



Le migliori pratiche per la distribuzione di SQL Server su Amazon EC2

AWS Guida prescrittiva



AWS Guida prescrittiva: Le migliori pratiche per la distribuzione di SQL Server su Amazon EC2

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

I marchi e il trade dress di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, in qualsiasi modo che possa causare confusione tra i clienti o in qualsiasi modo che denigri o discreditì Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà delle rispettive aziende, che possono o meno essere associate, collegate o sponsorizzate da Amazon.

Table of Contents

Introduzione	1
Configurazione delle impostazioni di calcolo e archiviazione	2
Usa un tipo di istanza ottimizzato per Amazon EBS	2
Ottimizza il layout del disco o la distribuzione dei file	3
Imposta la dimensione dell'unità di allocazione NTFS su 64 KB	3
Posiziona tempdb in un instance store	5
Spostamento di tempdb in un instance store	5
Inizializzazione dell'instance store	8
Utilizzo dell'estensione del buffer pool	10
Evita le discrepanze tra i core della CPU	10
Verifica le prestazioni del disco	11
Abilitare l'inizializzazione istantanea dei file	12
Bloccare le pagine in memoria	14
Disattiva l'offload TCP e le impostazioni RSS	16
Determina i tuoi requisiti di IOPS e di throughput	17
Utilizzate lo striping per aggirare i limiti di IOPS e di throughput	18
Escludere i file di SQL Server dal software antivirus	18
Configurazione di SQL Server	19
Configura tempdb per ridurre i conflitti	19
Imposta MAXDOP per prestazioni ottimali	20
Modificare la soglia di costo del parallelismo	22
Ottimizza per carichi di lavoro ad hoc	22
Usa i trace flag per migliorare le prestazioni	23
Installa le patch più recenti	24
Limita la memoria massima del server per evitare la pressione della memoria	24
Utilizza il massimo livello di compatibilità del database	26
Controlla il numero di VLFs	26
Controlla le impostazioni di crescita automatica del database	27
Configurazione dei gruppi di disponibilità Always On	31
Imposta RegisterAllProviders IP su true quando usi i gruppi di disponibilità Always On	31
Imposta HostRecord TTL su 60 o meno quando usi i gruppi di disponibilità Always On	32
Disattiva il fallback automatico per il gruppo di cluster Always On	32
Configurazione dei backup	33
Miglioramento dell'ottimizzazione del database	34

Ricostruisci gli indici	34
Aggiorna le statistiche	35
Ottimizzazione delle distribuzioni di SQL Server su Amazon EC2	36
Passaggi successivi	37
Risorse aggiuntive	38
Cronologia dei documenti	39
Glossario	41
#	41
A	42
B	45
C	47
D	50
E	54
F	56
G	58
H	59
I	61
L	63
M	64
O	69
P	71
Q	74
R	75
S	78
T	82
U	83
V	84
W	84
Z	86

Le migliori pratiche per la distribuzione di Microsoft SQL Server su Amazon EC2

Abhishek Soni e Sagar Patel, Amazon Web Services ()AWS

Dicembre 2023 ([cronologia dei documenti](#))

Lo scopo di questa guida è garantire un'esperienza coerente dopo la distribuzione o la migrazione di Microsoft SQL Server su Amazon Elastic Compute Cloud (Amazon EC2) sul cloud Amazon Web Services (AWS). Fornisce le migliori pratiche per configurare database e server, per ottimizzare l'infrastruttura, ottimizzare le prestazioni ed evitare di incorrere in problemi imprevisti dopo la distribuzione o la migrazione.

Questa guida è destinata agli architetti di database, ai responsabili di sistemi e database e agli amministratori che intendono migrare Microsoft SQL Server dal proprio ambiente locale ad Amazon EC2 o che desiderano ottimizzare la nuova distribuzione di SQL Server su Amazon EC2.

