

ETL-1 to 1 relation

asd asd

ETL-map/reduce

```
udf @panda s_u df( 'long') def pandas _pl us_ one (se rries: pd.Series) -> pd.Series:
    # Simply plus one by using pandas Series.
    .s how()
```

ETL - N to 1

groupby df.groupby('color').avg()

ETL - streaming

I/O

Local_CSV dataset = spark.read.csv("BostonHousing.csv",inferSchema=True, header =True) inferSchema=- Guess data type from csv

Local_Json

Cloud_s3

To SQL df.createOrReplaceTempView("tableA")
table

Efficiency

repartition

shuffle

collect

Spark All kind of handler

SparkC- Old man
ontext

SparkS- Young boy, that's only entry point got to know for late
ession spark

SparkConf <https://towardsdatascience.com/sparksession-vs-sparkcontext-vs-sqlcontext-vs-hivecontext-741d50c9486a>

spark.sql spark.sql("SELECT * FROM p left join df.query() -> e on p.name = e.name") Dataframe

RDD

EDA-Get the information for debugging/coding

printSchema DataFrame.printSchema()

columns:List- DataFrame.columns
[str]

show() df.show(1) head() A action, force the process to finish

take df.take(1)



By ChesterHsieh

Not published yet.

Last updated 15th November, 2021.

Page 1 of 1.

Sponsored by **Readable.com**

Measure your website readability!

<https://readable.com>