Database Compression

Kelly Jiang, Jax Luo

Sqlite

json

↓ spark

SQLite

↓ vertically partition

SQLite
↓

Columnar
Compression

SQLite

↓

Dictionary encoding

↓

Bitpacking

↓

zopfli

Dictionary Encoding

values

"abc"

"def"

"abc"

"abc"

"efg"

$\rightarrow$

keys

0

1

0

0

2

Dictionary

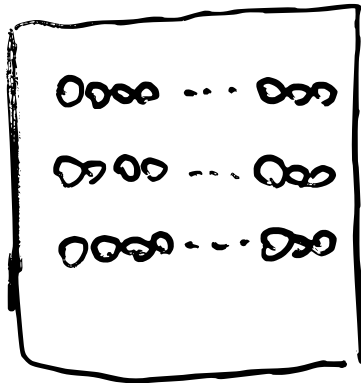| keys	values
0	"abc"
1	"def"
2	"efg"

Bitpacking

Compute Max & Min to figure out bits needed

Truncate the bits representation for each value

in that column

Concatenate bit strings together

```
0000 ... 000        0110
0000 ... 000        0111        →    Output        0110 0111 1000
0000 ... 000        1000
```

Discard

String Compression

Zepfli: Made by Google

Supports Deflate, gzip, or zlib encoding

3-8% more compression than zlib max,
~80x slower