springone 2GX

DALLAS

2014

Spring Batch Performance Tuning

By Chris Schaefer & Gunnar Hillert

springone 2GX

Agenda

- Spring Batch
- Spring Integration
- Spring Batch Integration
- Scaling Spring Batch
- Spring XD

Spring Batch

http://projects.spring.io/spring-batch/

> Batch processing ... is defined as the processing of data without interaction or interruption.

Michael T. Minella, Pro Spring Batch

Batch Jobs

- Generally long-running
- Non-interactive
  - Often include logic for handling errors and restartability options
- Process large volumes of data
  - More than what may fit in memory or a single transaction

Batch and offline processing

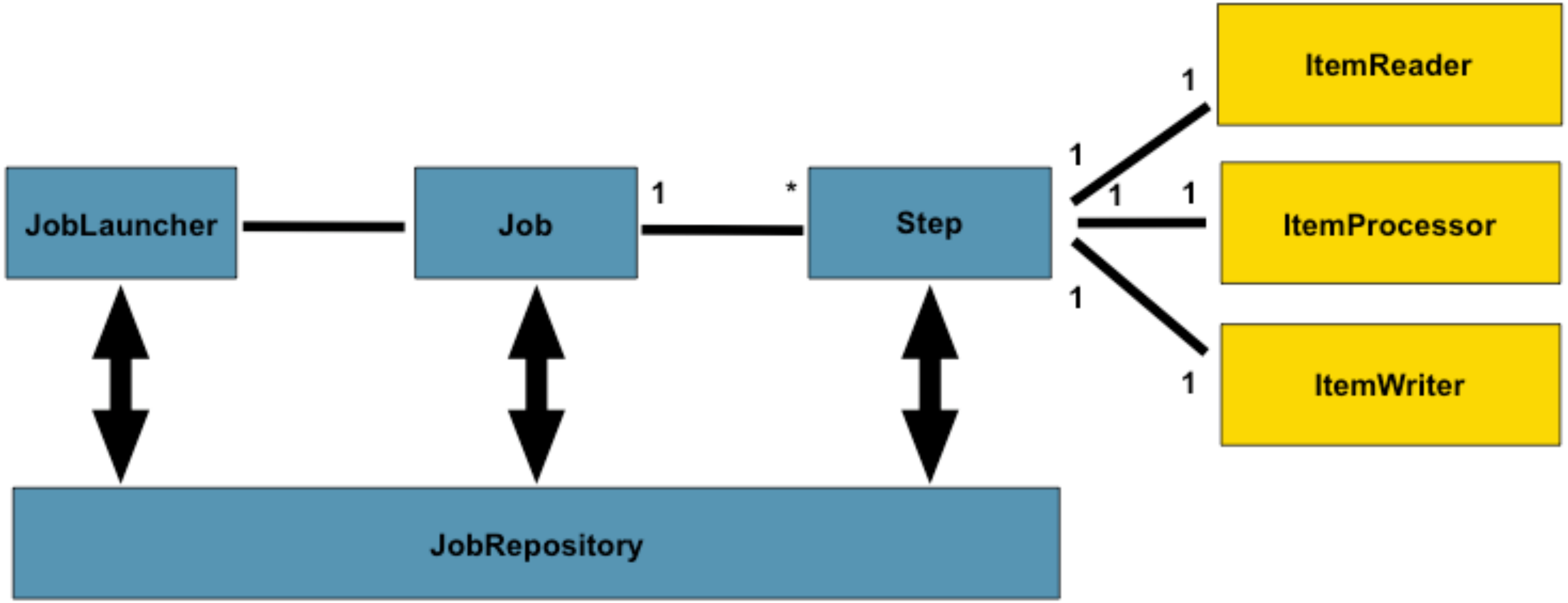
- Close of business processing
  - Order processing, Business reporting, Account reconciliation, Payroll
- Import / export handling
  - a.k.a. ETL jobs (Extract-Transform-Load)
  - Data warehouse synchronization
- Large-scale output jobs
  - Loyalty program emails, Bank statements
- Hadoop job orchestration

Features

- Transaction management
- Chunk based processing
- Schema and Java Config support
  - Annotations for callback type scenarios such as Listeners
- Start/Restart/Skip capabilities
- Based on the Spring framework
- JSR 352: Batch Applications for the Java Platform