[Amazon EC2](#) offre capacità di elaborazione scalabile nel AWS cloud. L'utilizzo di SQL Server su Amazon EC2 è simile all'esecuzione di SQL Server in locale. Amazon ti EC2 offre il pieno controllo della tua infrastruttura e del tuo ambiente di database. Benefici della scalabilità, delle prestazioni e dell'elasticità del AWS cloud, ma sei responsabile della configurazione e dell'ottimizzazione di tutti i componenti, inclusi EC2 istanze, volumi di archiviazione, file system, reti e sicurezza. Questa guida fornisce informazioni per aiutarti a ottimizzare la configurazione e massimizzare le prestazioni di SQL Server su AWS. Descrive in dettaglio le impostazioni e le best practice di server e storage. Spiega inoltre come automatizzare le impostazioni laddove applicabile e illustra le modifiche alla configurazione a livello di database.

Note

AWS offre anche opzioni per spostare il database SQL Server locale su un servizio gestito come Amazon Relational Database Service (Amazon RDS) per SQL Server. Per una discussione sulle opzioni di migrazione, consulta [Strategia di migrazione per i database relazionali](#) sul sito Web Prescriptive Guidance AWS

Configurazione delle impostazioni di calcolo e archiviazione

Prima di migrare o distribuire SQL Server su Amazon EC2, puoi configurare le impostazioni dell' EC2 istanza e dello storage per migliorare le prestazioni e ridurre i costi. Le sezioni seguenti forniscono suggerimenti e best practice per l'ottimizzazione.

Argomenti

- [Usa un tipo di istanza ottimizzato per Amazon EBS](#)
- [Ottimizza il layout del disco o la distribuzione dei file](#)
- [Imposta la dimensione dell'unità di allocazione NTFS su 64 KB](#)
- [Posiziona tempdb in un instance store](#)
- [Evita le discrepanze tra i core della CPU](#)
- [Verifica le prestazioni del disco](#)
- [Abilitare l'inizializzazione istantanea dei file](#)
- [Bloccare le pagine in memoria](#)
- [Disattiva l'offload TCP e le impostazioni RSS](#)
- [Determina i tuoi requisiti di IOPS e di throughput](#)
- [Utilizzate lo striping per aggirare i limiti di IOPS e di throughput](#)
- [Escludere i file di SQL Server dal software antivirus](#)

Usa un tipo di istanza ottimizzato per Amazon EBS

Se il tuo database SQL Server gestisce carichi di lavoro a uso intensivo di I/O, il provisioning di istanze ottimizzate per [Amazon Elastic Block Store \(Amazon EBS\)](#) contribuirà a migliorare le prestazioni.

Un'istanza ottimizzata per Amazon EBS utilizza uno stack di configurazione ottimizzato e fornisce capacità aggiuntiva dedicata per l'I/O di Amazon EBS. Questa ottimizzazione offre le migliori prestazioni per i tuoi volumi EBS riducendo al minimo la contesa tra l'I/O di Amazon EBS e altro traffico proveniente dall'istanza.

Ottimizza il layout del disco o la distribuzione dei file

Utilizza un volume per dati e file di registro, un altro volume per carichi di lavoro tempdb e volumi Cold HDD (sc1) o Throughput Optimized HDD () per i backup. st1

Se riscontri un problema relativo all'I/O e desideri separare i carichi di lavoro per i dati dai file di registro, prendi in considerazione l'utilizzo di volumi diversi. Se il carico di lavoro richiede la separazione di database specifici, valuta la possibilità di utilizzare un volume dedicato per ogni database.

In genere, tempdb è l'obiettivo dell'I/O più elevato, quindi se quel carico di lavoro non è separato, potrebbe diventare un collo di bottiglia. Questa separazione aiuta anche a isolare tempdb dai dati e dai file di registro del database utente. È possibile utilizzare uno storage relativamente più economico per i backup per ottimizzare i costi.

