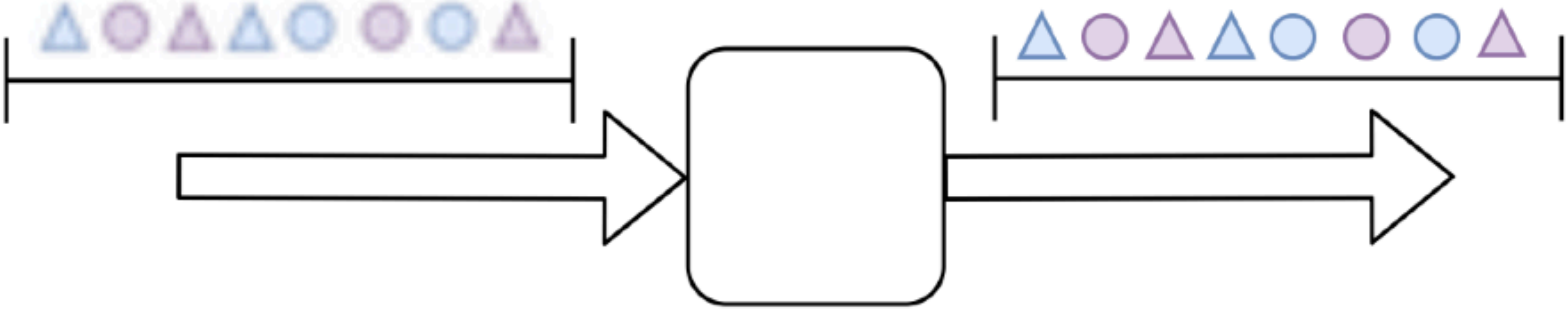
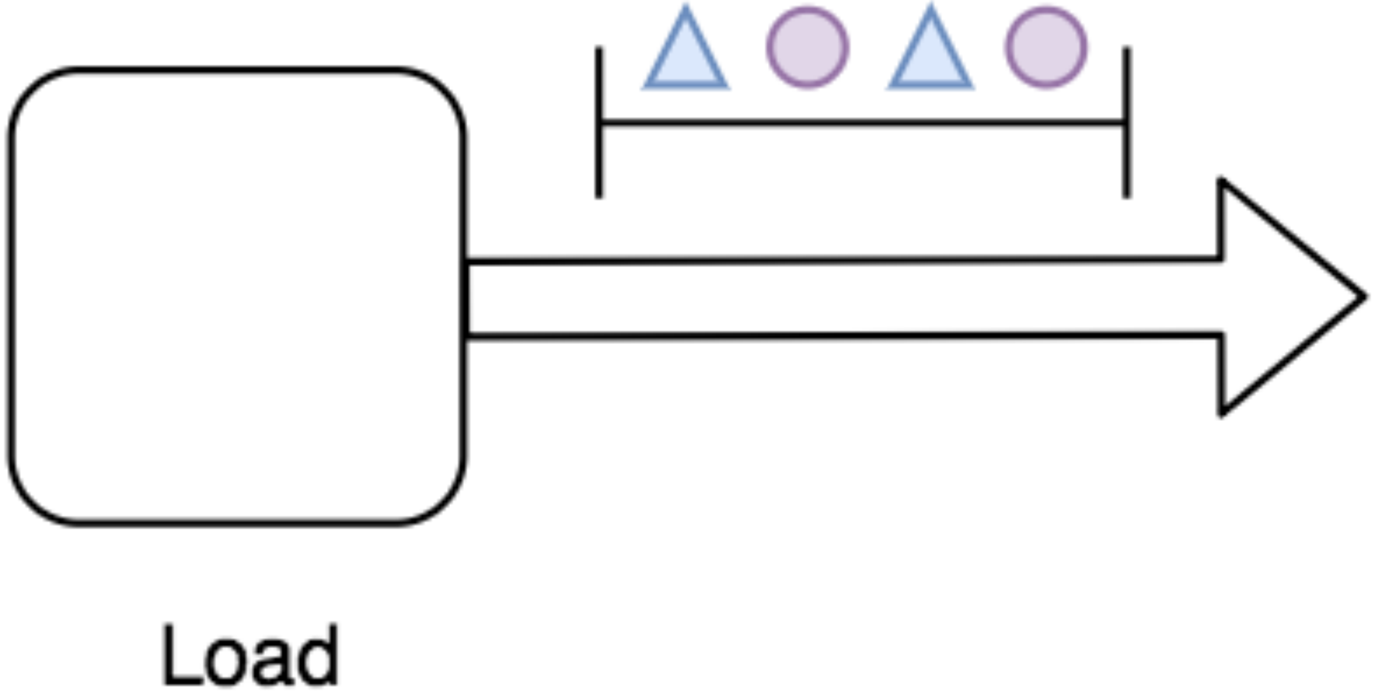


