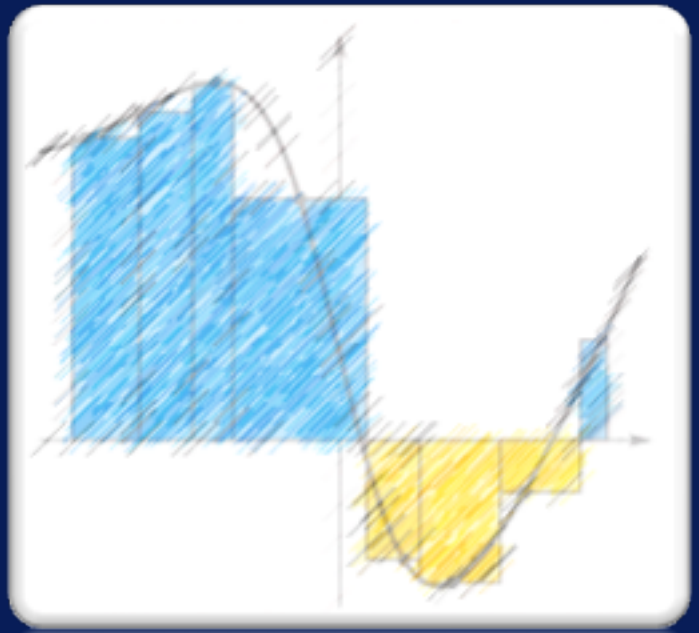
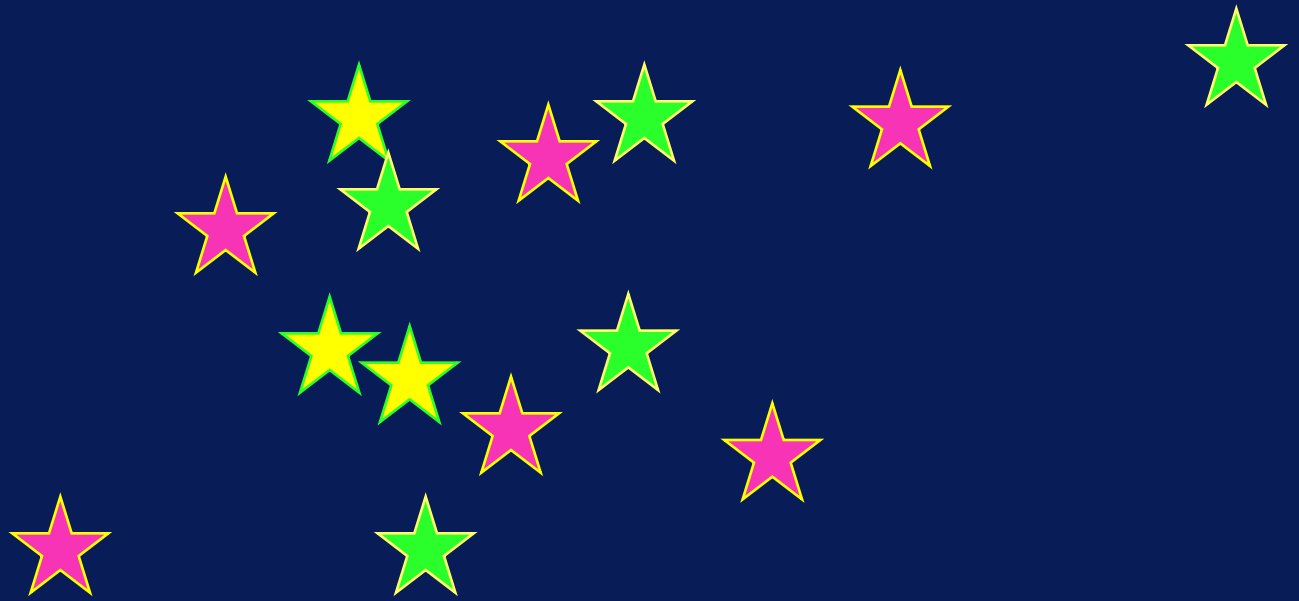
Aggregations in Big Data Pipelines



After this video you will be able to..