Imposta la dimensione dell'unità di allocazione NTFS su 64 KB

L'unità di archiviazione atomica in SQL Server è una pagina con una dimensione di 8 KB. Otto pagine fisicamente contigue costituiscono un'estensione (che ha una dimensione di 64 KB). SQL Server utilizza le estensioni per archiviare i dati. Pertanto, su un computer SQL Server, la dimensione dell'unità di allocazione NTFS per l'hosting dei file di database SQL (incluso tempdb) deve essere di 64 KB.

Per verificare la dimensione del cluster (allocazione NTFS) delle unità, è possibile utilizzare o la riga di comando. PowerShell

Utilizzando: PowerShell

```
Get-wmiObject -Class win32_volume | Select-object Label, BlockSize | Format-Table -AutoSize
```

La seguente illustrazione mostra un esempio di output da PowerShell.

```
PS C:\Users\Administrator> Get-wmiObject -Class win32_volume | Select-object Label, BlockSize | Format-Table -AutoSize
Label BlockSize
----- -----
4096
```

Oppure usa:

```
$wmiQuery = "SELECT Name, Label, BlockSize FROM win32_volume WHERE FileSystem='NTFS'"
```

```
Get-wmiObject -Query $wmiQuery -ComputerName '.' | Sort-Object Name | Select-Object Name, Label, BlockSize
```

Utilizzando la riga di comando:

```
$ fsutil fsinfo ntfsinfo C:
```

L'illustrazione seguente mostra un esempio di output dalla riga di comando. Il valore Bytes Per Cluster visualizza la dimensione del formato in byte. L'output di esempio mostra 4096 byte. Per le unità che ospitano i file di database di SQL Server, questo valore deve essere 64 KB.

```
C:\Users\Administrator>fsutil fsinfo ntfsinfo C:
NTFS Volume Serial Number :          0x2492618592615bf4
NTFS Version :                     3.1
LFS Version :                      2.0
Number Sectors :                   0x000000000063fefff
Total Clusters :                  0x00000000000c7fdff
Free Clusters :                   0x0000000000045698f
Total Reserved :                 0x00000000000001591
Bytes Per Sector :                512
Bytes Per Physical Sector :       512
Bytes Per Cluster :               4096
Bytes Per FileRecord Segment :    1024
Clusters Per FileRecord Segment : 0
Mft Valid Data Length :           0x00000000121c0000
Mft Start Lcn :                  0x000000000000c0000
Mft2 Start Lcn :                 0x0000000000000002
Mft Zone Start :                 0x000000000005f2760
Mft Zone End :                   0x000000000005f95e0
Max Device Trim Extent Count :   0
Max Device Trim Byte Count :     0x0
Max Volume Trim Extent Count :   62
Max Volume Trim Byte Count :     0x400000000
```

In alcuni casi, le prestazioni di SQL Server non dipendono dalla dimensione del blocco quando utilizzi lo storage SSD su Amazon EC2. Per ulteriori informazioni, consulta il post del blog [AWS I clienti traggono vantaggio dalla dimensione dei blocchi di 64 KB per lo storage di SQL Server?](#)

Posiziona tempdb in un instance store

Quando utilizzi un Amazon EC2 instance store, usa il volume dell'instance store per tempdb. Un instance store fornisce uno storage temporaneo (temporaneo) a livello di blocco per la tua istanza. Ti consigliamo di inserire tempdb su un volume di instance store per due motivi: velocità e costi. Tempdb è in genere il database più utilizzato, quindi trae vantaggio dall'unità più veloce disponibile. Un altro vantaggio dell'inserimento di tempdb in un instance store è il risparmio sui costi, in quanto non viene addebitato separatamente l'I/O in base all'instance store.

Tempdb viene ricreato ogni volta che si riavvia SQL Server, quindi l'arresto o la chiusura di un'istanza non causeranno la perdita di dati. Tuttavia, un volume dell'Instance Store viene perso quando la macchina virtuale viene avviata su un altro host, poiché il disco temporaneo è collegato localmente alla macchina, quindi pianifica con attenzione.

