CS244b

Dynamo (2007)

- Eventual consistency
- Quorum Systems
- Tail Latency
- Trade-offs in messy real-world systems
key-value put/get
no SQL
Availability response time
TAIL LATENCY

Atomicity
Consistency
Isolation
Durability
Figure 1: Service-oriented architecture of Amazon’s
get (key) -> (context, list of values)
put (key, context, value) -> void

- Replicate on N
- Heterogeneous node
- Minimize Churn
S = #nodes, T = tokens/node, Q = #partitions (#2-3)
Figure 6: Fraction of nodes that are out-of-balance (i.e., nodes
Figure 8: Comparison of the load distribution efficiency of
Quorum Systems

- write to \( W \) servers
- read from \( R \) servers
- \( R + W > N \)
- Write
  - Ask R replicas for vers#
  - Pick new vers# higher
  - Send new version w. # to servers
  - return if/when hear from W

- Read
  - ask for latest vers. (value, vers)
  - received & matching replies → done
  - Re-broadcast latest (val, vers)
Malkhi & Reider
Byzantine Faulty Systems

Reader needs \( R - \ell + 1 \) identical replies

\( R + W - N > 2F \)
Dynamo Request Flow

load-balancers

opt. 1

opt. 2
Nodes B, C and D store keys in range (A,B) including K.
Table 2: Performance of client-driven and server-driven coordination approaches.

<table>
<thead>
<tr>
<th></th>
<th>99.9th percentile read latency (ms)</th>
<th>99.9th percentile write latency (ms)</th>
<th>Average read latency (ms)</th>
<th>Average write latency (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server-driven</td>
<td>68.9</td>
<td>68.5</td>
<td>3.9</td>
<td>4.02</td>
</tr>
<tr>
<td>Client-driven</td>
<td>30.4</td>
<td>30.4</td>
<td>1.55</td>
<td>1.9</td>
</tr>
</tbody>
</table>
Node Failure

Key K

Nodes B, C and D store keys in range (A,B) including K.
Conflicts

Network partition
Concurrent client updates
Consistency

Syntactic if

Vector clock

\[ u_1 < (A-1, B-2) \]
\[ u_2 < (A-1, B-3) \]

\[ u_1 \leq u_2 \text{ iff } \forall s \ u_1(s) \leq u_2(s) \]
\langle A^{-1}, B^{-1} \rangle

\langle A^{-2}, B^{-1} \rangle \; \langle A^{-1}, B^{-2} \rangle

\underline{keep timestamps on U.C. entries
truncated at 10 entries (discard oldest)}

\langle A^{-1}, B^{-1}, C^{-1} \rangle \; \langle A^{-1}, B^{-1}, C^{-1}, D^{-1} \rangle