Set Types


Load

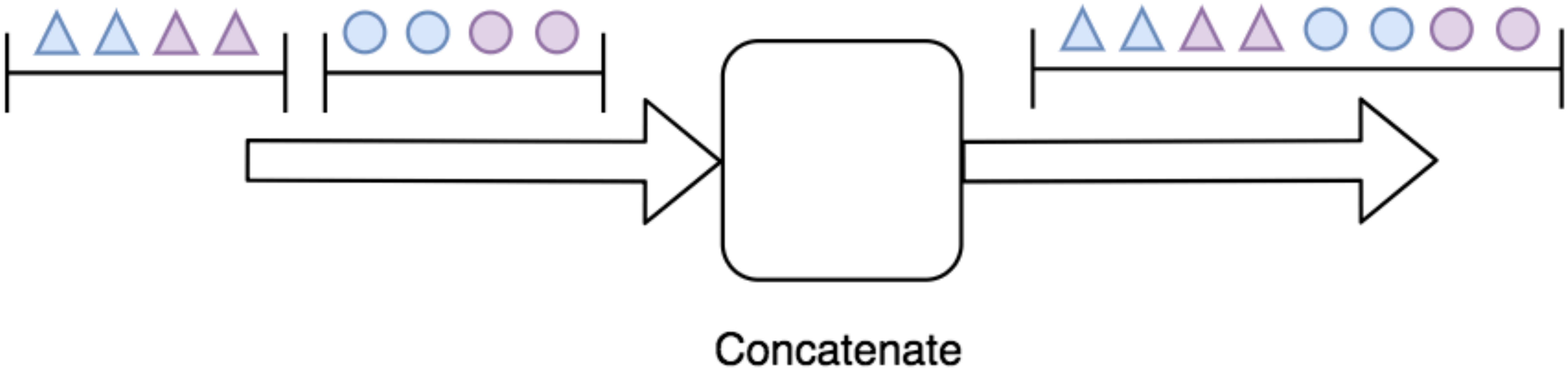
It's better to work with native types - dates, numbers, booleans - than with strings.
Set constraints on the data to automatically validate it -
e.g. min value, max length, pattern, not null etc. etc.

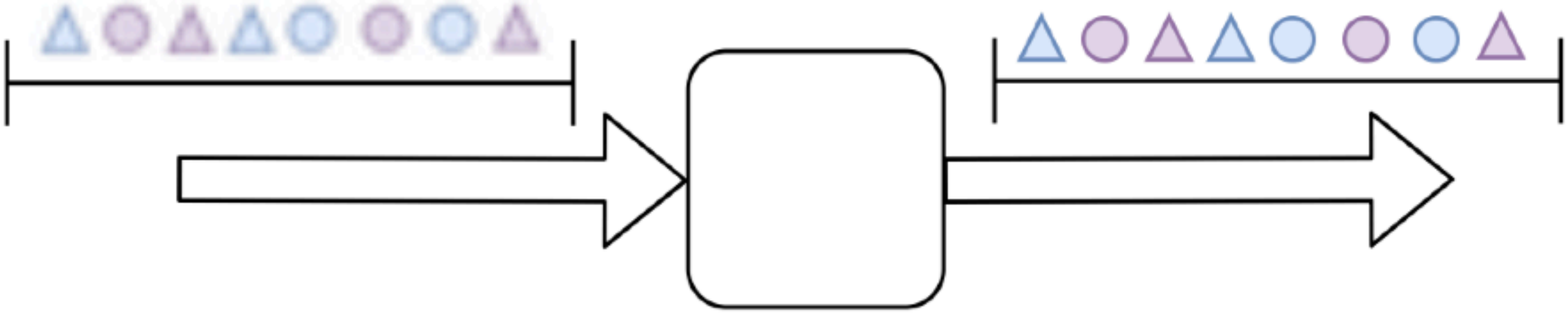
Can load from local file, remote URL or database

Standard Building Blocks

Each pipeline step is handled by a processor.