Quando si utilizza un volume di instance store:

- Inizializza il volume prima dell'avvio del servizio SQL Server. In caso contrario, la procedura di avvio di SQL Server avrà esito negativo.
- Concedi le autorizzazioni (controllo completo) sul volume dell'instance store in modo esplicito all'account di avvio di SQL Server.

Spostamento di tempdb in un instance store

Per spostare tempdb in un volume di instance store:

1. Da Windows, esegui `diskmgmt.msc` come amministratore per aprire l'utilità di sistema Disk Management.
2. Inizializza un nuovo disco.
3. Fate clic con il pulsante destro del mouse sul disco, quindi scegliete Nuovo volume semplice.
4. Completa le istruzioni utilizzando queste impostazioni per formattare il volume:
 - File system: NTFS
 - Dimensione dell'unità di allocazione: 64K
 - Etichetta del volume: tempdb

Per ulteriori informazioni, consulta la [documentazione sulla gestione del disco](#) sul sito Web di Microsoft.

5. Connect all'istanza di SQL Server ed esegui il comando seguente per annotare il nome di file logico e fisico del database tempdb:

```
$ sp_helpdb 'tempdb'
```

La schermata seguente mostra il comando e il relativo output.

The screenshot shows the SQL Server Management Studio interface with three tabs: SQLQuery16.sql, SQLQuery15.sql, and SQLQuery3.sql. The SQLQuery15.sql tab is active and contains the command: sp_helpdb 'Tempdb'. The Results tab displays two tables of data. The first table shows database properties for tempdb:

	name	db_size	owner	dbid	created	status	compatibility_level
1	tempdb	520.00 MB	sa	2	Jul 12 2020	Status=ONLINE, Updateability=READ_WRITE, UserAcc...	140

The second table shows file and filegroup details for tempdb:

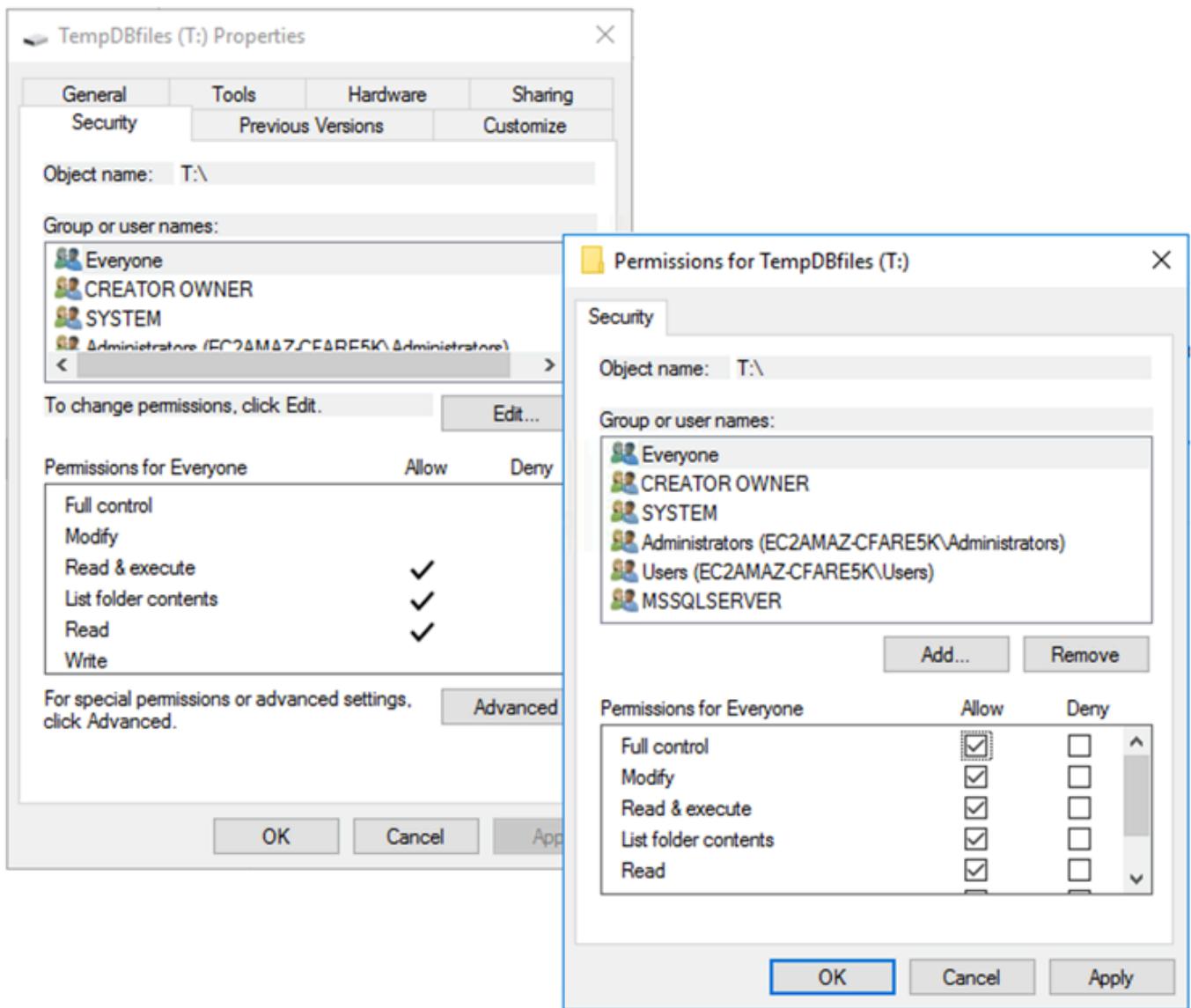
	name	fileid	filename	filegroup	size	maxsize	growth	usage
1	tempdev	1	C:\Program Files\Microsoft SQL Server\MSSQL14.MSS...	PRIMARY	524288 KB	Unlimited	65536 KB	data only
2	templog	2	C:\Program Files\Microsoft SQL Server\MSSQL14.MSS...	NULL	8192 KB	Unlimited	65536 KB	log only

