

catalog-synchronizer

This Spring Boot 3 Software reads a TSV format google product list file (default from resource directory /input, configurable with `path.input`), validates it and merge it in a sqlite database (`dbfilenamewithpath`). it uses apache kafka to handle the readed data. at the end it generates an XUni-type report (`path.output`) about how many record was readed and how long does it take. This report s easy to integrate in a CI Tool (e.g. Jenkins) because it there is an error, the job can be easily switched to "unsuccessfull". More details about the readed data and what exactly was done in the database, is stored in the kafka. every run has an unique requestid. Default the kafka-stream flushed 1000 records (`flushSize`) and Spring-JPA handles 100 entities (`hibernate.jdbc.batch_size`). Before every run, this software makes a backup about the sqlite database.

How to run this Software

Start your own Kafka-Server (on localhost:9092) or go to the kafka folder and start a Kafka in docker with make start-kafka

Than run `mvn spring-boot:run -Dspring-boot.run.arguments="file1.txt"` where file1.txt is under resource directory /input when not configured otherwise (`path.input`).

Example output report file: <https://github.com/Csini/catalog-synchronizer/blob/main/doc/report-72b8bde0-342b-4511-bf52-a565d9841b23.xml>

Problems Console Progress JUnit X

REQUEST: 72b8bde0-342b-4511-bf52-a565d9841b23

Runs: 16/16 (9 skipped) Errors: 0 Failures: 0

REQUEST: 72b8bde0-342b-4511-bf52-a565d9841b23 (100.934 s)

input: file3.txt

output: src/main/resources/products.db

Readed Valid Products:18244 (9.421 s)

TSV:9122 (9.421 s)

Valid:9122

Invalid:0

DB:9122 (6.473 s)

Processed Products: 9122

Processed Events: 9122 (47.002 s)

INSERT: 0

UPDATE: 9000

DELETE: 0

NOCHANGE: 9122

Invalid EXAMPLES: (0.431 s)

eOMtThyHvNLWUZNRCBaQKxl yedUsFwdkelQbxTeQOvaScfqlOOmaa JxkyvRnL PREORDER USED -4571.589078663

gNfZBdyFGRajVfJNonEnOinZj UfzQhdgLLfDTDGspDb QyBQYuxiXXVytGCxzVllpgTKhRQq IN_STOCK USED 8000.788

Oxm Qln ylvpRgmgQsYEKk BACKORDER USED 9600.541788329743 null -7717.821787758737 HUF AAAnyjCRhLTuhn1

jNBgpTmxx ploQM ODRHfGEFX OUT_OF_STOCK USED 3167.1582361716487 null 9492.924717417216 null IoTtOmcBei

OV dwprLNBgNvZUMpqAXLlBje TiQaUNWftniOAotnzWvkQt OUT_OF_STOCK NEW 2913.396158707964 null -1062.4

oBeMdQkAoYZDjfWhbzkmA NBaxYoMSiMNxLXFLxlok teOlrImKiDpJbLDnvK PREORDER USED 7643.421783862093 r

yybxVLsNIaELWVhniIULZAylBms LcdeWTdXPIQgjMVXbpRYzBT jzlgIRKAeamYUmWJtnJZLqwakeYcea PREORDER Nf

12345678910111213141516171819202122232425262728293031323334353637383940 WtKhlfZzDZ xlpKnXmsjfwUTYgff

zQeAydAJlwGuKWIVZLRcElhreuxvww YVeSekWdEtYtJHV etgDUdmb BACKORDER USED -922.7952371680925 null -1

Failure Trace

price.amount: -4571.589078663439 - must be greater than or equal to 0

image_link: UMaAIKkknjWEXJUfPxxQHeWKEJ - must be a valid URL

sale_price.currency: null - currency is required

link: RYtGKGbicZaHCBRQDSx - must be a valid URL

price.currency: null - currency is required

sale_price.amount: -6180.041442026416 - must be greater than or equal to 0

googleProductCategoryValid: false - google_product_category should be ir

Configuration properties

application.properties

```
spring.kafka.bootstrap-servers=localhost:9092
topic.name.readedFromDb=readedFromDb
topic.name.validProduct=validProduct
topic.name.invalidProduct=invalidProduct
topic.name.flushed=flushed
topic.name.product=productEvents
topic.name.dbevent=dbEvents
topic.name.runs=runs

flushSize=1000

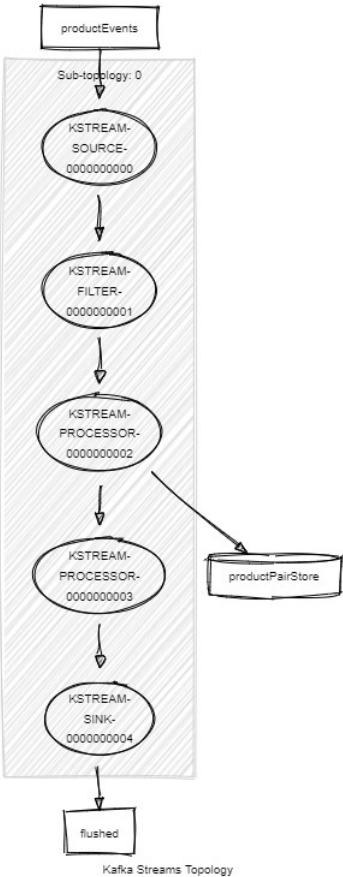
store.name.productPair=productPairStore

path.input=/input
path.output=output
```

persistence.properties

```
url=jdbc:sqlite:${dbfilenamewithpath}?date_class=TEXT
dbfilenamewithpath=src/main/resources/products.db
spring.jpa.database-platform=org.hibernate.community.dialect.SQLiteDialect
hibernate.jdbc.batch_size=100
hibernate.jdbc.batch_versioned_data=true
```

Kafka Topology



Kafka Producers and Consumers

<https://github.com/Csini/catalog-synchronizer/blob/main/doc/springwolf.yaml>

Product DB Schema

PRODUCT	
id	
age_group	
availability	
brand	
condition	
created	
description	
google_product_category	
image_link	
link	
priceAmount	123
priceCurrency	
salePriceAmount	123
salePriceCurrency	
title	
updated	
requestid	

RUN	
requestid	
created	
filename	
updated	