

Consideraciones generales para presentar el examen práctico de la certificación de ingesta.

El sustentante requiere cumplir con las siguientes consideraciones para presentar adecuadamente el examen práctico del módulo 2.

Antes del examen:

1. Contar con un correo y usuario BBVA.
2. La computadora deberá estar debidamente ambientada. En caso contrario, podrá realizar la ambientación en silencio y sin interrumpir a los formadores y/o asistentes, de no ser así se te pedirá que te retires del examen.
3. Recuerda trabajar los cambios en una rama con el siguiente formato **[Iniciales del nombre completo]-examen**. Por ejemplo, el participante Luis Ernesto Romero Robles deberá tener la rama **lerr-examen** y trabajará su solución en esta rama

Durante el examen:

4. Sólo se podrá hacer uso de los servicios de mensajería para la comunicación entre aplicadores y participantes. En caso de hacer uso de estos servicios para cualquier otro intercambio de información entre los asistentes, se cancelará el examen a los involucrados.
5. Contará con acceso a internet y a sus notas durante el examen.
6. Se creará un proyecto para la evaluación, del cual deberá realizar un fork y clone.
7. No se permitirán dispositivos de almacenamiento conectados en las computadoras.
8. No se permitirá el uso de dispositivos móviles; **hacer caso omiso de esto será motivo de cancelación del examen.**
9. No se responderán dudas que involucren respuesta a la solución o flujo de trabajo a evaluar.
10. En caso de suplantar la identidad de un asistente se cancelará el examen y se reportará el incidente.

Durante la calificación del examen:

11. No se podrán enviar soluciones después de la hora de cierre del examen. **Hacer caso omiso de esto será motivo de cancelación.**

12. Si no concluye la implementación completa de las reglas de negocio, se podrá hacer la escritura de la tabla hasta la etapa que se haya implementado y se te tomará en cuenta para la calificación.
13. Si durante la revisión del examen se detecta una copia, **No se tomará en cuenta la PR con el commit más reciente y se reportará el incidente.**
14. Proyectos que no compilen (**mvn clean install**) no serán tomados en cuenta.
15. Sólo se evaluarán salidas generadas en tiempo de ejecución.

Examen práctico

Módulo 2: Desarrollo de aplicaciones Big Data en Datio

Instrucciones:

A continuación realizarás el ejercicio en los que deberás aplicar los conocimientos adquiridos en el módulo 2 de procesamiento en Scala y Spark dentro de Datio.

Caso de uso

En esta época del año Amazon usualmente libera su almacén de celulares. John Wick, jefe de ventas y marketing del área de smartphones ha solicitado un reporte a sus big data developers, el reporte servirá para la campaña de ventas del siguiente Black Friday.

El área de sistemas cuenta con las tablas **t_fdev_customers** y **t_fdev_phones** para realizar el reporte solicitado por el Sr. Wick.

Lo que se requiere:

Justo antes de lanzar la campaña, el Sr. Wick recibió una terrible noticia. Los proveedores de los teléfonos de marca Dell, Coolpad, Chea, BQ y BLU dejarán de ser proveedores de Amazon, por lo que el Sr. Wick decidió no incluirlas en la siguiente campaña de ventas.

Además le han notificado que los países Chile (CH), Italia (IT), República Checa (CZ) y Dinamarca (DK) están dentro de una crisis económica, por tal motivo la campaña publicitaria no aplicará en dichos países. Otro punto importante a destacar, es que el Sr. Wick quisiera incluir sólo a los clientes cuya tarjeta de crédito sea menor a 17 dígitos.

1. Filtrar la tabla **t_fdev_phones**:

Por fecha diaria, tomando únicamente las fechas que comprenden del 2020-03-01 hasta el 2020-03-04 en el campo **cutoff_date**, recuperando

únicamente los celulares que **no** sean de las marcas “Dell”, “Coolpad”, “Chea”, “BQ” y “BLU”, también deberán omitirse los países señalados anteriormente que tienen como **country_code** “CH”, “IT”, “CZ” y “DK”.

2. Filtrar la tabla **t_fdev_customers**:

Por fecha diaria, tomando únicamente las fechas que comprenden del 2020-03-01 hasta el 2020-03-04 en el campo **gl_date**, recuperando únicamente a los clientes cuya tarjeta de crédito (**credit_card_number**) sea menor a 17 dígitos.

3. Tendrán que encontrar la relación entre los clientes y los teléfonos que adquirieron, descartando a los clientes que no compraron ningún teléfono y a los teléfonos que no fueron adquiridos. Las llaves a usar son **customer_id** y **delivery_id** en ambas tablas (**inner_join**).
4. John Wick quiere identificar los clientes vip mediante un nuevo campo (**customer_vip**), para que un cliente sea identificado como **customer_vip** deberá ser miembro prime (**prime**) y el costo (**price_product**) del teléfono adquirido deberá ser de 7500.00 en adelante, los clientes vip deben de clasificarse de la siguiente manera:
 - “Yes” -> En caso de ser cliente vip.
 - “No” -> En caso de no ser cliente vip.
5. Los teléfonos tendrán descuentos extra (**discount_extra**) si son miembros prime, el stock (**stock_number**) es menor a 35 y sus celulares **no** son de las siguientes marcas (**brand**): “XOLO”, “Siemens”, “Panasonic” y “BlackBerry”.
El descuento extra es el 10% del valor original (**price_product**) del teléfono, en caso de no tener descuento deberá colocarse un 0.00 en dicho campo.
6. También se desea saber el precio final (**final_price**) del teléfono celular, para ello deberás considerar el precio original (**price_product**), el descuento (**discount_amount**), el descuento extra (**discount_extra**) y los impuestos (**taxes**).
El Sr. Wick sabe que a algunos de sus empleados no se le dan bien las matemáticas, así que ha colocado la fórmula del precio final de la siguiente manera:

$$final\ price = price\ product + taxes - discount\ amount - discount\ extra$$

7. Actualmente no se cuenta con la edad de los clientes, por lo que el Sr. John Wick desea saberlo a partir de su fecha de nacimiento (**birth_date**), dato que se encuentra en la tabla **t_fdev_customers**. El Sr. John Wick ha sido muy específico, ha pedido que los años sean calculados a partir del día de hoy, es un jefe muy exigente y necesita conocer las

edades cumplidas, no las edades por cumplir.