- Compare and select the Aggregation operation that you require to solve your problem
- Explain how you can use Aggregations to compact your dataset and reduce volume (in many cases)
- Design complex operations in your pipeline using a series of Aggregations

What is Aggregation ?

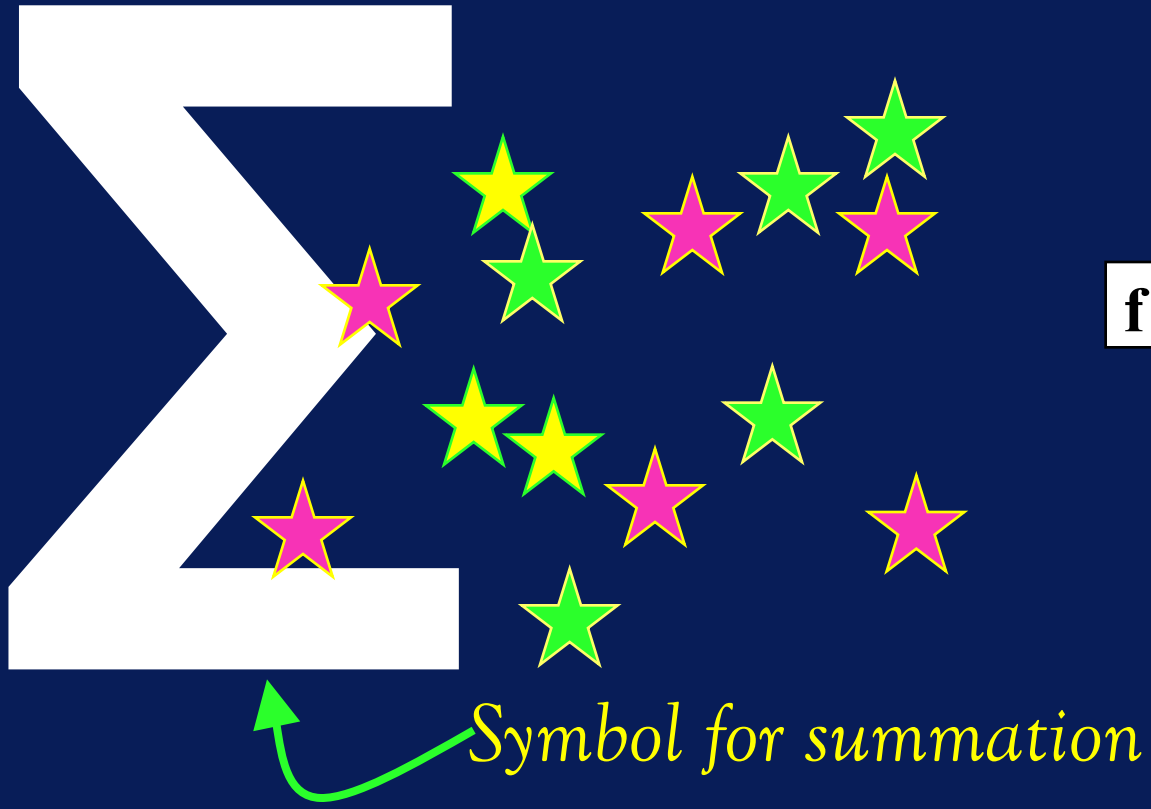


Symbol for any transformation

f

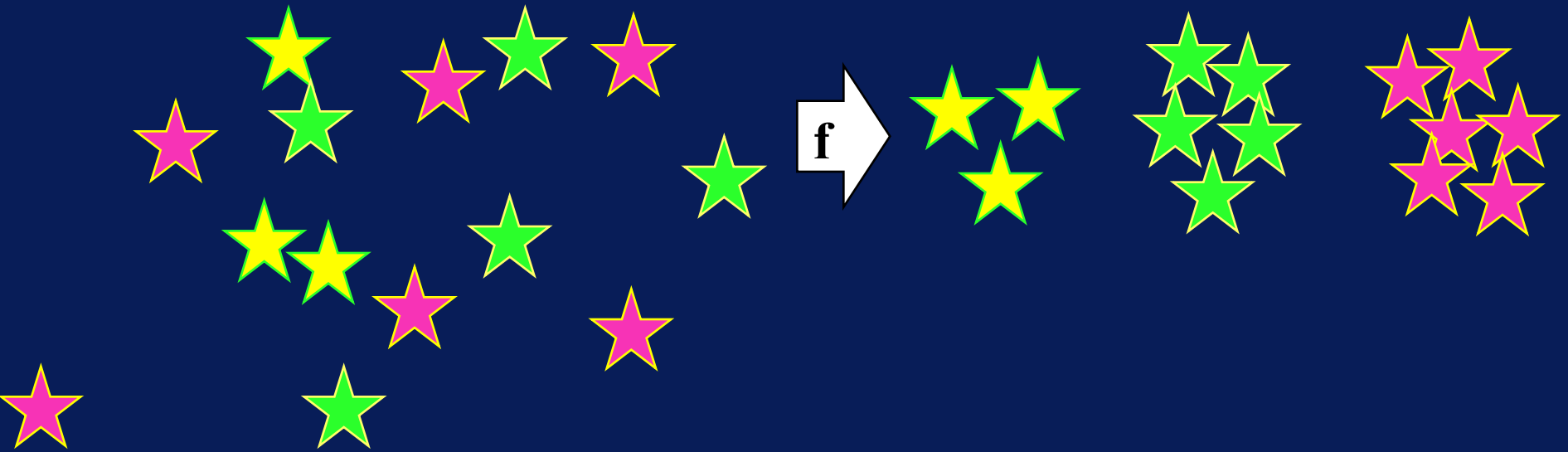
Aggregation → f (all elements)

Aggregation → f (all elements)



