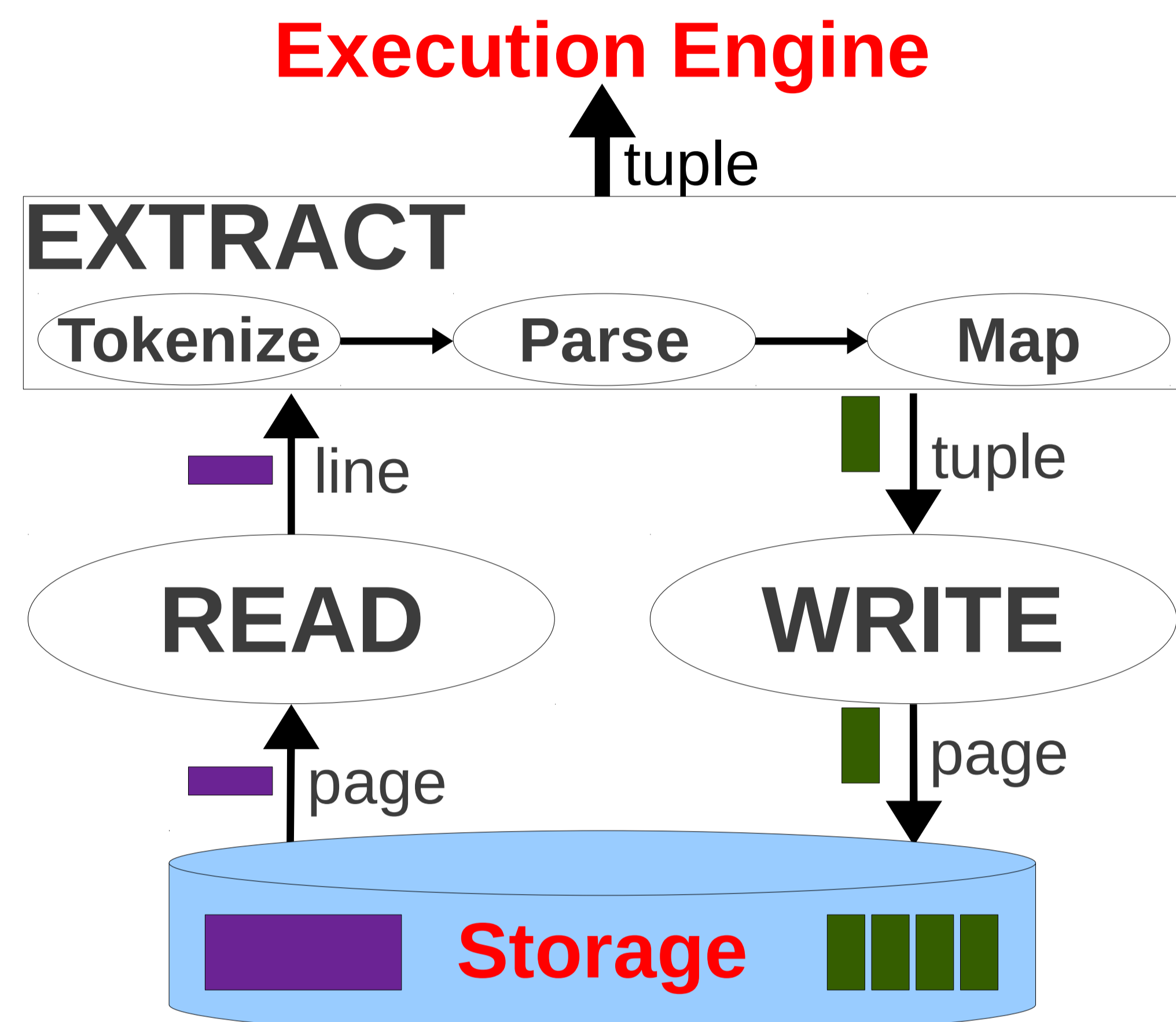


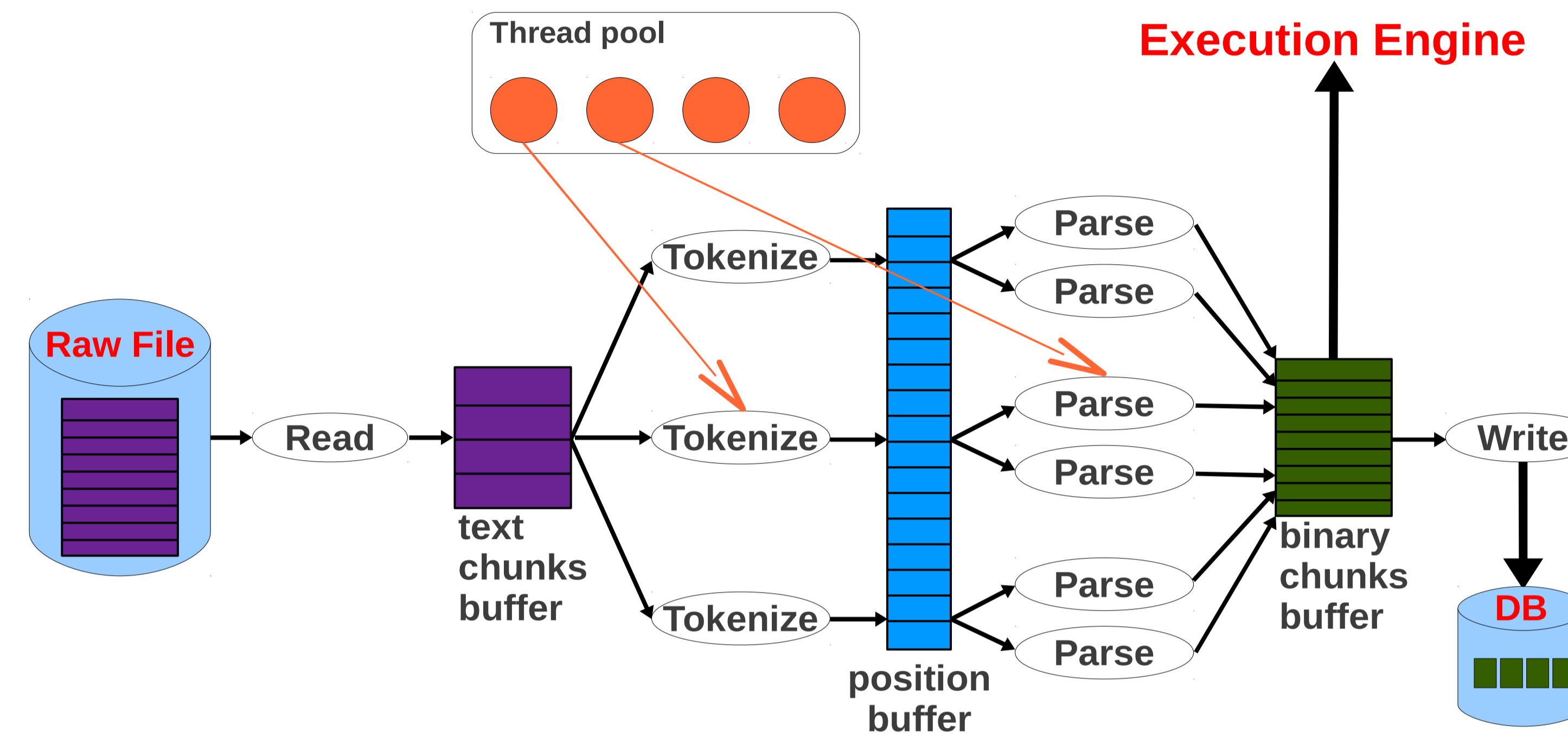
Raw File Process

The generic procedure to