Cache, cache, cache - Visionary's Guide to configuring Tableau Cache

Mark Wu (Tableau Visionary)
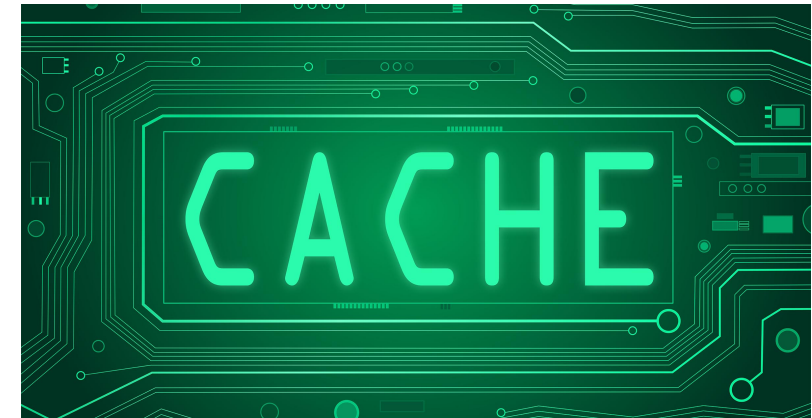
✉ markwu2000@gmail.com   🌐 enterprisetableau.com
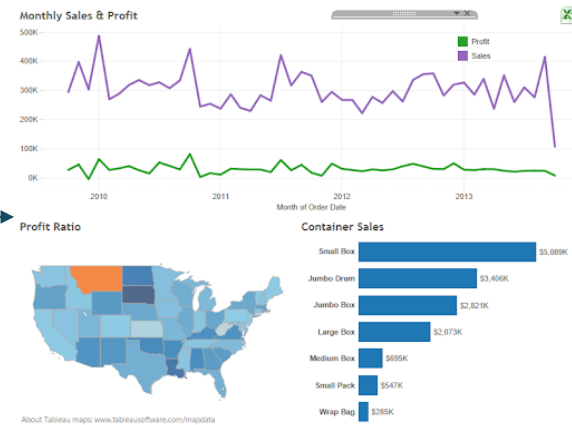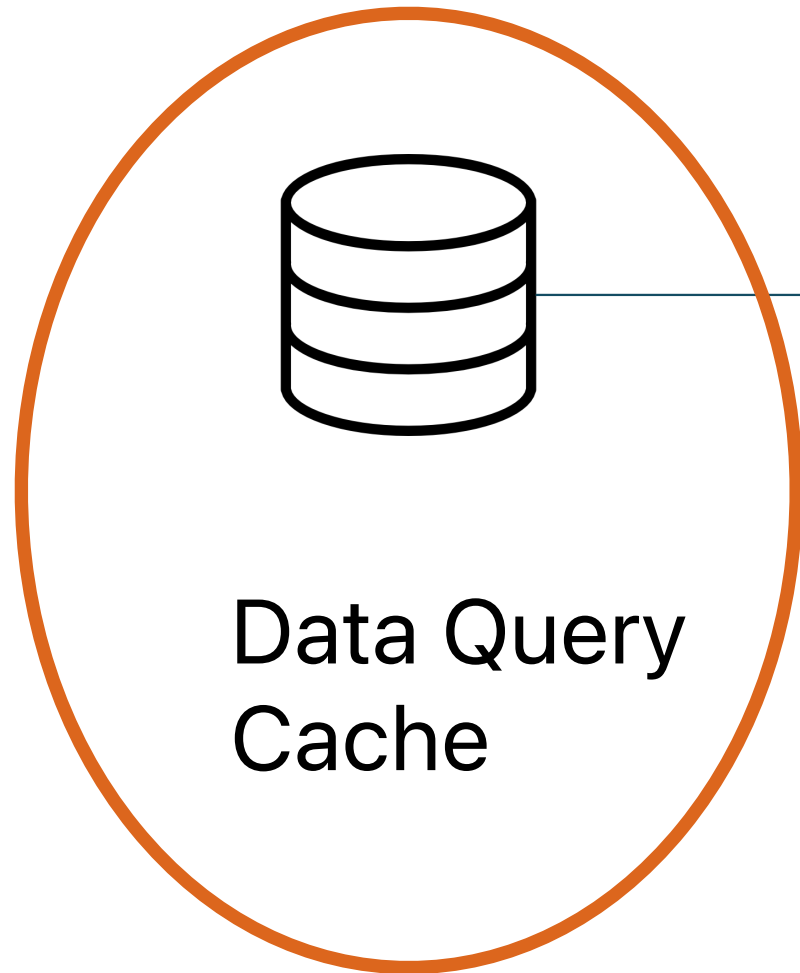
What is Cache

