

## Tutorial 3 d – Flatten and Partition Transform in Apache Beam:

### Partition:

- **Partition** is a Beam transform for **PCollection** objects that store the same data type. It splits a single **PCollection** into a fixed number of smaller collections.
- **Partition** divides the elements of a **PCollection** according to a partitioning function that you provide.
- The partitioning function contains the logic that determines how to split up the elements of the input **PCollection** into each resulting partition **PCollection**.
- The number of partitions must be determined at graph construction time.
- **Partition** accepts a function that receives the number of partitions, and returns the index of the desired partition for the element. The number of partitions passed must be a positive integer, and it must return an integer in the range 0 to num\_partitions-1.

### Flatten:

- **Flatten** is a Beam transform for **PCollection** objects that store the same data type.
- Merges multiple **PCollection** objects into a single logical **PCollection**.

### Resources:

- <https://beam.apache.org/documentation/programming-guide/#flatten>
- <https://beam.apache.org/documentation/programming-guide/#partition>
  - <https://beam.apache.org/documentation/transforms/python/elementwise/partition/>

Cloud & AI Analytics