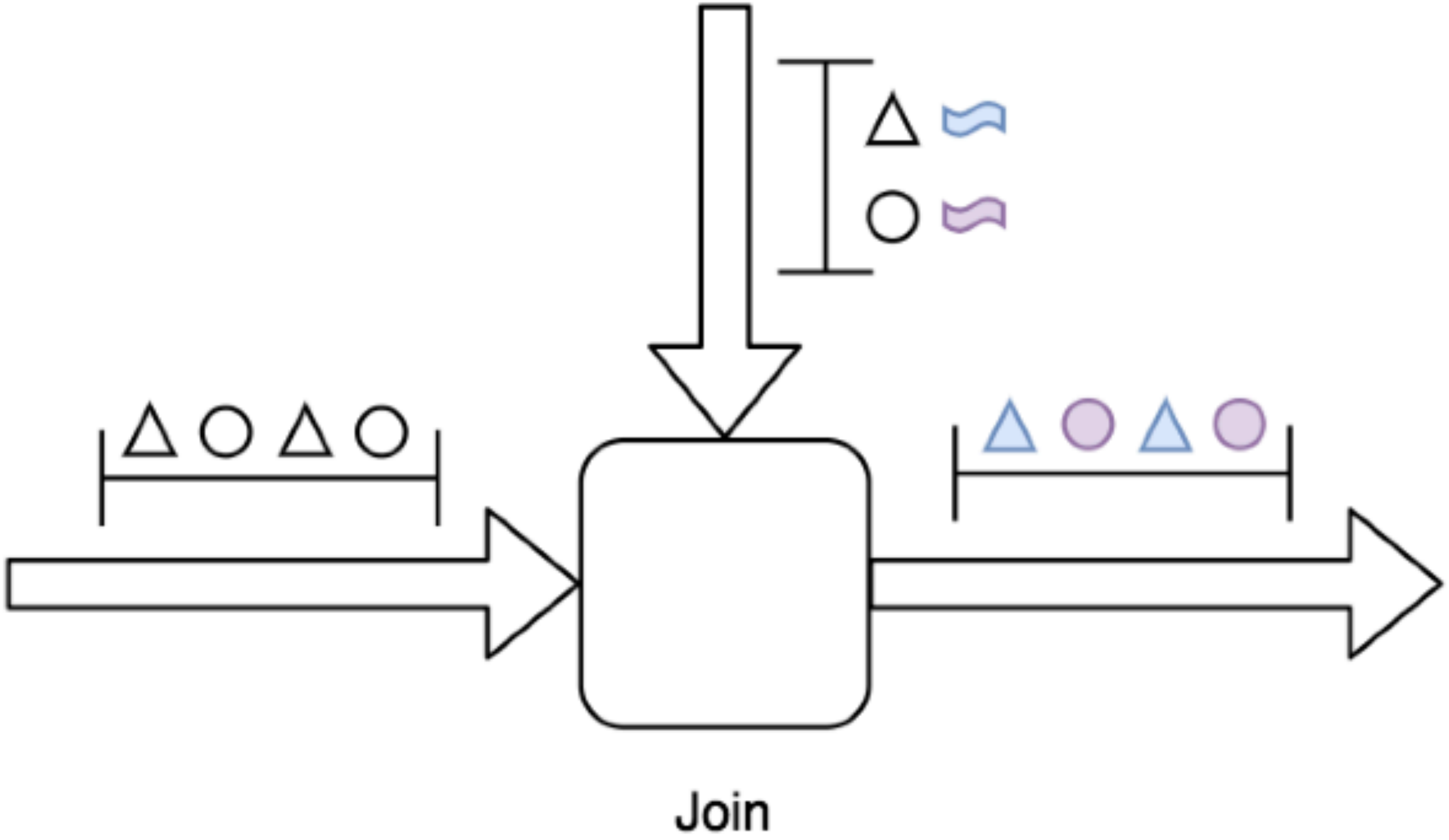
Most common tasks can be done by a **built-in processor**.