8. Además el Sr. John Wick ha dado la orden directa de que la promoción sólo aplicará para los teléfonos que estén dentro del top 50 de precios (**los más caros**) finales (**final_price**) de cada marca (**brand**).
9. El Sr. Wick es muy vanidoso y egocéntrico, por lo que desea que la tabla sea particionada por el campo John Wick date (**jwk_date**). Dicho campo tendrá que ser creado e ir parametrizado dentro del archivo de configuración, ya que el Sr. Wick ama cambiar la fecha de los reportes por pura maldad.

BONUS: No quieres hacer enojar al señor Wick

10. Al Sr. Wick los valores **null** le recuerdan a su perrito, por lo que ha pedido a sus desarrolladores encargados del reporte, que los registros que sean nulos en el campo nfc (**nfc**), sean sustituidos por un **"No"**.

Llegado a este punto la tabla deberá ser escrita cumpliendo con el esquema "t_fdev_customersphones.output.schema"

Realiza la implementación de los puntos anteriores, GOOD LUCK!!



Consideraciones para el diseño de la solución

1. La forma de lectura y escritura deberá de hacerse con la librería AmLib.
2. Características de entrada del job:
 - El parámetro **jwk_date** deberán de estar parametrizados.
 - La configuración únicamente deberá de contemplar lo que la ejecución requiere, en caso de que se agregue configuración extra innecesaria bajará la puntuación del examen.
 - Al enviar la P.R. es necesario colocar el nombre de los parámetros y sus valores con los que fueron ejecutados.
3. Características de la salida:
 - Cumplir con la estructura del esquema **t_fdev_customersphones.output.schema** (se especifica en los requerimientos) al validar con AmLib.
 - El campo de particionamiento deberá ser **jwk_date**. Además, la tabla es de naturaleza incremental, por lo que deberá de ser posible el reproceso de los datos en caso de que sean inconsistentes desde el origen.
 - Escritura en modo coalesce, en un solo archivo de tipo parquet.

Ejemplo de la escritura:

Muestra de la tabla **INPUT** t_fdev_phones:

Link: [Tabla t_fdev_phones](#)

Muestra de la tabla **INPUT** t_fdev_customers:

Link: [Tabla t_fdev_customers](#)

Muestra de la tabla **OUTPUT** t_fdev_customersphones:

Link: [Tabla t_fdev_customersphones](#)

4. Características de las pruebas unitarias

Se deberá de cumplir al menos el 80% de cobertura en los test unitarios y aplicar pruebas sobre los siguientes escenarios:

- Validación total de los registros de salida
- Validación de la regla de negocio aplicada

- Manejo de excepción
5. Características de las pruebas de aceptación.

En los test de aceptación deberá de validarse lo siguiente en la tabla **OUTPUT t_fdev_customersphones**:

- Que no existan valores “CH”, “IT”, “CZ” y “DK”, dentro del campo(**country_code**).
- El conteo del total de registros con fecha (**jwk_date**) **2020-03-04**.
- Que existan valores “MX”, “CA”, “BR”, “US”, “PE”, “JP”, “NZ”, “CO” y “AR”, dentro del campo (**country_code**).
- El conteo del total de registros filtrados con la fecha (**jwk_date**) **2020-03-04** y con marca (**brand**) “Gigabyte”.
- El conteo del total de registros filtrados con la fecha (**jwk_date**) **2020-03-04** y con marca (**brand**) “Palm”.

Ejemplo de test de aceptación:

Validando el número de registros de salida filtrados con la fecha (**jwk_date**) **2020-03-04** y con nfc (**nfc**) con el valor de “Carrier dependent”.

```
When I filter t_fdev_customersphones dataframe with filter
jwk_date='2020-03-04' AND nfc='Carrier dependent' and save it as
t_fdev_customersphones_filtered
Then t_fdev_customersphones_filtered dataframe should has 1 records
```

Rúbrica de calificación

Valor del examen para la certificación: 40 %

Criterios de evaluación

Criterio 1: Resolución de los requerimientos

Valor: 70 %

Regla de negocio aplicada	Validación de esquema	Uso de Kirby para la lectura	Parametrización de archivos de configuración	Uso de kirby en la escritura	Escritura con coalesce	Output con un archivo	Escritura en parquet
45 %	3 %	5 %	3 %	5 %	3 %	3 %	3 %

Criterio 2: Pruebas unitarias con cobertura del 80 %

Valor: 15 %

Redacción de las pruebas unitarias	Total de registros	Validación de la regla de negocio aplicada	Manejo de excepción
4 %	3 %	4 %	4 %

Criterio 3: Pruebas de aceptación

Valor 15 %

Validación previa la ejecución	Validación posterior a la ejecución
5 %	10 %