TPC-W

a Benchmark for E-commerce

Daniel A. Menascé
Dept. of Computer Science
George Mason University

Workload Characterization
Results for E-commerce

- Workload characterized at the session level: Customer Behavior Model Graph.
- Some findings:
  - Session lengths are heavy-tailed
    - most sessions last less than 1000 sec
    - 88% of the sessions have less than 10 requests
  - Growing agent activity:
    - 33% of the requests generated by robots
  - Power law distributions:
    - popularity of search terms follows Zipf’s Law
  - Predictability on fine time scales:
    - long range dependency between 4 and 4096 sec
Customer Behavior Model Graph (CBMG)

Session Length (in requests to execute e-business functions)
Robot Characterization

- Robots generate a non-negligible part of the traffic served by e-business sites
- Different types of robots: crawlers, meta-search robots, etc.
- Trend: agents will play an increasingly important role in the Web.
- Question: how can we identify robot’s presence in the workload?
Robot Characterization: multi-layer criteria

- **Session layer**
  - type of functions in a session
  - session length
- **Function layer**
  - entry point
  - unlikely functions
  - embedded files
- **Request layer**
  - inter-arrival time
  - self-identification

IAT for Human Generated Robots

Shopbots: Exponential IAT
IAT for Automatically Generated Robots

Crawlers:
Lognormal IAT

Robot Activity Characterization for E-commerce

- Robots can consume considerable resources.
- Crawlers consume more resources than shopbots.
- Utilization peaks increase in intensity with finer time scales.
- Robots significantly increase the miss ratio of server-side caches.
- Crawlers have a reference pattern that completely disrupt reference locality assumptions.
- Caches and servers should treat human- and robot-generated requests differently.
Robot Disk Utilization at Different Time Scales
bins of 14400 sec

Disk utilization (%)

peaks: 15%, 20%, 25%.

Robot Disk Utilization at Different Time Scales
bins of 1800 sec

Disk utilization (%)

peaks: 20%, 30%, 40%, 45%, and 65%.
Robot Disk Utilization at Different Time Scales

bins of 60 sec

Disk utilization (%)

peaks: 40%, 65%, 80%, and 100%.

TPC-W: an E-commerce benchmark by the TPC

- www.tpc.org
- Designed to mimic operation of an e-commerce site (e-tailer).
- Scalable in number of concurrent users and in the database size.
- Transactions generated by TPC-W include:
  - Browsing activities (e.g., browse, search, select, view product detail)
  - Product order activities (e.g., shopping cart, login, register, buy request, and buy confirm)
- Database transactions must be ACID.
- Security through SSL is used for authentication.
TPC-W Metrics: throughput and cost/throughput

- **WIPS** (Web Interactions Per Second) during shopping mix sessions. Specified as \( \text{WIPS} @ \text{number\_items} \)
- **WIPSb** - Web Interactions Per Second during browsing mix sessions.
- **WIPSo** - Web Interactions Per Second during ordering mix sessions.
- **Cost/Performance**
  \[
  \frac{\text{Hdw Cost} + \text{Softw Cost} + \text{Maint. Cost}}{\text{WIPS}}
  \]

Example of TPC-W for 10,000 Items in the Catalog

<table>
<thead>
<tr>
<th>Rank</th>
<th>System</th>
<th>WIPS</th>
<th>$/WIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>5,745</td>
<td>$69.00</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>3,130</td>
<td>$67.50</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>3,008</td>
<td>$81.77</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>1,262</td>
<td>$277.08</td>
</tr>
</tbody>
</table>

- the total price of System A is $396,405, i.e., 5,745 \times $69.00.
- System D costs almost the same, i.e., $349,675 but can only deliver 22% of the maximum throughput measured in WIPS.
TPC Emulated Browsers (EBs)

- Each EB starts a session and generates all requests of that session.
- The minimum duration of a session (USMD) is exponentially distributed with mean 15 minutes, truncated at 60 minutes.

Requests are separated by user think times (Z), which are exponentially distributed with mean 7 sec truncated at 70 sec.

Response Time Law:
\[ R = \frac{\text{(No. EBs)}}{\text{WIPS}} - Z \]
Use of Response Time Law to TPC-W Results

<table>
<thead>
<tr>
<th>Rank</th>
<th>System</th>
<th>WIPS</th>
<th>$/WIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>5,745</td>
<td>$69.00</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>3,130</td>
<td>$67.50</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>3,008</td>
<td>$81.77</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>1,262</td>
<td>$277.08</td>
</tr>
</tbody>
</table>

$R = \frac{\text{No. EBs}}{\text{WIPS}} - Z$

- Assume 50,000 concurrent users.
- System A: $R = \frac{50,000}{5,745} - 7 = 1.7$ sec.
- System D: $R = \frac{50,000}{1,262} - 7 = 32.6$ sec.
TPC-W Types of Sessions

- Browsing mix: 95% browse interactions and 5% ordering interactions – 0.69% buy to visit ratio.
- Shopping mix: 80% browse interactions and 20% ordering interactions – 1.2% buy to visit ratio.
- Ordering mix: 50% browse interactions and 50% ordering interactions – 10.18% buy to visit ratio.

TPC-W Database

- Minimum of 8 tables with defined minimum number of fields:
  - Customer: Customer name and ID information,
  - Address: Customer address data,
  - Country: Country name and exchange rate information,
  - Order: Order total and shipping information,
  - Order line: Order line item data,
  - Credit card: Credit card data,
  - Item: Book information, and
  - Author: Author data.
- TPC-W provides a function (WGEN) to generate the item title and the author last name in the database. We wrote a program to populate all other fields.
TPC-W Transactions

- **Browse**: typical browsing behavior.
- **Shopping Cart**: medium weight R/W database transaction; high frequency; stringent response time requirement.
- **Buy**: medium weight R/W database transaction; low frequency; fast response time requirements. SSL is used.
TPC-W Transactions (cont’d)

- Register: post user data through HTML forms.
- Search: heavy read only transaction; low frequency; relaxed response time and relaxed consistency.

TPC-W Scalability

Number of Emulated Browsers (1, 2, …, M)

CUSTOMER = Number of EBs x 2,880

ORDER = Number of Customers x 0.9

ADDRESS = Number of Customers x 2

ORDER_LINE = Number of Orders x 3

CC_XACTS = Number of Orders

Number of Items (1k, 10k, 100k, 1M, 10M)

Author = Item ^ 0.25

Assume 100 EBs:
- 288,000 rows in CUSTOMER table
- 259,200 rows in the ORDER table
- 576,000 rows in the ADDRESS table
- 777,600 rows in the ORDER_LINE table
- 259,200 rows in the CC_XACTS table