Database A Database B Übertragung in Data pre Lesen der .csv-Data pre DataSet mit processing Dateien processing Person-Erstellung der n-Aufnahme der Indexing Gramme Tokens in Bloom (Tokens) Filter Finden von Berechnung der Kandidaten-Comparison LSH-Keys Paaren Entfernung von doppelten Kandidaten-Classification Paaren Berechnung der Ausgabe aller Ähnlichkeiten für Paare mit Evaluation alle Kandidatenhinreichendem Ähnlichkeitswert Paare