



## Curso de Big Data en AWS

## ¡No te rindas!

Necesitas una **calificación mínima de 9.0** para aprobar.

Vuelve a intentarlo en 05 horas, 30 minutos, 45 segundos

# 8.67

Calificación

# 26 / 30

Aciertos

1. ¿Cómo se llama la arquitectura que no tiene capa Batch?

Kappa



2. ¿Cuál servicio en AWS podemos utilizar para IDE?

Cloud9



3. ¿Cómo se llama la librería de Python para interactuar con los servicios de AWS ?

Boto3



4. ¿Cuál servicio de AWS puede actuar como una "Puerta Frontal" para recibir cientos de miles de llamadas ?

API Gateway



5. ¿En cuál de los siguientes casos usaría storage gateway?

Cuando quiera llevar información de mi app onpremise a la nube.



6. ¿Qué componen una agrupación de data records?

Un Shard



7. Por defecto al configurar kinesis data stream, ¿Cuál es el periodo de retención?

24 Horas



8. ¿Cuál es el principal beneficio de realizar despliegues de infraestructura utilizando CloudFormation?

Alta Disponibilidad.



9. ¿Qué servicio deberías utilizar para alimentar un servicio como ElasticSearch?

Kinesis Firehose



10. ¿Cuál servicio utilizarías en un flujo de BigData si cuentas con Kinesis Data Stream y quisieras alimentar Splunk?

Kinesis Firehose



11. Tú como arquitecto tienes un clúster de Apache Kafka instalado en tu datacenter, te han pedido llevarlos a AWS ¿Cuál servicio seleccionarías para esta migración?

AWS MSK



12. Al desplegar un clúster de MSK, por defecto crea un nodo de: \_\_\_\_\_

Zookeeper



13. ¿Cuál es la unidad de procesamiento usada por GLUE?

DPU



14. Te han pedido que implementes una solución para probar los ETL de forma gráfica con seguridad e integración a directorio activo. Para cumplir con este requerimiento ¿Qué herramienta utilizarías?

Dev Endpoint de GLUE.



15. Dentro del servicio de Developer Endpoint de Glue , ¿Qué tipos de notebooks puedes crear?

Zeppelin y Sage Maker.



---

16. Tu como arquitecto de soluciones en Big Data, te han pedido que ejecute un ETL para convertir información que viene de la app móvil de la empresa y posteriormente puedan analizarla. ¿Qué pasos deberías seguir dentro del servicio Glue para lograr este objetivo?

Crawler - ETL - Crawler.



---

17. Estas desplegando un Clúster de EMR y necesitas ejecutar un script de inicio con unas configuraciones específicas para el clúster. ¿Qué deberías utilizar para lograr este objetivo?

Bootstrap.



---

18. ¿Qué tipos de nodos conforman un Clúster de EMR ?

Master Nodes, Core Nodes y Task Nodes.



---

19. Al intentar conectarte a Zeppelin en EMR ves que no aparece habilitada la interfaz web ¿Cuál podría ser la causa del error?

Habilitar el puerto 8890 en el grupo de seguridad para el Master Node.



---

20. ¿En qué servicio de AWS puedo utilizar notebooks de Apache Zeppelin ?

EMR y Glue.



---

21. ¿Cuáles aspectos debes tener en cuenta usando Lambda en proyectos de Big Data?

Concurrencia y Errores.



22. ¿Cuál arquitectura es recomendada para proyectos de Big Data con despliegue de servicios en AWS ?

Multi-Account.



23. ¿Qué granularidad agrega Lake formation en los permisos sobre tablas?

Permisos a nivel de columnas.



24. ¿De qué tamaño es recomendado un shard en ElasticSearch?

50GB - 150GB.



25. ¿Qué debes tener en cuenta en el momento de dimensionar un clúster de ElasticSearch?

Cantidad de Shards distribuidos en los nodos.



26. ¿Cuál alerta utilizarías para identificar que un usuario ha subido información confidencial de tarjetas de crédito a un bucket de S3?

Credential.



27. ¿Cómo se llama la estructura básica dentro de Apache Airflow?

Operator.



28. ¿Cuál es la forma de establecer las dependencias dentro de un DAG de Cloud Composer?

Utilizando "">>""



29. Se te ha pedido un reporte sobre diferente información que se tiene en formato CSV en S3, debes realizar esta consulta sobre varios archivos y buckets, encuentras que ya el Glue Catalog esta completamente creado ¿Cuál servicio utilizarías de la manera más eficiente para realizar esta consulta en relación al costo?

Athena.



30. ¿Cuál servicio en AWS utilizarías para hacer consultas SQL complejas sobre grandes cantidades de datos en S3?

RedShift.



**REGRESAR**