

Benefits of Datomic “Sub-pulling”

Note: this document is available on line at:

<https://docs.google.com/document/d/1SdZTNaNRScdEj1iJbYDCI-IooPBQEXWb1QGBy0I-Z0M/edit?usp=sharing>

Sub-pulling allows us in datomic to reach across DB ref type fields into the next object and down through its object graph. This is similar to the sub-select in relational systems however, Datomic has a special query syntax allowing this.

If we have the following reference structure within our schema:

Patient => [Session] => [Answer] => [Question] => CTemplate => [CScale] => [Scale]

This represents a 7 layer deep hierarchy of data objects, referencing downward into lists which then reference single elements which again reference lists and finally single objects which represent the data that we actually need to get to in order to perform our calculation. When we write single or even multi-object linking queries in datomic, the last nodes on the tree always return :db/id's, either as a single value or as an array of :db/ids for the foreign-object references. The nature of the return values then encourages us to write multi-step queries that are then sub-processed with code in order to drive the next query, such as the following:

```
(d/q '[:find (pull ?s [*])
      :in $ ?patient ?earliest-completion
      :where [?s :session/patient ?patient]
              [?s :session/status :session-state/completed]
              [?s :session/patient-last-activity ?patient-last-activity]
              [(>= ?patient-last-activity ?earliest-completion)]]
      db (Long/parseLong patient-id) earliest-completion-date)
```

This will return us a list of completed sessions for a given patient that occurred after a specific date. Then to retrieve the answers for each session, we would need to execute the following query:

```
(d/q '[:find (pull ?a [*])
      :in $ ?session-id ?answer-ids
      :where [?session-id :session/patient ?a]]
db session-id (:session/answer session))
```

We would then need to use the results from the answer query to then retrieve our next data-set and so on until we have reached down the hierarchy. However, performing even just a few queries in this fashion is incredibly slow, adding many seconds and perhaps even as much as a minute onto our execution time.

Thankfully, Datomic provides a Pull-Query syntax that offers us a way around this. We can rewrite what would have been perhaps 5 or more queries and subqueries into a single operation using what Datomic calls sub-pulling. Initially the query above took 110 seconds to retrieve all the necessary information to calculate the automated-answers for a patient within a range of time, however, with sub-pulling, it is always under a second, at least for that individual operation.

```
(d/q '[:find (pull ?s [* {:session/answer
                          [* {:answer/question
                              [* {:question/ctemplate [
                                  :db/id
                                  :ctemplate/name
                                  :ctemplate/method {:ctemplate/cscale [*]]}]})]
                          {:session/programstep [:programstep/name]})]
      :in $ ?patient ?earliest-completion
      :where [?s :session/patient ?patient]
              [?s :session/status :session-state/completed]
              [?s :session/patient-last-activity ?patient-last-activity]
              [(>= ?patient-last-activity ?earliest-completion)])]
db (Long/parseLong patient-id) earliest-completion-date)
```

The syntax for the sub-pull is essentially field-specification as a list and then sub-specifications on specific fields as a hash-keyword pair specifically naming the field for which the sub-specification is to be applied to.

Unfortunately, Datomic only mentions this operation and the syntax driving it only once that I am able to find in their documentation. It is on this page for Datomic Cloud <https://docs.datomic.com/cloud/query/query-pull.html> and examples are given in the sections **“Map Specification Title”** and **“Map Specification Nesting Title”**. <https://docs.datomic.com/cloud/query/query-pull.html#orgc735b42>, <https://docs.datomic.com/cloud/query/query-pull.html#nesting>.

