Tuning Tableau Server for High Performance

“I wanna go fast”

PRESENTED BY
Francois Ajenstat
Alan Doerhoefer
Daniel Meyer
Agenda

• What are the things that can impact performance?

• Tips and tricks to improve Server Performance

• How do you troubleshoot performance issues?

• What did we do in 7.0 to improve scalability and performance
What do you mean by “it’s slow”?

- Is a workbook loading slowly?
- Are interactions slow?
- Does the system slow down over time?
- Can you not support the # of desired users?
Key Characteristics of Performance

Workbook

Deployment Scenario

Environment
Not all views are created equal

1. Fast
   - Fast query
   - Fast rendering

2. Query-bound
   - Slow query
   - Fast Rendering
   - Same as Fast once query is cached

3. CPU-bound
   - Render-Intensive
   - Uses the most CPU
   - Example: text tables with thousands of cells
Deployment scenarios with different performance profiles

• Single vs. multi-user (cache warm up)
• Fixed size vs. non-fixed size dashboards
• User filtering
• Ratio of users to distinct workbooks
• Level of interaction (none, tooltip, selection, filtering)
Your Client and Server environment matters

- Browser type
- Network speed
- Mobile
- Hardware
How To Improve Server Performance

1. Learn the architecture
2. Use appropriate hardware
3. Use distributed server
4. Use extracts
5. Tune
6. Monitor
7. Troubleshoot
8. Understand the load test results
1. Learn the Architecture

- Web Browser
- Web Applications
- Tableau Professional
- Extract Storage
- HTTP(S) Server
- VizQL Server
- Web App
- Active Cache
- User Mgmt (AD)
- Search
- Repository
- Data Sources

Interactions:
- Web Browser interacts with server content
- Tableau Professional connects directly to data sources
- Extract Storage

Diagram: 2011 Tableau Customer Conference
2. Use Appropriate Server Hardware

Good* Server for performance:

• Windows Server 2008 64 bit
• 16 GB+ RAM
• Fast RAID-5 Drives
• Gigabit Network Card
• 8 Fast CPU cores

*But Tableau Server will also work on lower class hardware
3. Use distributed server

Tableau Server – 1 Primary, Any configuration of Workers

Primary Server
- Internal Storage
- Scheduler
- Load Balancer
- User Mgmt (AD)
- Search
- Repository

Workers
- HTTP Server
- VizQL Server
- Active Cache
- Web App

Workers - VizQL
- VizQL Server
- Active Cache

External Storage

Data Sources

Internal Storage

External Storage

Load Balancer

Scheduler

User Mgmt (AD)

Search

Repository

HTTP Server

VizQL Server

Active Cache

Web App

Active Cache

VizQL Server

Active Cache

VizQL Server

Active Cache

Tableau Server – 1 Primary, Any configuration of Workers

2011 TABLEAU CONFERENCE
3. Use distributed server

Determining the number of server processes

Use $2 \times$ the number of processor cores

Example

3 server cluster (2 workers)
8 cores per machine

• Use $2 \times 8 = 16$ App processes on one worker
• Use $2 \times 8 = 16$ VizQL processes on the other worker
Avoid ‘out of memory’ errors

\[ \text{maximum_throughput} = \frac{\text{session_capacity} \times \text{num_procs}}{\text{session_expiry}} \]

1. Find your peak throughput from maintenance server activity page

2. Worst case session capacity is 60 sessions / process

3. Balance shorter sessions against having more processes and hence more cache misses.

4. Parameters:
   - vizqlserver.session.expiry.timeout: 120

5. Ensure you have the same number of wgserver and vizqlserver procs. Do not exceed 2 * n where n is number of cores.

6. worker0.wgserver.procs: 8
   - worker0.vizqlserver.procs: 8
Optimize Cache Hits

- Caching is per process and cache cannot be shared amongst processes.

- Load balancing over processes is round robin. If you have 16 processes you will have to repeatedly load a particular workbook 16 times to populate the cache.

To maximize cache hits:
- Minimize number of processes.
- Balance need for speed with demands of memory and concurrency.
4. Use extracts – Speed up your queries

Usage and query performance
• The queries are very ad-hoc.
• Response time needs to be measured in seconds.
• Need to avoid “performance cliffs”.

Data sizes
• Data size is millions of rows rather than billions or trillions.
• The working set will typically fit into physical memory.

Hardware
• Needs to run effectively on a business user laptop or desktop and on a corporate server.
• Need to analyze data offline.
4. Use Extracts:
Tableau Server and Data Engine

• Data Engine is the default Extract Storage for Server

• When documents with Extracts are uploaded to Server, the data is loaded into the Data Engine Server.

