🙂 Inclusion Dependency Discovery using Akka

📖 System Overview

This system discovers inclusion dependencies (INDs) in databases using AKKA's actor model for distributed processing. The implementation focuses only on discovering unary INDs.

🔎 1: Data Reading

The system builds a lookup structure mapping string values to bit arrays, implemented using HashMap and BigInteger. Each column position is tracked using bit operations, allowing efficient storage and comparison.

Key features:
- Hash-based distribution of work
- Parallel processing capability
- Pre-sorting at input level

📊 2: Dependency Analysis

The comparison phase determines column relationships through bit array analysis. Workers process BigInteger value sets independently and report qualified INDs to the Miner.

Features:
- Highly parallelizable
- Efficient bit-level operations
- Reduced memory footprint
- Scalable worker distribution
- implemented shutdown sequence for proper shutdown