> *Resilient Distributed Datasets (RDDs) are a fundamental data structure in Apache Spark, a distributed computing framework designed for large-scale data processing and analysis.*

RDDs provide an immutable, fault-tolerant, and parallelized collection of data that can be processed in parallel across a cluster of machines.

Key characteristics of RDDs include:

1. Resilience

RDDs are resilient, meaning they can recover from node failures.

2. Distributed

RDDs are distributed across a cluster of machines in a parallel and fault-tolerant manner. Each partition of an RDD can be processed independently on different nodes, enabling parallel execution and efficient utilization of cluster resources.

3. Immutable

RDDs are immutable, meaning their content cannot be changed once created. Instead of modifying an RDD in place, transformations on RDDs create new RDDs.

4. Lazy Evaluation

RDDs support lazy evaluation, meaning transformations on RDDs are not executed immediately. Spark only computes the result when an action is called, allowing for optimizations and the avoidance of unnecessary computations.

5. Parallel Processing

RDDs allow for parallel processing of data by dividing it into partitions, each of which can be processed independently on different nodes of the cluster.

6. Type Information

RDDs can carry type information, allowing Spark to apply optimizations based on the data types involved.

7. Transformations and Actions

RDDs support two types of operations: transformations and actions. Transformations create a new RDD from an existing one (e.g., map, filter), while actions return a value to the driver program or write data to an external storage system (e.g., reduce, collect).