



EBook Gratuito

APPENDIMENTO

spark-dataframe

Free unaffiliated eBook created from
Stack Overflow contributors.

#spark-
dataframe

Sommario

Di.....	1
Capitolo 1: Iniziare con spark-dataframe.....	2
Osservazioni.....	2
Examples.....	2
Installazione o configurazione.....	2
Caricamento di dati in un DataFrame.....	2
Titoli di coda.....	4

You can share this PDF with anyone you feel could benefit from it, downloaded the latest version from: [spark-dataframe](#)

It is an unofficial and free spark-dataframe ebook created for educational purposes. All the content is extracted from [Stack Overflow Documentation](#), which is written by many hardworking individuals at Stack Overflow. It is neither affiliated with Stack Overflow nor official spark-dataframe.

The content is released under Creative Commons BY-SA, and the list of contributors to each chapter are provided in the credits section at the end of this book. Images may be copyright of their respective owners unless otherwise specified. All trademarks and registered trademarks are the property of their respective company owners.

Use the content presented in this book at your own risk; it is not guaranteed to be correct nor accurate, please send your feedback and corrections to info@zzzprojects.com

Capitolo 1: Iniziare con spark-dataframe

Osservazioni

Questa sezione fornisce una panoramica di ciò che scintilla dataframe è, e perché uno sviluppatore potrebbe voler usarlo.

Dovrebbe anche menzionare tutti i soggetti di grandi dimensioni all'interno di spark-dataframe e collegarsi agli argomenti correlati. Poiché la documentazione per spark-dataframe è nuova, potrebbe essere necessario creare versioni iniziali di tali argomenti correlati.

Examples

Installazione o configurazione

Istruzioni dettagliate su come installare o installare spark-dataframe.

Caricamento di dati in un DataFrame

In Spark (scala) possiamo ottenere i nostri dati in un DataFrame in diversi modi, ognuno per diversi casi d'uso.

Crea DataFrame da CSV

Il modo più semplice per caricare i dati in un DataFrame è caricarlo dal file CSV. Un esempio di questo (tratto dalla [documentazione ufficiale](#)) è:

```
import org.apache.spark.sql.SQLContext

val sqlContext = new SQLContext(sc)
val df = sqlContext.read
  .format("com.databricks.spark.csv")
  .option("header", "true") // Use first line of all files as header
  .option("inferSchema", "true") // Automatically infer data types
  .load("cars.csv")
```

Creare DataFrame da RDD implicitamente

Molto spesso nelle applicazioni spark abbiamo dati in un RDD, ma è necessario convertirli in un DataFrame. Il modo più semplice per farlo è utilizzare la funzione RDD `.toDF()`, che determinerà implicitamente i tipi di dati per il nostro DataFrame:

```
val data = List(
  ("John", "Smith", 30),
  ("Jane", "Doe", 25)
)

val rdd = sc.parallelize(data)
```

```
val df = rdd.toDF("firstname", "surname", "age")
```

Crea DataFrame da RDD in modo esplicito

In alcuni scenari l'utilizzo dell'approccio `.toDF()` non è l'idea migliore, poiché è necessario definire esplicitamente lo schema del nostro DataFrame. Questo può essere ottenuto utilizzando un `StructType` che contiene una matrice di `StructField`.

```
import org.apache.spark.sql.types._
import org.apache.spark.sql.Row

val data = List(
  Array("John", "Smith", 30),
  Array("Jane", "Doe", 25)
)

val rdd = sc.parallelize(data)

val schema = StructType(
  Array(
    StructField("firstname", StringType, true),
    StructField("surname", StringType, false),
    StructField("age", IntegerType, true)
  )
)

val rowRDD = rdd.map(arr => Row(arr :_*))

val df = sqlContext.createDataFrame(rowRDD, schema)
```

Leggi Iniziare con spark-dataframe online: <https://riptutorial.com/it/spark-dataframe/topic/8988/iniziare-con-spark-dataframe>

Titoli di coda

S. No	Capitoli	Contributors
1	Iniziare con spark-dataframe	Ben Horsburgh , Community