Concepts

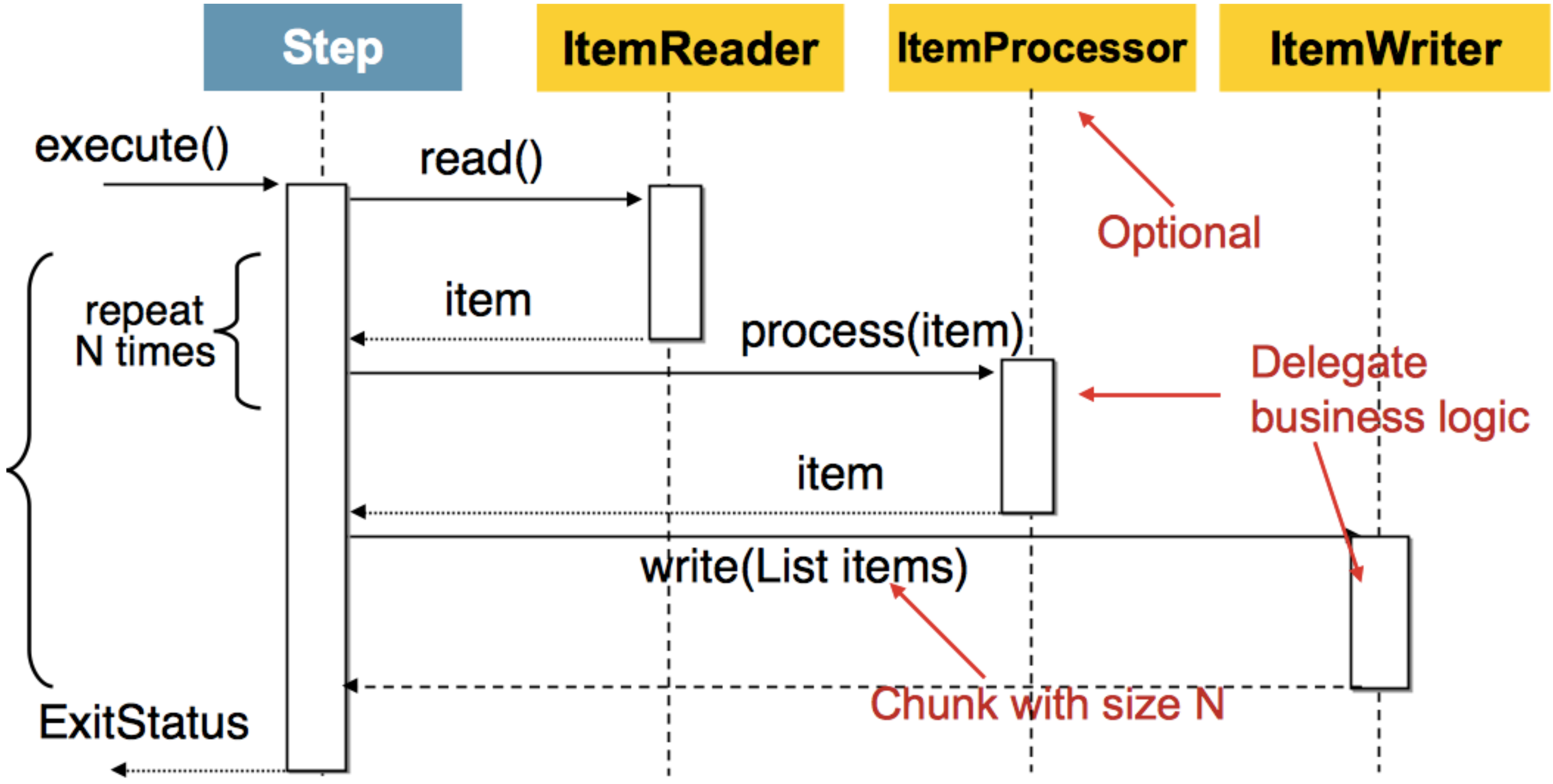
- Job
- Step
- Chunk
- Item



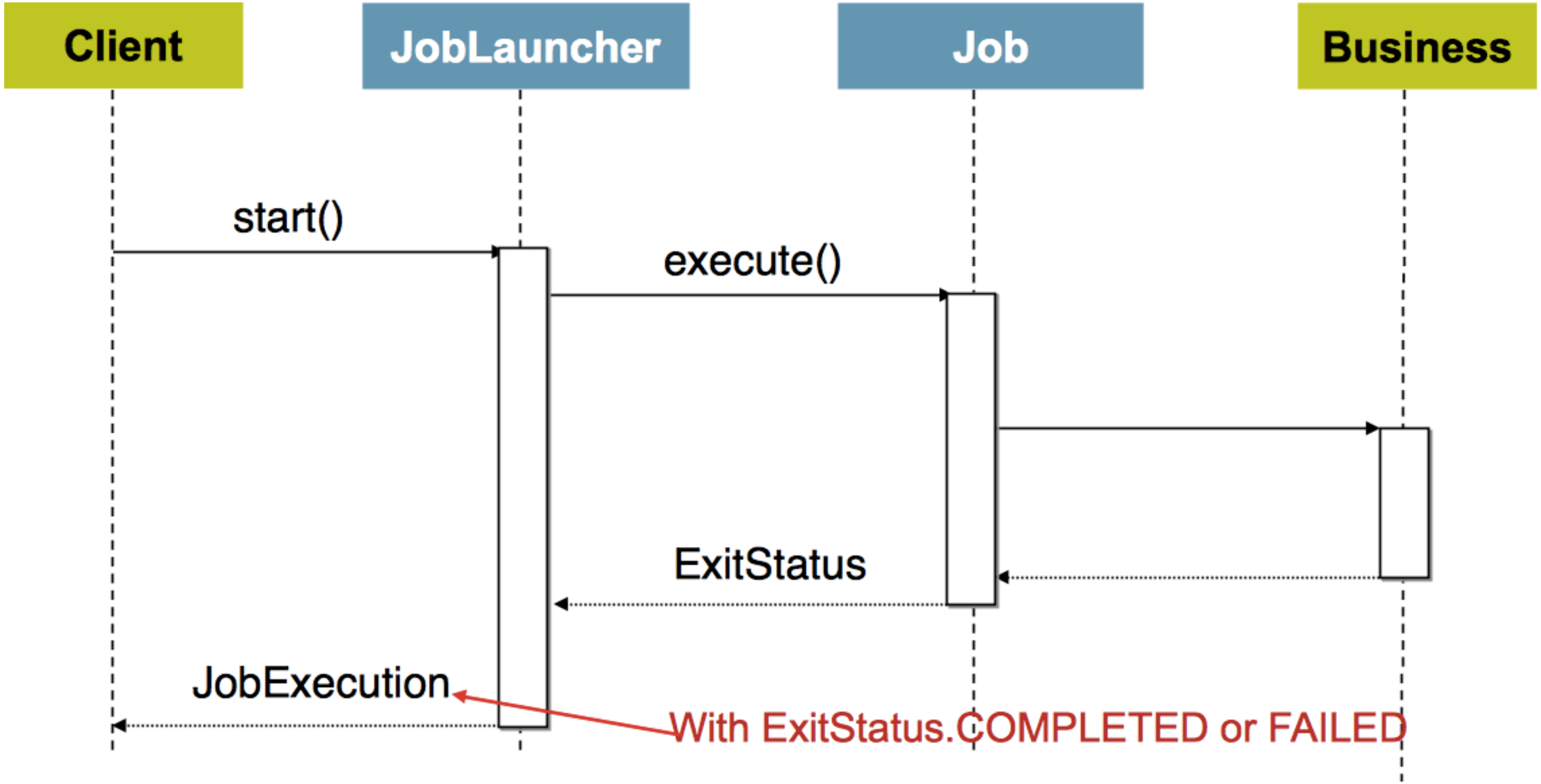
Repeat | Retry | Skip | Restart

Chunk-Oriented Processing

- Read data, optionally process and write out the “chunk” within a transaction boundary.



JobLauncher

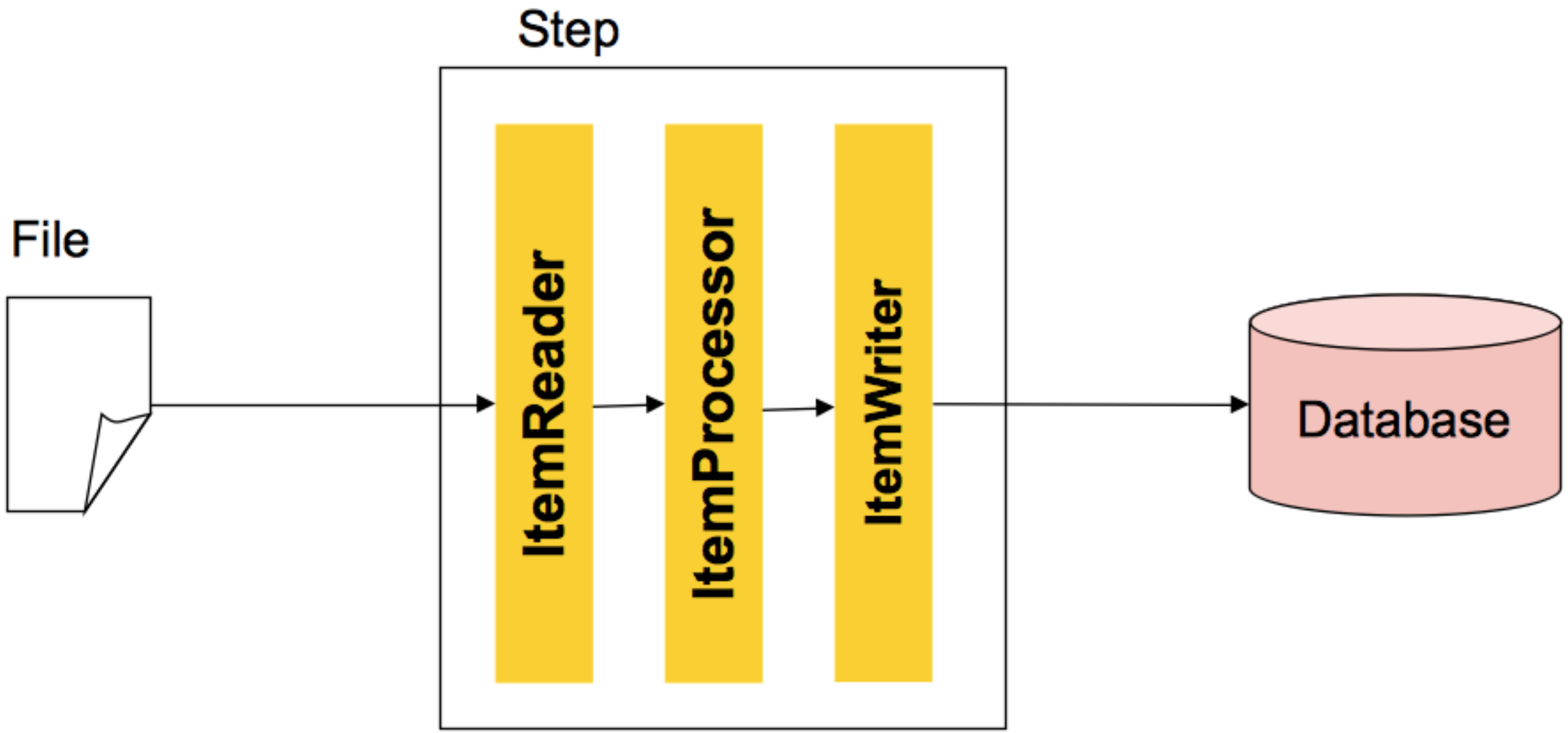


With ExitStatus.COMPLETED or FAILED

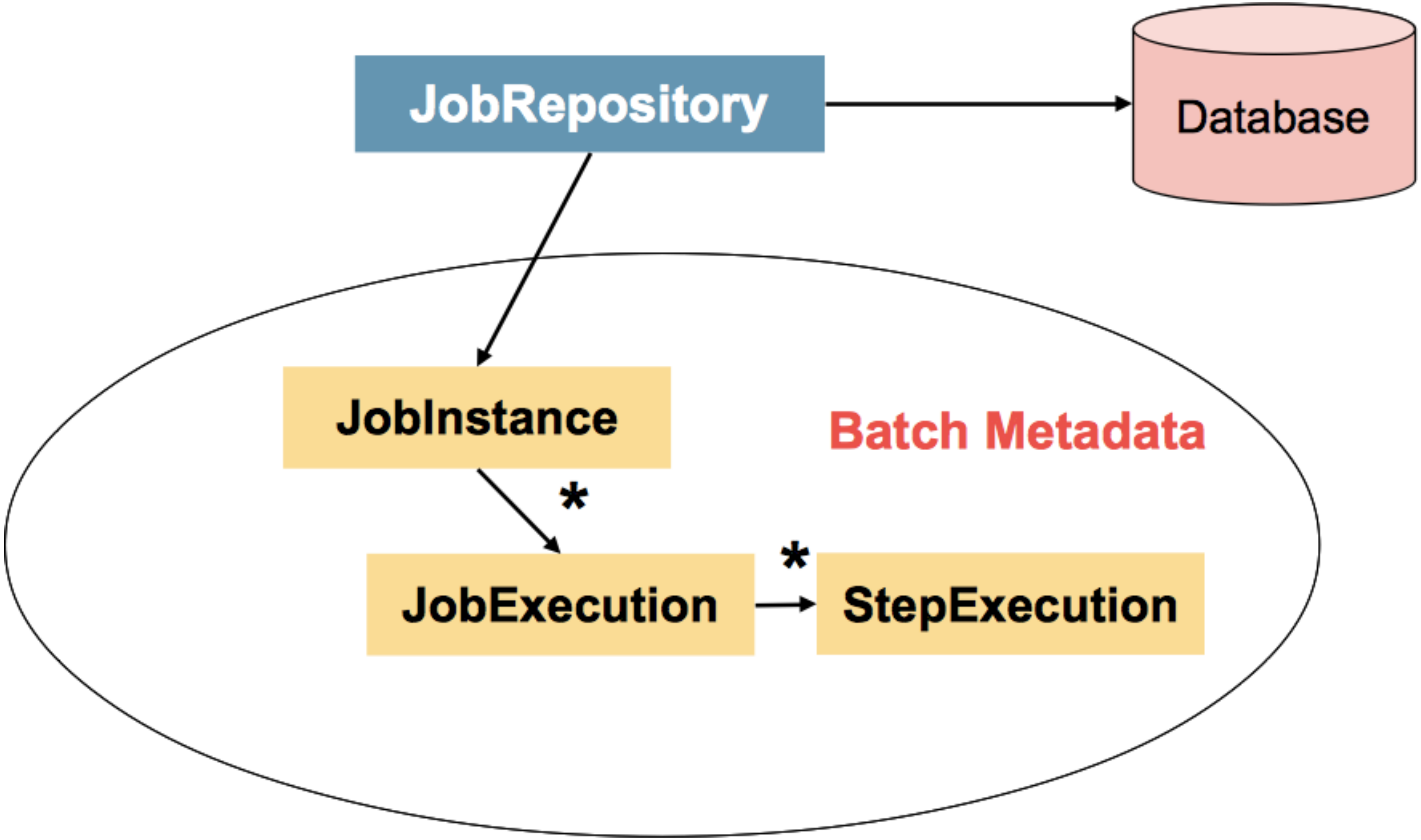
ItemReaders and ItemWriters

- Flat File
- XML (StAX)
- Multi-File Input
- Database
- JDBC, JPA/Hibernate, Stored Procedures, Spring Data
- JMS
- AMQP
- Email
- Implement your own...

