Explanation of all PySpark RDD, DataFrame and SQL examples present on this project are available at Apache PySpark Tutorial, All these examples are coded in Python language and tested in our development environment.

**Table of Contents (Spark Examples in Python)**

**PySpark Basic Examples**

* How to create SparkSession
* PySpark – Accumulator
* PySpark Repartition vs Coalesce
* PySpark Broadcast variables
* PySpark – repartition() vs coalesce()
* PySpark – Parallelize
* PySpark – RDD
* PySpark – Web/Application UI
* PySpark – SparkSession
* PySpark – Cluster Managers
* PySpark – Install on Windows
* PySpark – Modules & Packages
* PySpark – Advantages
* PySpark – Features
* PySpark – What is it? & Who uses it?

**PySpark DataFrame Examples**

* PySpark – Create a DataFrame
* PySpark – Create an empty DataFrame
* PySpark – Convert RDD to DataFrame
* PySpark – Convert DataFrame to Pandas
* PySpark – StructType & StructField
* PySpark Row using on DataFrame and RDD
* Select columns from PySpark DataFrame
* PySpark Collect() – Retrieve data from DataFrame
* PySpark withColumn to update or add a column
* PySpark using where filter function
* PySpark – Distinct to drop duplicate rows
* PySpark orderBy() and sort() explained
* PySpark Groupby Explained with Example
* PySpark Join Types Explained with Examples
* PySpark Union and UnionAll Explained
* PySpark UDF (User Defined Function)
* PySpark flatMap() Transformation
* PySpark map Transformation

**PySpark SQL Functions**

* PySpark Aggregate Functions with Examples
* PySpark Window Functions

**PySpark Datasources**

* PySpark Read CSV file into DataFrame
* PySpark read and write Parquet File