Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

File Sizes

	File	Size (KB)
1	Original Data Dump	4,529,194
2	Extracted Desired Data	1,669,496
3	Delta Compressed Data by Int Values	1,664,451
4	Delta Compressed Data by Bytes	1,669,496
5	Variable Byte Compressed Data	1,239,426

Performance Analysis

	File	Time Elapsed (sec)
1	Extracted Desired Data	234.168
2	Delta Compressed Data by Int Values	315.033
3	Delta Compressed Data by Bytes	272.397
4	Variable Byte Compressed Data	296.672
5	Variable Byte Compressed Data (On Delta)	185.977

Analysis Reports

File 1: **No Compression**

```
1   Running
2   Opening json file
3   File opened, loading data from json
4   Main dictionary created
5   Cleaning/Deleting the old database
6
7
8   About to populate database with 10M pairs - batch writing (Asynchronous)
9
10  ('    Working . . . Currently at:', 0)
11  ('    Working . . . Currently at:', 999999)
12  ('    Working . . . Currently at:', 1999998)
13  ('    Working . . . Currently at:', 2999997)
14  ('    Working . . . Currently at:', 3999996)
15  ('    Working . . . Currently at:', 4999995)
16  ('    Working . . . Currently at:', 5999994)
17  ('    Working . . . Currently at:', 6999993)
18  ('    Working . . . Currently at:', 7999992)
19  ('    Working . . . Currently at:', 8999991)
20  ('    Working . . . Currently at:', 9999990)
21  ('    Working . . . Currently at:', 10999989)
22  ('    Working . . . Currently at:', 11999988)
23  ('    Working . . . Currently at:', 12999987)
24  ('    Working . . . Currently at:', 13999986)
25  ('    Working . . . Currently at:', 14999985)
26  ('    Working . . . Currently at:', 15999984)
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
27  ('    Working . . . Currently at:', 16999983)
28  ('    Working . . . Currently at:', 17999982)
29  ('    Working . . . Currently at:', 18999981)
30  ('    Working . . . Currently at:', 19999980)
31  ('    Working . . . Currently at:', 20999979)
32  ('    Working . . . Currently at:', 21999978)
33  ('    Working . . . Currently at:', 22999977)
34  ('    Working . . . Currently at:', 23999976)
35  ('    Working . . . Currently at:', 24999975)
36  ('    Working . . . Currently at:', 25999974)
37  ('    Working . . . Currently at:', 26999973)
38  ('    Working . . . Currently at:', 27999972)
39  ('    Working . . . Currently at:', 28999971)
40  ('    Working . . . Currently at:', 29999970)
