System-Level Metrics (CPU, Memory, Disk)

Metric	Description	Why Monitor?	Notes/Thresholds
CPU Utilization	Percentage of CPU used by SQL Server processes.	High usage indicates inefficient queries, missing indexes, or overload.	Aim for <70-80% sustained; spikes may need investigation.
Memory Usage (e.g., Buffer Cache Hit Ratio, Page Life Expectancy - PLE)	Amount of memory allocated to SQL Server; hit ratio shows cache efficiency; PLE indicates how long pages stay in memory (in seconds).	Memory pressure leads to more disk I/O and slower performance.	Buffer hit ratio >95%; PLE >300-1000 seconds (depending on system RAM).
Disk I/O (Reads/Writes per Sec, Latency, Queue Length)	Rate of read/write operations, time taken for I/O, and pending requests.	High latency or queues signal storage bottlenecks.	Latency <20ms for reads/writes; queue length <2 per disk.
Checkpoint Pages/Sec	Number of pages written to disk during checkpoints.	Excessive checkpoints can increase I/O load.	Monitor for spikes during heavy writes.

Query and Workload Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
Batch Requests/Sec	Rate of T-SQL batches processed per second.	Measures overall throughput; high values indicate heavy load.	Varies by hardware; baseline and watch for drops.
SQL Compilations/Recompilations/Sec	Number of query compilations or recompilations per second.	Excessive recompilations waste CPU and indicate plan instability.	<10% of batch requests; aim low.
Query Execution Time (e.g., last_elapsed_time)	Time taken for queries to run (in microseconds or seconds).	Identifies slow queries for optimization.	Flag queries >1-5 seconds; use averages over time.
Wait Statistics	Types and durations of waits (e.g., CPU, I/O, locks).	Pinpoints root causes like resource contention.	Focus on top waits; e.g., PAGEIOLATCH_XX for I/O issues.

Metric	Description	Why Monitor?	Notes/Thresholds
Plan Cache Hit Ratio	Percentage of query plans reused from cache.	Low ratios mean more compilations and CPU overhead.	>90% ideal.

Concurrency and Locking Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
Lock Waits/Sec	Number of lock requests causing waits per second.	Indicates contention from concurrent transactions.	Keep low; investigate if >0 sustained.
Processes Blocked / Blocking Sessions	Count of blocked processes or sessions.	Blocking leads to delays; deadlocks cause failures.	Zero ideal; alert on sustained blocking >30 seconds.
Deadlocks	Occurrences of deadlocks (mutual blocks resolved by killing a process).	Disrupts transactions; frequent deadlocks need query redesign.	Aim for zero; log and analyze.

Index and Storage Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
Index Fragmentation (avg_fragmentation_in_percent)	Percentage of out-of-order pages in indexes.	Fragmented indexes slow scans and increase I/O.	<10% ideal; rebuild if >30%.
Page Splits/Sec	Number of index page splits due to overflows per second.	High splits cause fragmentation and I/O.	<20-50 per sec; monitor trends.
TempDB Usage/Contention	Space used, growth, and latch waits in TempDB.	TempDB bottlenecks affect sorting, hashing, and temp objects.	Keep free space >20%; watch for PAGELATCH_XX waits.
Out-of-Date Statistics	Age and accuracy of table/index statistics.	Stale stats lead to poor query plans.	Update if rows changed >20%; automate jobs.

Connection and Other Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
User Connections	Number of active user connections.	Too many can exhaust resources or indicate leaks.	Monitor peaks; set limits if needed.
Network Latency	Time for data transfer between server and clients.	High latency slows applications.	<50ms ideal; check for spikes.
Database File Sizes (Data/Log Files)	Used and free space in database files.	Growth can lead to auto-growth events and pauses.	Alert when >80% full; plan expansions.
Replication/Availability Metrics (e.g., Lag, Queue Sizes)	Lag in replication or Always On groups (redo/log queues).	Ensures high availability; lag causes data inconsistency.	Lag <1 minute; monitor queues.

Database Availability and State Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
Database State (sys.databases.state_desc)	Current state of databases (e.g., ONLINE, OFFLINE, RECOVERING, SUSPECT).	Detects unavailable or corrupted databases needing intervention.	Aim for all ONLINE; alert on SUSPECT or EMERGENCY.
Uptime (@@CPU_BUSY, server start time via sys.dm_os_sys_info)	Time since last restart or CPU busy time indicating activity.	Tracks overall server availability and unexpected restarts.	Monitor for low uptime post-restart; correlate with error logs.
Active Transactions (sys.dm_tran_active_transactions)	Number and duration of ongoing transactions.	Identifies long-running transactions that could affect availability during failovers or backups.	Alert on transactions >1-5 minutes; check for orphans.

