Meeting 11/17/17

Friday, November 17, 2017 11:00 AM

Agenda: Attendees: Everyone

- Are we meeting next week?
    - Yes, 9am instead of 11
- Go over experiments
- Discuss future work

Notes:

Experiments:

- Column order
    - Why does it choose the column order it does?
    - How many columns in a table before re-ordering not optimal?
- Basic vs Advanced Block Structure

- Basic vs Advanced - looks like it's doing the same thing format wise
    - Tokens look the same
    - Nothing's different until you start inserting stuff
- Column order same?
    - Column order is the same, but it combined tokens
- Column order is same between blocks

Does it choose the same column order every time?

- No
- In a table with 6 columns, and columns 1 and 6 contain the same data, it does not move them right next to one another
    - However, it does tokenize the data (combines tokens to link same data)
- It does move next to one another when entered next to one another
- When column position 1 and 4 (out of 6) contain identical data, it only moves it up 1
- How many tokens does it create?

Pre-order columns make a difference?

- How well does it work with a lot of columns?
    a. Create a table with 100 columns, reorder columns, see if compression puts them in a specific order every time
    b. How many columns in a table before re-ordering not optimal?
- BASIC VS ADVANCED BLOCK STRUCTURE
    a. Didn't re-order, but only used 3 columns

Both compression algorithms are deduplication

- Create table with all unique values - see if it's compressed
- Semi-similar values - see how it's compressed

When you create a table, it doesn't always enter the data the same way, or at least you cannot assume it does

For next week:
Figure out why compression chooses the column order it chooses

- Only make 1 block's worth of data (we know what the data is)
- 10 rows, only 1 block, would it put everything next to one another?
- Use order by to see if it overwrites that

Ideally, we want to be able to write up some data on the board and accurately predict how it will compress

What's the best way to order columns