41  ('    Working . . . Currently at:', 30999969)
42  ('    Working . . . Currently at:', 31999968)
43  ('    Working . . . Currently at:', 32999967)
44  ('    Working . . . Currently at:', 33999966)
45  Timer Ended - Batch Writing
46  Randomizing the keys
47
48
49  About to read all data - Random Order
50  ('    Reading . . . Currently at:', 0)
51  ('    Reading . . . Currently at:', 999999)
52  ('    Reading . . . Currently at:', 1999998)
53  ('    Reading . . . Currently at:', 2999997)
54  ('    Reading . . . Currently at:', 3999996)
55  ('    Reading . . . Currently at:', 4999995)
56  ('    Reading . . . Currently at:', 5999994)
57  ('    Reading . . . Currently at:', 6999993)
58  ('    Reading . . . Currently at:', 7999992)
59  ('    Reading . . . Currently at:', 8999991)
60  ('    Reading . . . Currently at:', 9999990)
61  ('    Reading . . . Currently at:', 10999989)
62  ('    Reading . . . Currently at:', 11999988)
63  ('    Reading . . . Currently at:', 12999987)
64  ('    Reading . . . Currently at:', 13999986)
65  ('    Reading . . . Currently at:', 14999985)
66  ('    Reading . . . Currently at:', 15999984)
67  ('    Reading . . . Currently at:', 16999983)
68  ('    Reading . . . Currently at:', 17999982)
69  ('    Reading . . . Currently at:', 18999981)
70  ('    Reading . . . Currently at:', 19999980)
71  ('    Reading . . . Currently at:', 20999979)
72  ('    Reading . . . Currently at:', 21999978)
73  ('    Reading . . . Currently at:', 22999977)
74  ('    Reading . . . Currently at:', 23999976)
75  ('    Reading . . . Currently at:', 24999975)
76  ('    Reading . . . Currently at:', 25999974)
77  ('    Reading . . . Currently at:', 26999973)
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
78 ('    Reading . . . Currently at:', 27999972)
79 ('    Reading . . . Currently at:', 28999971)
80 ('    Reading . . . Currently at:', 29999970)
81 ('    Reading . . . Currently at:', 30999969)
82 ('    Reading . . . Currently at:', 31999968)
83 ('    Reading . . . Currently at:', 32999967)
84 ('    Reading . . . Currently at:', 33999966)
85
86
87 Database populated with batch writes
88     Time Started: 1496724754.58sec
89     Time Ended: 1496730692.1sec
90     Time Elapsed: 5937.52049994sec
91 Database read randomly
92     Time Started: 1496730782.03sec
93     Time Ended: 1496731016.19sec
94     Time Elapsed: 234.168370962sec
95
96 --- Finished Program ---
```

