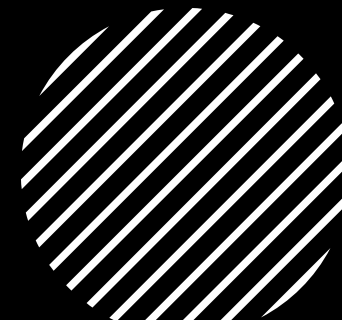


TRANSFORMATIONS



map ()

Applies a function to each element of an RDD or DataFrame.



groupByKey()

Groups elements of an RDD or DataFrame based on a key.



join()

Performs an inner or outer join operation between two RDDs or DataFrames.



reduceByKey()

Performs a reduction operation on the values of each key in an RDD or DataFrame.

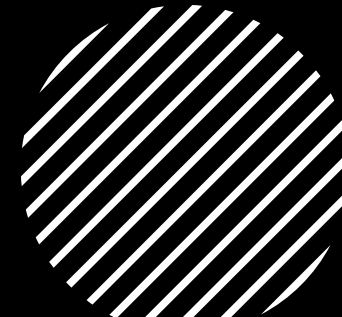


sortBy()

Sorts the elements of an RDD or DataFrame based on a given key.



ACTIONS



collect()

Retrieves all elements from an RDD or DataFrame to the driver program



count()

Returns the number of elements in an RDD or DataFrame



take()

Returns the first n elements from an RDD or DataFrame



display()

Returns the few element from an RDD or DataFrame



foreach()

Applies a function to each element of an RDD or DataFrame.



DATAFRAME AND SQL FUNCTIONS



select()

Selects specific columns from a DataFrame.



where()

Filters rows of a DataFrame based on a condition.



groupBy()

Groups rows of a DataFrame based on one or more columns.



agg()

Performs aggregation operations on a DataFrame.



join()

Performs joins between two DataFrames.



orderBy()

Orders the rows of a DataFrame based on one or more columns.



SPARK SQL FUNCTIONS



sql()

Executes SQL queries on registered tables or DataFrames.



createOrReplaceTempView()

Creates or replaces a temporary view from a DataFrame.