Cache stores recently used query results so that it can be quickly rendered at a later time without querying source data again

Cache always happens

- access the view
- run subscription
- API ....

Cache happens all layers

Data Query Cache

Render Tile Cache

Browser (client side) Cache

Default Cache policy on Tableau Cloud

12 hrs (not customizable)

Default Cache policy on Tableau server

Cache and reuse **as long as possible**

```
(tsm data-access caching set -r low)
```

Tableau Server Cache Config

If mostly Live Connections

tsm data-access caching set -r always

Always get the latest data : cache refreshes each time a page is reloaded.

Server: **Mixed** Live & Extract (most common)

💡 **Recommend: Longer Cache Policy (12-24 hrs) with cache warmup**

General   Extensions

(1) **tsm data-access caching set -r 720**

12+ Hrs

Workbook Performance after a Scheduled Refresh

Recently viewed workbooks with scheduled refreshes can be pre
faster. Learn more

(2) ☑ Pre-compute recently viewed workbooks

(3) Live: Workbook level data refresh
policy **by owner**   (v2021.3)

(4) Extract: View Acceleration **by owner**
          (v2022.1)

#1: `tsm data-access caching set -r <value>`

- **low or empty string ("")**. Default. Aways use cached data
- → <value>. "<value>" maximum minutes data should be cached
- **always** or **0** (zero). Always get the latest data

1. *Server (not site)*
2. *Tableau Cloud 720 mins (12 hrs)*
3. *Mark's server recommendation : 720-1440 mins*
4. *All workbooks (live or extract) except ...*
   a. *Row Level Security*
   b. *Data fresh policy on live workbooks*

#2: Pre-compute Query Cache After Extract

Workbook Performance after a Scheduled Refresh

Workbooks with scheduled refreshes can be pre-computed to open faste

✓ Pre-compute workbooks viewed recently.

warm up cache only for frequently used workbooks (not all workbooks)

#2: Pre-compute Query Cache After Extract

By default, $$\frac{\text{\# of views of the workbook last 7 days}}{\text{\# of refresh in next 7 days}} \geq 2$$

