Diagnosing Memory-related Instance Issues

Glenn Berry
PRINCIPAL CONSULTANT – SQLSKILLS.COM
@GlennAlanBerry
www.sqlskills.com/blogs/glenn

Module Overview

Memory-related instance queries

Interpreting the results of these queries

System Memory

System memory amount is a good indicator of the size of the system

Is Lock Pages In Memory enabled?

System memory state is an indicator of external memory pressure

“Available physical memory is high”

System Memory

Provides total and available memory at operating-system level

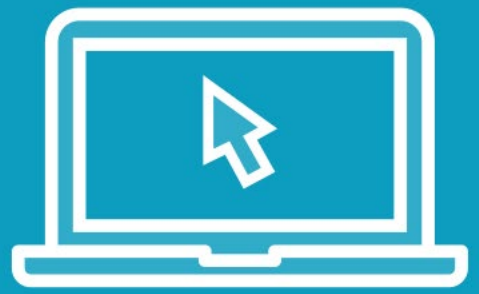
System Memory State shows whether OS is under memory pressure

Desired state is “Available physical memory is high”

If you see any other value, consider reducing max server memory value

Adding additional memory to physical or virtual machine can also help

Demo



System Memory

Process Memory

Physical memory usage by SQL Server process

Is Lock Pages in Memory enabled?

Process memory flags

Process Memory

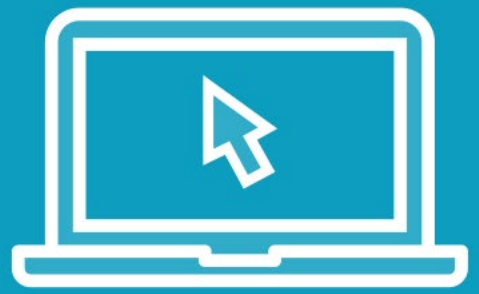
Shows actual SQL Server memory usage; don't believe Task Manager!

Shows whether Locked Pages in Memory is enabled

Shows whether the SQL Server process is low on physical memory

Shows whether the SQL Server process is low on virtual memory

Demo



Process Memory

Total Buffer Usage by Database

Returns buffer usage by database

Helps understand which databases are using memory

Helps evaluate effects of data compression, columnstore indexes

Helps evaluate effectiveness of index tuning

Total Buffer Usage by Database

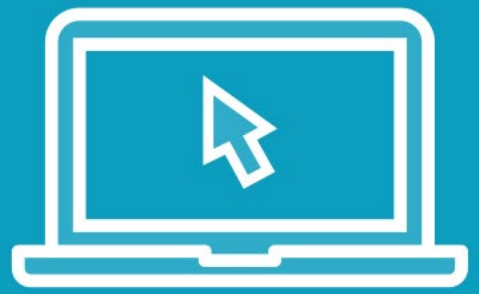
Helps determine source of memory pressure

Lets you focus index and query tuning efforts in right direction

Validates effectiveness of index and query tuning efforts

Data compression and columnstore indexes can be very useful

Demo



Total Buffer Usage by Database

Page Life Expectancy (PLE) by NUMA Node

How long data stays in buffer pool

Returns separate values for each NUMA node

PLE is a good measure of internal memory pressure

Watch the PLE trend over time

Page Life Expectancy by NUMA Node

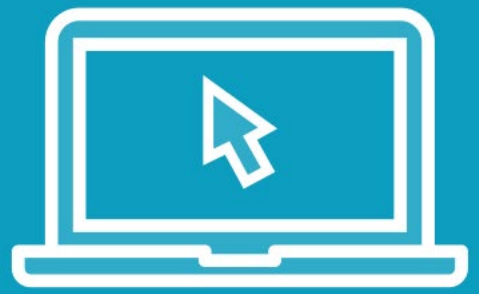
Higher PLE values are always better than lower PLE values

