

Spark

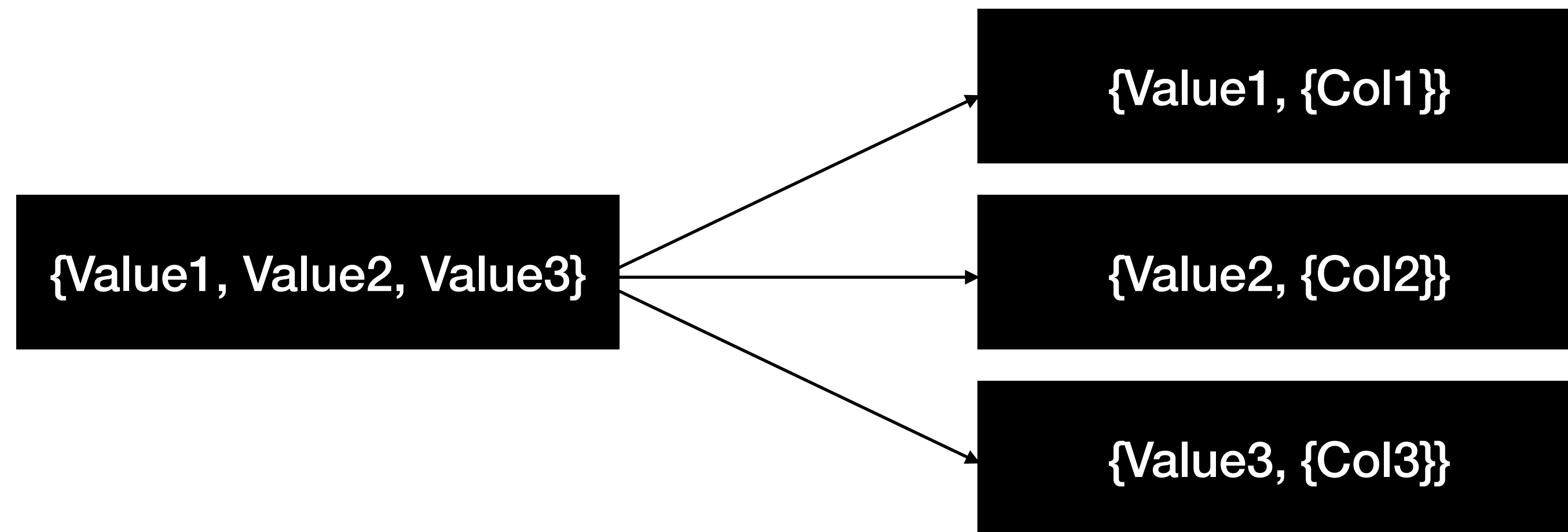
Team: BAM

- Read all data into one DataFrame

Spark

Team: BAM

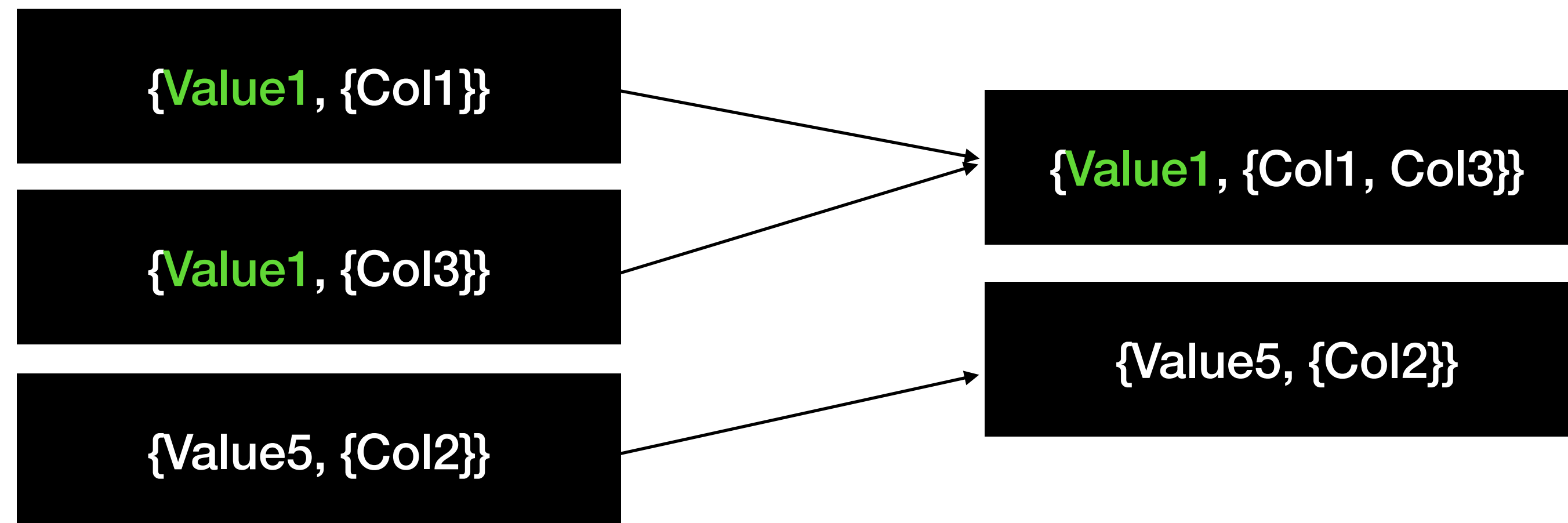
- Split cells and append column names



Spark

Team: BAM