extract tuples with the given schema from the raw file



SCANRAW Architecture

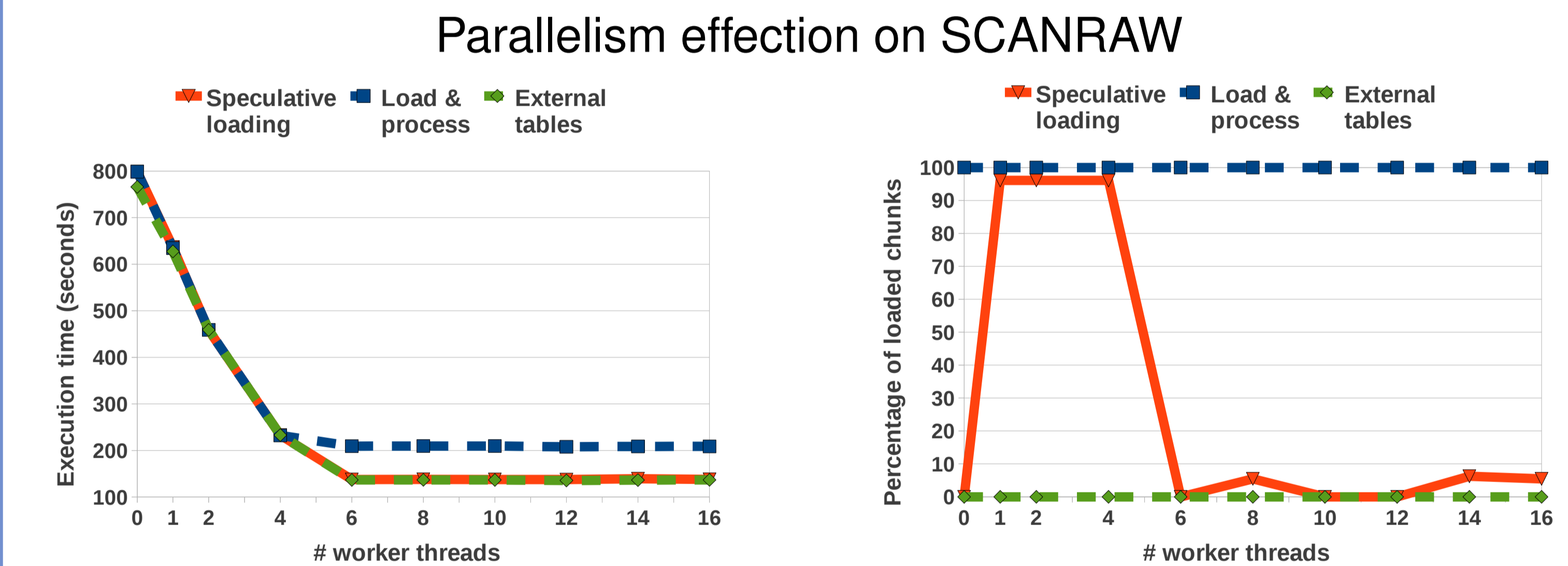
