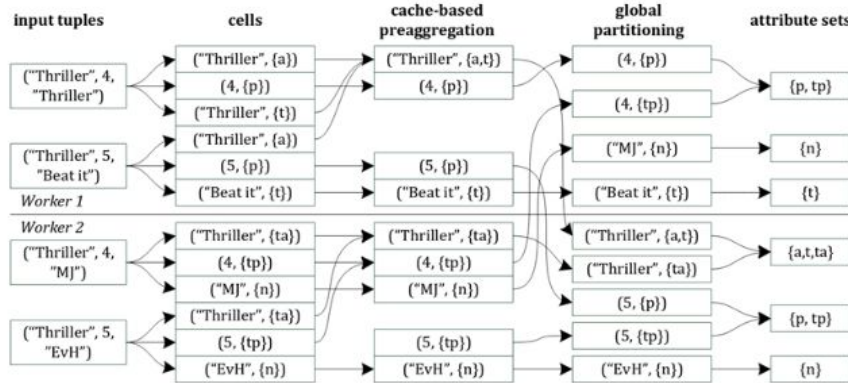We did →

Figure 1: Dataflow for a distributed full outer join on all columns of a dataset. The attribute names and some values are abbreviated for the purpose of lucidity.

Figure 2: Dataflow for deriving INDs from the outer join product.

Scaling Out the Discovery of Inclusion Dependencies
S. Kruse, T. Papenbrock, F. Naumann,
Proceedings of the conference on Database Systems for Business, Technology, and Web (BTW). pp. 445-454 (2015).

Figure 3: Stratosphere plan for unary IND detection.

Quelle:
https://hpi.de/fileadmin/user_upload/fachgebiete/naumann/lehre/SS2021/DDM/12_DDM_Hands-on_Spark_Batch_Processing.pdf

Done