Designing with Best Practices

Stacia Varga
CONSULTANT, INSTRUCTOR, AUTHOR

@_StaciaV_ datainspirations.com

Overview

General data flow

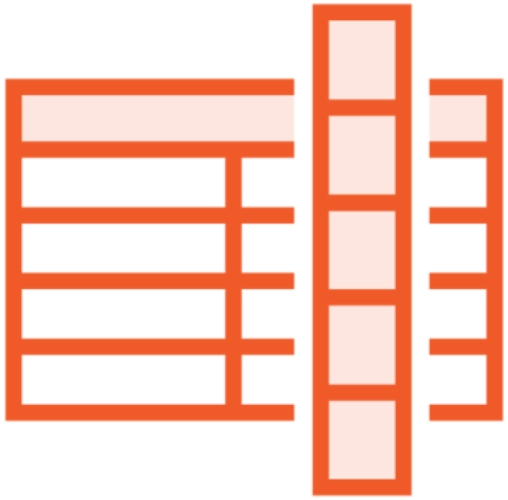
Data sources

Data destinations

Transformations

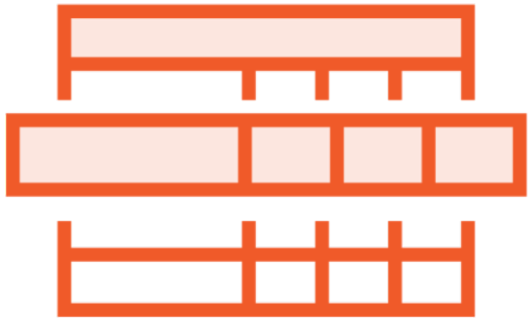
Security

General Data Flow



Reduce columns

Unneeded columns consume memory



Reduce rows

Unneeded rows can create backpressure

Use small data types

Data type influences memory allocation

Data Sources

OLE DB

Use SQL Command data access mode

Avoid SELECT * in source query

Flat files

Set data types for each column

Use FastParse on numeric and date columns if possible

Data Destinations

OLE DB

Use fast load option for SQL Server

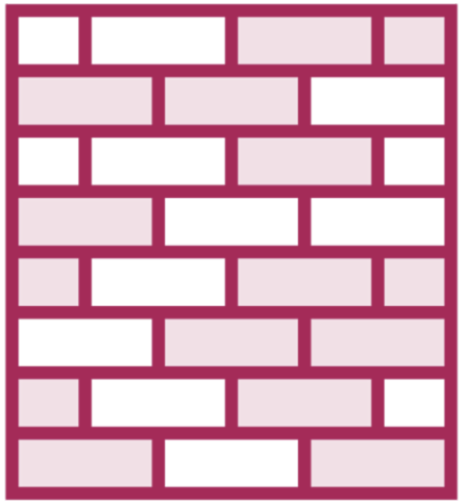
SQL Server

Avoid if data flow can lag or server is remote

Indexes

Drop and recreate to avoid fragmentation

Transformations

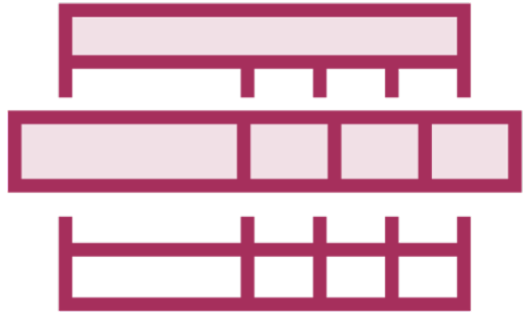


Blocking

All rows must be read into memory to continue

