Get real:
How Benchmarks Fail to Represent the Real World

Adrian Vogelsgesang, Michael Haubenschild
Jan Finis, Alfons Kemper, Viktor Leis, Tobias Muehlbauer, Thomas Neumann, Manuel Then
{avogelsgesang, mhaubenschild, jfinis, …}@tableau.com

June 15th 2018
Tableau®

(and other BI tools)
The actual data crunching…

… is delegated to an actual database
The actual data crunching…

… is delegated to an actual database

This could be your system
Tableau Public

- Free cloud hosting for visualizations
- Including both visual specification and raw data

For us: a huge repository of test data

This talk: statistics about 60k visualizations only from Public

Biased towards small datasets, but we can share our findings with you 😊

Over 1 million queries
Our Insights
Many meta data queries

- Column names, data types, … (“SELECT tablename FROM pg_tables;”)
- Current server time (“SELECT NOW;”)
- Feature testing (“Let’s see, can I create a temporary table?”)

All in all: 75% of the queries

Make metadata queries efficient!
Data set sizes

Number of Tuples

<10^1  2,591
<10^2  14,639
<10^3  15,642
<10^4  14,734
<10^5  10,618
<10^6  4,535
<10^7  553
<10^8  45
<10^9  1
Strings are everywhere

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Percentage</th>
<th>Number of Tuple Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>48.8%</td>
<td>8e10</td>
</tr>
<tr>
<td>Integer</td>
<td>25.6%</td>
<td>4e10</td>
</tr>
<tr>
<td>Real</td>
<td>20.5%</td>
<td>2e10</td>
</tr>
<tr>
<td>Datetime</td>
<td>2.6%</td>
<td>2.6%</td>
</tr>
<tr>
<td>Date</td>
<td>1.5%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Boolean</td>
<td>0.9%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>
Strings are everywhere

- ISO country codes, IANA airport codes, ISBNs for books, UUIDs
- Boolean encoded as "0"/"1" (60% of single-character-strings!)
- "male"/"female"

People don’t care about a clean schema – but: they do care about performance
International strings & collation support

- 0.64% of the strings contain non-ASCII characters
- Small fraction, but nevertheless must be supported
- Not covered by benchmarks

Even worse: collations \( (\text{"A"} = \text{"â"}) \)
- 85% of the string columns have a collation
- 70% case- or accent-insensitive
- Makes query optimization harder
- Collations are expensive to evaluate

HashJoin? Anyone?
The queries

- Most are small: Only 0.5% larger than 5KB
- But: Huge outliers
The queries

• Most are small: Only 0.5% larger than 5KB

• But: Huge outliers

• Largest query in our data set: 6.7MB
• Largest query I saw so far: 27MB

And that’s not all due to constant strings…
Expression-heavy queries

Number of scalar expressions

1. 157,128
   - Arithmetics: 56,640
   - String handling: 56,640
   - Cast: 25,640
   - Comparison: 6,640
   - NULL handling: 3,640
   - Boolean: 1,640

2. 18,119
   - Arithmetics: 3,840
   - String handling: 3,840
   - Cast: 2,634
   - Comparison: 1,634
   - NULL handling: 1,634
   - Boolean: 1,634

3. 12,223
   - Arithmetics: 2,585
   - String handling: 2,585
   - Cast: 3,168
   - Comparison: 3,168
   - NULL handling: 3,168
   - Boolean: 3,168

4. 15,758
   - Arithmetics: 3,961
   - String handling: 3,961
   - Cast: 3,168
   - Comparison: 3,168
   - NULL handling: 3,168
   - Boolean: 3,168

5. 3,961
   - Arithmetics: 4,746
   - String handling: 4,746
   - Cast: 4,746
   - Comparison: 4,746
   - NULL handling: 4,746
   - Boolean: 4,746

6. 4,745
   - Arithmetics: 4,745
   - String handling: 4,745
   - Cast: 4,745
   - Comparison: 4,745
   - NULL handling: 4,745
   - Boolean: 4,745

7. 4,381
   - Arithmetics: 4,381
   - String handling: 4,381
   - Cast: 4,381
   - Comparison: 4,381
   - NULL handling: 4,381
   - Boolean: 4,381

8. 3,616
   - Arithmetics: 3,616
   - String handling: 3,616
   - Cast: 3,616
   - Comparison: 3,616
   - NULL handling: 3,616
   - Boolean: 3,616

9. 3,616
   - Arithmetics: 3,616
   - String handling: 3,616
   - Cast: 3,616
   - Comparison: 3,616
   - NULL handling: 3,616
   - Boolean: 3,616

10. 3,616
    - Arithmetics: 3,616
    - String handling: 3,616
    - Cast: 3,616
    - Comparison: 3,616
    - NULL handling: 3,616
    - Boolean: 3,616
Incomplete queries

- Interactive exploration
- Not all queries make sense
- Missing filters, missing join conditions, …
And Benchmarks?
What do benchmarks do? (TPC-H/DS)

• Meaningful queries returning useful results
• Handwritten queries
• Well-designed schema
• Scale the data set size
What does that mean for benchmarks?

1. Include meta-data queries
2. Do bad schema design, use strings more often
3. Include Unicode & collations
4. Benchmark on tiny data sets, too
5. Scale query complexity, not only data size
6. Take into account incomplete/incorrect queries
Questions?
Get real:
How Benchmarks Fail to Represent the Real World

Adrian Vogelsgesang, Michael Haubenschild
Jan Finis, Alfons Kemper, Viktor Leis, Tobias Muehlbauer, Thomas Neumann, Manuel Then
{avogelsgesang, mhaubenschild, jfinis, …}@tableau.com

June 15th 2018
Hyper

(Tableau) + (Hyper) = BOOM!