• Data Engine can be hosted on a single node deployment, on a worker, or on its own node.
5. Tune the OS

• Disable other processes

• Consider using Server Core

• On Windows Server 2003
  Change TCP Setting:
  Set SynAttackProtect to zero:
  [HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters]
  "SynAttackProtect"=dword:00000000
5. Tune Tableau Server

Session Timeouts

Can be changed in the configuration file

C:\ProgramData\Tableau\Tableau Server\config\tabsvc.yml

• WG Session
Default is 4 hours
To change it to 1 hour add this to tabsvc.yml:
wgserver.session.idle_limit: 60

• VizQL Session
Default is 2 hours
To change it to 30 minutes add this to tabsvc.yml:
vizqlserver.session.expiry.timeout: 30
6. Monitor Tableau Server

Server Status

Maintenance

Status

<table>
<thead>
<tr>
<th>Machine</th>
<th>Repository</th>
<th>Server Web Application</th>
<th>VisQL Server</th>
<th>Background Tasks</th>
<th>Web Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>172.16.17.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.16.17.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key:
- ✔️ Service up, waiting for request
- ✗ Service up, currently handling request
- ⚠ Service down

Analysis

Click on the views below to display the corresponding analysis.

<table>
<thead>
<tr>
<th>View</th>
<th>Analysis</th>
</tr>
</thead>
</table>
6. Monitor Tableau Server

Server Activity

Sessions and Requests

Activity by user

Activity by server process
6. Monitor Tableau Server

User Activity

<table>
<thead>
<tr>
<th>User</th>
<th>IP Address</th>
<th>Last Activity Time</th>
<th>Currently Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>adoerhoef</td>
<td>172.16.16.249</td>
<td>07/17/2009 02:54:32 PM</td>
<td>Currently Active</td>
</tr>
<tr>
<td>rmajedd</td>
<td>172.16.16.235</td>
<td>07/17/2009 11:49:04 AM</td>
<td>Idle</td>
</tr>
<tr>
<td>iheath</td>
<td>172.16.16.245</td>
<td>07/17/2009 11:35:10 AM</td>
<td>Idle</td>
</tr>
<tr>
<td>pdavis</td>
<td>172.16.16.238</td>
<td>07/17/2009 11:34:47 AM</td>
<td>Idle</td>
</tr>
<tr>
<td>hboehm</td>
<td>127.0.0.1</td>
<td>07/17/2009 10:52:20 AM</td>
<td>Idle</td>
</tr>
</tbody>
</table>

Session Activity for User adoerhoef

<table>
<thead>
<tr>
<th>Service</th>
<th>Action</th>
<th>Action Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>workgroup</td>
<td>create</td>
<td>errors, index, quicksort, select, setportsize, show, size, tooltip, update</td>
</tr>
</tbody>
</table>

Select rows in User Activity
6. Monitor Tableau Server

- Apache logs –
  C:\ProgramData\Tableau\Tableau Server\logs

  Contains information about every request to server at the HTTP request level
6. Monitor System

- Monitor resources with perfmon/logman
- Use the admin server activity page to find peak sessions/minute.
- Correlate peak traffic with resource consumption.
- Correlation will provide route for scaling with workers. For example if vizqlserver is pushing 90% CPU utilization during peak throughput you should add vizqlserver procs to a worker machine. If tdeserver is bottlenecking adding vizqlserver processes will not help.
7. Troubleshoot Server

WG logs and VizQL logs:
Contain details of each operation performed by the server

C:\ProgramData\Tableau\Tableau Server\data\tabsvc\wgserver
C:\ProgramData\Tableau\Tableau Server\data\tabsvc\vizqlserver\Logs
7. Troubleshoot Server

*tabadmin ziplogs* command

collects all logs – even on a distributed cluster – into one archive
Understand the Load Test Results (v7)

- 8 core server; 16G RAM
- Single view viz
- Initial load
- # of concurrent users*

*marketing says that concurrency = 10% of actual named users
Understand the Load Test Results (v7)

- Applying a filter
Summary

1. Learn the architecture
2. Use appropriate hardware
3. Use distributed server
4. Use extracts
5. Tune
6. Monitor
7. Troubleshoot
8. Understand the load test results
Please evaluate this session (TCC11 402)
Tuning Tableau Server for High Performance

Text to **32075**
In the body of the message, type: **TCC11 <space> 402** then letters from the table below to indicate each response.
Provide additional comments after an asterisk “*”
Sample text: **TCC11 402a**ho*That was great!*

<table>
<thead>
<tr>
<th>Please give your response to the following:</th>
<th>Excellent</th>
<th>Great</th>
<th>Good</th>
<th>Average</th>
<th>Poor</th>
<th>Bad</th>
<th>Very Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>What was the value of this session to you?</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>g</td>
</tr>
<tr>
<td>What are the chances you will apply what you learned in this session in your work?</td>
<td>h</td>
<td>i</td>
<td>j</td>
<td>k</td>
<td>l</td>
<td>m</td>
<td>n</td>
</tr>
<tr>
<td>What are the chances you would recommend this session to a colleague?</td>
<td>o</td>
<td>p</td>
<td>q</td>
<td>r</td>
<td>s</td>
<td>t</td>
<td>u</td>
</tr>
</tbody>
</table>

Each text evaluation you send enters you into a drawing for an iPad!