Semi-blocking

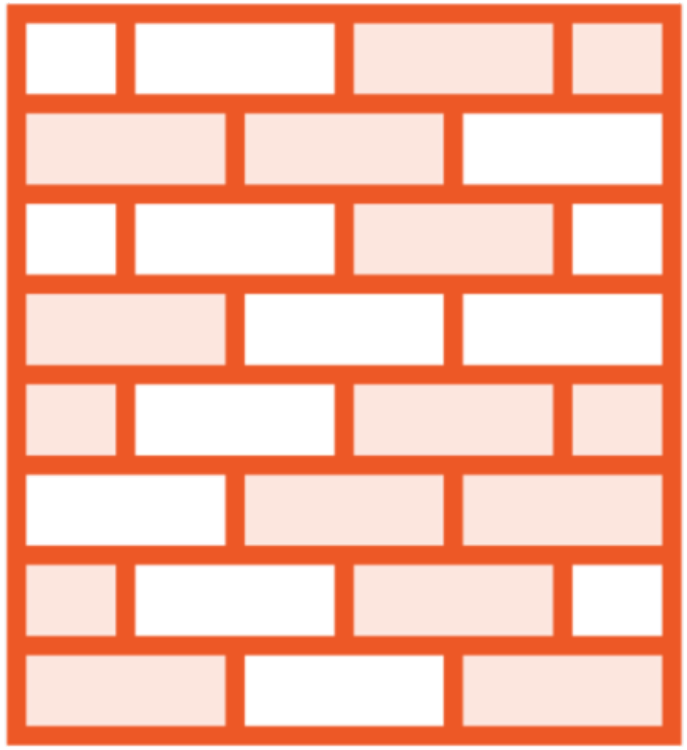
Row sampling or data structure change requires new buffers



Row-based

External resource called to process one row at a time

Blocking Transformations

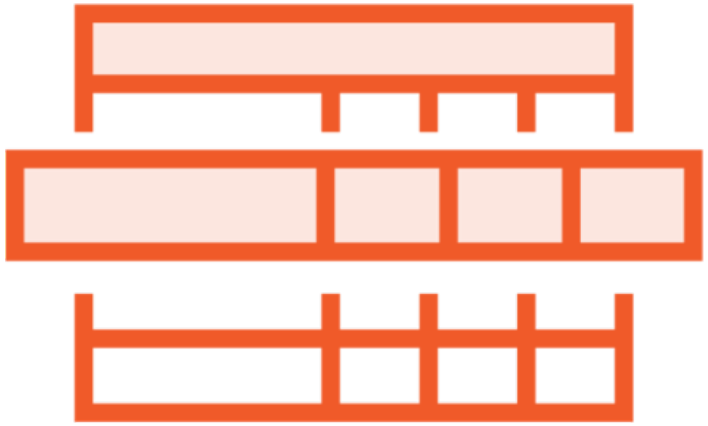


- Aggregate
- Fuzzy Grouping
- Fuzzy Lookup
- Row Sampling
- Sort
- Term Extraction

Semi-blocking Transformations

- Merge
- Merge join
- Pivot
- Term Lookup
- Unpivot
- Union All

Row-based Transformations



- Export Column
- Import Column
- Lookup (no cache)
- OLE DB Command
- Slowly Changing Dimension Wizard

Lookup Transformation

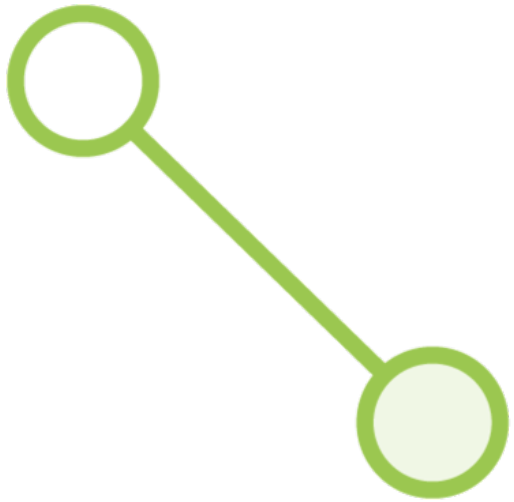
Use partial or full cache to load reference data into memory

Filter the reference data to load only the needed data and use a SQL query to limit the number of columns in the cache

For non-cached lookups, add an index to the reference table with the lookup key as first column and add columns to return to lookup

