

Big Data Systems Project

DEPENDENCY INCLUSION DISCOVERY

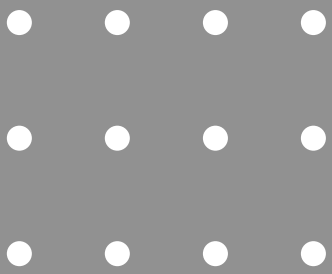
Work done by:

Chiheb Eddine Saidi

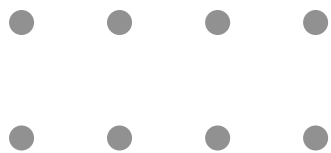
Shady Abdou

Jayesh

Data Profiling Actors



**Dependency
miner**



**Dependency
Worker**

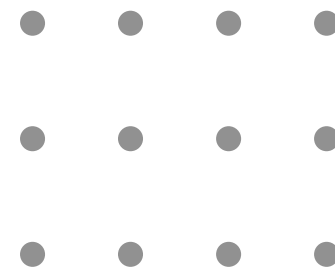
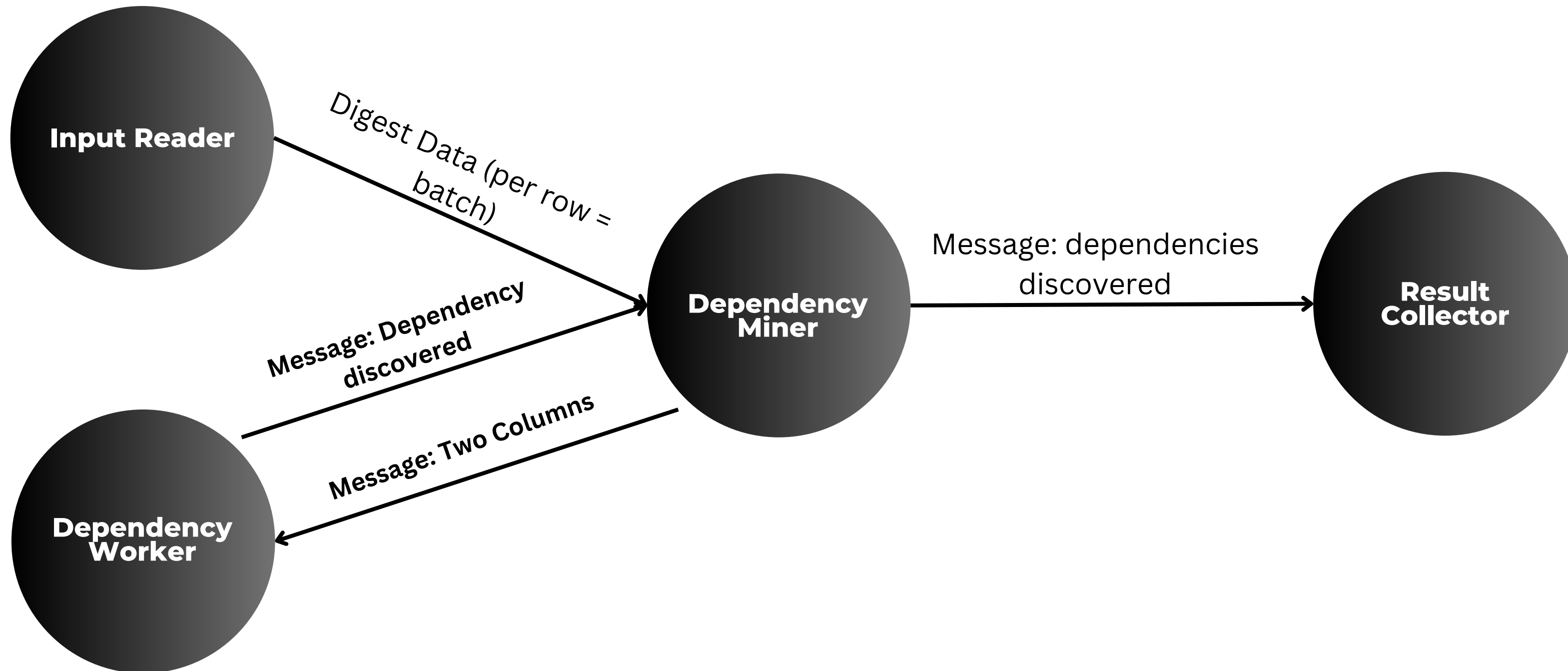