File 2: **Delta Compression on Ints**

```
 1 Running
 2 Opening json file
 3 File opened, loading data from json
 4 Main dictionary created
 5 Cleaning/Deleting the old database
 6
 7
 8 About to populate database with 10M pairs - batch writing (Asynchronous)
 9
10 ('    Working . . . Currently at:', 0)
11 ('    Working . . . Currently at:', 999999)
12 ('    Working . . . Currently at:', 1999998)
13 ('    Working . . . Currently at:', 2999997)
14 ('    Working . . . Currently at:', 3999996)
15 ('    Working . . . Currently at:', 4999995)
16 ('    Working . . . Currently at:', 5999994)
17 ('    Working . . . Currently at:', 6999993)
18 ('    Working . . . Currently at:', 7999992)
19 ('    Working . . . Currently at:', 8999991)
20 ('    Working . . . Currently at:', 9999990)
21 ('    Working . . . Currently at:', 10999989)
22 ('    Working . . . Currently at:', 11999988)
23 ('    Working . . . Currently at:', 12999987)
24 ('    Working . . . Currently at:', 13999986)
25 ('    Working . . . Currently at:', 14999985)
26 ('    Working . . . Currently at:', 15999984)
27 ('    Working . . . Currently at:', 16999983)
28 ('    Working . . . Currently at:', 17999982)
29 ('    Working . . . Currently at:', 18999981)
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
30 ('    Working . . . Currently at:', 19999980)
31 ('    Working . . . Currently at:', 20999979)
32 ('    Working . . . Currently at:', 21999978)
33 ('    Working . . . Currently at:', 22999977)
34 ('    Working . . . Currently at:', 23999976)
35 ('    Working . . . Currently at:', 24999975)
36 ('    Working . . . Currently at:', 25999974)
37 ('    Working . . . Currently at:', 26999973)
38 ('    Working . . . Currently at:', 27999972)
39 ('    Working . . . Currently at:', 28999971)
40 ('    Working . . . Currently at:', 29999970)
41 ('    Working . . . Currently at:', 30999969)
42 ('    Working . . . Currently at:', 31999968)
43 ('    Working . . . Currently at:', 32999967)
44 ('    Working . . . Currently at:', 33999966)
45 Timer Ended - Batch Writing
46 Randomizing the keys
47
48
49 About to read all data - Random Order
50 ('    Reading . . . Currently at:', 0)
51 ('    Reading . . . Currently at:', 999999)
52 ('    Reading . . . Currently at:', 1999998)
53 ('    Reading . . . Currently at:', 2999997)
54 ('    Reading . . . Currently at:', 3999996)
55 ('    Reading . . . Currently at:', 4999995)
56 ('    Reading . . . Currently at:', 5999994)
57 ('    Reading . . . Currently at:', 6999993)
58 ('    Reading . . . Currently at:', 7999992)
59 ('    Reading . . . Currently at:', 8999991)
60 ('    Reading . . . Currently at:', 9999990)
61 ('    Reading . . . Currently at:', 10999989)
62 ('    Reading . . . Currently at:', 11999988)
63 ('    Reading . . . Currently at:', 12999987)
64 ('    Reading . . . Currently at:', 13999986)
65 ('    Reading . . . Currently at:', 14999985)
66 ('    Reading . . . Currently at:', 15999984)
67 ('    Reading . . . Currently at:', 16999983)
68 ('    Reading . . . Currently at:', 17999982)
69 ('    Reading . . . Currently at:', 18999981)
70 ('    Reading . . . Currently at:', 19999980)
71 ('    Reading . . . Currently at:', 20999979)
72 ('    Reading . . . Currently at:', 21999978)
73 ('    Reading . . . Currently at:', 22999977)
74 ('    Reading . . . Currently at:', 23999976)
75 ('    Reading . . . Currently at:', 24999975)
76 ('    Reading . . . Currently at:', 25999974)
77 ('    Reading . . . Currently at:', 26999973)
78 ('    Reading . . . Currently at:', 27999972)
79 ('    Reading . . . Currently at:', 28999971)
80 ('    Reading . . . Currently at:', 29999970)
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
81 ('     Reading . . . Currently at:', 30999969)
82 ('     Reading . . . Currently at:', 31999968)
83 ('     Reading . . . Currently at:', 32999967)
84 ('     Reading . . . Currently at:', 33999966)
85
86
87 Database populated with batch writes
88      Time Started: 1496984931.5 sec
89      Time Ended: 1496991614.47 sec
90      Time Elapsed: 6682.9688251 sec
91 Database read randomly
92      Time Started: 1496991841.62 sec
93      Time Ended: 1496992156.66 sec
94      Time Elapsed: 315.033441067 sec
95
96 --- Finished Program ---
```