Security

Packages



Connection managers

Data sources and destinations

Package Security

DontSaveSensitive

EncryptSensitive WithPassword

EncryptSensitive WithUserKey

EncryptAll WithPassword

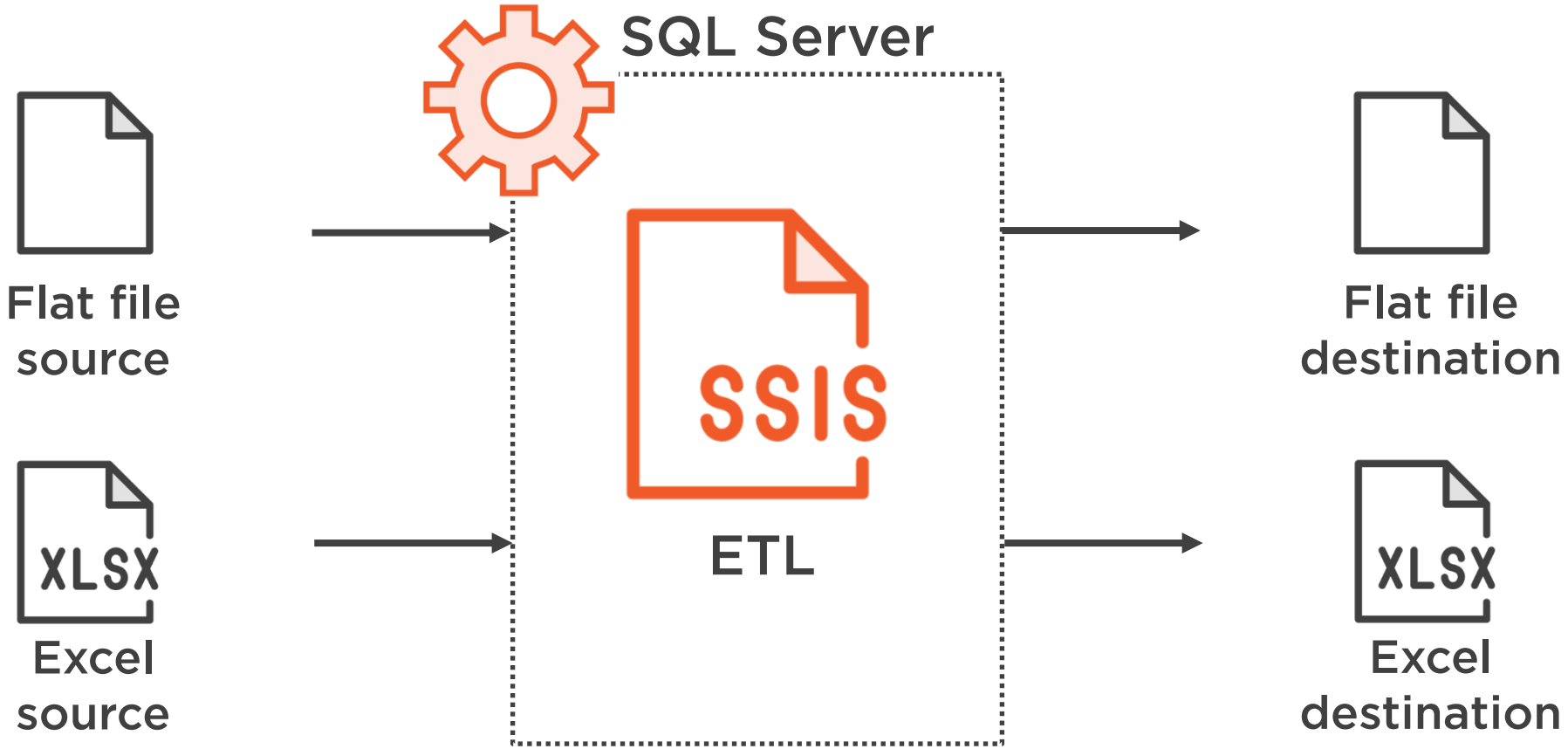
EncryptAll WithUserKey

Connection Manager Security

Windows authentication

Sensitive parameters

Data Sources and Destinations



Summary

- General data flow
- Data sources
- Data destinations
- Transformations
- Security