6. Sposta il file tempdb nella nuova posizione. Ricorda di impostare tutti i file del database tempdb sulla stessa dimensione iniziale. Lo script SQL Server di esempio seguente sposta i file tempdb nell'unità T e imposta i file di dati sulla stessa dimensione.

```
USE master
GO
ALTER DATABASE TempDB MODIFY FILE (NAME = tempdev, FILENAME = 'T:\tempdb.mdf',SIZE
= 524288KB)
GO
ALTER DATABASE TempDB MODIFY FILE (NAME = temp2, FILENAME = 'T:
\tempdb_mssql_2.ndf',SIZE = 524288KB)
GO
ALTER DATABASE TempDB MODIFY FILE (NAME = temp3, FILENAME = 'T:
\tempdb_mssql_3.ndf',SIZE = 524288KB)
GO
ALTER DATABASE TempDB MODIFY FILE (NAME = temp4, FILENAME = 'T:
\tempdb_mssql_4.ndf',SIZE = 524288KB)
GO
ALTER DATABASE TempDB MODIFY FILE (NAME = templog, FILENAME = 'T:\templog.ldf')
```

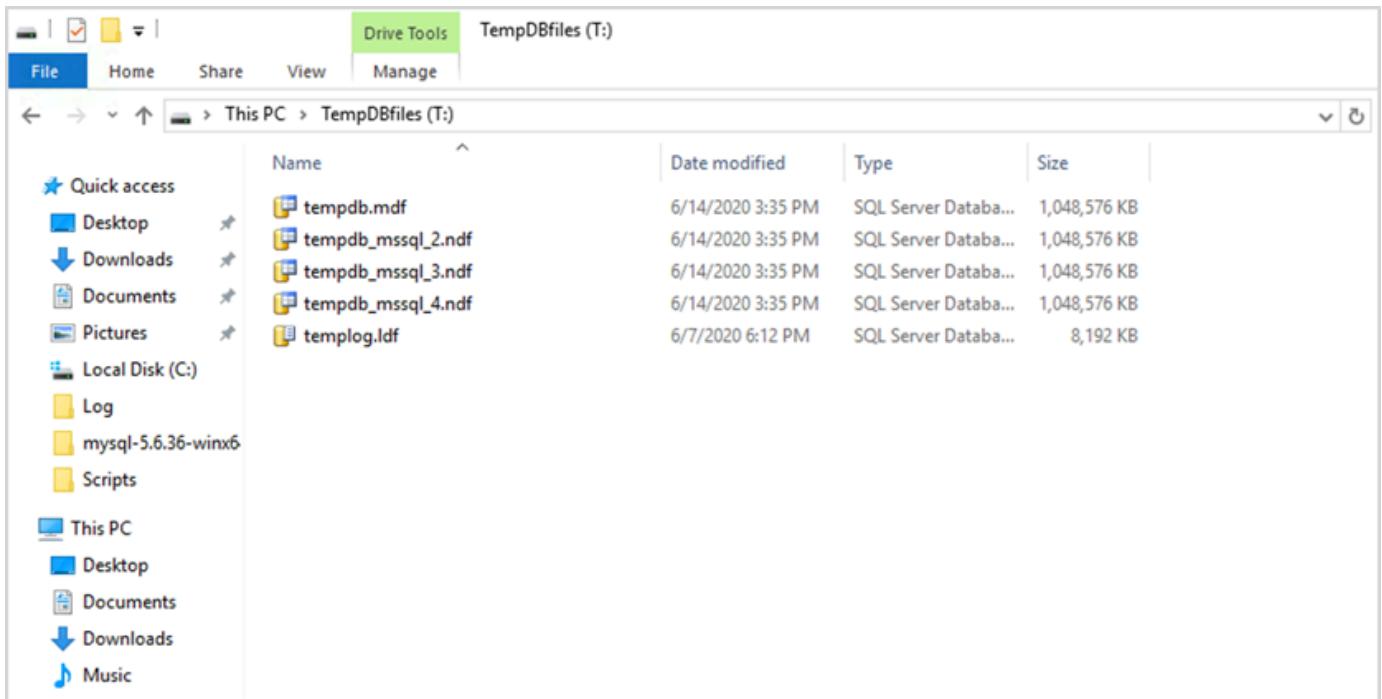
GO

7. Concedi all'account di avvio di SQL Server le autorizzazioni per la nuova posizione del database tempdb, in modo che possa creare i file tempdb, come mostrato nella schermata seguente.



8. Riavviare SQL Server per utilizzare la nuova posizione per tempdb.

Verranno visualizzati i file tempdb creati nella nuova posizione, come mostrato nella schermata seguente.



9. Eliminare i file tempdb dalla vecchia posizione.

Per assicurarsi che il volume dell'Instance Store sia inizializzato prima dell'avvio di SQL Server in caso di riavvio o avviamento dell'istanza, segui i passaggi nella sezione successiva. In caso contrario, l'avvio di SQL Server fallirà perché tempdb non è stato inizializzato.

Inizializzazione dell'instance store

Per inizializzare il data store:

1. Aprire Windows Services Manager (`services.msc`) e impostare SQL Server e i relativi servizi dipendenti (ad esempio, SQL Server Agent) in modo che vengano avviati manualmente. (Utilizzerai uno script per aviarlo quando il volume dell'instance store sarà pronto.)
2. Crea uno PowerShell script da passare all' EC2 istanza Amazon come dati utente. Lo script svolge le funzioni seguenti:
 - Rileva lo storage temporaneo e ne crea un'unità tempdb (l'unità T nell'esempio).
 - Aggiorna il disco temporaneo se l'istanza si arresta e si riavvia. EC2
 - Concede all'account di avvio di SQL Server il controllo completo del volume tempdb appena inizializzato. L'esempio presuppone un'istanza predefinita, quindi utilizza. NT SERVICE

\MSSQLSERVER Per un'istanza denominata, questa è in genere l'impostazione NT SERVICE \MSSQL\$*<InstanceName>* predefinita.