Workbook Performance

CPU and memory

#2: Pre-compute Query Cache After Extract

By default, $\dfrac{\text{\# of views of the workbook last 7 days}}{\text{\# of refresh in next 7 days}}$ >= 2
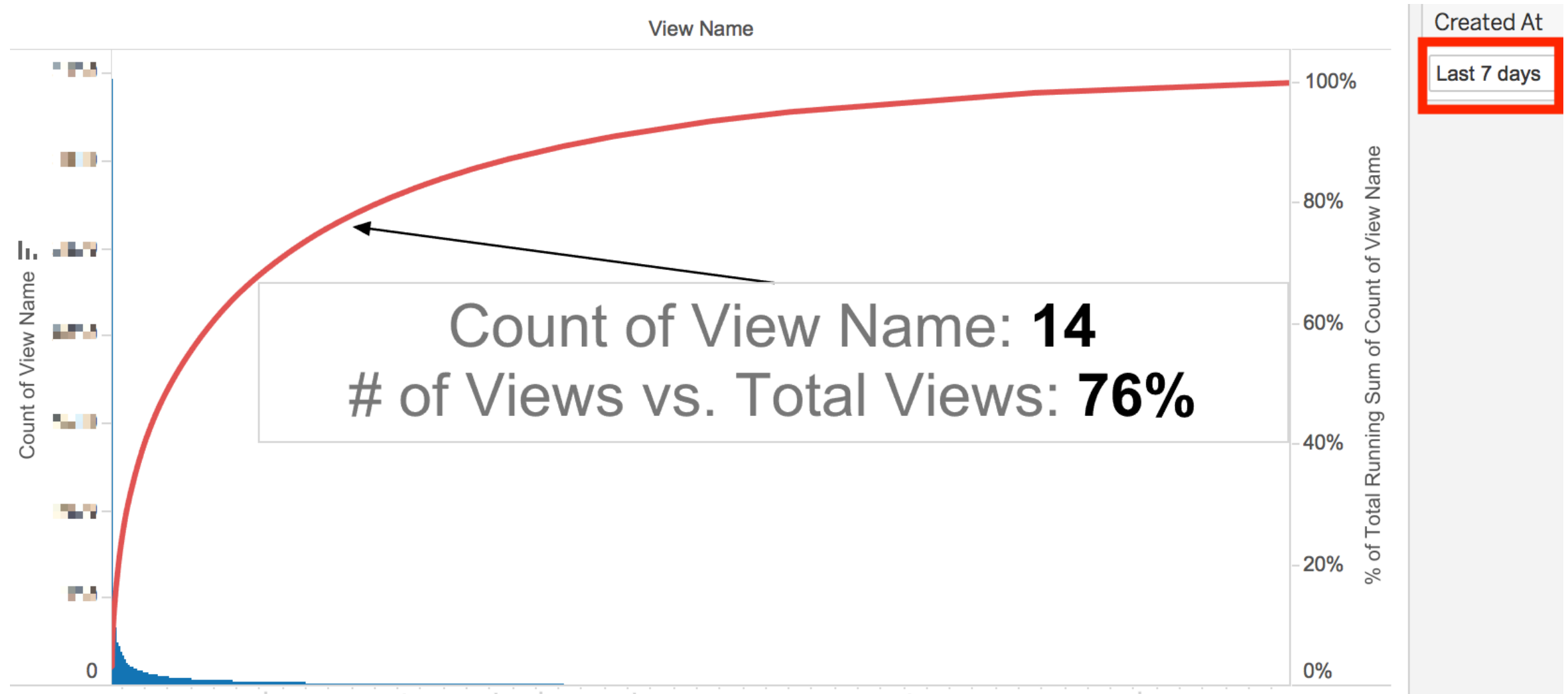
backgrounder.externalquerycachewarmup.view_threshold 2

Implement Pre-compute Query Cache After Extract

1. **Enabled by default**: *backgrounder.externalquerycachewarmup.enabled* true

2. **Config**: *backgrounder.externalquerycachewarmup.view_threshold* 2
   (# of views of the workbook last 7 days / # of refresh in next 7 days)

3. **Observe & Adjust**

Deciding Cachewarmup Threshold?

Count of View Name: **14**
# of Views vs. Total Views: **76%**

Created At

Last 7 days

Download workbook   https://enterprisetableau.com/cachewarmup/

#3: Live – Set Data Refresh Policy **by owner**
(v2021.3)

