Branko Zečević | 2023-10

Processing spreadsheets in Go

Utilization of the Go programming language in the implementation of integration services in business application environments

Why spreadsheets?

Spreadsheets are everywhere !

Common uses:

- Data/business analytics and computing
- Data storage
- Data import/export
- Data verification and cleanup
- Data mining and manipulation
- Data visualization (charts)
- Reporting
- Business administration (invoices, receipts, inventory, …)
- …

Data formats: CSV, DSV

Features:

- text file format
- stores tabular data, each line represents one data record

Comma-separated values (CSV):

- uses commas to separate values
- variant of DSV

Delimiter-separated file format (DSV):

- two-dimensional array of data, w/ values in row separated with specific delimiter
- common delimiters: comma, semicolon, pipe, …

Standards:

- RFC 4180, Common Format and MIME Type for Comma-Separated Values (CSV) Files

Data formats: MS Excel

Features:

- file format for spreadsheets (worksheets)
- stores data in the form of a matrix
- stores tabular data, each row includes cells

Office Open XML :

- developed by Microsoft for representing spreadsheets

Standards:

- ECMA-376
- ISO/IEC 29500

Spreadsheets and ETL

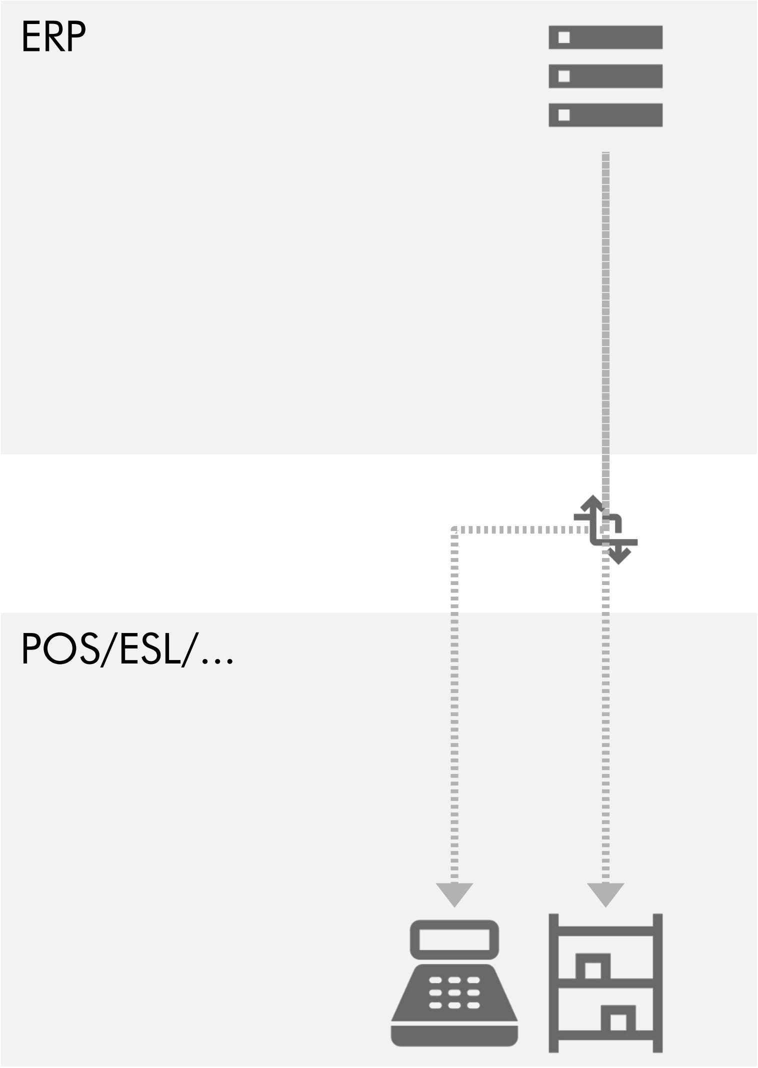
ETL (Extract, Transform, Load)

- process of data extraction/transformation/loading into output data container
- data can be loaded/collated from multiple sources
- data output can target multiple destinations

Why spreadsheets ?

  - used as source of data
  - used as target for output
  - used as format for information exchange

