Meeting 1/9/18

Tuesday, January 9, 2018 9:30 AM

First 199 tokens are represented as single bytes Attendees: everyone

Column length 1-50 represented by a single byte

fa and fb bytes indicate it's a multibyte column
0xfa - read next two bytes that's length
0xfb - read next two bytes as token number

0x00-0xC7 = tokens

0xC8 = special

0xC9-0xF9 = length

0xfA-0xff = special /multibyte token / null

==========

Outline for presentation:

1) Intro
2) Overview
3) How block compression works
    a. Quick overview of a block
    b. Token table
    c. Multi-column tokens
    d. Column reordering
4) Compression Experiment Results (Ratios)
    a. Row ordering (should improve)
    b. Column ordering (should be equal)
        i. Manual vs block reordering
    c. Parent / child
        i. Separate vs combined
    d. Normalized vs denormalized
    e. Block size (8/16/32)
5) IO/CPU experiment results
    a. Parent/child
    b. Normalized/denormalized

==========

Tables for testing:

Parent/child denormalized

Parent: parent-device, 1 million rows

Device	(10 devices)	varchar16	
Timestamp	(100,000 timestamps)	varchar16	Ideally no repeating values (milliseconds in 2017)

Child: results

Side	(2 options)	varchar16	
Location	(5 options)	varchar16	
Value	(rand)	number	10 million random numbers (10m to 20m-1)

After initial testing, side and location may be too insignificant to see a difference
Changing to 5 sides, 20 locations

==========

For next meeting:
(step 3 in outline) Explain blocks in a clean, simple, easy way to understand