PLE will fluctuate based on query workload and recent activity

You should know what your normal PLE range is

Index and query tuning along with data compression can improve PLE

Demo



Page Life Expectancy by NUMA Node

Memory Grants Pending

Windows Perfmon counter exposed through DMV

Number of processes waiting on a memory grant

Desired value is zero

Memory Grants Pending

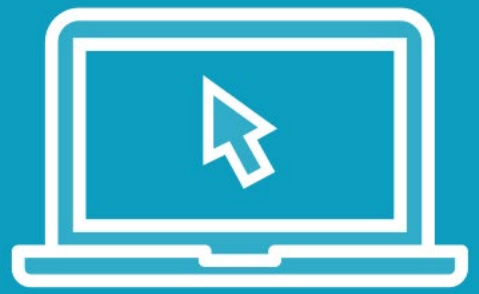
“Memory Grants Pending” confirms memory pressure

If value is above zero, that means processes are waiting for memory grants

You will also see low PLE when this is happening

Fairly unusual (and bad) to see value above zero for sustained periods

Demo



Memory Grants Pending

Memory Clerk Usage

Shows memory clerks using most memory

Helps understand memory usage

Helps spot cache bloating issues

Memory Clerk Usage

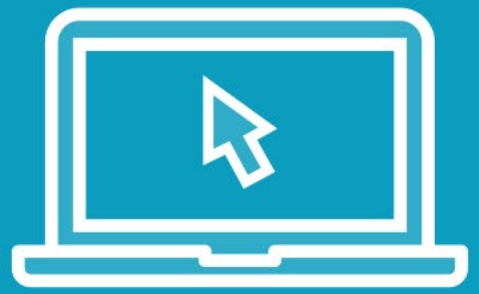
MEMORYCLERK_SQLBUFFERPOOL should be your top consumer

Watch for high CACHESTORE_SQLCP values

Optimize for ad hoc workloads setting helps control cache bloating

Often necessary to also run DBCC FREESYSTEMCACHE ('SQL Plans')

Demo



Memory Clerk Usage

Ad Hoc Queries

Shows cached single-use ad hoc and prepared query plans

These query plans waste memory

This memory would be better used to cache data

Take standard measures to control plan cache bloating

Ad Hoc Queries

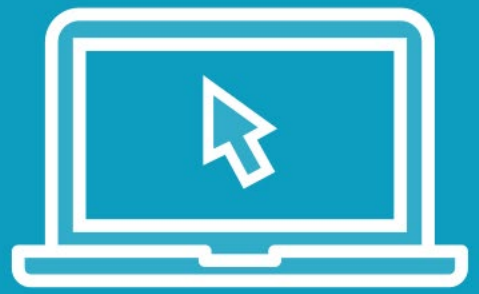
Make sure “optimize for ad hoc workloads” is enabled at instance level

Encourage developers to use stored procedures or parameterized queries

Periodically run DBCC FREESYSTEMCACHE (‘SQL Plans’) to flush plan cache

Minimizing plan cache bloating helps reduce internal memory pressure

Demo



Ad Hoc Queries

Top Logical Reads Queries

Top logical reads queries for entire instance

Shows queries and stored procedures from plan cache

Shows which queries are causing memory pressure

Top Logical Reads Queries

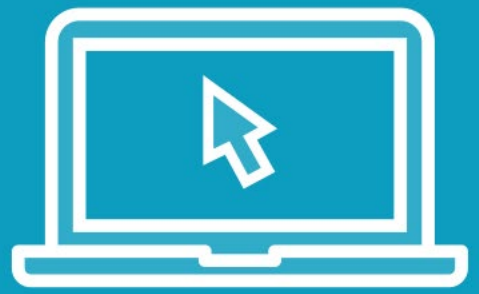
Helps identify which queries are generating the most logical reads

Helps find opportunities for query and index tuning in all databases

Data compression can be very effective for reducing logical reads

Columnstore indexes can also be very effective for reducing logical reads

Demo



Top Logical Reads Queries

What We Covered

Memory-related instance queries

Interpreting the results of these queries