Meeting 1/16/18

Tuesday, January 16, 2018 9:30 AM

Agenda:
- Go over slides for Block composition
- Go over Kirby's scripts / column reordering
- Measuring IO/CPU

Attendees: everyone

Snapper - can us for IO and CPU

Slides review:
- Basic Table Compression also uses Multi-column (3)
- If you set %free to 0 for advanced, you get the exact same compression (4)
- Slide 6 - go into how the clever encoding works.
- Include the Oracle Data Block Binary Dumps: Tokens and Column Lengths
- Simplify slide 7 - too much text
- Slide 8 - doesn't re-order nulls

Visualize block dumps
- Look at hex / token table
- Grab header and say here are the pieces that matter

Random vs. sorted
Sorting decreases size of table

Dump a normalized table and see if the columns are also ordered from low to high cardinality:
- They are

Look at size of token table to figure out why there's a drop at 800x with 32k blocks (sorted)
- Creating more tokens and replacing more data?

  It's making tokens with the parent key
  1450 rows
  1250 parent keys
  200 duplicates of parent keys

-- Comes up with a lot of tokens because of the way the test was designed
Repeating number in parent key values. Takes up 6 characters, is able to compress it
Ordered by primary, secondary, then parent-key, so there will be multiple parent-keys

For a 16k block with 800x, how many tokens are in token table?

- First block: one
- Second block: three. One token for the first column, then multi column tokens
    - Makes token for first column, then multi column token for the other two permutations of the second column

==========

IO - removing data from disk
CPU - joining the parent/child table and the math

Denormalize is another join to get rid of (storing parent/child together)

Can manipulate tables (join, unjoin, remove keys, etc)

==========

Future work:

- Use snapper to measure IO / CPU
- Kirby will write queries we want to measure
- Look for the scripts

Get all slides together by end the 30th