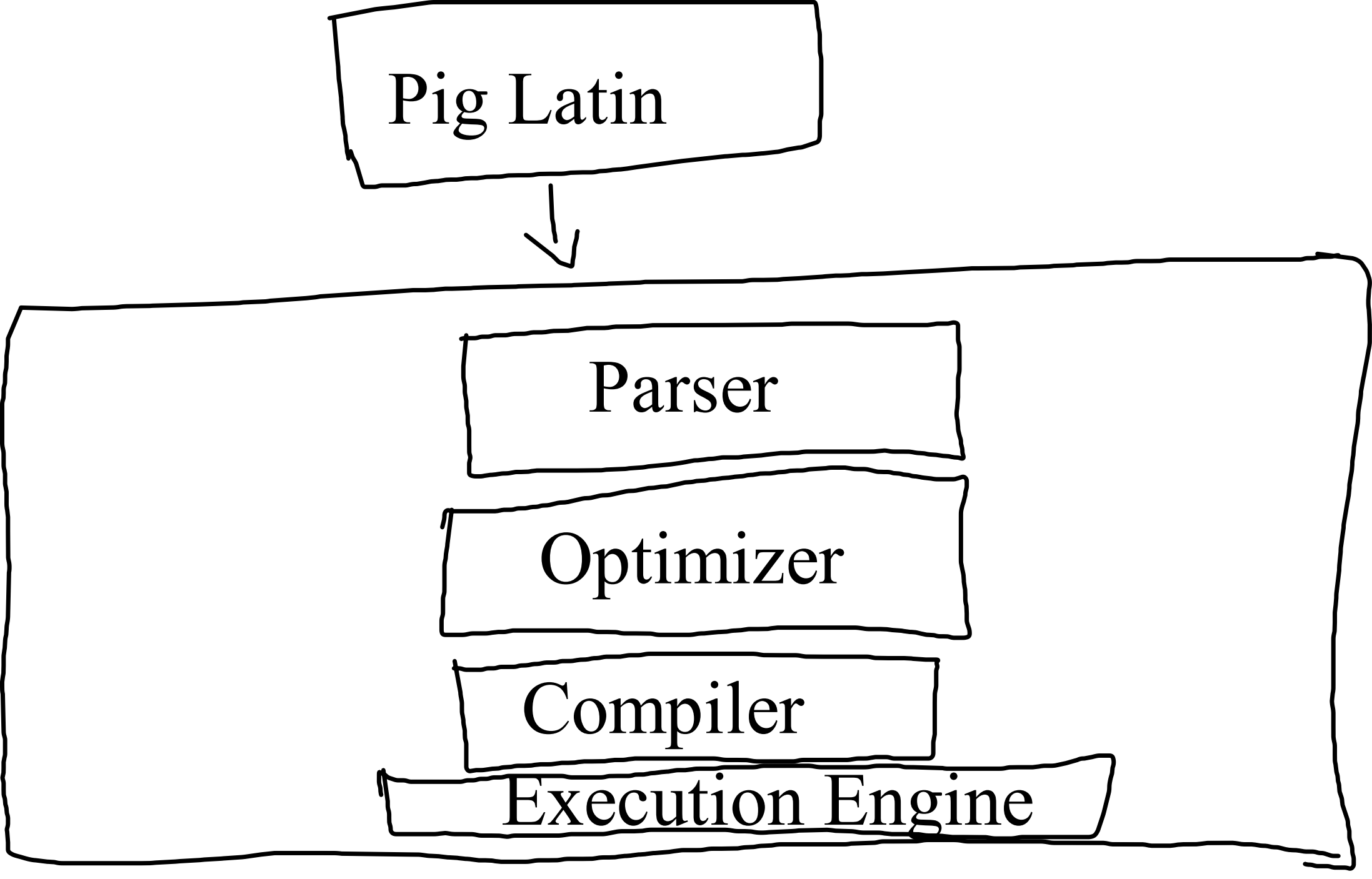
HDFS Master Slave Architecture
Advantages and Disadvantages of Hadoop
Apache Hive
Components and Architecture of Hive
Data Model of Hive
Apache Pig


Pig Latin
Parser
Optimizer
Compiler
Execution Engine

Data Model in Pig

Scalar Data Types

Integer, Long, Float, Double, chararray, bytearray

Complex Data Types

Maps: Key Value Pairs

['name'#'Jane', 'Age'#35]

Tuple: Multiple fields

(37, 17, 'Jane', 'Krish', 23)