GROUP BY

AVERAGE # PER COLOR



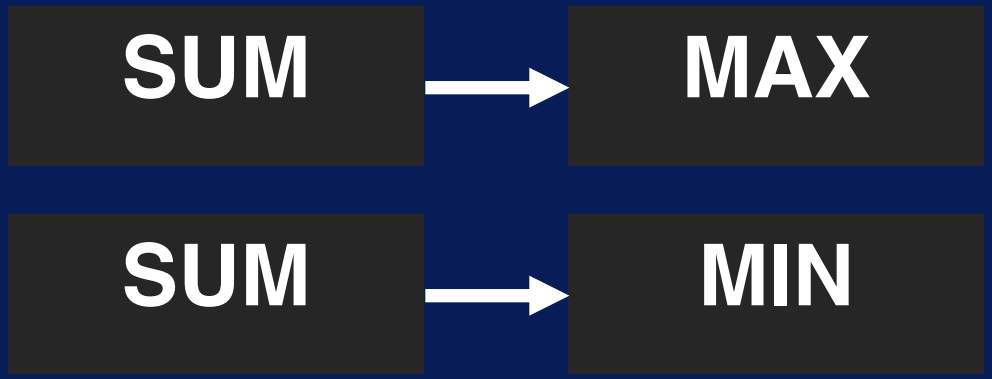
OTHER TRANSFORMATIONS

MAX

MIN

STANDARD DEVIATION

Connecting Aggregations



BOOLEAN AGGREGATION

AND 10110100110101101011011010111010  0

OR 10110100110101101011011010111010  1

SETS

STRINGS

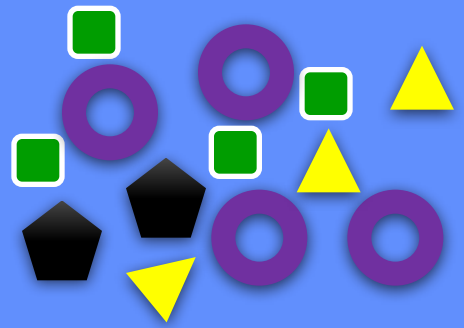
UNION

CONCATENATION

INTERSECTION

DIFFERENCE

Aggregations → Organized & Compact Data



f

AGGREGATED OUTPUT

Variety & Volume

Actionable Insights