

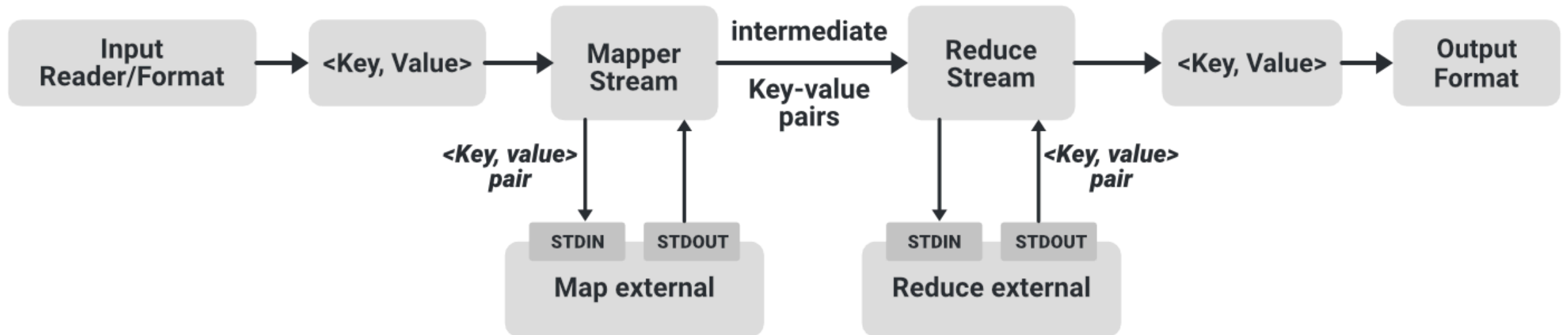
Hadoop Streaming

- <https://hadoop.apache.org/docs/current/>
 - <https://hadoop.apache.org/docs/current/streaming.html#How+Streaming+W>
 - Use ***any language*** that supports 'stdin' and 'stdout' I/O
 - Standard Job inc
 - Command line
- ```
$HADOOP_HOME/bin/hadoop jar
$HADOOP_HOME/hadoop-streaming.jar \
-input myInputDirs \
-output myOutputDir \
-mapper <mapper program> \
-reducer <reducer program>
```

## Some Hadoop Streaming Commands

<https://www.geeksforgeeks.org/what-is-hadoop-streaming/>

# Hadoop Streaming



# Example: word count in the linux shell

```
$HADOOP_HOME/bin/hadoop jar \
$HADOOP_HOME/hadoop-streaming.jar \
-input myInputDirs \
-output myOutputDir \
-mapper /bin/cat \
-reducer /bin/wc
```

## Example: word count in python\*

```
$HADOOP_HOME/bin/hadoop jar \
$HADOOP_HOME/hadoop-streaming.jar \
-input myInputDirs \
-output myOutputDir \
-mapper mapper.py \
-reducer reducer.py
```