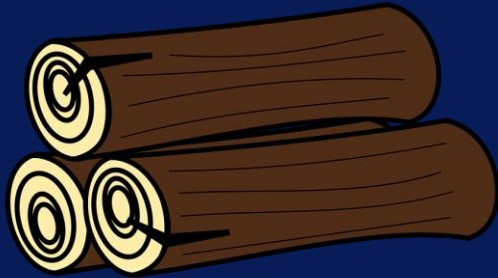


Data Transformations



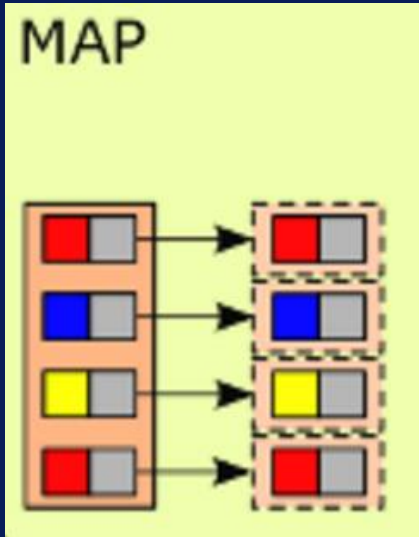
After this video you will be able to..

- List common data transformations within big data pipelines
- Design a conceptual data processing pipeline using the basic data transformations

Transformation are Tools to shape your data



Map

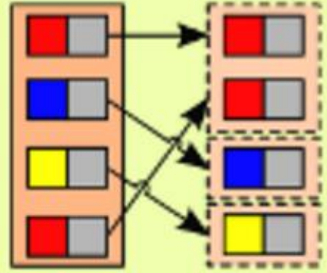


Apply same operation to each member of a collection

- Color **each** member of a set
- Discount **each** product's price by 5%
- Apply formatting to **each** document in a folder

Reduce

REDUCE



‘Collecting’ things that have same ‘key’

Key: Colors

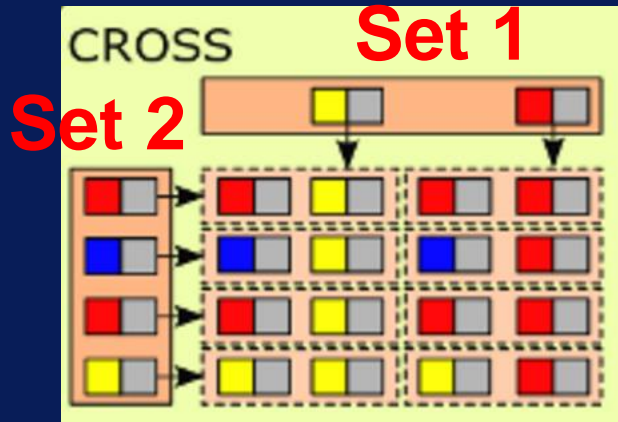
- Collect blocks as per their colors

Word Count Example

Key: Words

- Sum frequency counts of words

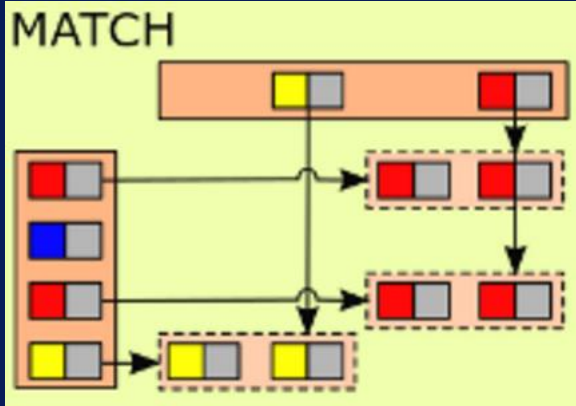
Cross/ Cartesian



→ Multiplication

Do some process to
each pair from two
sets

Match/Join

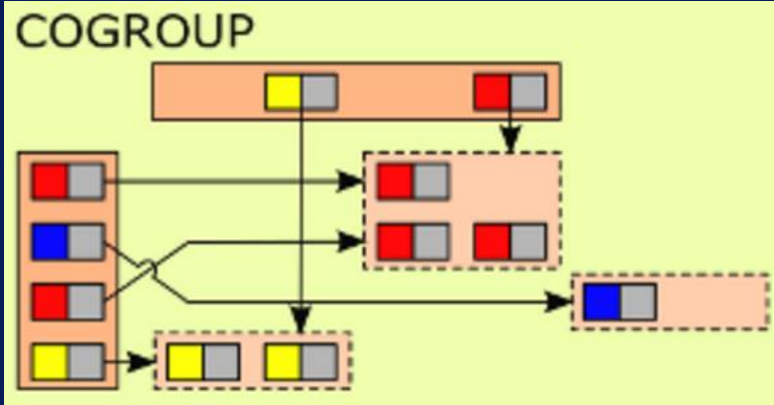


→ Selective Multiplication

Do some process to each pair from two sets – **which have same 'key'**

Co-Group

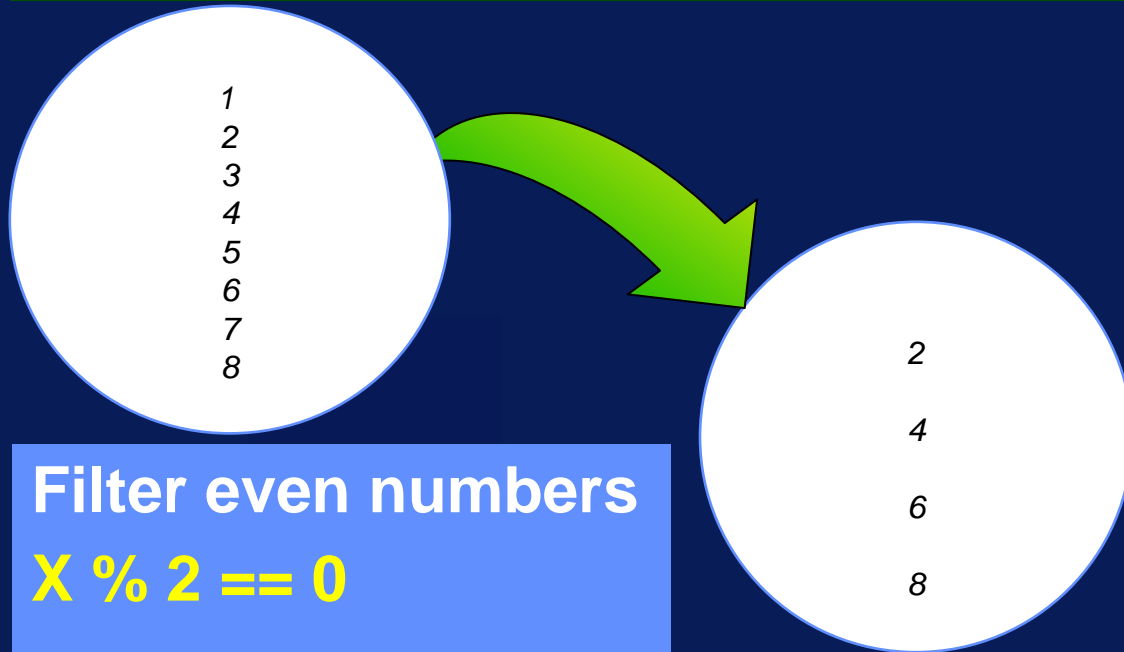
→ Group common items



- Collect similar things first
- **Apply a process to each collection**

Filter

Select elements that match a criteria



Basic Transformations → Get Results