Simple File Load Job



Job Repository



Spring Integration

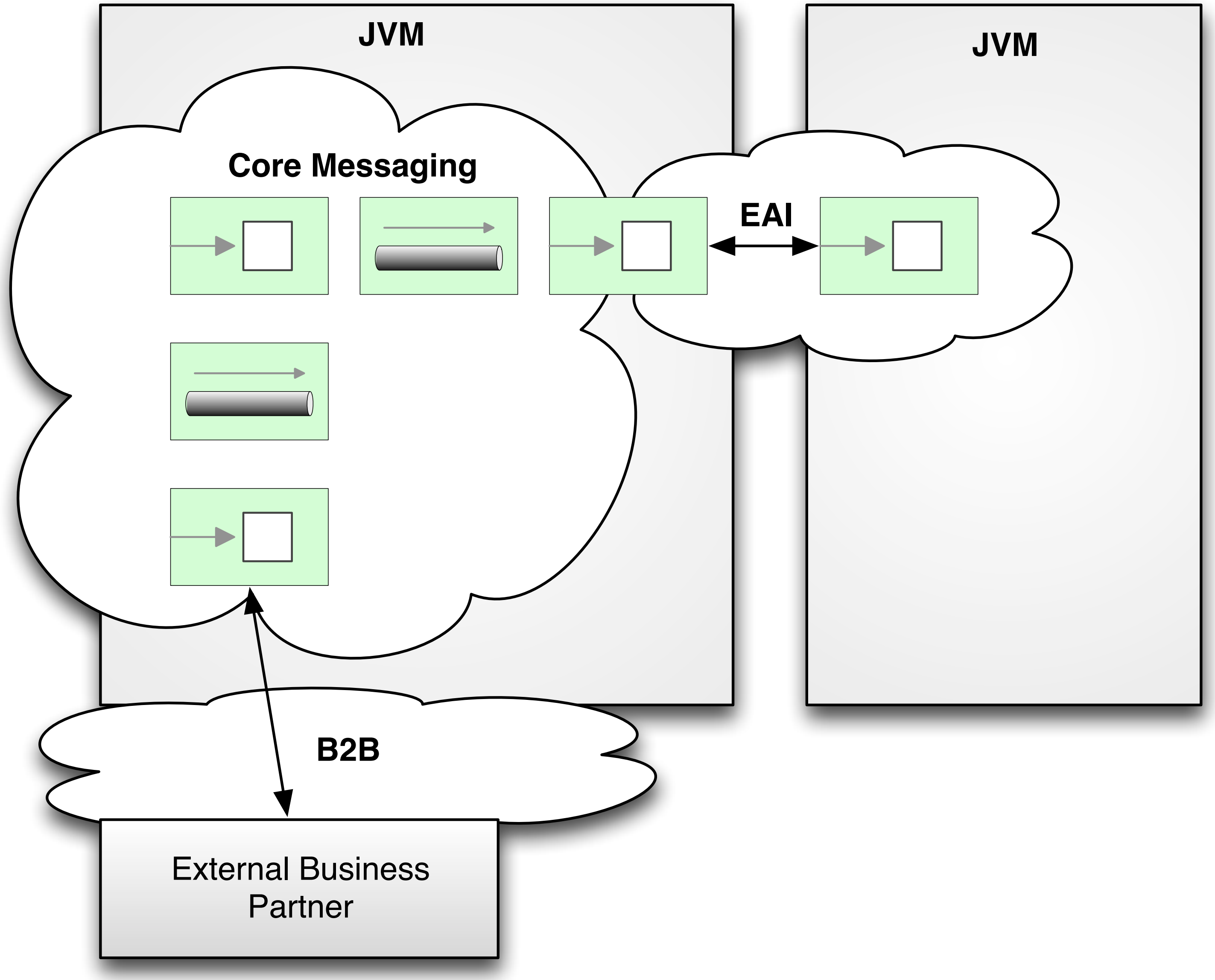
http://projects.spring.io/spring-integration/

Integration Styles

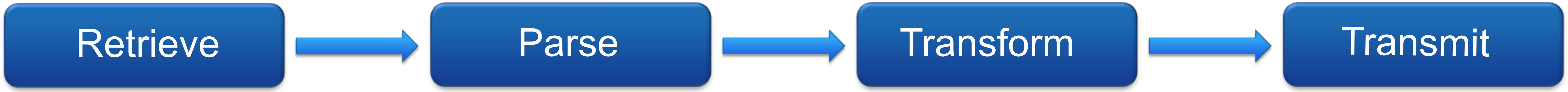
- File Transfer
- Shared Database
- Remoting
- Messaging

Integration Styles

- Business to Business Integration (B2B)
- Inter Application Integration (EAI)
- Intra Application Integration

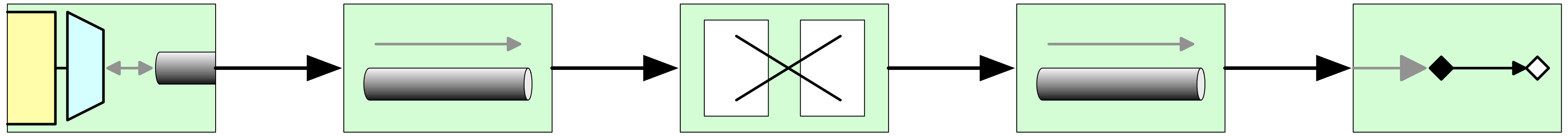


Common Patterns

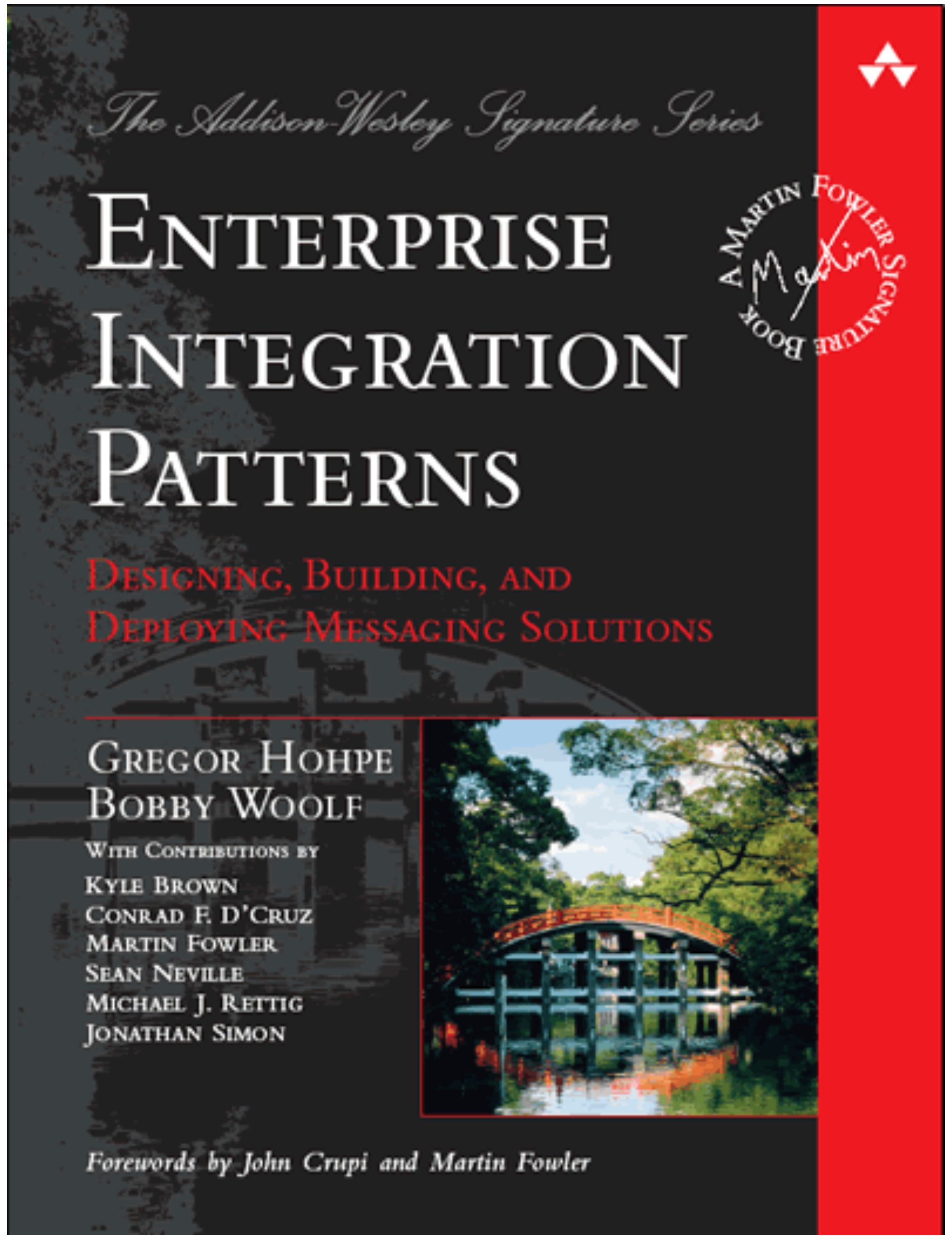


Enterprise Integration Patterns

- By Gregor Hohpe & Bobby Woolf
- Published 2003
- Collection of well-known patterns
- Icon library provided