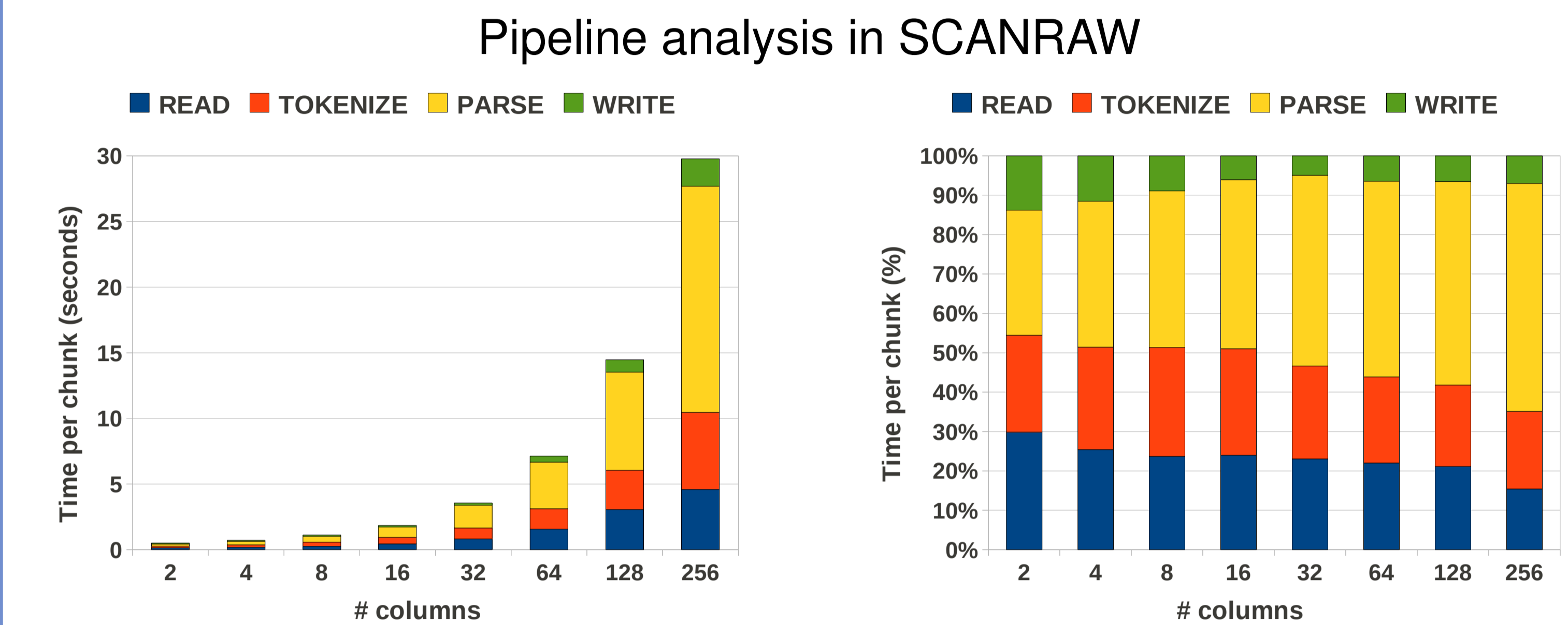
A parallel in-situ data processing operator targeted at the current many- and multi-core processors