Input Reader

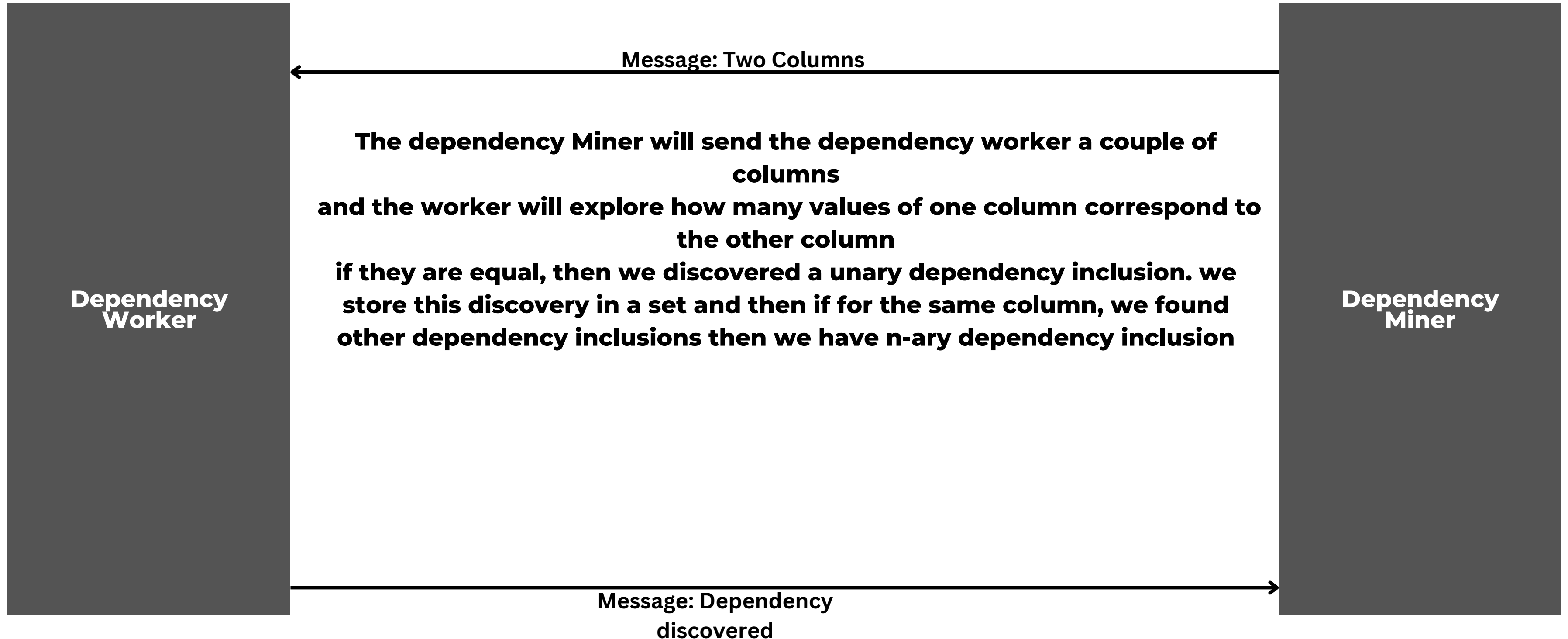


Result Collector

Messages Exchange



Dependency inclusion discovery approach



Dependency inclusion discovery approach



Input Reader

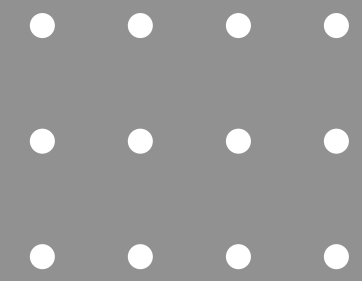
Reads Files Batchwise



Dependency
Miner

- **The Input Reader** takes on the initial role in the workflow, managing the intake of data by reading files in batches. This approach ensures the efficient processing of large datasets, handling them in manageable chunks to prevent overwhelming the system.

Dependency inclusion discovery approach



Dependency
Miner

Send all Inclusion Dependencies

Result
collector

- The last phase of the workflow is the **Result Collector**. Gathering the processed data from the Dependency Miner is its main duty. Next, the results are saved to the result.txt file.