
pyspark Documentation

Release master

Author

Apr 17, 2023

Contents

1	pyspark package	3
1.1	Subpackages	3
1.2	Contents	4
2	Core classes:	5
3	Indices and tables	7

Contents:

1.1 Subpackages

1.1.1 pyspark.sql module

Module Context

pyspark.sql.types module

pyspark.sql.functions module

1.1.2 pyspark.streaming module

Module contents

pyspark.streaming.kafka module

1.1.3 pyspark.ml package

ML Pipeline APIs

pyspark.ml.param module

pyspark.ml.feature module

pyspark.ml.classification module

pyspark.ml.clustering module

pyspark.ml.recommendation module

pyspark.ml.regression module

pyspark.ml.tuning module

pyspark.ml.evaluation module

1.1.4 pyspark.mllib package

pyspark.mllib.classification module

pyspark.mllib.clustering module

pyspark.mllib.evaluation module

pyspark.mllib.feature module

pyspark.mllib.fpm module

pyspark.mllib.linalg module

pyspark.mllib.linalg.distributed module

pyspark.mllib.random module

pyspark.mllib.recommendation module

pyspark.mllib.regression module

pyspark.mllib.stat module

pyspark.mllib.tree module

pyspark.mllib.util module

1.2 Contents

CHAPTER 2

Core classes:

`pyspark.SparkContext`

Main entry point for Spark functionality.

`pyspark.RDD`

A Resilient Distributed Dataset (RDD), the basic abstraction in Spark.

`pyspark.sql.SQLContext`

Main entry point for DataFrame and SQL functionality.

`pyspark.sql.DataFrame`

A distributed collection of data grouped into named columns.

CHAPTER 3

Indices and tables

- search