- A physical operator for query processing over raw data.
- Containing a parallel super-scalar pipeline
- Could optimally use the available system resources

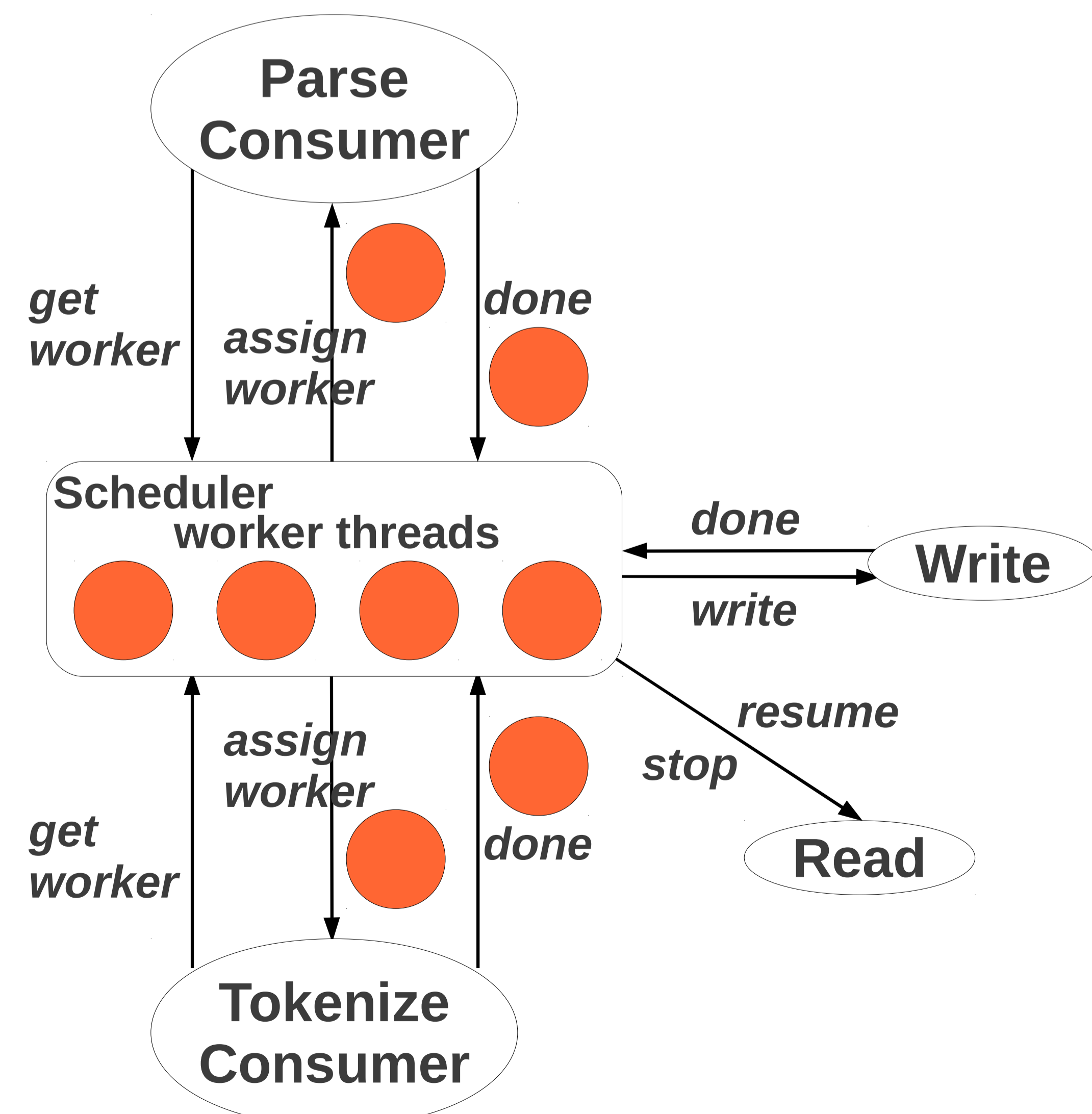
Experiments

Data. CSV files of 40 GB having **64** columns with 2^{26} lines, each column is a random unsigned integer smaller than 2^{31} .
System. 2 AMD Opteron 6128 - 16 cores - 40 GB of memory, and four 2 TB storage with Ubuntu 12.04.3.
Query. $SELECT \sum_{i=1}^K C_i FROM FILE$ where K columns C_i are projected.



Pipeline Structure

A series of asynchronous stand-alone threads



Speculative Loading