#4: Extract - View Acceleration **by owner**
(v2022.1)
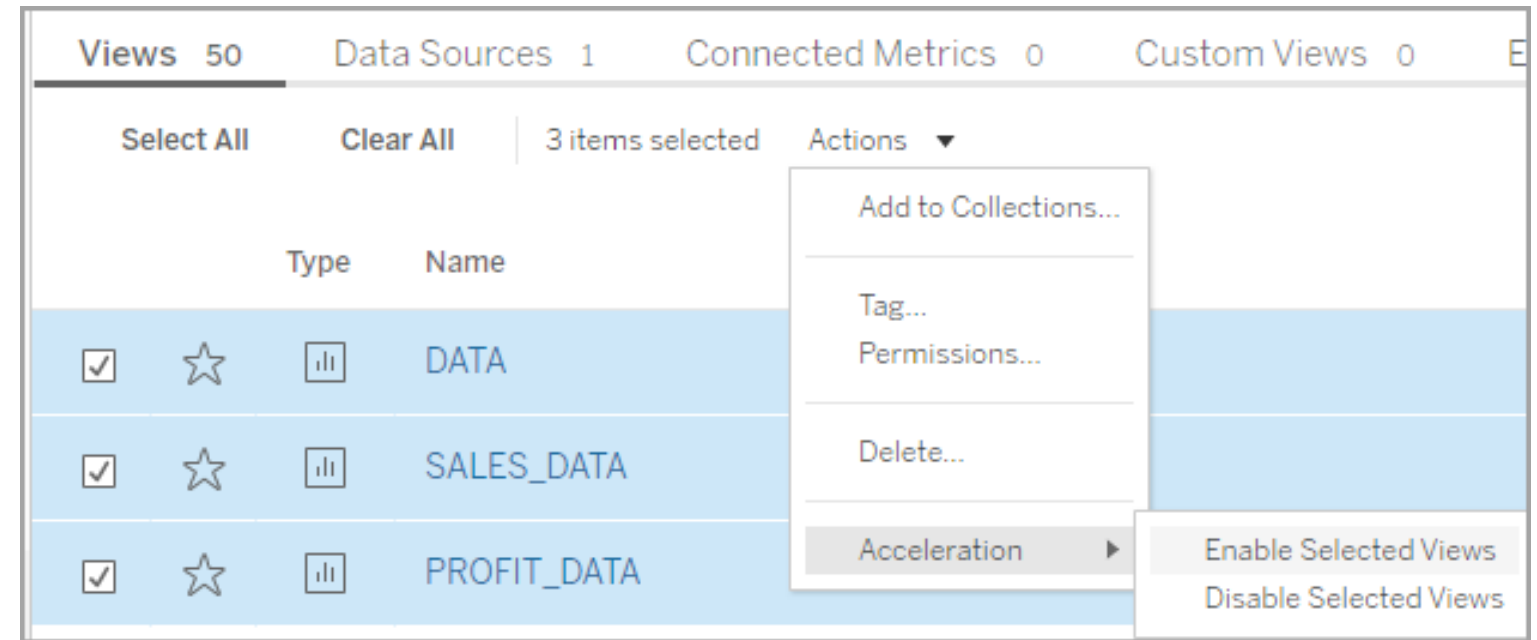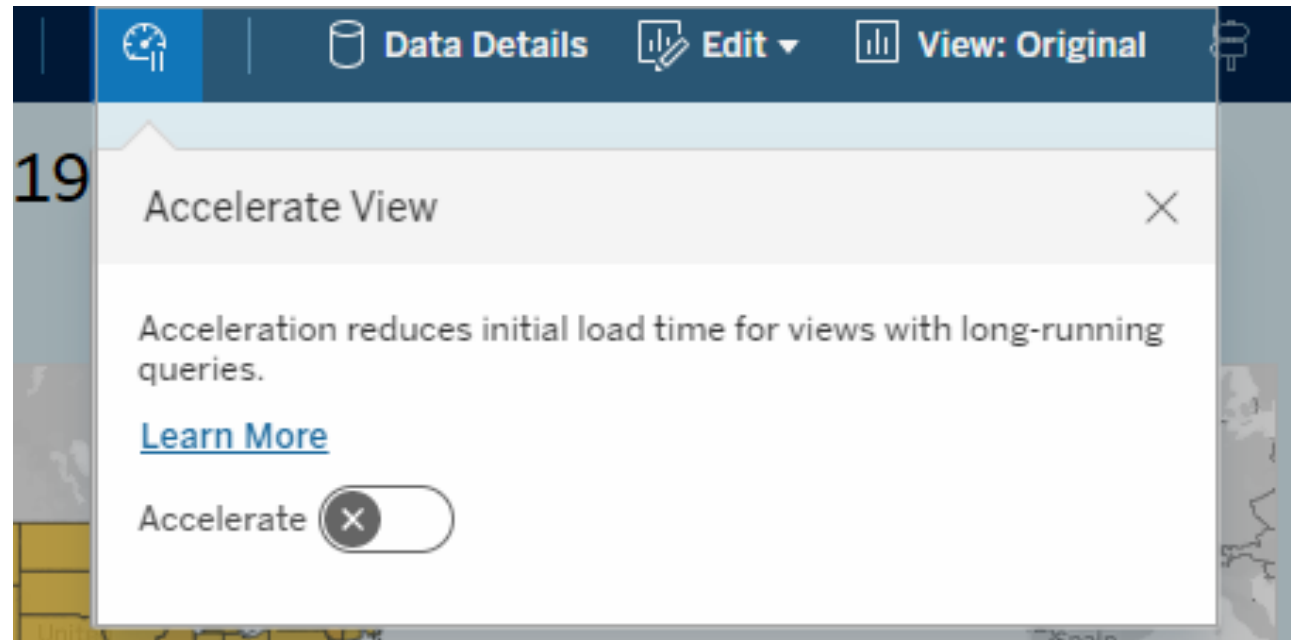
View Acceleration : Force pre-compute cache after extract refresh for the specific queries used in the view only

