

Spring Batch

¿Qué es el batch processing?

El batch processing o procesamiento por lotes es un método para ejecutar "tareas" repetitivas de alto volumen de datos con poca o ninguna interacción del usuario, por ejemplo, una tarea de envíos de emails masivos.

¿Cuándo necesitamos el batch processing?

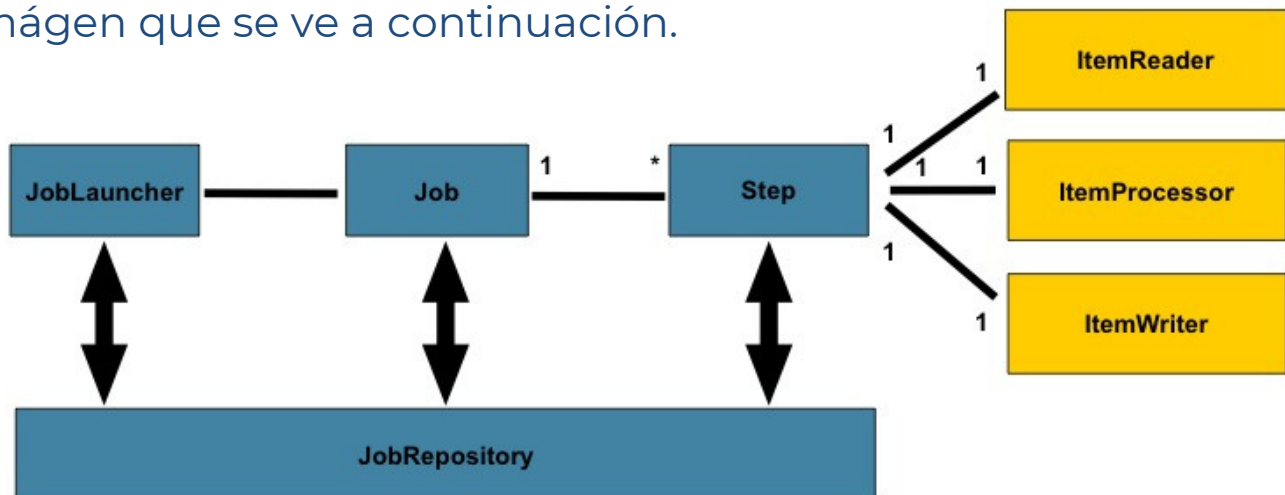
- Cuando se tiene la necesidad de tomar mucha información para procesar día con día.
- Cuando se implementa un batch processing se deben cubrir las siguientes preguntas
 - ¿Quién está enviando la tarea a ejecutar?
 - ¿Qué se va a ejecutar?
 - ¿Cuáles son las entradas y salidas de información?
 - ¿Cuándo se debe ejecutar?

¿Qué es spring batch?

Es un framework ligero de spring que está enfocado específicamente a la creación de procesos batch.



Este framework esta compuesto por los componentes de la imagen que se ve a continuación.



Job Launcher: Es el componente encargado de lanzar los procesos, subministrando los parámetros de entrada requeridos.

Job Repository: Es un repositorio que se utiliza para almacenar toda la información de los procesos que se ejecutan. Se puede utilizar para almacenar los detalles de los jobs fallidos, incluso puede ser consultado para comprobaciones de ficheros.

Job y Steps: Un job es un bloque de trabajo que se va a ejecutar, está compuesto por pasos o steps, tantos pasos se necesiten para completar el job.

Cada uno de los steps consta de tres partes:

itemReader: Se encarga de la lectura del procesamiento por lotes. Puede leer de una base de datos por ejemplo o de un fichero, etc.

ItemProccessor: Se encarga de transformar los items previamente leídos, en este punto se puede incluir lógica de negocio.

ItemWriter: Se encargara de la escritura de items. Puede ser inserciones a base de datos o a un fichero un envío de correos, etc.

En el batch processing, es posible definir el número de ítems en el que se organizará el procesamiento, es decir, si configuramos un tamaño de 10, se leerán, procesarán y escribirán de 10 en 10. A este bloque de ítems se le denomina **chunk**.

Al implementar estos procesos se recomienda:

- Simplificar la lógica, de forma que los procesos queden fragmentados, sean pequeños para facilitar las lecturas y escrituras.
- Optimizar las sentencias sql para mejorar el rendimiento de la base de datos y el tiempo de ejecución de los jobs.