http://www.eaipatterns.com/eaipatterns.html



The Addison-Wesley Signature Series

ENTERPRISE INTEGRATION PATTERNS

DESIGNING, BUILDING, AND DEPLOYING MESSAGING SOLUTIONS

A MARTIN FOWLER SIGNATURE BOOK

GREGOR HOHPE
BOBBY WOOLF

WITH CONTRIBUTIONS BY
KYLE BROWN
CONRAD F. D'CRUZ
MARTIN FOWLER
SEAN NEVILLE
MICHAEL J. RETTIG
JONATHAN SIMON

Forewords by John Crupi and Martin Fowler

> Spring Integration provides an extension of the Spring programming model to support the well-known enterprise integration patterns.

Spring Integration Website

Adapters

AMQP/RabbitMQ
AWS
File/Resource
FTP/FTPS/SFTP
GemFire
HTTP (REST)
JDBC
JMS
JMX
JPA

MongoDB
POP3/IMAP/SMTP
Print
Redis
RMI
RSS/Atom
SMB
Splunk
Spring Application Events

Stored Procedures
TCP/UDP
Twitter
Web Services
XMPP
XPath
XQuery

Custom Adapters

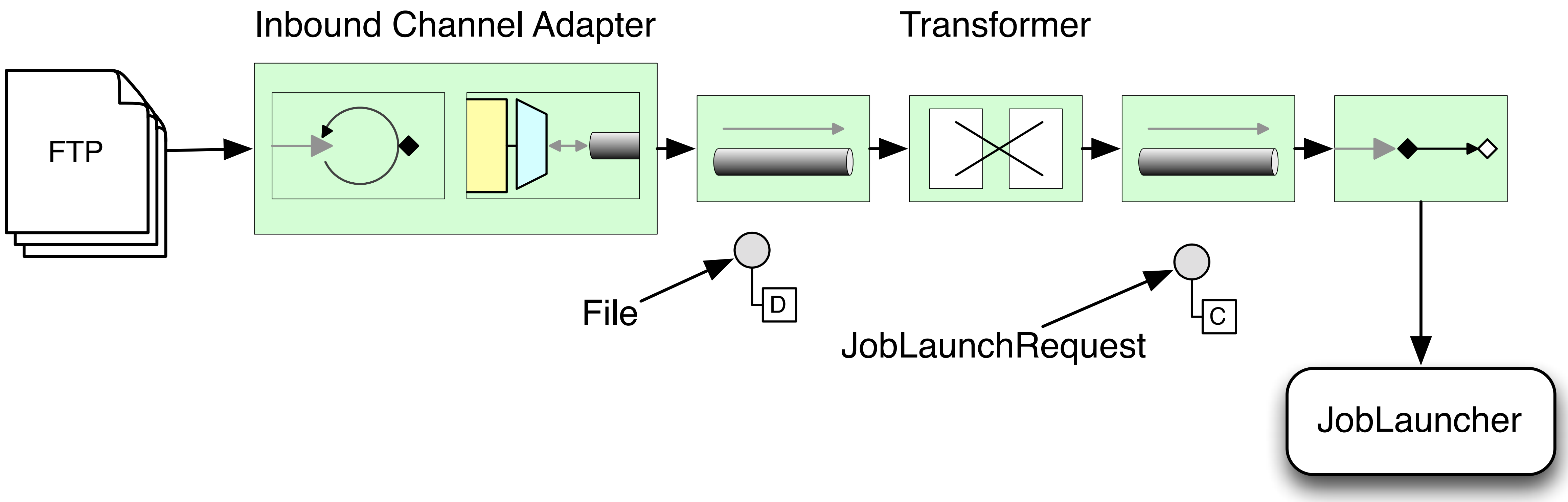
Samples

- https://github.com/spring-projects/spring-integration-samples
- Contains 50 Samples and Applications
- Several Categories:
  - Basic
  - Intermediate
  - Advanced
  - Applications

Spring Batch Integration

Launching batch jobs through messages

- Event-Driven execution of the JobLauncher
- Spring Integration retrieves the data (e.g. file system, FTP, ...)
- Easy to support separate input sources simultaneously



JobLaunchRequest

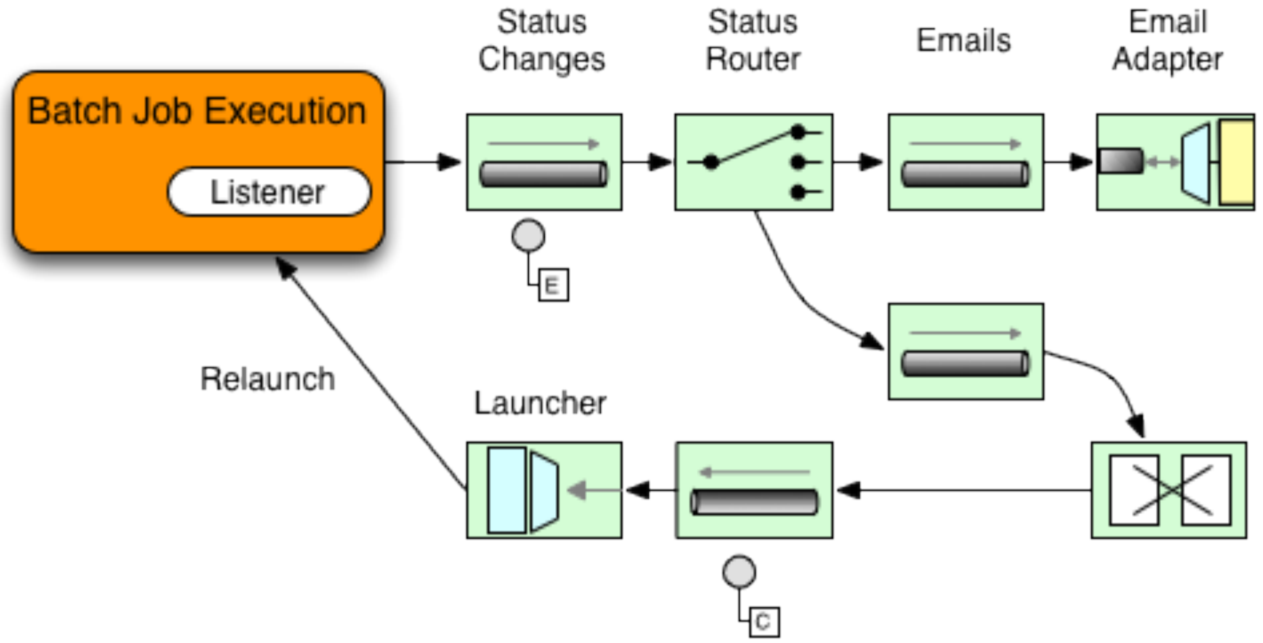
```
public class FileMessageToJobRequest {
  private Job job;
  private String fileParameterName;
    ...
  @Transformer
  public JobLaunchRequest toRequest(Message<File> message) {
    JobParametersBuilder jobParametersBuilder = new JobParametersBuilder();
    jobParametersBuilder.addString(fileParameterName,
                                   message.getPayload().getAbsolutePath());
      return new JobLaunchRequest(job, jobParametersBuilder.toJobParameters());
  }
}
```

JobLaunchRequest

```
<batch-int:job-launching-gateway  request-channel="requestChannel"
                                  reply-channel="replyChannel"
                                  job-launcher="jobLauncher"/>
```

Get feedback with informational messages

- Spring Batch provides support for listeners:
  - StepExecutionListener
  - ChunkListener
  - JobExecutionListener



Get feedback with informational messages

```
<batch:job id="importPayments">
    ...
    <batch:listeners>
        <batch:listener ref="notificationExecutionsListener"/>
    </batch:listeners>
</batch:job>

<int:gateway id="notificationExecutionsListener"
    service-interface="o.s.batch.core.JobExecutionListener"
    default-request-channel="jobExecutions"/>
```