Concatenate

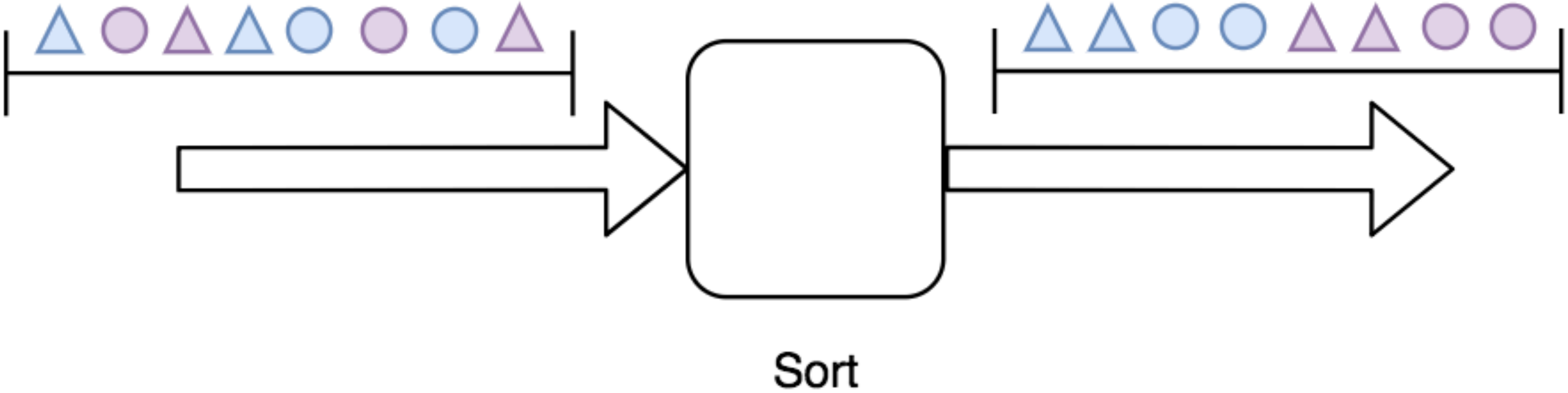


Set Types

e.g. combine per-year data files into one big stream

