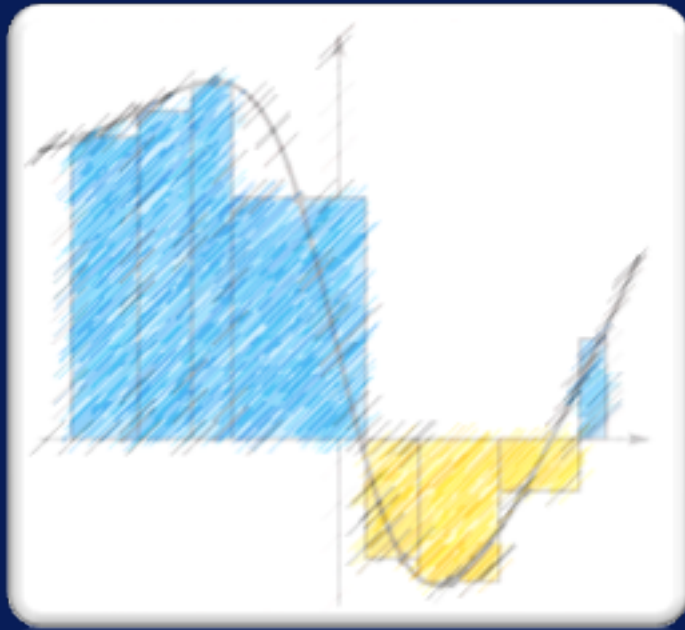


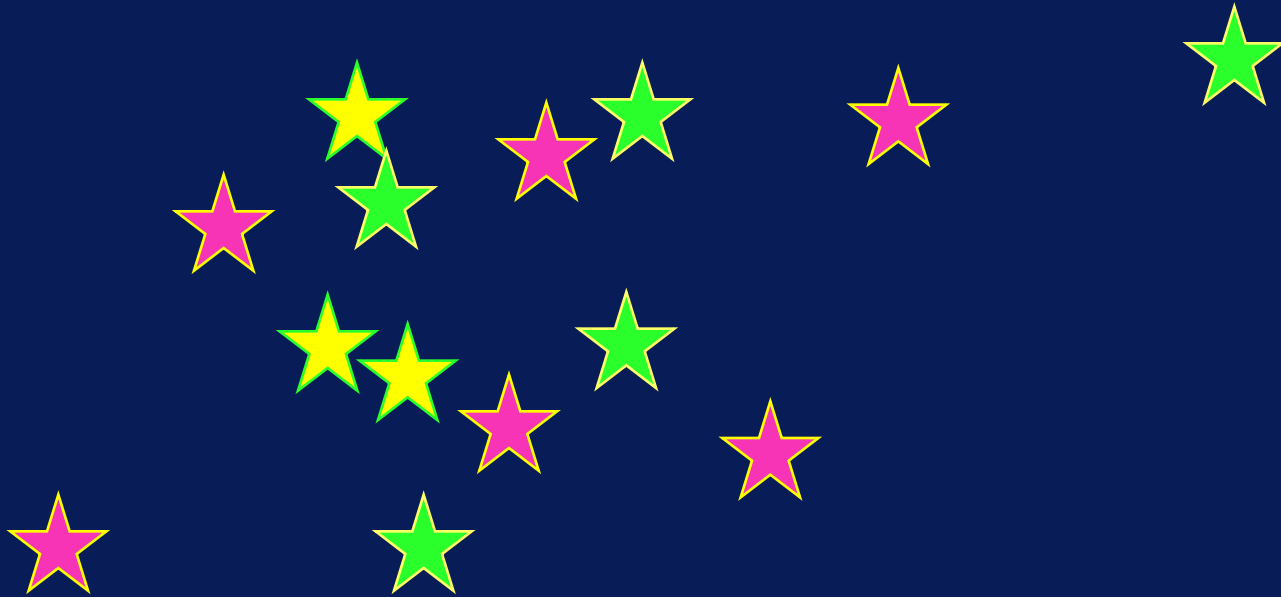
Aggregations in Big Data Pipelines



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

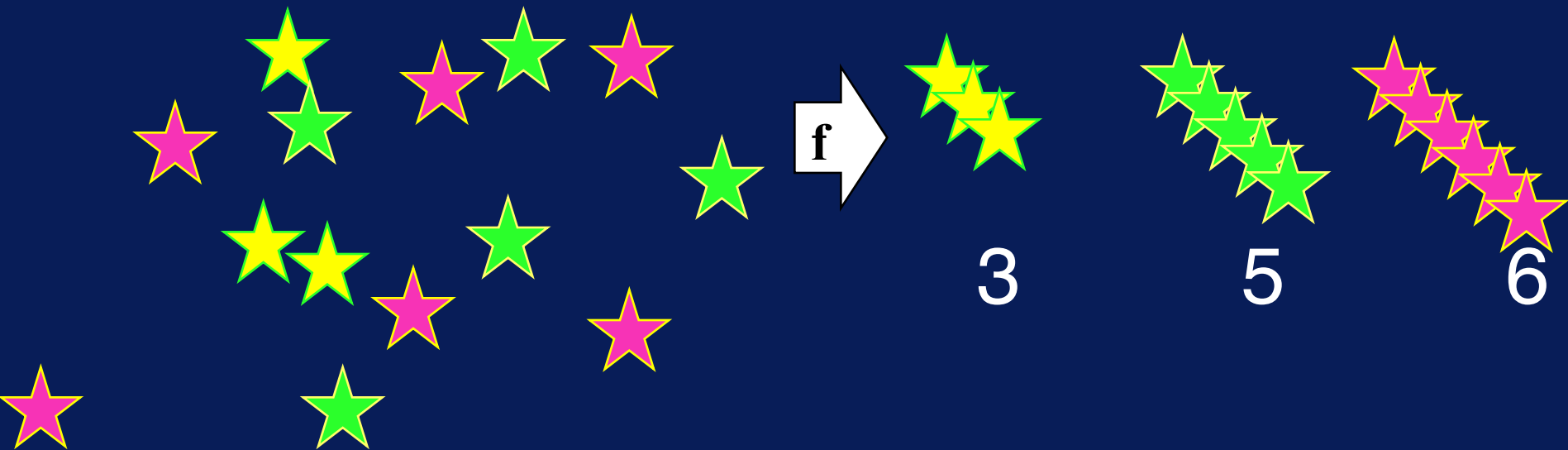
Aggregation $\rightarrow f(\text{all elements})$



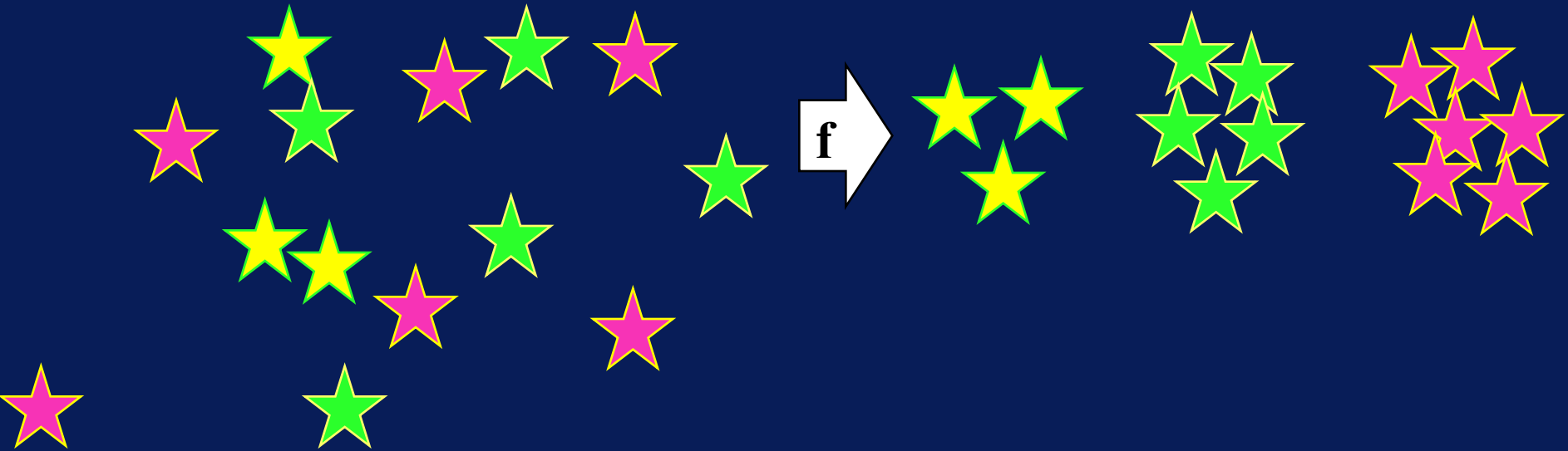
Aggregation $\rightarrow f(\text{all elements})$



GROUP BY



AVERAGE # PER COLOR



OTHER TRANSFORMATIONS

MAX

MIN

STANDARD DEVIATION

Connecting Aggregations

SUM



MAX

→ MAX(SUM)

SUM



MIN

→ MIN(SUM)

BOOLEAN AGGREGATION

AND

1011010011010110101101101101011101010



0

OR

1011010011010110101101101101011101010



1

SETS

STRINGS

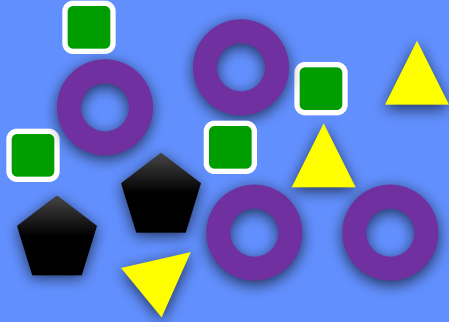
UNION

CONCATENATION

INTERSECTION

DIFFERENCE

Aggregations → Organized & Compact Data



f

**AGGREGATED
OUTPUT**

Variety & Volume

Actionable Insights