

# Refactoring Large Graphs



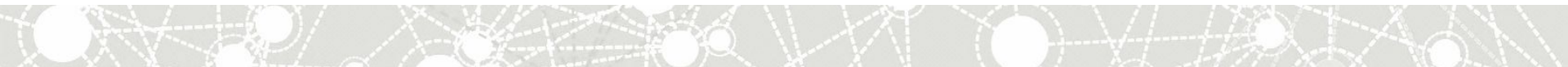
# What are we going to do?



- Why do we need to batch?
- The batch refactoring workflow
- Automate batch refactorings with the apoc library



# Why do we need to batch?

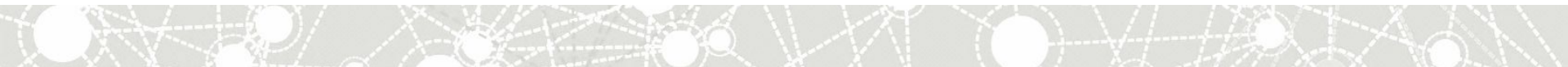


## Transaction State



Cypher keeps **all transaction state in memory** while running a query, which is fine most of the time.

When refactoring the graph, however, this state can get **very large** and may result in an **OutOfMemory** exception.



# The batch refactoring workflow



# The batch refactoring workflow



- tag all the nodes we need to process with a temporary label  
e.g. `Process`
- iterate over a subset of nodes flagged with that label (using `LIMIT`)  
and execute the refactoring
- remove the tag from the node
- return a count of how many rows were processed
- once the count reaches 0 then we've finished.

# The batch refactoring workflow



```
MATCH (itemToProcess:Process)
WITH itemToProcess
LIMIT 1000
REMOVE itemToProcess:Process
WITH itemToProcess
// do the refactoring
```

# Start playing the next guide....



...if you aren't playing it already

▶ Refactoring large graphs

:play [http://guides.neo4j.com/modeling\\_airports/05\\_refactoring\\_large\\_graphs.html](http://guides.neo4j.com/modeling_airports/05_refactoring_large_graphs.html)





# End of Module Refactoring Large Graphs

Questions ?

