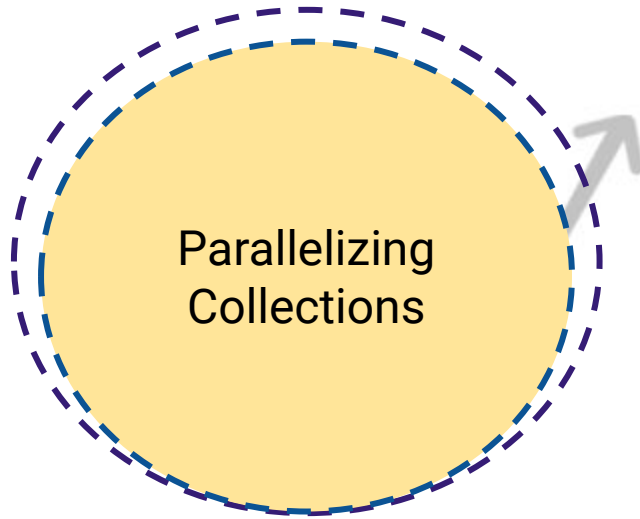


How to Create RDDs?



How to Create RDDs?



Analytics
Vidhya

How to Create RDDs?

- Parallelizing Collections

```
# Import SparkContext
from pyspark import SparkContext

# Create SparkContext object
sc = SparkContext()
```

Import and create
Spark Context

```
# Create a List
my_list = [1,2,3,4,5,6,7,8,9,10]
```

Collection

```
# Create an RDD
rdd = sc.parallelize(my_list, numSlices=4)
```

Number of Partitions

Parallelize function

How to Create RDDs?



How to Create RDDs?

- External Dataset

```
# Import SparkContext
from pyspark import SparkContext

# Create SparkContext object
sc = SparkContext()

# Read an external file
rdd_movies = sc.textFile("data/movies.csv", minPartitions=5)
```

TextFile function

File path

Number of Partitions

Practical !!

A large, faint, light gray watermark logo is centered in the background. It features a stylized line graph with an upward-pointing arrow on the right side, followed by the text "Analytics Vidhya" in a sans-serif font.