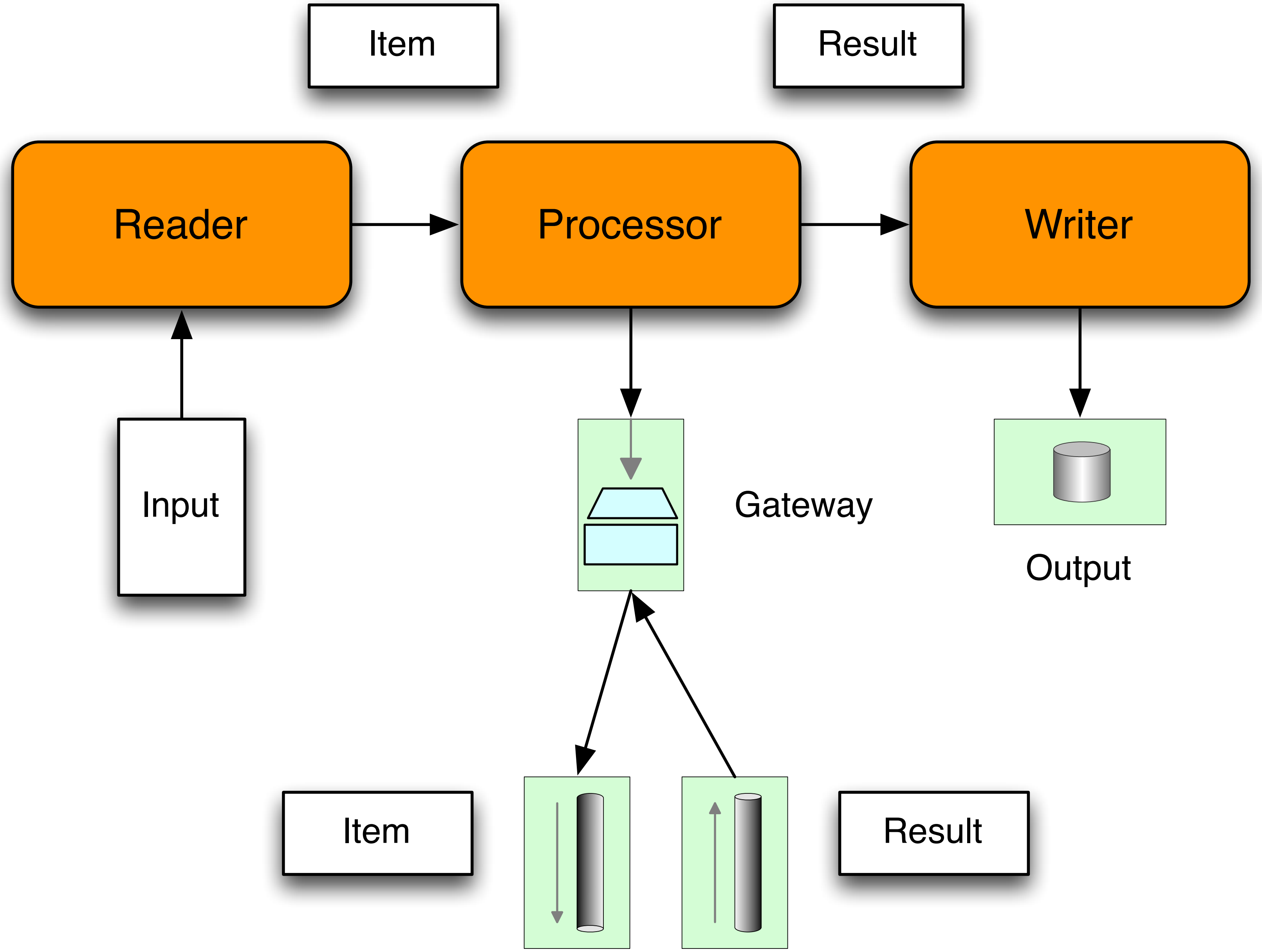
Launching and information messages demo in next section

Scaling Spring Batch

Scaling and externalizing batch process execution

- Utilization of Spring Integration for multi process communication
- Distribute complex processing
  - Single process
    - Multi-threaded steps
    - Parallel steps
    - Local partitioning
  - Multi process
    - Remote chunking
    - Remote partitioning
- Asynchronous Item processing support
  - AsyncItemProcessor
  - AsyncItemWriter