- Salva lo script su un volume locale (c:\scriptsnell'esempio) e gli assegna un nome di file (InstanceStoreMapping.ps1).
- Crea un'attività pianificata utilizzando Windows Task Scheduler. Questa attività esegue lo PowerShell script all'avvio.
- Avvia SQL Server e SQL Server Agent dopo le azioni precedenti.

Lo script seguente proviene dal secondo laboratorio del [MS-SQL Availability Group Workshop](#) con alcune modifiche. Copia lo script nel campo Dati utente all'avvio dell' EC2 istanza e personalizzalo se necessario.

```
<powershell>
# Create pool and virtual disk for TempDB using the local NVMe, ReFS 64K, T: Drive
$NVMe = Get-PhysicalDisk | ? { $_.CanPool -eq $True -and $_.FriendlyName -eq "NVMe Amazon EC2 NVMe"}
New-StoragePool -FriendlyName TempDBPool -StorageSubsystemFriendlyName "Windows Storage*" -PhysicalDisks $NVMe
New-VirtualDisk -StoragePoolFriendlyName TempDBPool -FriendlyName TempDBDisk -ResiliencySettingName simple -ProvisioningType Fixed -UseMaximumSize
Get-VirtualDisk -FriendlyName TempDBDisk | Get-Disk | Initialize-Disk -Passthru
| New-Partition -DriveLetter T -UseMaximumSize | Format-Volume -FileSystem ReFS -AllocationUnitSize 65536 -NewFileSystemLabel TempDBfiles -Confirm:$false
# Script to handle NVMe refresh on start/stop instance
$instanceStoreMapping = {
if (!(Get-Volume -DriveLetter T)) {
    #Create pool and virtual disk for TempDB using mirroring with NVMe
    $NVMe = Get-PhysicalDisk | ? { $_.CanPool -eq $True -and $_.FriendlyName -eq "NVMe Amazon EC2 NVMe"}
    New-StoragePool -FriendlyName TempDBPool -StorageSubsystemFriendlyName "Windows Storage*" -PhysicalDisks $NVMe
    New-VirtualDisk -StoragePoolFriendlyName TempDBPool -FriendlyName TempDBDisk -ResiliencySettingName simple -ProvisioningType Fixed -UseMaximumSize
    Get-VirtualDisk -FriendlyName TempDBDisk | Get-Disk | Initialize-Disk -Passthru
    | New-Partition -DriveLetter T -UseMaximumSize | Format-Volume -FileSystem ReFS -AllocationUnitSize 65536 -NewFileSystemLabel TempDBfiles -Confirm:$false
    #grant SQL Server Startup account full access to the new drive
    $item = gi -literalpath "T:\"
    $acl = $item.GetAccessControl()
```

```
$permission="NT SERVICE\MSSQLSERVER","FullControl","Allow"
$rule = New-Object System.Security.AccessControl.FileSystemAccessRule
$permission
    $acl.SetAccessRule($rule)
    $item.SetAccessControl($acl)
#Restart SQL so it can create tempdb on new drive
Stop-Service SQLSERVERAGENT
Stop-Service MSSQLSERVER
Start-Service MSSQLSERVER
Start-Service SQLSERVERAGENT
}
}

New-Item -ItemType Directory -Path c:\Scripts
$instanceStoreMapping | set-content c:\Scripts\InstanceStoreMapping.ps1
# Create a scheduled task on startup to run script if required (if T: is lost)
$action = New-ScheduledTaskAction -Execute 'Powershell.exe' -Argument 'c:\scripts
\InstanceStoreMapping.ps1'
$trigger = New-ScheduledTaskTrigger -AtStartup
Register-ScheduledTask -Action $action -Trigger $trigger -TaskName "Rebuild
TempDBPool" -Description "Rebuild TempDBPool if required" -RunLevel Highest -User
System
</powershell>
```

Utilizzo dell'estensione del buffer pool

Se hai intenzione di utilizzare l'estensione del buffer pool, potresti anche prendere in considerazione l'idea di posizionarla su un volume effimero. Tuttavia, consigliamo vivamente di testarla a fondo prima dell'implementazione. Evita di usare lo stesso volume per l'estensione del buffer pool e tempdb.

