

### 1- Reading Data

The readData function reads each input file as a DataFrame using the SparkSession. It configures options such as schema inference, header, quote character, and delimiter.

### 2- Data Transformation:

The discoverINDs function maps each input file path to a DataFrame using the readData function. It then performs a series of transformations on each DataFrame.

For each DataFrame, it retrieves the column names and uses them to iterate over the rows.

It flattens the rows by generating tuples containing the column name and corresponding value using a for-comprehension.

### 3- Union and Grouping:

The transformed data from different input files is combined into a single dataset using the union operation, resulting in the fullData dataset.

The fullData dataset is then grouped by the value (the second element of each tuple) using groupByKey.

### 4- IND Discovery:

The grouped dataset is further transformed to discover inclusion dependencies.

The flatMap operation is used to generate tuples containing the current attribute and a set of attributes that do not match the current attribute.

The dataset is then grouped by the attribute using groupByKey.

Finally, the sets of attributes for each attribute are reduced by taking the intersection using the reduce operation.

### 5- Sorting and Printing:

The discovered inclusion dependencies are collected and stored in the INDs array.

The sortedINDs array is created by sorting the INDs array based on the attribute.

Each tuple in the sortedINDs array is iterated over, and the attribute and corresponding values are extracted.

If the set of values is not empty, the attribute and values are printed.

