

# Architecture

More Information in ReadMe

## Dependency Miner:

- 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

## Worker:

- Creates unique values of columns
- Checking for cardinality of columns (**Pruning**)
- **IND-Algorithm:** Check if all values of table with smaller cardinality is in table with bigger value
- Results are sent in Completionmessage to Dependency Miner

# Datastructure

*More Information in ReadMe*

- Inputreader sends rows of the tables batchwise to the DependencyMiner
- The tables are stored as columns in the DependencyMiner
- Distinct values will be assigned an index and all values will be stored under those indices

Dog	0
Horse	1
Cat	2
Cat	2

- Columns are sent batchwise to the DependencyWorker