File 3: **Byte Delta Compression**

```
1  Running
2  Opening json file
3  File opened, loading data from json
4  Main dictionary created
5  Cleaning/Deleting the old database
6
7
8  About to populate database with 10M pairs - batch writing (Asynchronous)
9
10 ('     Working . . . Currently at:', 0)
11 ('     Working . . . Currently at:', 999999)
12 ('     Working . . . Currently at:', 1999998)
13 ('     Working . . . Currently at:', 2999997)
14 ('     Working . . . Currently at:', 3999996)
15 ('     Working . . . Currently at:', 4999995)
16 ('     Working . . . Currently at:', 5999994)
17 ('     Working . . . Currently at:', 6999993)
18 ('     Working . . . Currently at:', 7999992)
19 ('     Working . . . Currently at:', 8999991)
20 ('     Working . . . Currently at:', 9999990)
21 ('     Working . . . Currently at:', 10999989)
22 ('     Working . . . Currently at:', 11999988)
23 ('     Working . . . Currently at:', 12999987)
24 ('     Working . . . Currently at:', 13999986)
25 ('     Working . . . Currently at:', 14999985)
26 ('     Working . . . Currently at:', 15999984)
27 ('     Working . . . Currently at:', 16999983)
28 ('     Working . . . Currently at:', 17999982)
29 ('     Working . . . Currently at:', 18999981)
30 ('     Working . . . Currently at:', 19999980)
31 ('     Working . . . Currently at:', 20999979)
32 ('     Working . . . Currently at:', 21999978)
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
33  ('    Working . . . Currently at:', 22999977)
34  ('    Working . . . Currently at:', 23999976)
35  ('    Working . . . Currently at:', 24999975)
36  ('    Working . . . Currently at:', 25999974)
37  ('    Working . . . Currently at:', 26999973)
38  ('    Working . . . Currently at:', 27999972)
39  ('    Working . . . Currently at:', 28999971)
40  ('    Working . . . Currently at:', 29999970)
41  ('    Working . . . Currently at:', 30999969)
42  ('    Working . . . Currently at:', 31999968)
43  ('    Working . . . Currently at:', 32999967)
44  ('    Working . . . Currently at:', 33999966)
45  Timer Ended - Batch Writing
46  Randomizing the keys
47
48
49  About to read all data - Random Order
50  ('    Reading . . . Currently at:', 0)
51  ('    Reading . . . Currently at:', 999999)
52  ('    Reading . . . Currently at:', 1999998)
53  ('    Reading . . . Currently at:', 2999997)
54  ('    Reading . . . Currently at:', 3999996)
55  ('    Reading . . . Currently at:', 4999995)
56  ('    Reading . . . Currently at:', 5999994)
57  ('    Reading . . . Currently at:', 6999993)
58  ('    Reading . . . Currently at:', 7999992)
59  ('    Reading . . . Currently at:', 8999991)
60  ('    Reading . . . Currently at:', 9999990)
61  ('    Reading . . . Currently at:', 10999989)
62  ('    Reading . . . Currently at:', 11999988)
63  ('    Reading . . . Currently at:', 12999987)
64  ('    Reading . . . Currently at:', 13999986)
65  ('    Reading . . . Currently at:', 14999985)
66  ('    Reading . . . Currently at:', 15999984)
67  ('    Reading . . . Currently at:', 16999983)
68  ('    Reading . . . Currently at:', 17999982)
69  ('    Reading . . . Currently at:', 18999981)
70  ('    Reading . . . Currently at:', 19999980)
71  ('    Reading . . . Currently at:', 20999979)
72  ('    Reading . . . Currently at:', 21999978)
73  ('    Reading . . . Currently at:', 22999977)
74  ('    Reading . . . Currently at:', 23999976)
75  ('    Reading . . . Currently at:', 24999975)
76  ('    Reading . . . Currently at:', 25999974)
77  ('    Reading . . . Currently at:', 26999973)
78  ('    Reading . . . Currently at:', 27999972)
79  ('    Reading . . . Currently at:', 28999971)
80  ('    Reading . . . Currently at:', 29999970)
81  ('    Reading . . . Currently at:', 30999969)
82  ('    Reading . . . Currently at:', 31999968)
83  ('    Reading . . . Currently at:', 32999967)
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
84 ('   Reading . . . Currently at:', 33999966)
85
86
87 Database populated with batch writes
88       Time Started: 1496995923.99 sec
89       Time Ended: 1497001728.99 sec
90       Time Elapsed: 5805.00258112 sec
91 Database read randomly
92       Time Started: 1497001857.33 sec
93       Time Ended: 1497002129.73 sec
94       Time Elapsed: 272.397305012 sec
95
96 --- Finished Program ---
```

