Architecture

Dependency Miner:

Worker:

- Stores tables batchwise that are sent by Input-reader

Master-Worker-Pattern:

- Generates tasks:
  - A task contains a number of columns of two tables
  - Number of columns in one message is limited by 80MB, to save storage. If one table is too big, the columns get split
  - All permutations of two tables and rows are created as tasks
- Taskmessages are sent to free workers over large message Proxy
- Results are processed to Resulthandler

- Checking for cardinality in columns **(Pruning)**
- IND-Algorithm is performed (N-Ary not yet implemented)
- Results are sent in Completionmessage to Dependency Miner

Datastructure

- Inputreader sends rows of the tables batchwise to the DependencyMiner

- The tables are stored columnwise in the DependencyMiner

- Optimized storage: Distinct values will be assigned an index and all values will be stored under those indices

Values stored in table

```
Dog
Horse
Cat
Cat
Horse
```

Distinct values

```
Dog
Horse
Cat
```

Indices

```
0
1
2
2
1
```