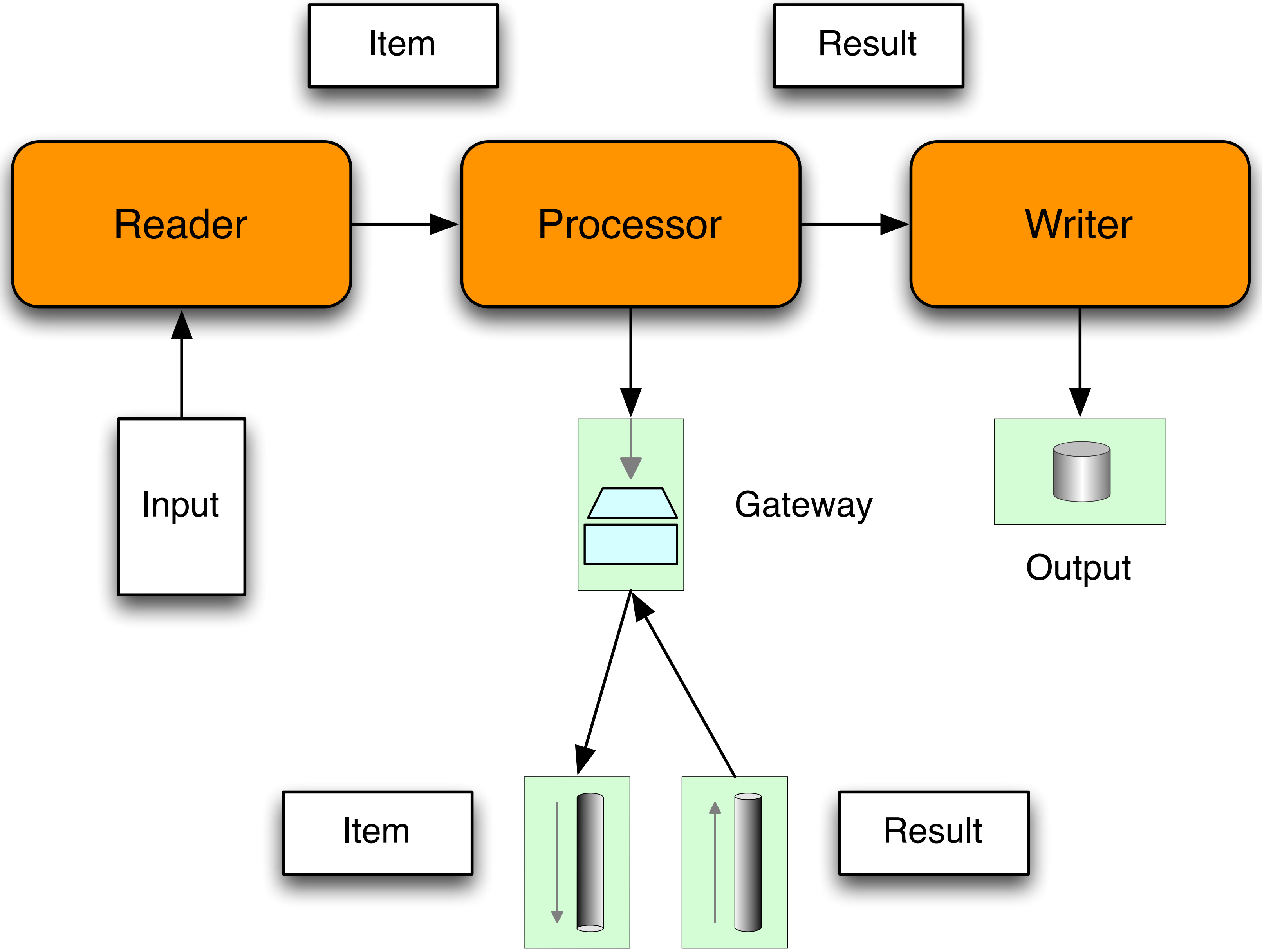
Single Thread



Single Thread - Demo

Multi-threaded

- Simply add a TaskExecutor to your Tasklet configuration



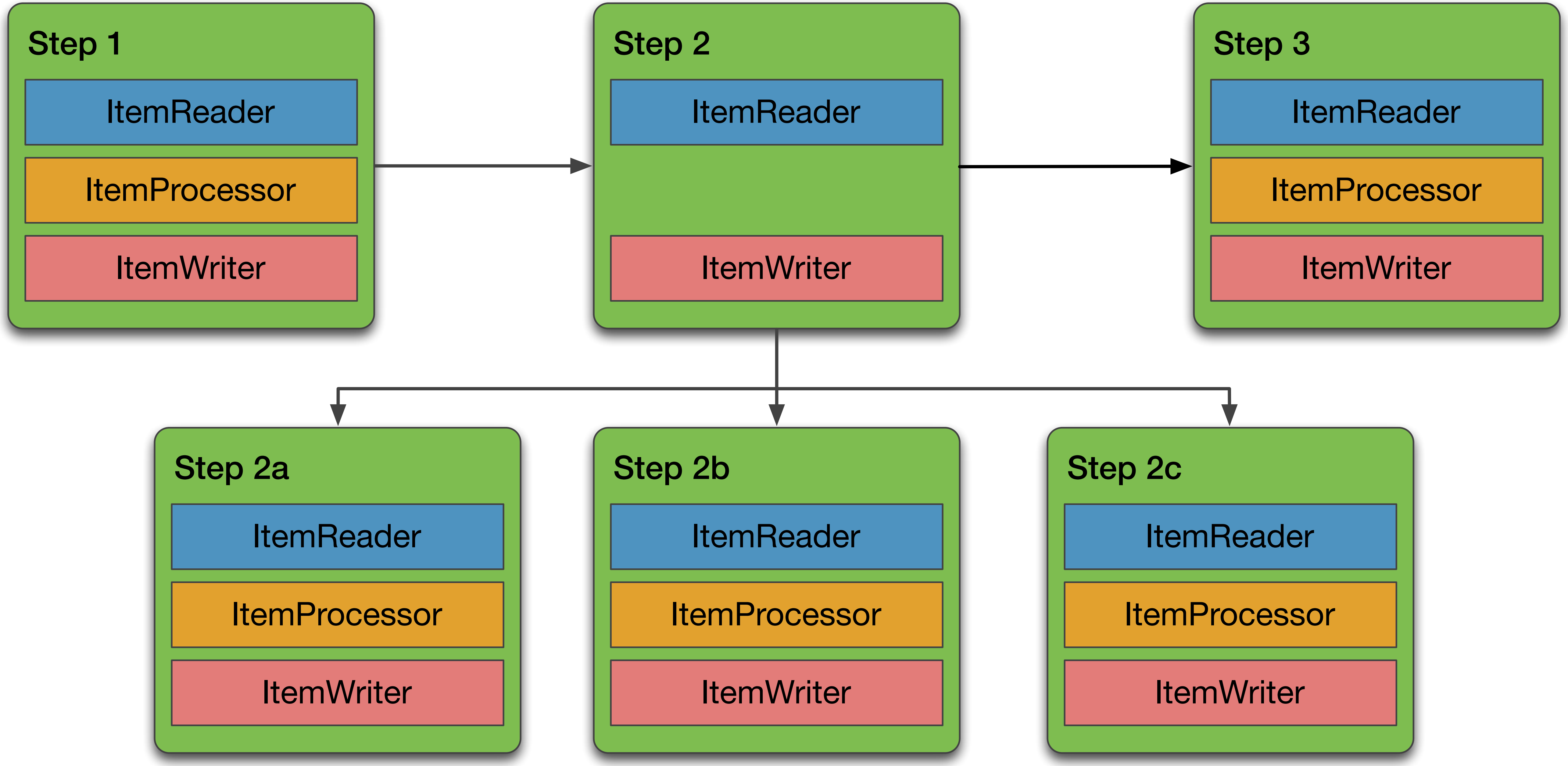
Multi-Threaded - Demo

Asynchronous Processors

- **AsyncItemProcessor**
  - Dispatches ItemProcessor logic on new thread, returning a Future to the AsyncItemWriter
- **AsyncItemWriter**
  - Writes the processed items after processing is complete

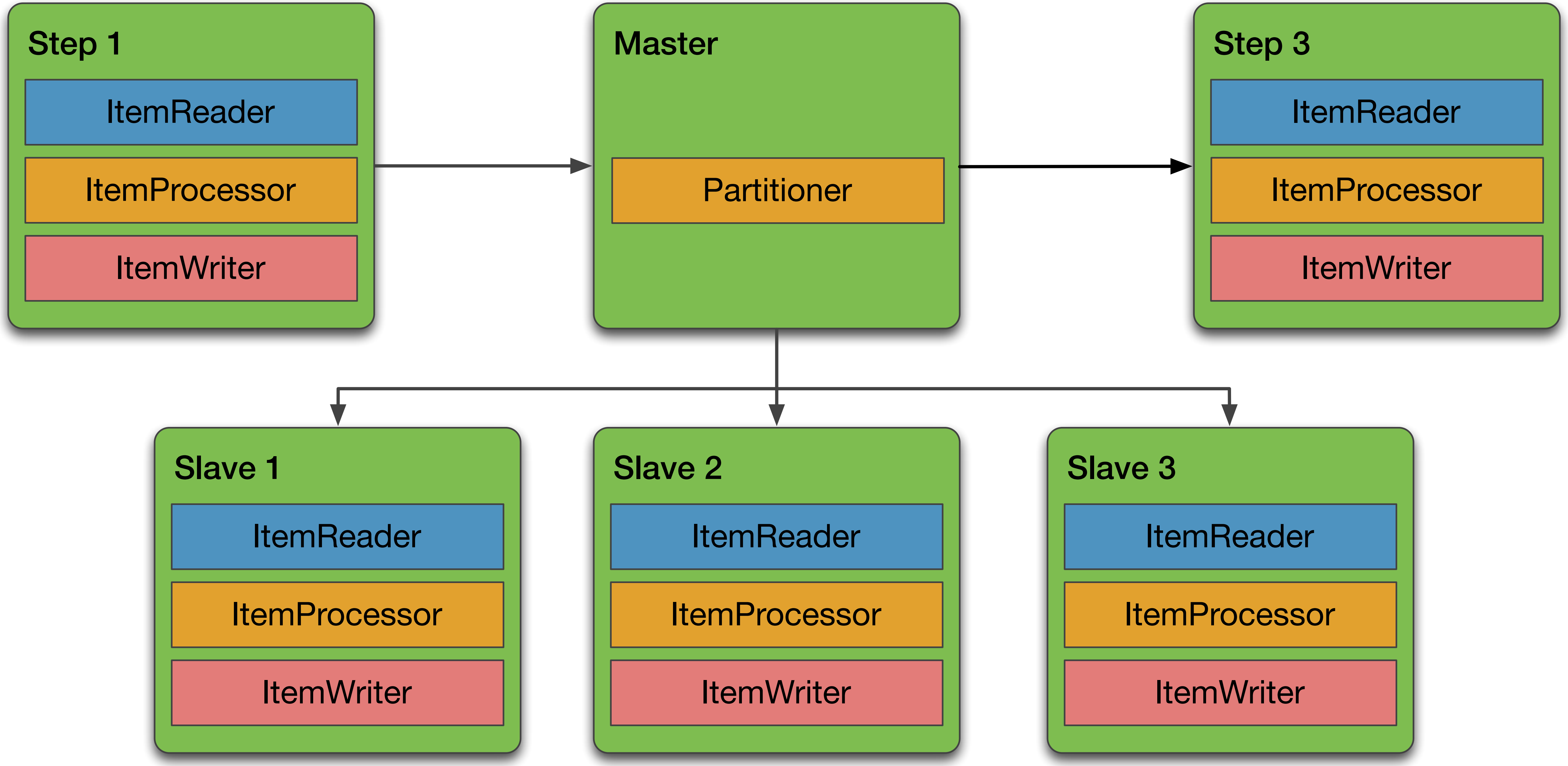
Asynchronous Processors - Demo

Remote Chunking



Remote Chunking - Demo

Remote Partitioning



Remote Partitioning - Demo

Demo - Launching via messages & informational messages

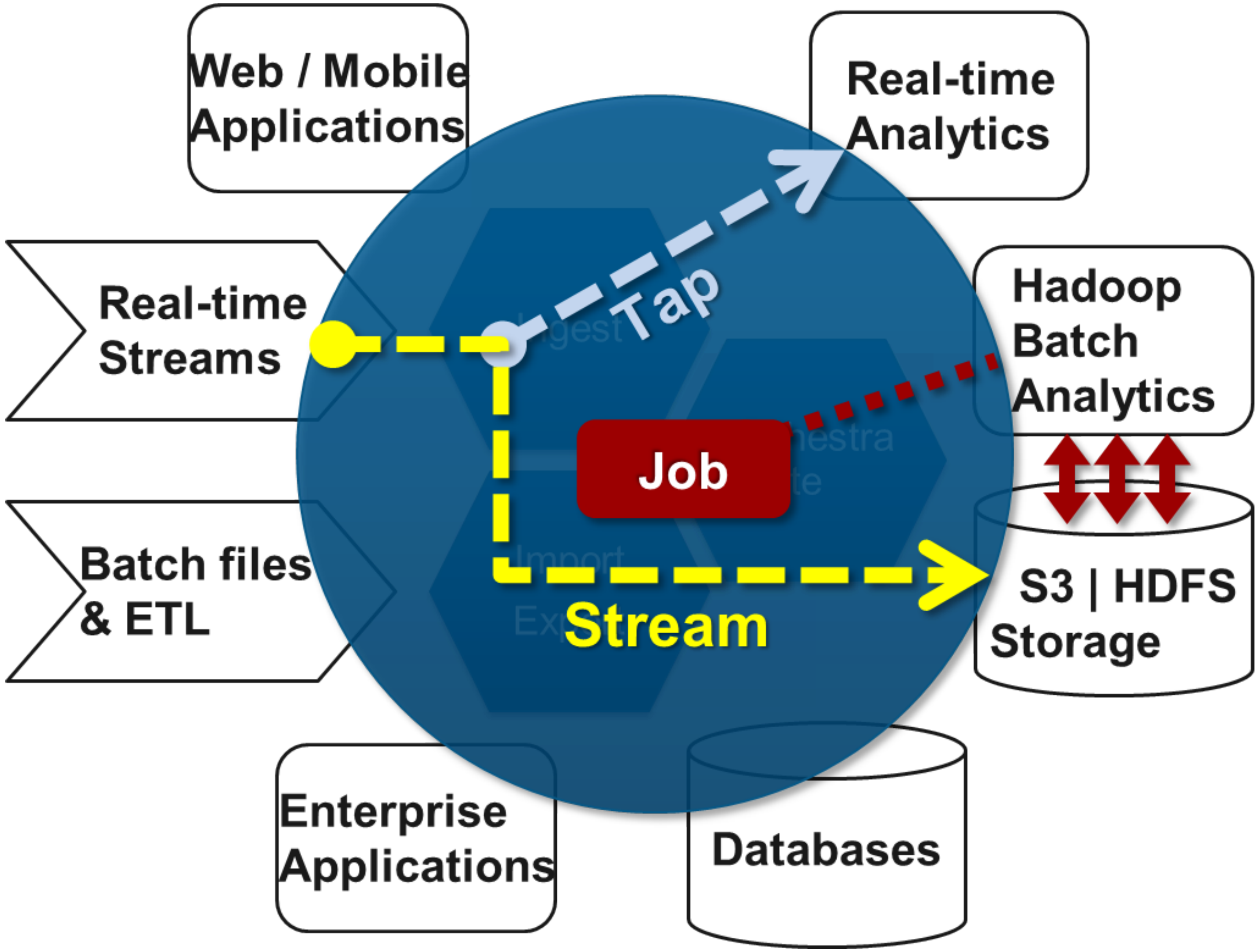
Does not provide scaling but demonstrates how launch job via messages and send information messages to integration points