File 4: **Variable Byte Compression**

```
1  Running
2  Opening json file
3  File opened, loading data from json
4  Main dictionary created
5  Cleaning/Deleting the old database
6
7
8  About to populate database with 10M pairs - batch writing (Asynchronous)
9
10 ('   Working . . . Currently at:', 0)
11 ('   Working . . . Currently at:', 999999)
12 ('   Working . . . Currently at:', 1999998)
13 ('   Working . . . Currently at:', 2999997)
14 ('   Working . . . Currently at:', 3999996)
15 ('   Working . . . Currently at:', 4999995)
16 ('   Working . . . Currently at:', 5999994)
17 ('   Working . . . Currently at:', 6999993)
18 ('   Working . . . Currently at:', 7999992)
19 ('   Working . . . Currently at:', 8999991)
20 ('   Working . . . Currently at:', 9999990)
21 ('   Working . . . Currently at:', 10999989)
22 ('   Working . . . Currently at:', 11999988)
23 ('   Working . . . Currently at:', 12999987)
24 ('   Working . . . Currently at:', 13999986)
25 ('   Working . . . Currently at:', 14999985)
26 ('   Working . . . Currently at:', 15999984)
27 ('   Working . . . Currently at:', 16999983)
28 ('   Working . . . Currently at:', 17999982)
29 ('   Working . . . Currently at:', 18999981)
30 ('   Working . . . Currently at:', 19999980)
31 ('   Working . . . Currently at:', 20999979)
32 ('   Working . . . Currently at:', 21999978)
33 ('   Working . . . Currently at:', 22999977)
34 ('   Working . . . Currently at:', 23999976)
35 ('   Working . . . Currently at:', 24999975)
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
36 ('    Working . . . Currently at:', 25999974)
37 ('    Working . . . Currently at:', 26999973)
38 ('    Working . . . Currently at:', 27999972)
39 ('    Working . . . Currently at:', 28999971)
40 ('    Working . . . Currently at:', 29999970)
41 ('    Working . . . Currently at:', 30999969)
42 ('    Working . . . Currently at:', 31999968)
43 ('    Working . . . Currently at:', 32999967)
44 ('    Working . . . Currently at:', 33999966)
45 Timer Ended - Batch Writing
46 Randomizing the keys
47
48
49 About to read all data - Random Order
50 ('    Reading . . . Currently at:', 0)
51 ('    Reading . . . Currently at:', 999999)
52 ('    Reading . . . Currently at:', 1999998)
53 ('    Reading . . . Currently at:', 2999997)
54 ('    Reading . . . Currently at:', 3999996)
55 ('    Reading . . . Currently at:', 4999995)
56 ('    Reading . . . Currently at:', 5999994)
57 ('    Reading . . . Currently at:', 6999993)
58 ('    Reading . . . Currently at:', 7999992)
59 ('    Reading . . . Currently at:', 8999991)
60 ('    Reading . . . Currently at:', 9999990)
61 ('    Reading . . . Currently at:', 10999989)
62 ('    Reading . . . Currently at:', 11999988)
63 ('    Reading . . . Currently at:', 12999987)
64 ('    Reading . . . Currently at:', 13999986)
65 ('    Reading . . . Currently at:', 14999985)
66 ('    Reading . . . Currently at:', 15999984)
67 ('    Reading . . . Currently at:', 16999983)
68 ('    Reading . . . Currently at:', 17999982)
69 ('    Reading . . . Currently at:', 18999981)
70 ('    Reading . . . Currently at:', 19999980)
71 ('    Reading . . . Currently at:', 20999979)
72 ('    Reading . . . Currently at:', 21999978)
73 ('    Reading . . . Currently at:', 22999977)
74 ('    Reading . . . Currently at:', 23999976)
75 ('    Reading . . . Currently at:', 24999975)
76 ('    Reading . . . Currently at:', 25999974)
77 ('    Reading . . . Currently at:', 26999973)
78 ('    Reading . . . Currently at:', 27999972)
79 ('    Reading . . . Currently at:', 28999971)
80 ('    Reading . . . Currently at:', 29999970)
81 ('    Reading . . . Currently at:', 30999969)
82 ('    Reading . . . Currently at:', 31999968)
83 ('    Reading . . . Currently at:', 32999967)
84 ('    Reading . . . Currently at:', 33999966)
85
86
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
87  Database populated with batch writes
88        Time Started: 1497007421.17 sec
89        Time Ended: 1497019697.8 sec
90        Time Elapsed: 12276.6317949 sec
91  Database read randomly
92        Time Started: 1497019802.41 sec
93        Time Ended: 1497020099.08 sec
94        Time Elapsed: 296.672063112 sec
95
96  --- Finished Program ---
```