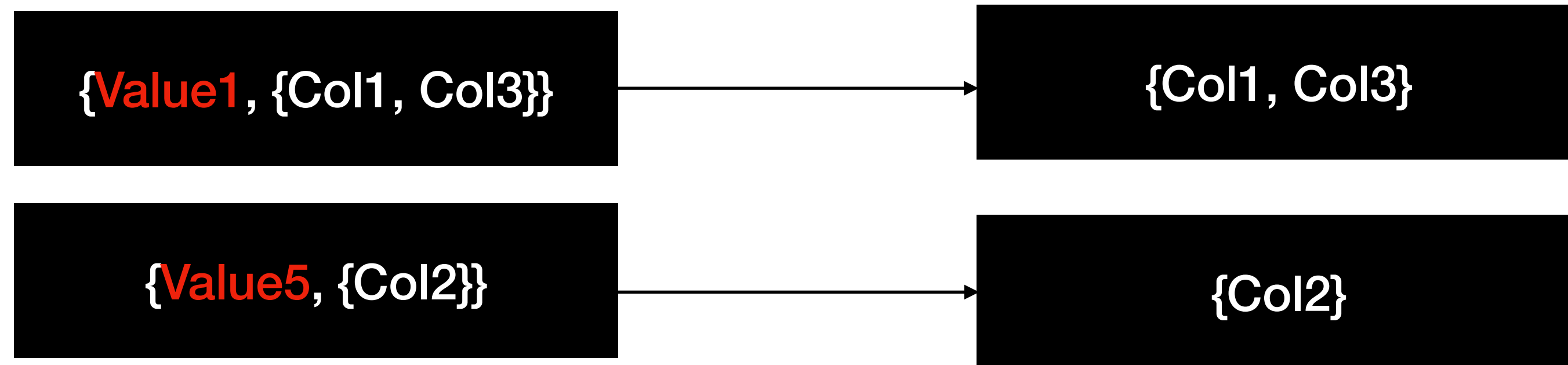
- Reduce tuples and union attributes



Spark

Team: BAM

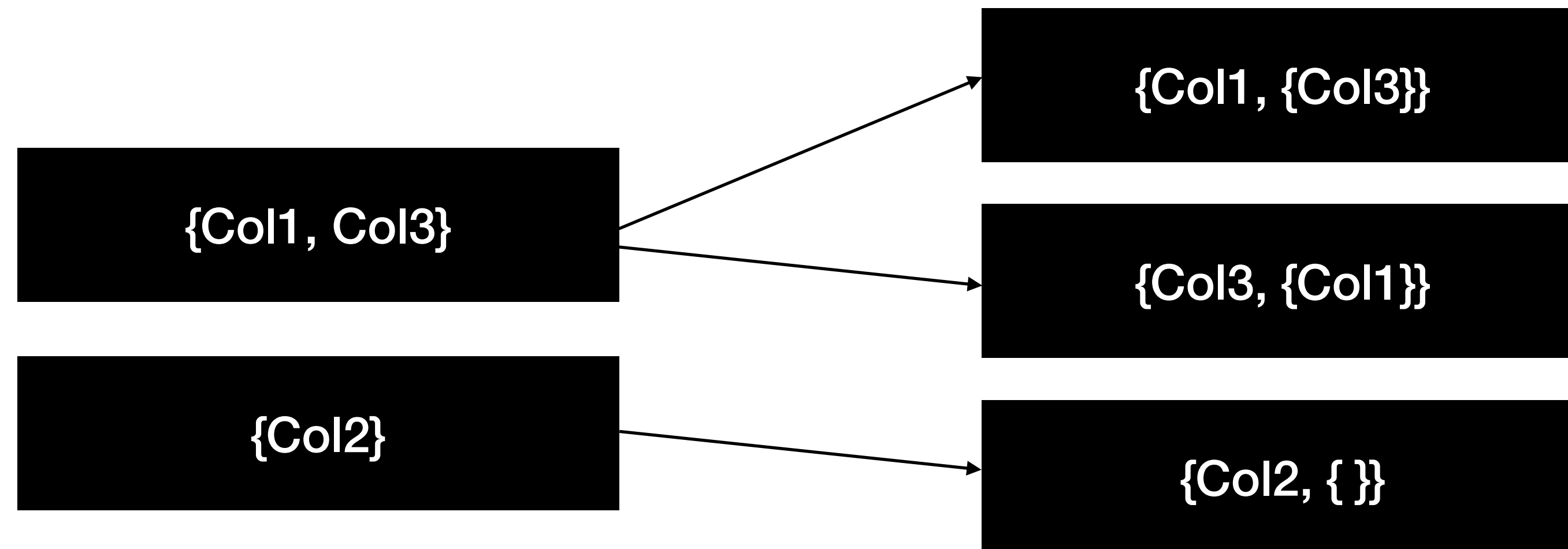
- Extract attributes



Spark

Team: BAM

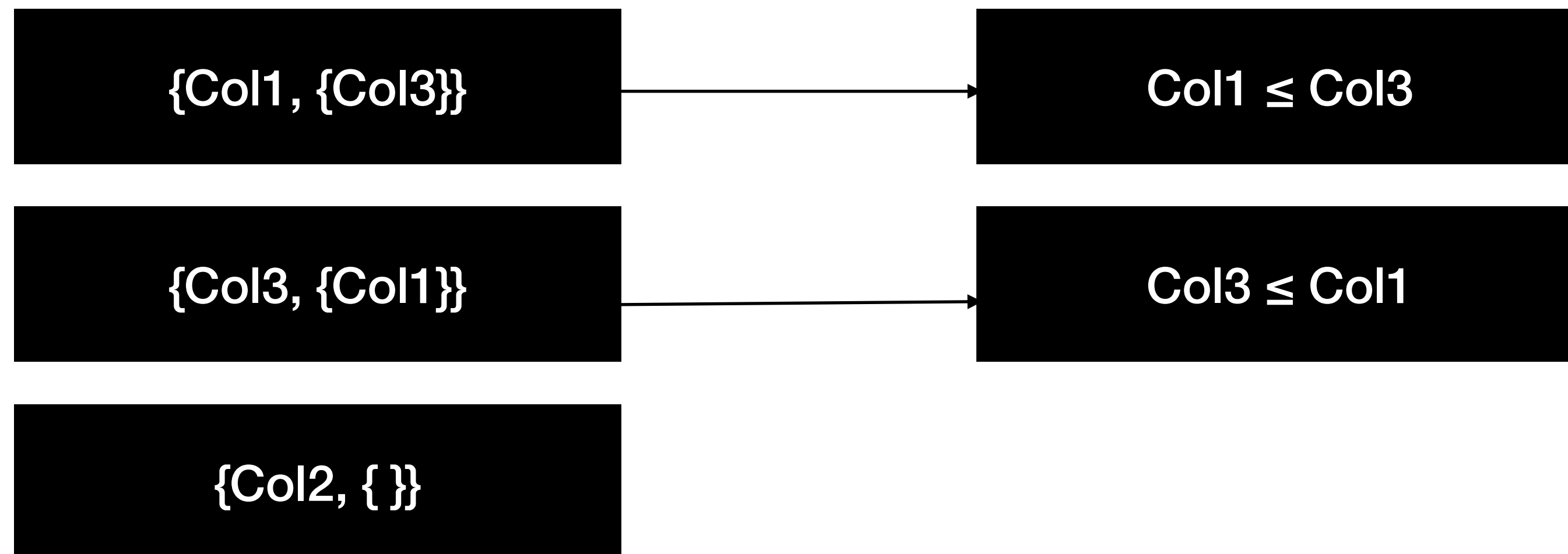
- Generate inclusion lists



Spark

Team: BAM

- Generate INDs



Spark

Team: BAM

- Problems with our solution
 - Couldn't manage to stay in Java Heap Space with big TPCCH Set
 - Some INDs are false positives