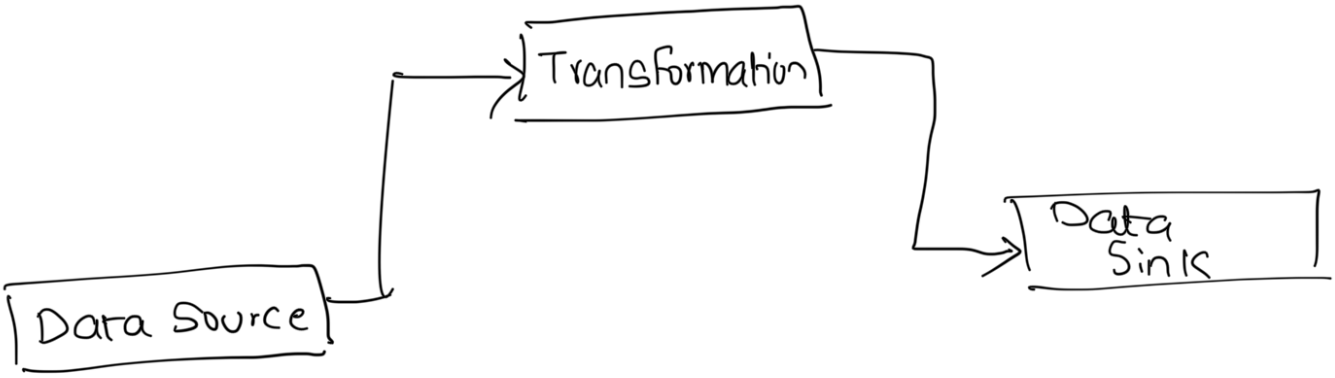
* Introduction to Data Transformation



⇒What do we mean by transformations?

- So, In spark we read the data and create one of the two things:

Data → DataFrames → Spark Program
OR
Data → Database Tables → Spark SQL

Both of these are same but two different interfaces.

The DataFrame is programmatic interface for your Data

The Database Table is SQL interface of your Data.

So, answering to the initial question

⇒What do we mean by transformations?

Transform DataFrames Using Join &

1] Combining DataFrames (using ... union)

2] Aggregating and Summarizing (using operations such as grouping, windowing, & rollups)

3] Applying functions and built-in transformatio (filtering, sorting, splitting, sampling & finding unique)

4] Use Built-in and column-level functions

5] Creating and using user defined functions

6] Referencing Rows / Columns

7] Creating Column Expression.

* Working with DataFrame Rows

We have three specific scenarios when we might have to work with the Row object:

1] Manually creating Rows and DataFrame

2] Collecting DataFrame rows to the Driver

3] Work with an individual row in

5] Work ...

Spark Transformations

[1] First work demo is in Databricks notebook

2] Second work demo is in local machine in Pycharm.

3] Third work demo is also in Pycharm]