDOS

Módulo 1: Plataforma de datos de Cloudera

- Tendencias de la industria para Big Data
- El desafío de estar basado en datos
- La nube de datos empresariales
- Descripción general de CDP
- Factores de forma de CDP
- Ejercicio práctico: configurar la red de ejercicios

Módulo 2: Instalación de la base de nube privada de CDP

- Descripción general de la instalación
- Instalación de Cloudera Manager
- Ejercicio práctico: instalación de Cloudera Manager Server
- Descripción general del tiempo de ejecución de CDP
- Introducción a Cloudera Manager
- Demostración dirigida por un instructor: Cloudera Manager
- Ejercicio práctico: instalación en clúster

Módulo 3: Configuración de clúster

- Visión general
- Ajustes de configuración
- Modificación de configuraciones de servicio
- Archivos de configuración
- Administrar instancias de roles
- Agregar nuevos servicios
- Agregar y quitar hosts
- Ejercicio práctico: configuración de un clúster de Hadoop

Módulo 4: Almacenamiento de datos

- Visión general
- Topología y roles de HDFS
- Rendimiento de HDFS y tolerancia a fallos
- Descripción general de seguridad de HDFS y Hadoop
- Trabajar con la interfaz de usuario de Namenode
- Demostración dirigida por un instructor: interfaz de usuario de Namenode
- Trabajando con HDFS
- Ejercicio práctico: trabajar con HDFS
- Descripción general de HBase
- Descripción general de Kudu
- Descripción general de almacenamiento en la nube
- Ejercicio práctico: almacenamiento de datos en Amazon S3

Módulo 5: Ingesta de datos

- Descripción general de la ingesta de datos
- Formatos de archivo
- Ingesta de datos mediante transferencia de archivos o interfaces REST
- Importación de datos de bases de datos relacionales con Apache Sqoop
- Ejercicio práctico: importación de datos con Sqoop
- Ingesta de datos con NiFi
- Demostración dirigida por un instructor: interfaz de usuario de NiFi
- Prácticas recomendadas para importar datos
- Ejercicio práctico: verificación de NiFi

Módulo 6: Flujo de datos

- Descripción general de Cloudera Flow Management y NiFi
- Arquitectura NiFi
- Gestión de flujo de Cloudera Edge y MiNiFi
- Demostración dirigida por un instructor: uso de NiFi
- Descripción general de Apache Kafka
- Arquitectura de clúster de Apache Kafka
- Herramientas de línea de comandos de Apache Kafka
- Ejercicio práctico: trabajar con Kafka

Módulo 7: Acceso y descubrimiento de datos

- Apache Hive
- Apache Impala
- Ajuste de Apache Impala
- Ejercicio práctico: instalar Impala y Hue
- Resumen de búsqueda
- Descripción general de Hue
- Administrar y configurar Hue
- Autenticación y autorización de Hue
- Descripción general de CDSW
- Ejercicio práctico: uso de Hue, Hive e Impala

Módulo 8: Computación de datos

- Descripción general de YARN
- Ejecución de aplicaciones en YARN
- Visualización de aplicaciones de YARN
- Registros de aplicaciones de YARN
- Aplicaciones MapReduce
- Configuración de CPU y memoria YARN
- Ejercicio práctico: ejecución de aplicaciones YARN
- Descripción general de Tez
- ACID para Hive
- Descripción general de Spark
- Cómo se ejecutan las aplicaciones Spark en YARN
- Supervisión de aplicaciones Spark
- Ejercicio práctico: ejecución de aplicaciones Spark

Módulo 9: Gestión de recursos

- Descripción general de la gestión de recursos
- Etiquetas de nodo
- Configuración de cgroups
- El programador de capacidad
- Administrar colas
- Programación de consultas de Impala
- Ejercicio práctico: uso del programador de capacidad

Módulo 10: Planificación de su clúster

- Consideraciones generales de planificación
- Elegir el hardware adecuado
- Consideraciones de red
- Consideraciones sobre la nube privada de CDP
- Configurar nodos

Módulo 11: Configuración avanzada de clústeres

- Configuración de puertos de servicio
- Ajuste de HDFS y MapReduce
- Gestionar el crecimiento del clúster
- Codificación de borrado
- Habilitación de alta disponibilidad para HDFS y YARN
- Ejercicio práctico: configuración de HDFS para alta disponibilidad

Módulo 12: Mantenimiento de clústeres

- Comprobación del estado de HDFS
- Copiar datos entre clústeres
- Reequilibrio de datos en HDFS
- Instantáneas del directorio HDFS
- Ejercicio práctico: creación y uso de una instantánea
- Mantenimiento de host
- Actualizar un clúster
- Ejercicio práctico: actualizar el clúster

Módulo 13: Monitoreo de clústeres

- Funciones de supervisión de Cloudera Manager
- Pruebas de salud
- Ejercicio práctico: Romper el grupo
- Eventos y alertas
- Gráficos e informes
- Recomendaciones de seguimiento
- Ejercicio práctico: confirmación de la reparación del clúster y configuración de alertas por correo electrónico

Módulo 14: Solución de problemas del clúster

- Visión general
- Herramientas de resolución de problemas
- Ejemplos de configuración incorrecta
- Ejercicio práctico: resolución de problemas de un clúster

Módulo 15: Seguridad

- Gobierno de datos con SDX
- Conceptos de seguridad de Hadoop
- Autenticación de Hadoop mediante Kerberos
- Autorización de Hadoop
- Cifrado Hadoop
- Asegurar un clúster de Hadoop
- Apache Ranger
- Atlas de Apache
- Copia de seguridad y recuperación

Módulo 16: Nube privada / Nube pública

- Descripción general de CDP
- Capacidades de nube privada
- Capacidades de nube pública
- ¿Qué es Kubernetes?
- Descripción general de Workload XM
- Escalado automático

Módulo 17: Conclusión

Módulo 18: Apéndice: API de Cloudera Manager

- API de Cloudera Manager
- Instalación y configuración
- Ejemplos de código

Módulo 19: Apéndice: Descripción general de Ozone

- Descripción general de Ozone
- Trabajar con Ozone

Plazas limitadas. Si te interesa, **inscríbete ya** para que podamos evaluar tu candidatura.

¿TIENES DUDAS?

Para cualquier consulta no dudes en contactarnos a través del siguiente correo electrónico: puetalent@pue.es

 pue | talent