Application examples

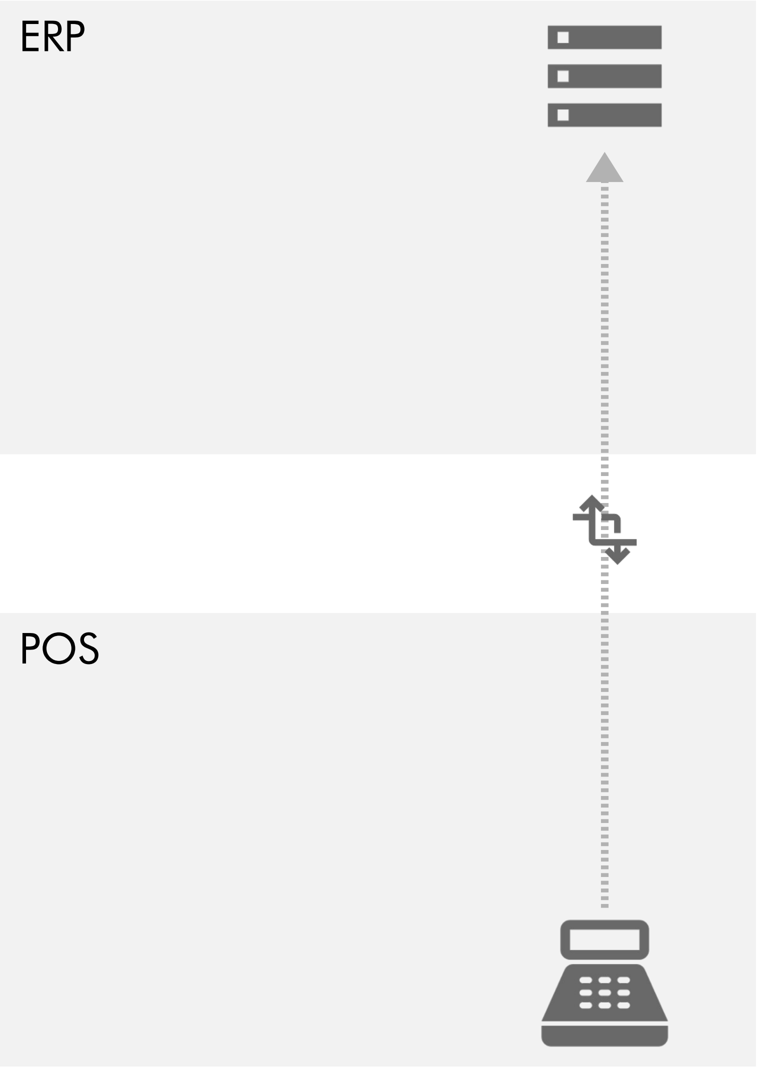


ERP-POS integration:

- Extract data from ERP database
- Transform to “spreadsheet” format
- Transport to POS
- Load to POS database

- Common data sets: sales master data, tax master data, employee data, ...

Application examples (2)



POS-ERP integration:

- Extract data from POS database
- Transform to “spreadsheet” format
- Transport to ERP
- (optional) Load to ERP database

- Common data sets: sales revenue data, invoices, taxes, ...
- Spreadsheet as “report” format for selected org. units (eg, Accounting, Controlling, ...)

Go and CSV/DSV

Package encoding/csv:

- https://pkg.go.dev/encoding/csv
- package reads and writes comma-separated values (CSV) files
- basic types: Reader, Writer

Reader features:

- field delimiter
- comment character (ie, lines beginning with comment are ignored)
- number of expected fields per record (optional, no checking)
- "LazyQuotes" (quote may appear in unquoted field, non-doubled quote may appear in a quoted field
- leading white space in a field can be ignored

Writer features:

- field delimiter

Go and Excel

Package Excelize:

- https://xuri.me/excelize/
- package reads and writes XLAM / XLSM / XLSX / XLTM / XLTX documents
- basic types: Workbook, Worksheet, Cell, Chart, Picture, Style

Features:

- not standard library/3rd-party
- pure Go
- requires Go version 1.16 or later
- reads and writes documents generated by MS Excel 2007 and later
- streaming API for generating or reading huge amounts of data

Go software

Miller: Command-line tool for querying, shaping, and reformatting data files in CSV, TSV, JSON, JSON Lines, ...

- https://github.com/johnkerl/miller

Dasel: Query and update data structures from command line in JSON, TOML, YAML, XML, CSV, ...

- https://github.com/TomWright/dasel

omniparser: Golang ETL streaming parser and transform library for CSV, JSON, XML, EDI, text, ...

- https://github.com/jf-tech/omniparser

Sources:

- Golang Weekly, https://golangweekly.com

Common tasks

CSV/DSV:

- decode/encode document according to character encoding (eg, utf8, cp852, win1250, …)
- setup delimiter, trimming
- verify number of columns
- format cell content according to locale

MS Excel:

- get/set workbook properties
- get/set document properties
- retrieve list of worksheets
- select worksheet
- format cell content according to locale

Sample programs

CSV/DSV, csvread:

- test 1
- test 2: encoding, language

CSV/DSV, csvwrite:

- test 1
- test 2: w/ encoding, language

MS Excel, excelread:

- test 1: get workbook props, get doc props, get sheet, read rows

MS Excel, excelwrite:

- test 1: new sheet, set row, set sheet name, set doc props

References

Comma-separated values (CSV):

- https://www.rfc-editor.org/info/rfc4180
- https://en.wikipedia.org/wiki/Comma-separated_values
- https://en.wikipedia.org/wiki/Delimiter-separated_values

Office Open XML:

- https://en.wikipedia.org/wiki/Office_Open_XML

Thank you!

Contacts:

Branko Zečević, branko.zecevic@pointer.hr