File 5: **Variable Byte Compression On Delta**

```
1   Opening json file
2   File opened, loading data from json
3   Main dictionary created
4   Finished compressing dictionary values, writing to file
5   Making Dict Values to Bytes
6   Finished writing
7
8
9
10  About to populate database with 10M pairs - batch writing (Asynchronous)
11
12      Working . . . Currently at: 0
13      Working . . . Currently at: 999999
14      Working . . . Currently at: 1999998
15      Working . . . Currently at: 2999997
16      Working . . . Currently at: 3999996
17      Working . . . Currently at: 4999995
18      Working . . . Currently at: 5999994
19      Working . . . Currently at: 6999993
20      Working . . . Currently at: 7999992
21      Working . . . Currently at: 8999991
22      Working . . . Currently at: 9999990
23      Working . . . Currently at: 10999989
24      Working . . . Currently at: 11999988
25      Working . . . Currently at: 12999987
26      Working . . . Currently at: 13999986
27      Working . . . Currently at: 14999985
28      Working . . . Currently at: 15999984
29      Working . . . Currently at: 16999983
30      Working . . . Currently at: 17999982
31      Working . . . Currently at: 18999981
32      Working . . . Currently at: 19999980
33      Working . . . Currently at: 20999979
34      Working . . . Currently at: 21999978
35      Working . . . Currently at: 22999977
36      Working . . . Currently at: 23999976
37      Working . . . Currently at: 24999975
38      Working . . . Currently at: 25999974
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
39      Working . . . Currently at: 26999973
40      Working . . . Currently at: 27999972
41      Working . . . Currently at: 28999971
42      Working . . . Currently at: 29999970
43      Working . . . Currently at: 30999969
44      Working . . . Currently at: 31999968
45      Working . . . Currently at: 32999967
46      Working . . . Currently at: 33999966
47  Timer Ended - Batch Writing
48  Randomizing the keys
49
50
51  About to read all data - Random Order
52      Reading . . . Currently at: 0
53      Reading . . . Currently at: 999999
54      Reading . . . Currently at: 1999998
55      Reading . . . Currently at: 2999997
56      Reading . . . Currently at: 3999996
57      Reading . . . Currently at: 4999995
58      Reading . . . Currently at: 5999994
59      Reading . . . Currently at: 6999993
60      Reading . . . Currently at: 7999992
61      Reading . . . Currently at: 8999991
62      Reading . . . Currently at: 9999990
63      Reading . . . Currently at: 10999989
64      Reading . . . Currently at: 11999988
65      Reading . . . Currently at: 12999987
66      Reading . . . Currently at: 13999986
67      Reading . . . Currently at: 14999985
68      Reading . . . Currently at: 15999984
69      Reading . . . Currently at: 16999983
70      Reading . . . Currently at: 17999982
71      Reading . . . Currently at: 18999981
72      Reading . . . Currently at: 19999980
73      Reading . . . Currently at: 20999979
74      Reading . . . Currently at: 21999978
75      Reading . . . Currently at: 22999977
76      Reading . . . Currently at: 23999976
77      Reading . . . Currently at: 24999975
78      Reading . . . Currently at: 25999974
79      Reading . . . Currently at: 26999973
80      Reading . . . Currently at: 27999972
81      Reading . . . Currently at: 28999971
82      Reading . . . Currently at: 29999970
83      Reading . . . Currently at: 30999969
84      Reading . . . Currently at: 31999968
85      Reading . . . Currently at: 32999967
86      Reading . . . Currently at: 33999966
87  Time Elapsed (for compresssion): 71.28672200000437
88
89
```

Research
Spring 2017

Inverted Index Searching
Compression and Read Perf.

Danish Vaid
danishvaid@umail.ucsb.edu

```
90  Database populated with batch writes
91       Time Started: 85876.857805sec
92       Time Ended: 86098.197818sec
93       Time Elapsed: 221.34001299999363sec
94  Database read randomly
95       Time Started: 86156.352479sec
96       Time Ended: 86342.329792sec
97       Time Elapsed: 185.97731300001033sec
98
99  --- Finished Program ---
```