

Spring Batch

Rodrigo Martínez Zambrano

Spring Batch es un framework de código abierto, parte del ecosistema de Spring, diseñado para crear aplicaciones de procesamiento por lotes robustas y de alto rendimiento.

Algunas de sus características clave incluyen:

1. Modelado de Trabajo: Permite definir trabajos, pasos y flujos de manera declarativa.
2. Gestión de Transacciones: Soporta transacciones, lo que garantiza la consistencia de los datos.
3. Reintentos y Salto de Errores: Ofrece mecanismos para reintentar pasos fallidos y saltar errores controlados.
4. Manejo de Recursos: Proporciona herramientas para leer y escribir datos desde y hacia múltiples fuentes como bases de datos, archivos, etc.
5. Escalabilidad y Paralelismo: Soporta la ejecución paralela de tareas, lo que permite procesar grandes volúmenes de datos de manera eficiente.

Spring Batch es especialmente útil en contextos donde se necesita procesar grandes volúmenes de datos de manera periódica, como en la generación de informes, migración de datos, procesos de integración, entre otros.

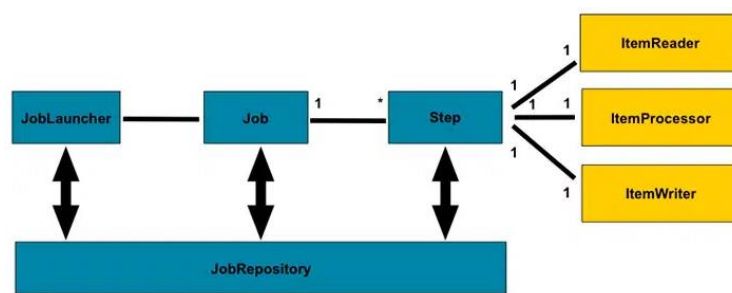
¿Qué es un proceso batch?

Un proceso batch o proceso por lote es un proceso pensado para trabajar con volúmenes muy grandes de datos y generalmente de una forma programada. Es decir, sin intervención humana.

Por ejemplo, la carga de un fichero enorme con millones de registros; o bien un proceso nocturno que, a partir de una serie de consultas, envía una gran cantidad de e-mails, sms, etc.

Componentes de Spring Batch

Este framework está compuesto por los siguientes componentes



Componentes de Spring Batch.

JobRepository

Spring Batch está pensando para que la información de los procesamientos quede almacenada en un repositorio persistente o bien en memoria.

Es responsabilidad de este repositorio almacenar información sobre cada job, step que se produzca, los parámetros del Job, los errores que tengan lugar, etc.

JobLauncher

Es el componente encargado de lanzar la ejecución de los Job.

Job y Step

Un *job* es un bloque de trabajo y está compuesto por uno o varios pasos o *steps*. Una vez se han llevado a cabo todos estos pasos, se considera el job como completado.

Cada uno de estos *steps* suele constar de tres partes:

1. **ItemReader**: se encarga de la lectura del procesamiento por lotes. Esta lectura puede ser, por ejemplo, de una base de datos; o también podría ser de un broker de mensajes o bien un fichero csv, xml, json, etc.
2. **ItemProcessor**: se encarga de transformar items previamente leídos. Esta transformación además de incluir cambios en el formato puede incluir filtrado de datos o lógica de negocio.
3. **ItemWriter**: este elemento es lo opuesto al itemReader. Se encarga de la escritura de los ítems. Esta puede ser inserciones en una base de datos, en un fichero csv, en un broker de mensajes, etc.

Spring Batch es una herramienta poderosa y flexible para el desarrollo de aplicaciones de procesamiento por lotes, ofreciendo una infraestructura sólida y escalable que facilita la gestión y ejecución de tareas repetitivas y de gran volumen de datos.

Su arquitectura modular permite a los desarrolladores definir trabajos complejos de manera declarativa y reutilizable, garantizando la consistencia y eficiencia a través de características como la gestión de transacciones, reintentos automáticos y manejo de errores.