BG Benchmark for Interactive Social Networking Actions

bgbenchmark.org

Shahram Ghandeharizadeh and Sumita Barahmand
Outline

- Big data
  - Classification
  - 3 Vs
- BG Benchmark
- BG Use Cases
- Conclusions
Big Data: Operations

- Simple
- Complex
- Ad-hoc
- Pre-specified
- Off-line
- Interactive
Big Data: Google Analytics

Objective:
1. Advertising ROI
2. Frequency of access to pages

1. Gather click stream data: Optimized for writes,
2. Compute aggregated data: MapReduce/Hadoop.
Big Data: Google Analytics

Objective:
1. Advertising ROI
2. Frequency of access to pages

1. Gather click stream data: Optimized for writes,
2. Compute aggregated data: MapReduce/Hadoop,
3. Enable users to view aggregated data.
Big Data: Facebook/Twitter

- Show profile page of Farah Fawcett
- Follow Barack Obama
- Friend Miley Cyrus
3 Vs: Facebook

- **High Volume:**
  - 1.2 billion user profiles, 150 billion friend connections, 1.13 trillion likes, 17 billion tagged locations, 240 billion photos, ...

- **High Velocity:**
  - 700 million active users daily, 4.5 billion likes daily, 350 million photos uploaded daily, ...

- **High Variety:**
  - Mix of data types: Structured records, multimedia content, text.

Motivated by a real need in 2011:

- Plug-and-play JDBC wrapper (KOSAR) that enhances performance of SQL solutions by more than a factor of 100.
  - Physical data independence: Transparent caching,
  - Strong consistency: ACID.
- Evaluate KOSAR for social networking workloads!
### Benchmark for interactive social networking actions

<table>
<thead>
<tr>
<th>Action</th>
<th>facebook</th>
<th>Google+</th>
<th>Twitter</th>
<th>LinkedIn</th>
<th>YouTube</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Profile (VP)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>List Friends (LF)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>View Friend Requests (VFR)</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Invite Friend (IF)</td>
<td>✓</td>
<td>Add to Circle</td>
<td>Follow</td>
<td>✓</td>
<td>Subscribe</td>
</tr>
<tr>
<td>Accept Friend Request (AFR)</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Reject Friend Request (RFR)</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Thaw Friendship (TF)</td>
<td>✓</td>
<td>Remove from Circle</td>
<td>Unfollow</td>
<td>✓</td>
<td>Unsubscribe</td>
</tr>
<tr>
<td>View Top-K Resources (VTR)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>View Comments on a Resource (VCR)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Post Comment on a Resource (PCR)</td>
<td>✓</td>
<td>✓</td>
<td>Reply to a tweet</td>
<td>Recommend a colleague's work</td>
<td>Post comment on a video</td>
</tr>
<tr>
<td>Delete Comment from a Resource (DCR)</td>
<td>✓</td>
<td>✓</td>
<td>Delete the reply for a tweet</td>
<td>Withdraw recommendation</td>
<td>Remove comment on a video</td>
</tr>
</tbody>
</table>

System Performance as a Single Number

SLA: 95% of requests observe a response time equal to or faster than 100 msecs with at most 0.001% of requests observing unpredictable data for 10 mins.
SoAR: Social Action Rating

SLA: 95% of requests observe a response time equal to or faster than 100 msecs with at most 0.001% of requests observing unpredictable data for 10 mins.
SLA: 95% of requests observe a response time equal to or faster than 100 msecs with at most 0.001% of requests observing unpredictable data for 10 mins.

“A good sketch is better than a long speech…”
— Napoleon Bonaparte
Quick & Automated Rating Process

- Uses intelligent techniques to rate the system in a reasonable amount of time.
  - Agile data loading techniques to reduce the overhead of constructing the benchmark database.
  - Heuristic search within 10% margin of error which potentially reduces the number of conducted experiments.

Quick & Automated Rating Process

- Uses intelligent techniques to rate the system in a reasonable amount of time.
  - Agile data loading techniques to reduce the overhead of constructing the benchmark database.
  - Heuristic search within 10% margin of error which potentially reduces the number of conducted experiments.

Quick & Automated Rating Process

- Uses intelligent techniques to rate the system in a reasonable amount of time.
  - Agile data loading techniques to reduce the overhead of constructing the benchmark database.
  - Heuristic search within 10% margin of error which potentially reduces the number of conducted experiments.

Quick & Automated Rating Process

- Uses intelligent techniques to rate the system in a reasonable amount of time.
  - Agile data loading techniques to reduce the overhead of constructing the benchmark database.
  - Heuristic search within 10% margin of error which potentially reduces the number of conducted experiments.

Use Cases: Evaluate

- **Novel hardware:**
  - Networking gear.
  - Multi-core CPUs: Vertical scalability.
  - Mass storage devices, e.g., SSD.

- **Novel software architectures:**
  - Cache augmented relational data stores.
  - NoSQL: Document stores, key-value stores, extensible stores.

- **Alternative system configurations.**
Example: Physical Data Design

- **Industrial strength SQL solution**
  - Basic: 300
  - Boosted: 11,820

- **Changes:**
  - Manage images/friendships effectively.

- **MongoDB as a document store**
  - Basic: 0
  - Boosted: 7,700

- **Changes:**
  - Inline thumbnail/profile image with a document.
  - Use arrays.

Consistency & Availability

- A performance metric, SoAR, as a function of the amount of unpredictable data. Investigate CAP theorem & alternative consistency models (BASE/ACID).
Consistency & Availability

- A performance metric, SoAR, as a function of the amount of unpredictable data. Investigate CAP theorem & alternative consistency models (BASE/ACID).

SoAR

Unpredictable data
BG: Summary

- BG is a scalable benchmark to rate alternative
  - Hardware,
  - Software,
  - Data store configurations/designs/architectures,
  for read and write of small amount of data (OLTP style) from big data.

- Target application maintains a social graph in support of interactive actions.
Visit [http://bgbenchmark.org](http://bgbenchmark.org) to download BG, review documentation, published papers, client code for MongoDB, VoltDB, Microsoft Azure, CouchBase.

Contact us for question and help:
- [barahman@usc.edu](mailto:barahman@usc.edu)
- [shahram@usc.edu](mailto:shahram@usc.edu)

Comments & suggestion are most welcomed!