🔴 View Acceleration vs Pre-compute Query Cache After Extract

What	Pre-compute Query Cache	View Acceleration
Who controls	Admin	Content Owner
Trigger	Extract refresh	Extract refresh
Scope	Extract workbooks only	Extract workbooks only
Exclude RLS workbooks	Yes	No but only for thumbnail generation user
All frequently used workbook	Yes	Selected workbook
Exclude fast workbooks	No	Yes
Exclude refresh often workbooks	In context of usage	Yes - 6 times or more refresh/ daily
Exclude inactive owner workbook	No	Yes
If embedded credentials expired	No warm up	No warm up
If extract failed	No warm up	No warm up
Site level control	Enable or not	site limit
Admin blacklist	Not possible	Yes

Test your understanding........

.......with a Quiz!!!

```
Quiz 1:

tsm data-access caching set -r <value>
```

Workbooks do not follow the above server setting

1.Row Level Security **- no cache (live or extract)**
2.Data freshness policy **(live only)**

Quiz 2: Cache & Extract

```
tsm data-access caching set -r 720(or 1440)
```

Workbook w/o RLS, if
- 7:00am: view accessed by an user
- 8:00am: extract refreshed

- **9:00am: John clicks the view, would John get the changed data?**

✓ YES

✗ NO

➡ ? MAYBE

if pre-compute workbooks flag set or not

Quiz 3: Cache & Extract

```
tsm data-access caching set -r 720(or 1440)
```

Workbook w/o RLS **and cache warm-up flag checked,** if
- 7:00am view accessed by an user,
- 8:00am: extract refreshed

- **9:00am: John clicks the view, would John get the changed data?**

Workbook Performance after a Scheduled Refresh

Workbooks with scheduled refreshes can be pre-computed to open faste

☑ Pre-compute workbooks viewed recently.

✓ YES

✗ NO

? MAYBE

if workbook is frequently used or not

Quiz 4: Cache & Extract

```
tsm data-access caching set -r 720(or 1440)
```

Workbook w/o RLS **and cache warm-up flag checked** **and workbook is the frequently used**, if
- 7:00am view accessed by an user,
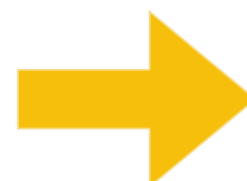- 8:00am: extract refreshed

- **9:00am: John clicks the view, would John get the changed data?**

Workbook Performance after a Scheduled Refresh

Workbooks with scheduled refreshes can be pre-computed to open faster

☑ Pre-compute workbooks viewed recently.

✓ **YES**

✕ **NO**

❓ **MAYBE**

possible browser cache, refresh

Quiz 5: Cache & Extract

`tsm data-access caching set -r 720(or 1440)`

Workbook warm-up:
- 7:00am ... an user,
- 8:00am: extract refreshed

- 9: ... view, ... ta?

