



## Posición Big Data Engineer

### Prueba técnica

---

Documento guión para poder desarrollar la prueba técnica cuyo objetivo es poder analizar la pericia técnica del candidato desde diferentes perspectivas y así poder optar por la posición de Big Data Engineer en la práctica de Data Technologies de SDG Group.



## Prueba técnica para la posición de Big Data Engineer en la práctica de Data Technologies de SDG Group.

El objetivo de la prueba técnica es que puedas demostrar tus capacidades de desarrollo con Spark Scala. En el cual tendras que diseñar e implementar un framework de ingesta, validación y transformación de datos dirigida por metadatos/configuración.

Las especificaciones son las siguientes:

1. **Descripcion:** El programa realizado en spark / scala debería, dirigido por los metadatos que se comentarán a continuación, ejecutar un proceso que cargará los orígenes descritos, ejecutar una serie de transformaciones (en este caso lo definido como dataflow: dos validaciones funcionales y técnicas de dos campos y añadir una columna) y escribir los datos resultado , tanto los que han pasado la validación, como por otro lado los que no las han pasado y el código de error que autoexplique porque no lo ha pasado.

Es muy importante que sea un programa que, o genera código leyendo **metadatos** que posteriormente se ejecuta, o que cambie su comportamiento dinámicamente **via los metadatos** propuestos **y que no sea un desarrollo adhoc.**

### 2. Datos de Entrada (fichero json ejemplarizante)

```
{"name":"Xabier","age":39,"office":""}  
{"name":"Miguel","office":"RIO"}  
{"name":"Fran","age":31,"office":"RIO"}
```

### 3. Datos de Salida STANDARD\_OK (fichero json ejemplarizante)

```
{"name":"Fran","age":31,"office":"RIO","dt":"2020-12-29 09:00:00" }
```

### 4. Datos de Salida STANDARD\_KO (fichero json)

```
{"name":"Xabier","age":39,"office":"","dt":"2020-12-29 09:00:00","arraycoderrorbyfield":{"...}}  
{"name":"Miguel","age":,"office":"RIO","dt":"2020-12-29 09:00:00","arraycoderrorbyfield":{"...}}
```

### 5. Metadatos que debe usar el programa

```
{  
  "dataflows": [  
    {  
      "name": "prueba-acceso",  
      "sources": [  
        {  
          "name": "person_inputs",
```

```
"path": "/data/input/events/person/*",
"format": "JSON"
},
],
"transformations": [
{
  "name": "validation",
  "type": "validate_fields",
  "params": {
    "input": "person_inputs",
    "validations": [
      {
        "field": "office",
        "validations": ["notEmpty"]
      },
      {
        "field": "age",
        "validations": ["notNull"]
      }
    ]
  }
}
],
{
  "name": "ok_with_date",
  "type": "add_fields",
  "params": {
    "input": "validation_ok",
    "addFields": [
      {
        "name": "dt",
        "function": "current_timestamp"
      }
    ]
  }
}
],
"sinks": [
{
  "input": "ok_with_date",
  "name": "raw-ok",
  "paths": [
    "/data/output/events/person"
  ],
  "format": "JSON",
  "saveMode": "OVERWRITE"
},
{
  "input": "validation_ko",
  "name": "raw-ko",
```

## Propuesta de prueba técnica

```
{
  "paths": [
    "/data/output/discards/person"
  ],
  "format": "JSON",
  "saveMode": "OVERWRITE"
}
]
```

La estructura de los metadatos/configuración mostrados anteriormente es una template/guía de lo que podría ser una versión de la misma, damos total libertad en el diseño de los metadatos y el tipo de archivo en el que se guarde estos metadatos mientras realice lo especificado anteriormente y genere el fichero STANDARD\_OK y STANDARD\_KO correctamente.

Si en algún momento tienes algo que comentar con nosotros, o necesitas que movamos la fecha o la hora de presentación por algún tema personal, no dudes en contactarnos a mí ([miguel.romero@sdggroup.com](mailto:miguel.romero@sdggroup.com) – 696371810) o a Xabier Gil de la Pisa ([xabier.gil@sdggroup.com](mailto:xabier.gil@sdggroup.com) – 680720805)

Sin más, muchas gracias por la entrevista y esperamos vernos contigo de nuevo en breve.

Por otro lado, por favor, confírmanos que has recibido la prueba técnica y si la aceptas.