Bags: Multiple Tuples

{('Jane', 35), ('Sally', 27), ('John', 33)}

Pig includes the concept of a data element being null

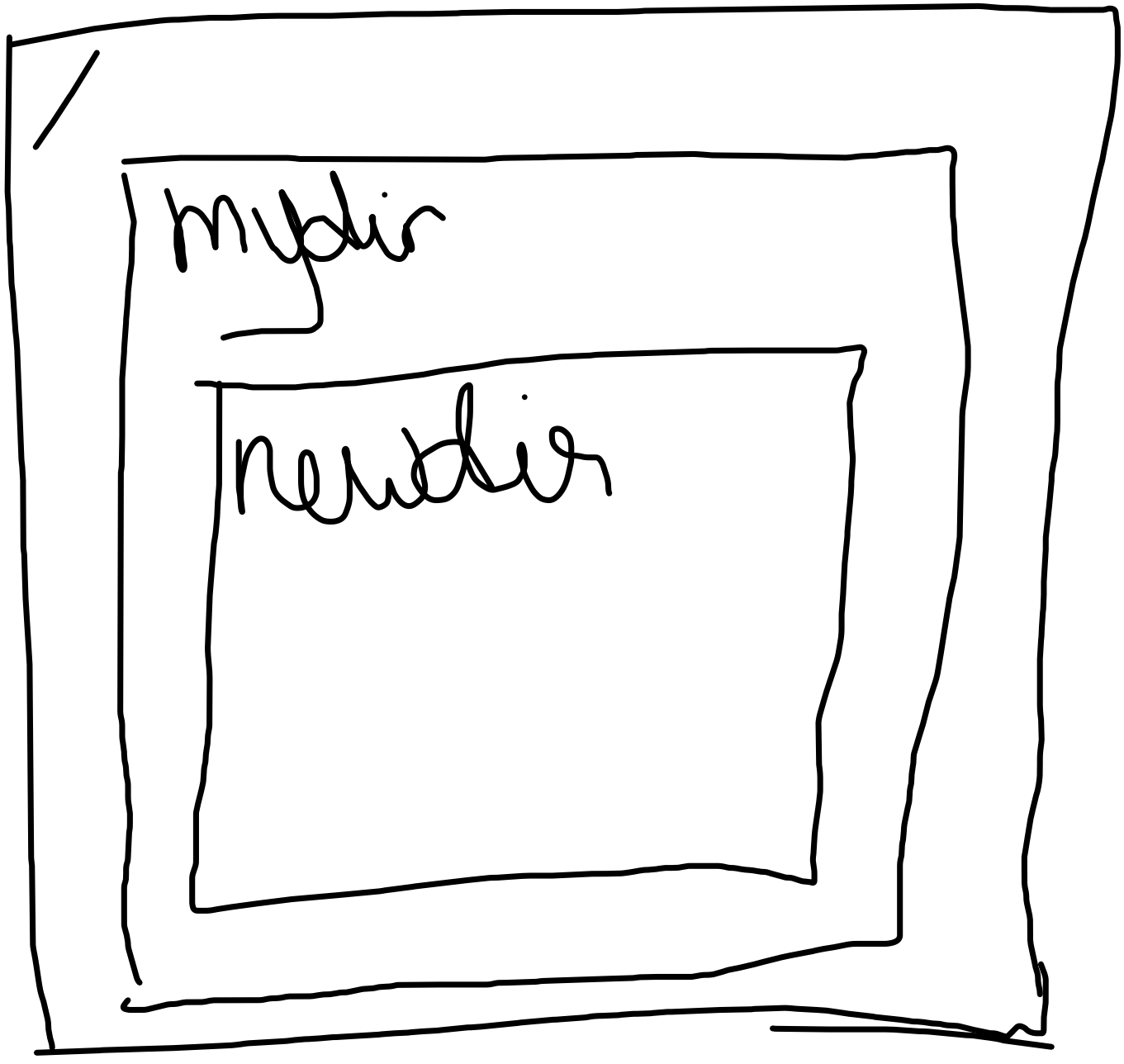
If a schema for the data is available, Pig will make use of it, but if no schema is available, Pig will still process the data making the best guesses it can based on how the script treats the data.

To create a new directory

hdfs dfs -mkdir <path to directory>

hdfs dfs -mkdir /mydir

hdfs dfs -mkdir /mydir/newdir


/
mydir
newdir

To send a file from Local File System to HDFS

hdfs dfs -put <source path> <destination path>
hdfs dfs -put /home/bigdata/samplefile.txt /mydir

To list the contents of directory:

hdfs dfs -ls <path to directory>
hdfs dfs -ls /mydir

To see the contents of a file:
hdfs dfs -cat <path to file>
hdfs dfs -cat /mydir/samplefile.txt

To send a file from HDFS to Local File System

hdfs dfs -get <source path> <destination path>
hdfs dfs -get /mydir/samplefile.txt /home/bigdata/result.txt

To delete a directory
hdfs dfs -rm -r <path to directory>
hdfs dfs -rm -r /mydir

hadoop jar <path to jar file> -mapper <path to mapper file> -reducer <path to reducer file> -input <path to input data> -output <path to output directory>