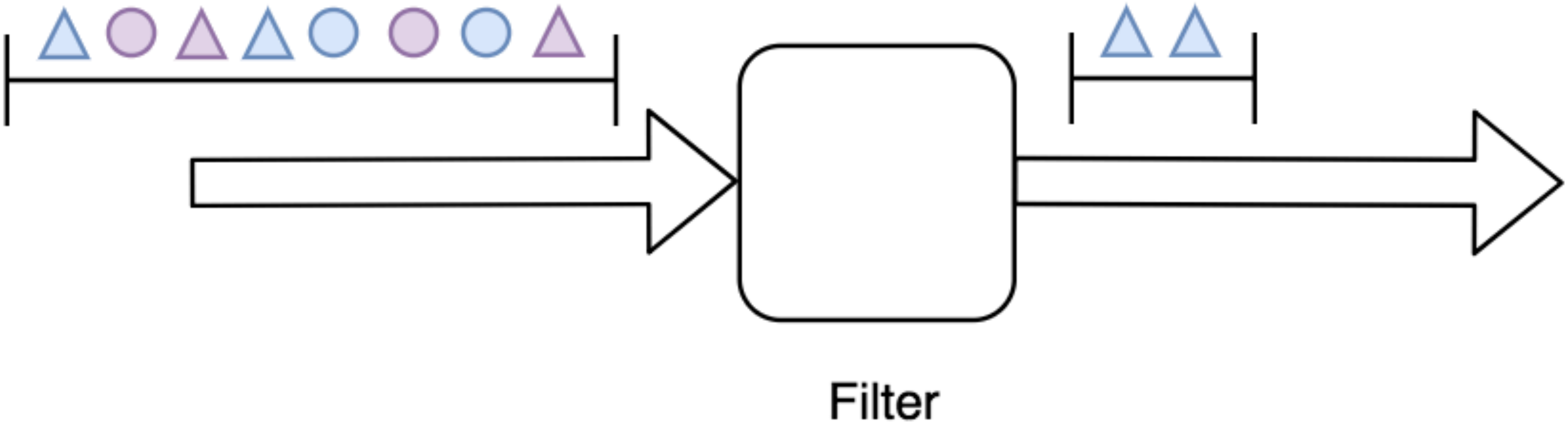

Join

Use data from one stream to augment another stream

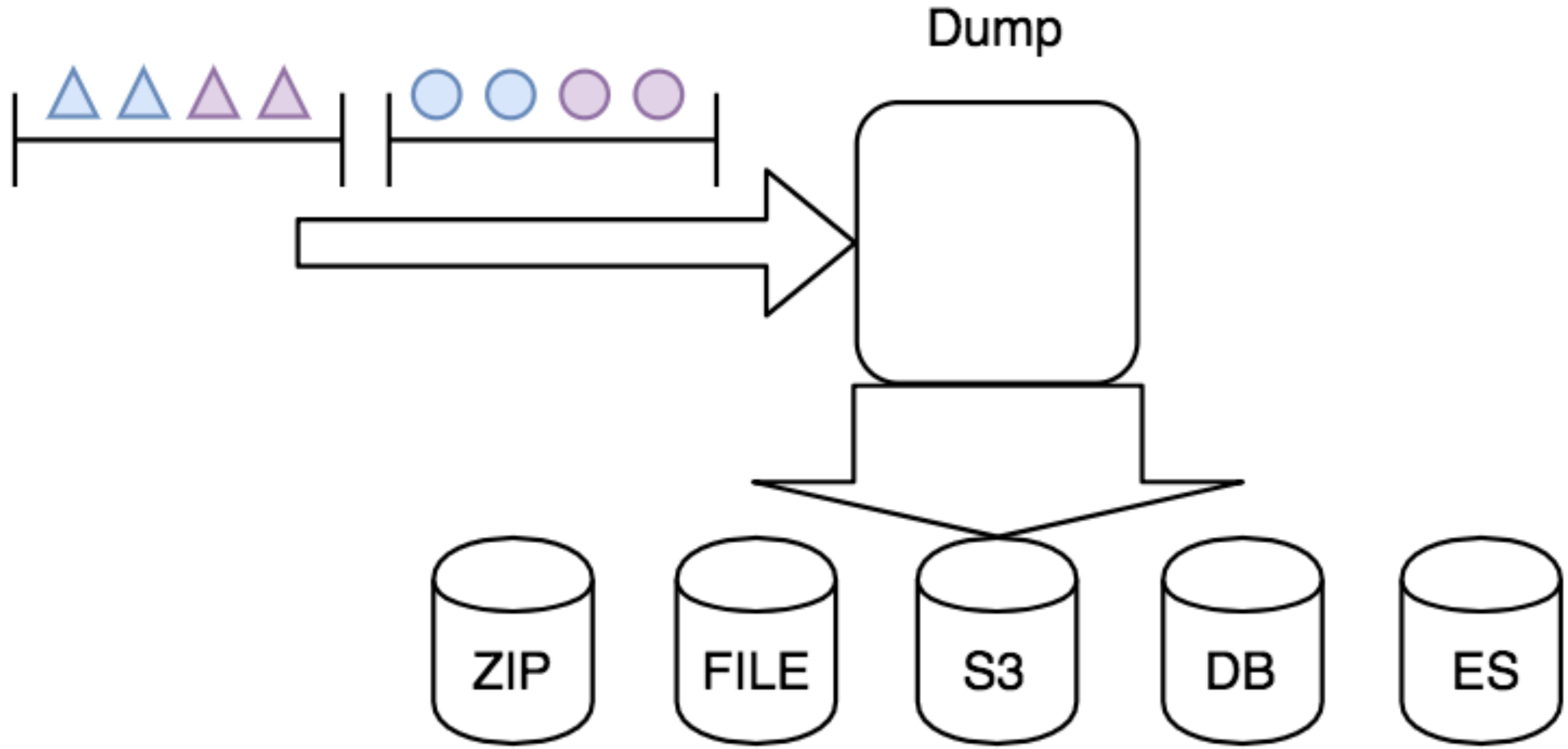

Sort

Well, sort.