Spring XD

http://projects.spring.io/spring-xd/

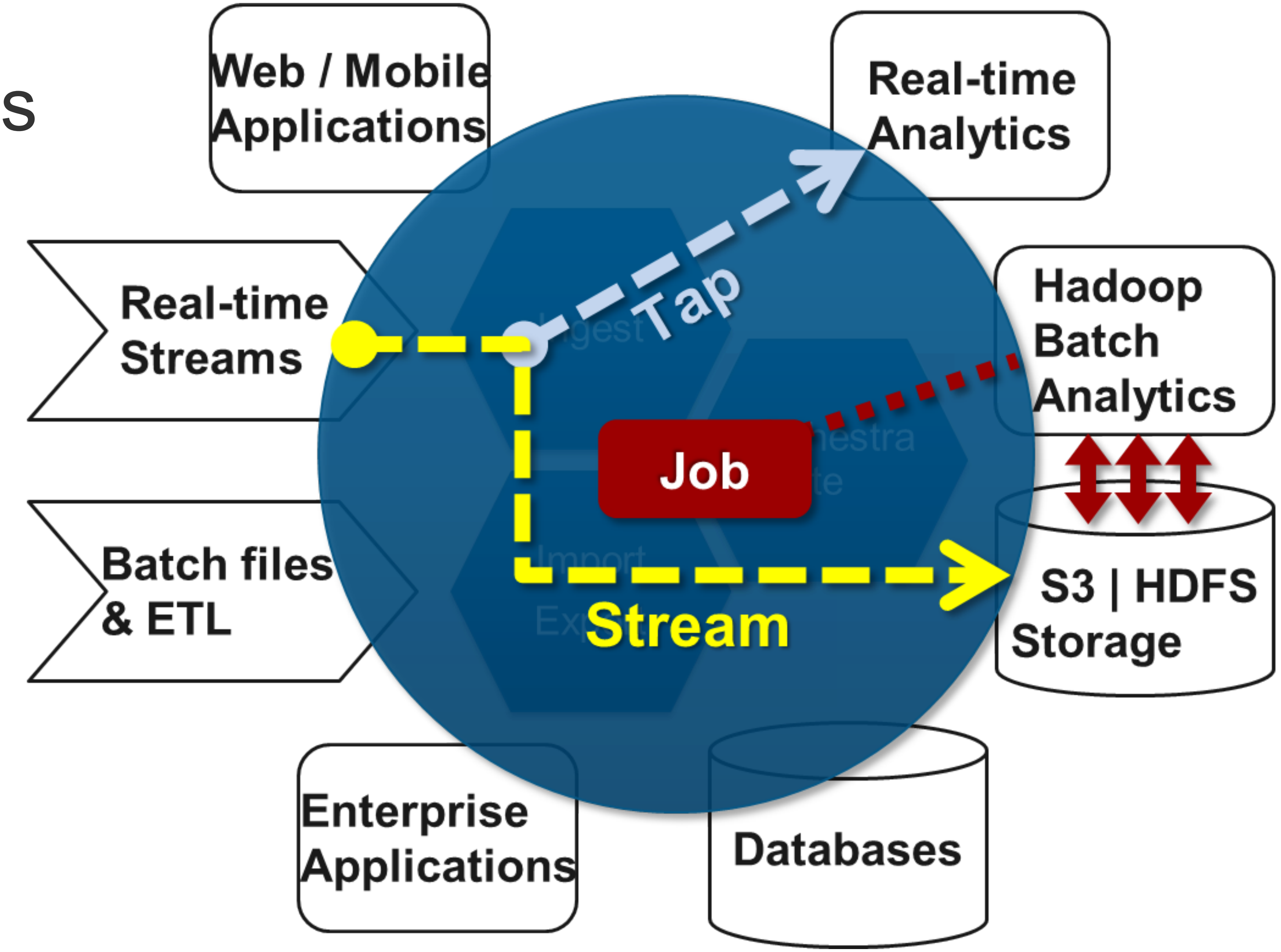
Tackling Big Data Complexity

- Data Ingestion
- Real-time Analytics
- Workflow Orchestration
- Data Export



Tackling Big Data Complexity cont.

- Built on existing Spring assets
  - Spring Integration
  - Spring Batch
  - Spring Data
  - Spring Boot
  - Spring for Apache Hadoop
  - Spring Shell
- Redis, GemFire, Hadoop



Data Ingestion Streams

- DSL based on Unix pipes and filters syntax

```
http | file
```

- Modules are parameterizable

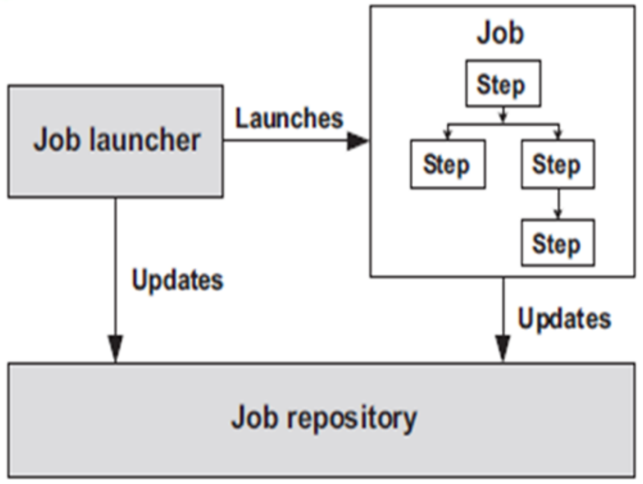
```
twittersearch --query=spring | file --dir=/spring
```

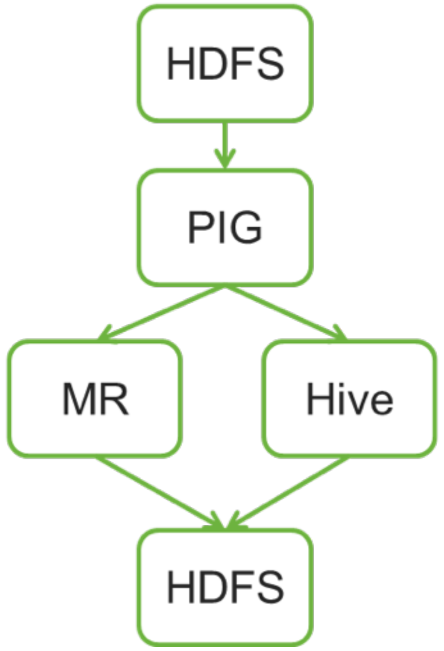
- Simple logic can be added via expressions or scripts

```
http | filter --expression=payload=='Spring' | hdfs
```