High Availability (Always On) Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
Synchronization State (sys.dm_hadr_availability_replica_states.synchronization_state_desc)	State of data sync between primary and secondary replicas (e.g., SYNCHRONIZED, SYNCHRONIZING).	Ensures replicas are ready for failover; desync can lead to data loss.	Alert if not SYNCHRONIZED; monitor during high load.
Failover Readiness (sys.dm_hadr_availability_replica_states.is_local, role_desc)	Indicates if a replica can failover (e.g., PRIMARY, SECONDARY roles).	Verifies HA setup health; prevents failed failovers.	Ensure at least one healthy secondary; test periodically.
Redo Queue Size (sys.dm_hadr_database_replica_states.redo_queue_size)	Amount of log records in KB waiting to be redone on secondary.	High queues indicate backlog, delaying recovery and availability.	<100 MB ideal; alert on growth trends.
Log Send Queue Size (sys.dm_hadr_database_replica_states.log_send_queue_size)	Size of unsent log records to secondary in KB.	Signals network or replica issues affecting sync and HA.	<10 MB; investigate spikes.
Lease Timeout (Extended Events for lease_renewal)	Time for lease renewal between SQL Server and WSFC;	Prevents false failovers from network glitches;	Alert on timeouts; adjust based on network stability.

Metric	Description	Why Monitor?	Notes/Thresholds
	defaults to 10 seconds.	monitors cluster health.	

Backup and Recovery Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
Backup Success Rate (msdb.dbo.backupset)	Percentage of successful backups; track last backup time and type.	Ensures recoverability; failed backups risk data loss during outages.	100% success; alert on failures or backups older than RPO (e.g., 24 hours).
Restore Test Results	Frequency and success of restore tests from backups.	Validates backups are usable for recovery, maintaining availability.	Test weekly/monthly; ensure RTO compliance.
Transaction Log Growth (DBCC SQLPERF(LOGSPACE))	Percentage of used space in transaction logs.	Prevents log full errors that halt operations and affect availability.	<70% used; auto-grow sparingly.

Error and Integrity Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
Error Log Entries (sp_readerrorlog, Extended Events)	Count and severity of errors/warnings in SQL Server error logs.	Early detection of issues like corruption, hardware failures, or misconfigurations.	Alert on severity >16; review daily.
Database Integrity (DBCC CHECKDB results)	Checks for physical/logical corruption in databases.	Maintains data health; undetected corruption can cause outages.	Run weekly; alert on any errors.
Deadlocks (sys.dm_os_performance_counters, Trace Flags)	Number of deadlocks per second or day.	Indicates contention issues that terminate	Aim for zero; log and analyze graphs.

Metric	Description	Why Monitor?	Notes/Thresholds
		transactions, affecting health.	
Index Fragmentation (sys.dm_db_index_physical_stats)	Average fragmentation percentage in indexes.	Fragmented indexes degrade health over time; rebuild to prevent issues.	<10% ideal; reorganize >10%, rebuild >30%.

Connectivity and Session Metrics

Metric	Description	Why Monitor?	Notes/Thresholds
User Connections (@@CONNECTIONS, sys.dm_exec_connections)	Total attempted and active connections.	Detects connection leaks or spikes that could exhaust resources and reduce availability.	Monitor peaks; set max connections limit.
Failed Logins (Error logs, Extended Events)	Number of failed login attempts.	Indicates security threats or configuration issues affecting access.	Alert on >5/minute; audit for patterns.
Network Packet Errors (@@PACKET_ERRORS)	Count of network errors on connections.	Signals network issues impacting availability and health.	Zero ideal; investigate any increase.
Blocked Processes (sys.dm_os_waiting_tasks, Activity Monitor)	Number of processes waiting due to blocks.	Reveals contention that could escalate to availability problems.	Zero sustained; alert >30 seconds.

Security and Auditing Metrics

1. Authentication and Login Metrics

2Access Control and Permission Metrics

3 Object and Configuration Change Metrics

4 Data Access and Modification Metrics