Needs to temporarily store the data on disk (as you can't sort a stream)


Filter

Choose some rows


Dump
ZIP
FILE
S3
DB
ES

Can dump to local file, S3, DB table, Elasticsearch...

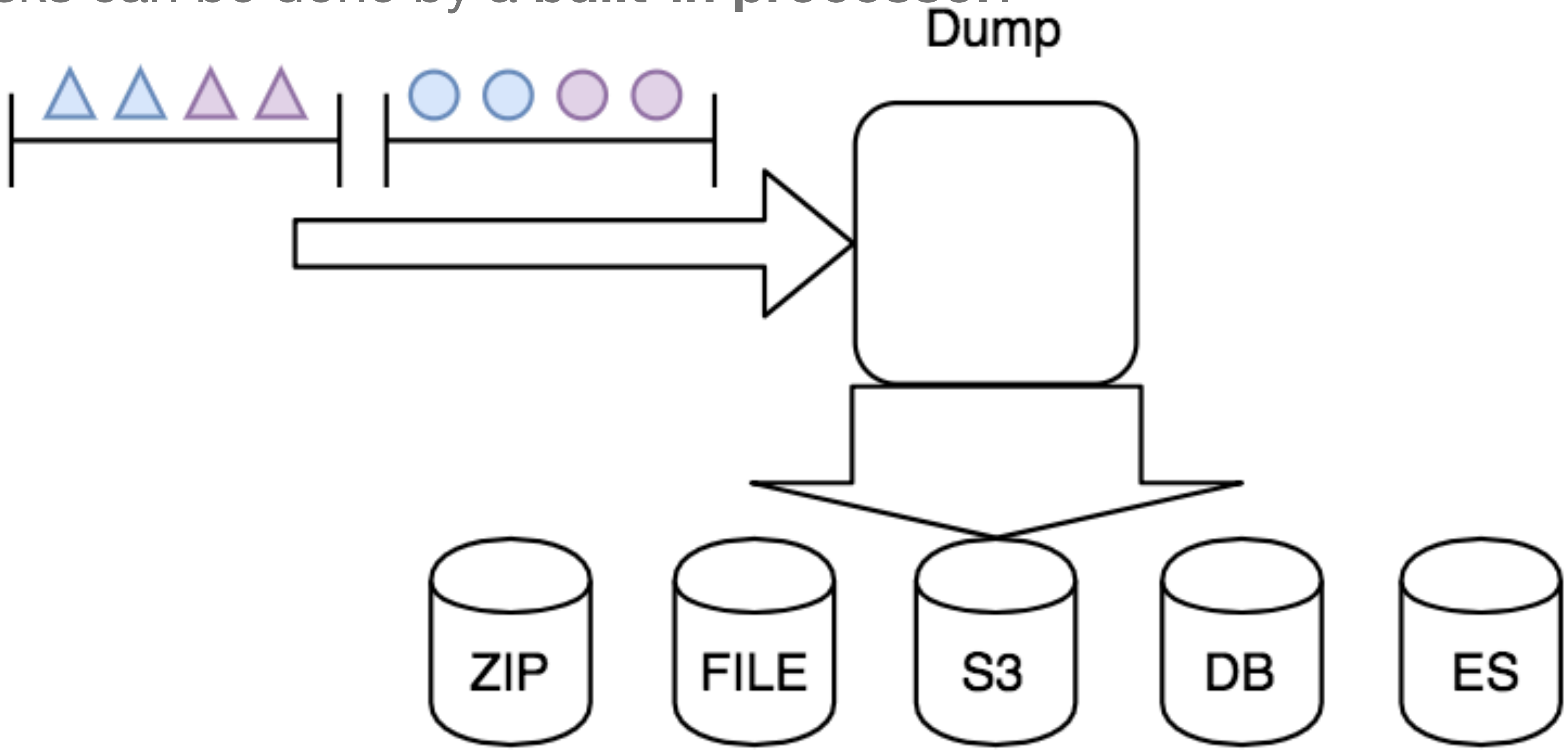
… And many more

- unpivot
- add computed field
- find-replace
- duplicate stream
- ...

Standard
Building Blocks

Each pipeline step is handled by a processor.

Most common tasks can be done by a **built-in processor**.



Can dump to local file, S3, DB table, Elasticsearch...