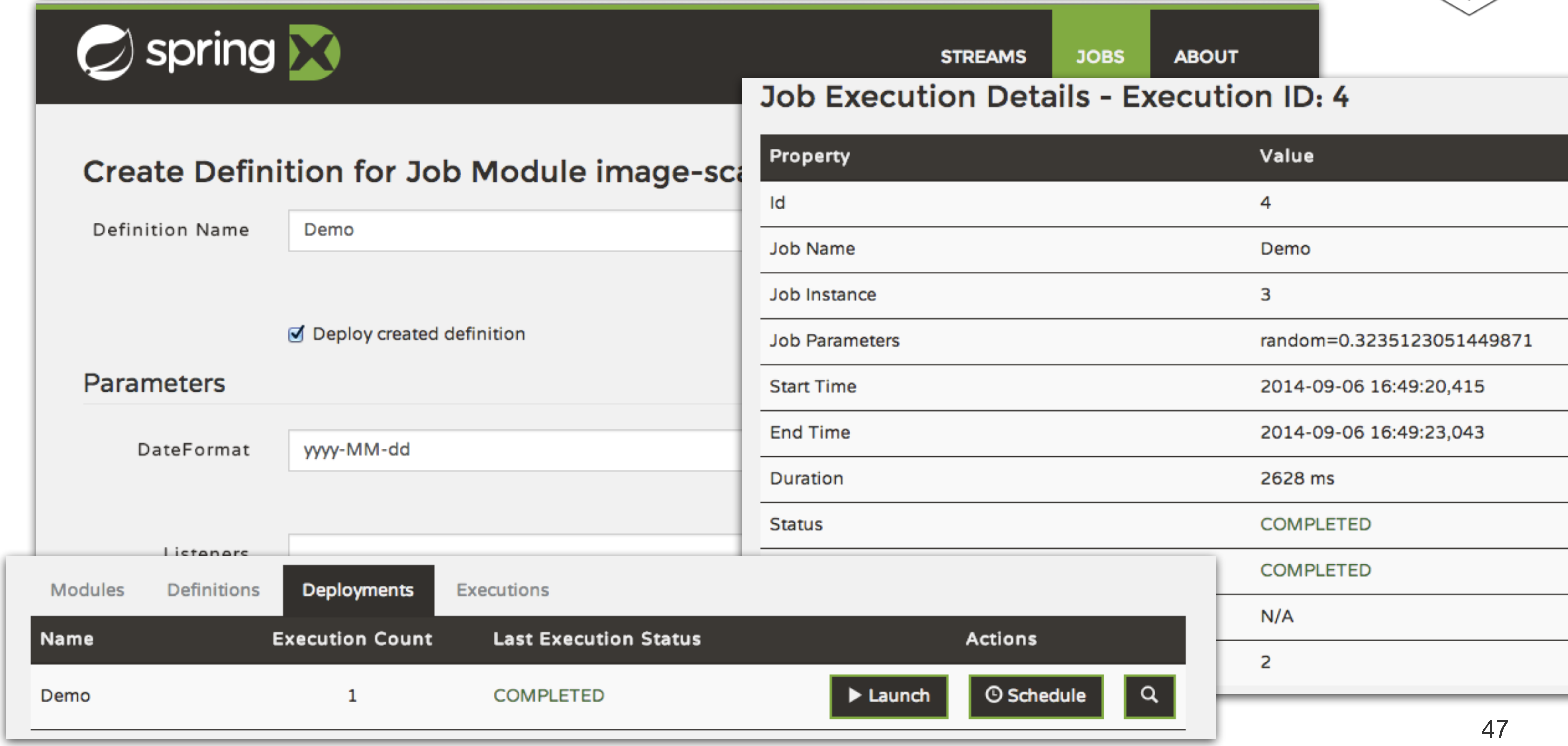
Hadoop workflow managed by Spring Batch

- Reuse Batch infrastructure and features to manage Hadoop workflows
- Job state management, launching, monitoring, restart/retry policies, etc.
- Step can be any Hadoop job type or HDFS script
- Can mix and match with other Batch readers/ writers, e.g. JDBC for import/export use-cases





Manage Batch Jobs with Spring XD



Create Definition for Job Module image-sca

Definition Name: Demo

☑ Deploy created definition

Parameters

DateFormat: yyyy-MM-dd

Listeners

Job Execution Details - Execution ID: 4

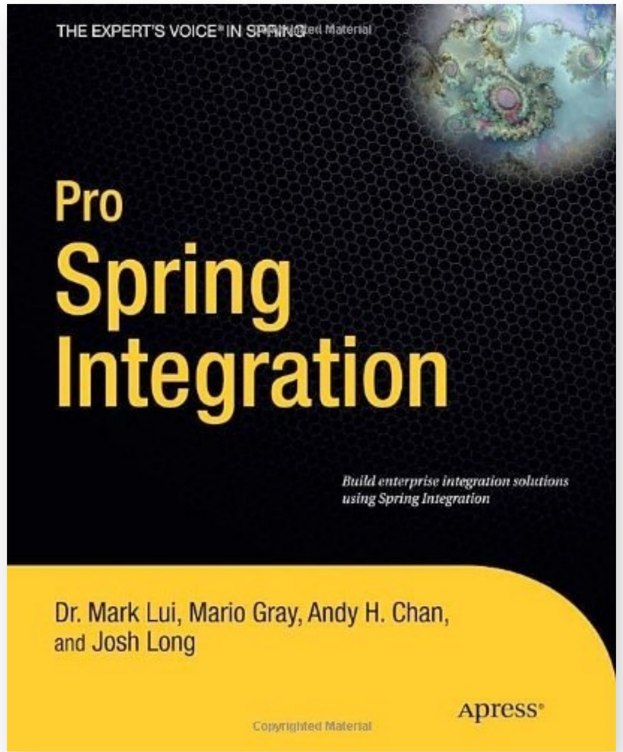
Property	Value
Id	4
Job Name	Demo
Job Instance	3
Job Parameters	random=0.3235123051449871
Start Time	2014-09-06 16:49:20,415
End Time	2014-09-06 16:49:23,043
Duration	2628 ms
Status	COMPLETED
	COMPLETED
	N/A
	2

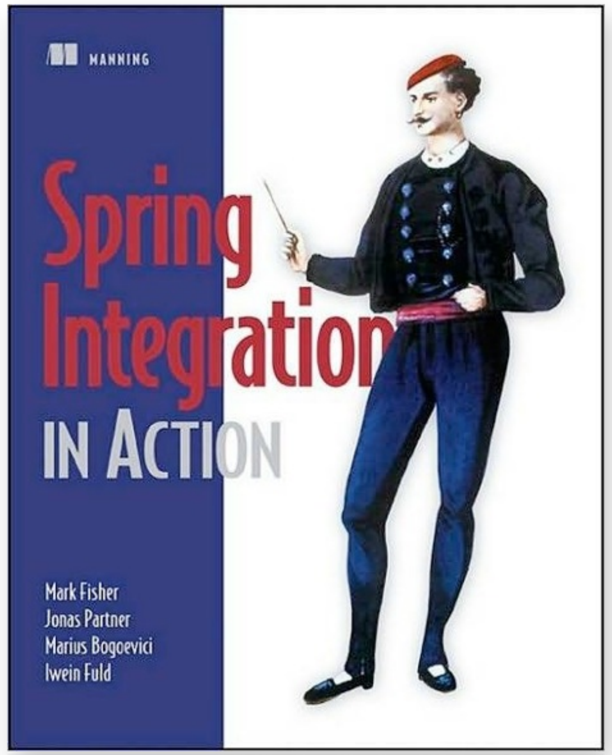
Name	Execution Count	Last Execution Status	Actions
Demo	1	COMPLETED	▶ Launch · Schedule · 🔍

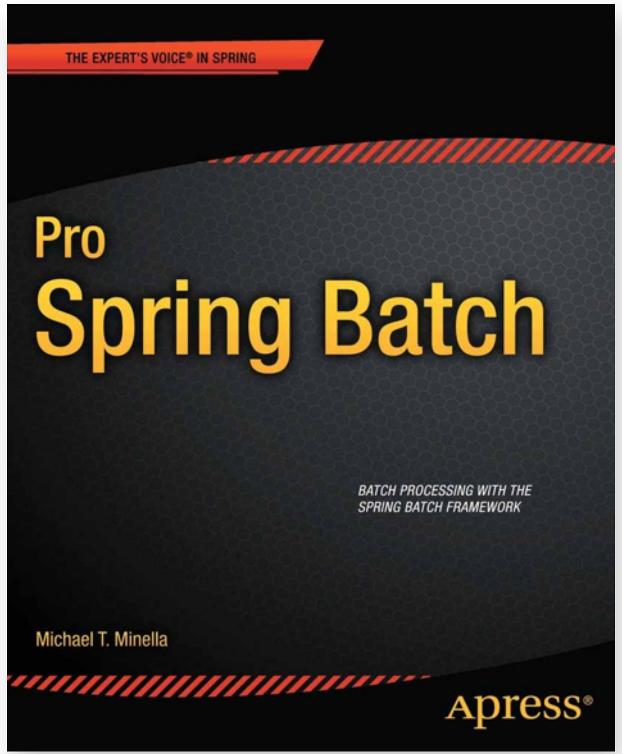
Spring XD - Demo

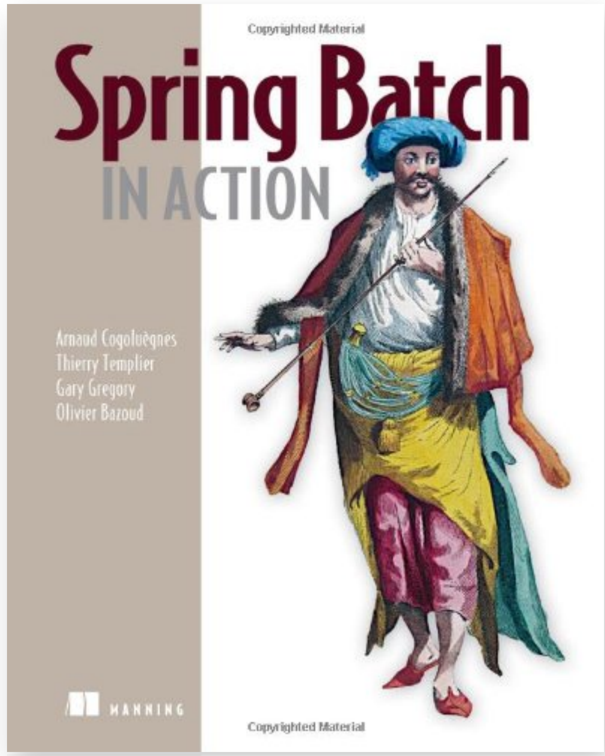
Books

Enterprise Application Patterns and Messaging
Just Spring Integration
O'REILLY
Madhusudhan Konda
Foreword by Josh Long


THE EXPERT'S VOICE® IN SPRING
Pro Spring Integration
Build enterprise integration solutions using Spring Integration
Dr. Mark Lui, Mario Gray, Andy H. Chan, and Josh Long
Apress
Copyrighted Material


MANNING
Spring Integration IN ACTION
Mark Fisher
Jonas Partner
Marius Bogoevici
Iwein Fuld


THE EXPERT'S VOICE® IN SPRING
Pro Spring Batch
BATCH PROCESSING WITH THE SPRING BATCH FRAMEWORK
Michael T. Minella
Apress


Copyrighted Material
Spring Batch IN ACTION
Arnaud Cogoluègnes
Thierry Templier
Gary Gregory
Olivier Bazoud
MANNING
Copyrighted Material

Learn More. Stay Connected.

THANK YOU!

Demo code and slides:

- https://github.com/SpringOne2GX-2014/spring-batch-performance-tuning