Use Case:
Critical workbook, get latest data always?

Answers:
1. **Live: Data refresh policy 2021.3**
2. **Extract: View Acceleration 2022.1**

Workbook Performance after a Scheduled Refresh

Workbooks with scheduled refreshes can be pre-computed to open faster

☑ Pre-compute workbooks viewed recently.

✓ YES — Frequently used

✗ NO

Less Frequently used

Re-cap: Tableau Tools Available for Cache

	Command	Notes
1	tsm data-access caching set -r <value>	Server level
2	Pre-compute recently viewed workbooks	Server & Site level
3	backgrounder.subscription_image_caching	Server level
4	No cache for Row Level Security workbook	Workbook level
5	Data acceleration (for less frequently used workbook)	Workbook level
6	Force a refresh • ?:refresh=yes • "Refresh" button • Reload browser	Workbook level
7	Set workbook level policy (Live Connection only)	Workbook level

Server CPU and Memory Control with 12-24 hrs Cache

Workbook Performance

CPU and memory

| 1 | Shorter VizQL Session Timeout | 3 mins (default 30 mins) |

vizqlserver.querylimit –v 180

3 mins

Avoid one badly designed workbook bringing whole server to its knees

Behavior of VizQL mem
80% limit : Reclamation
95-100% limit : Restart

Problem: Restart happens too late

Question: How to let VizQL restart earlier enough (60% memory) before damage made

Server Status

VizQL Server	✓ ✓ ✓ ✓ ✓ ✓
Cache Server	✓ ✓ ✓ ✓ ✓ ✓
Search & Browse	✓
Backgrounder	

native_api.memory_limit_enabled -v true -frc
native_api.memory_limit_per_process_gb -v xx -frc

Q: What is the value for host has 300GB with 4 VizQL w/o File Store?

A: 50 (50GB x 4 = 200GB is 66% of 300GB)

VizQL Server	Each of the 4 VizQL →	✓ ✓ ✓ ✓
Cache Server	has memory 50G	✓ ✓ ✓ ✓
Search & Browse	limit	✓
Backgrounder		
Data Server		✓ ✓ ✓ ✓

1. No measurable impact on rolling recycle VizQL process
2. Undocumented feature but works perfect!
3. This setting applies to Hyper process as well

Process

Cluster Controller ✅

Gateway

Application Server

VizQL Server

Cache Server

Standalone Hyper node

Search & Browse

Backgrounder

Data Server

Data Engine ✅

File Store ✅

Repository

Problem Statement: Only **ONE** 'Hyperd' process on each HoSS node so the memory_limit_per_process_gb cap HoSS not able to leverage system memory

Solution:
hyper.srm_memory_limit_per_process_gb -v 210 -frc

- **HoSS to consume 210G**
- **Recommend 70% of system memory (for example 300GB for HoSS only)**

Server Management Add-on feature

4 | **Hyper Session Memory Timeout** | **5-10G (default no limit)**

- hyper.session_memory_limit –v 10g –frc

10 GB

Avoid one badly designed workbook impacts many other users

• hyper.session_memory_limit -v 10g -frc

10 GB

Unable to connect to the data source.

Try connecting again. If the problem persists, disconnect from the data source and contact the data source owner.

Try Again

Hyper memory limit error

[6] hyper_execute_query: O Cannot allocate 5308416 bytes of memory: exceeding allocation limit of 104857600 bytes for local transaction memory limit

Re-cap: **Mixed** Live & Extract (most common)

💡 **Recommend: Longer Cache Policy (12-24 hrs) with cache warmup**

General   Extensions

(1) tsm data-access caching set -r 720

12+ Hrs

Workbook Performance after a Scheduled Refresh

Recently viewed workbooks with scheduled refreshes can be pre faster. Learn more

(2) ☑ Pre-compute recently viewed workbooks

(3) Live: Workbook level data refresh policy **by owner** (v2021.3)

(4) Extract: View Acceleration **by owner** (v2022.1)