Note

Sebbene l'estensione del buffer pool possa essere utile in alcuni casi, non sostituisce la RAM. Prima di decidere di utilizzarlo, consulta i [dettagli disponibili sul sito Web di Microsoft](#).

Evita le discrepanze tra i core della CPU

La scelta di un server con un numero di core superiore a quello coperto dalle licenze può causare distorsioni della CPU e spreco di potenza della CPU. Ciò è dovuto alla mappatura tra core logici e core effettivi. Quando si utilizza SQL Server con una Client Access License (CAL), alcuni scheduler lo

saranno VISIBLE ONLINE e gli altri sì. VISIBLE OFFLINE Ciò può causare problemi di prestazioni con le topologie di accesso non uniforme alla memoria (NUMA), a causa del mancato utilizzo ottimale dei nodi di pianificazione.

Ad esempio, se si esegue SQL Server su un'm5.24xlarge istanza, verranno rilevati due socket con 24 core e 48 processori logici per socket, il che si traduce in un totale di 96 processori logici. Se disponi di una licenza per soli 48 core, nel log degli errori di SQL Server verrà visualizzato un messaggio simile al seguente:

2020-06-08 12:35:27.37 Server SQL Server ha rilevato 2 socket con 24 core per socket e 48 processori logici per socket, 96 processori logici in totale; utilizzando 48 processori logici basati sulle licenze di SQL Server. Questo è un messaggio informativo; non è richiesta alcuna azione da parte dell'utente.

Se vedi una differenza tra il numero totale di core e il numero di core utilizzati da SQL Server, verifica lo squilibrio nell'utilizzo della CPU o usa un tipo di server con lo stesso numero di core supportato dalla licenza.

Inclinazione della CPU: per il tipo di istanza nel nostro esempio (m5.24xlarge), SQL Server crea otto nodi NUMA per impostazione predefinita. Solo quattro di questi nodi (ID nodo principale 0,1,2,3) dispongono di scheduler con lo stato. VISIBLE ONLINE Le pianificazioni rimanenti sono tutte. VISIBLE OFFLINE Questa disparità tra gli scheduler può portare a un peggioramento delle prestazioni.

Per controllare le informazioni e lo stato dello scheduler, usa:

```
$ select * from sys.dm_osSchedulers
```

Se desideri utilizzare un'istanza server con un numero di core superiore a quello supportato dalla licenza di SQL Server, valuta la possibilità di personalizzare il numero di core seguendo le istruzioni in [Specificare le opzioni CPU per l'istanza nella documentazione di Amazon. EC2](#)

Verifica le prestazioni del disco

Ti consigliamo di controllare le prestazioni del tuo disco utilizzando uno strumento come [DiskSpd](#). Questo strumento fornisce una stima della velocità del disco prima di eseguire test specifici di SQL Server. È molto importante valutare le prestazioni del disco, poiché un volume EBS funziona in modo diverso da un SAN tradizionale in un ambiente locale. La mancanza di test adeguati delle prestazioni

può portare a un rallentamento inaspettato delle prestazioni dopo la migrazione. Puoi anche eseguire [test personalizzati](#) con DiskSpd.

Abilitare l'inizializzazione istantanea dei file

In SQL Server, utilizza l'impostazione Esegui attività di manutenzione del volume per abilitare l'inizializzazione istantanea dei file, a meno che tu non stia seguendo le restrizioni normative. Questa opzione aumenta in modo significativo le prestazioni di crescita automatica dei file.

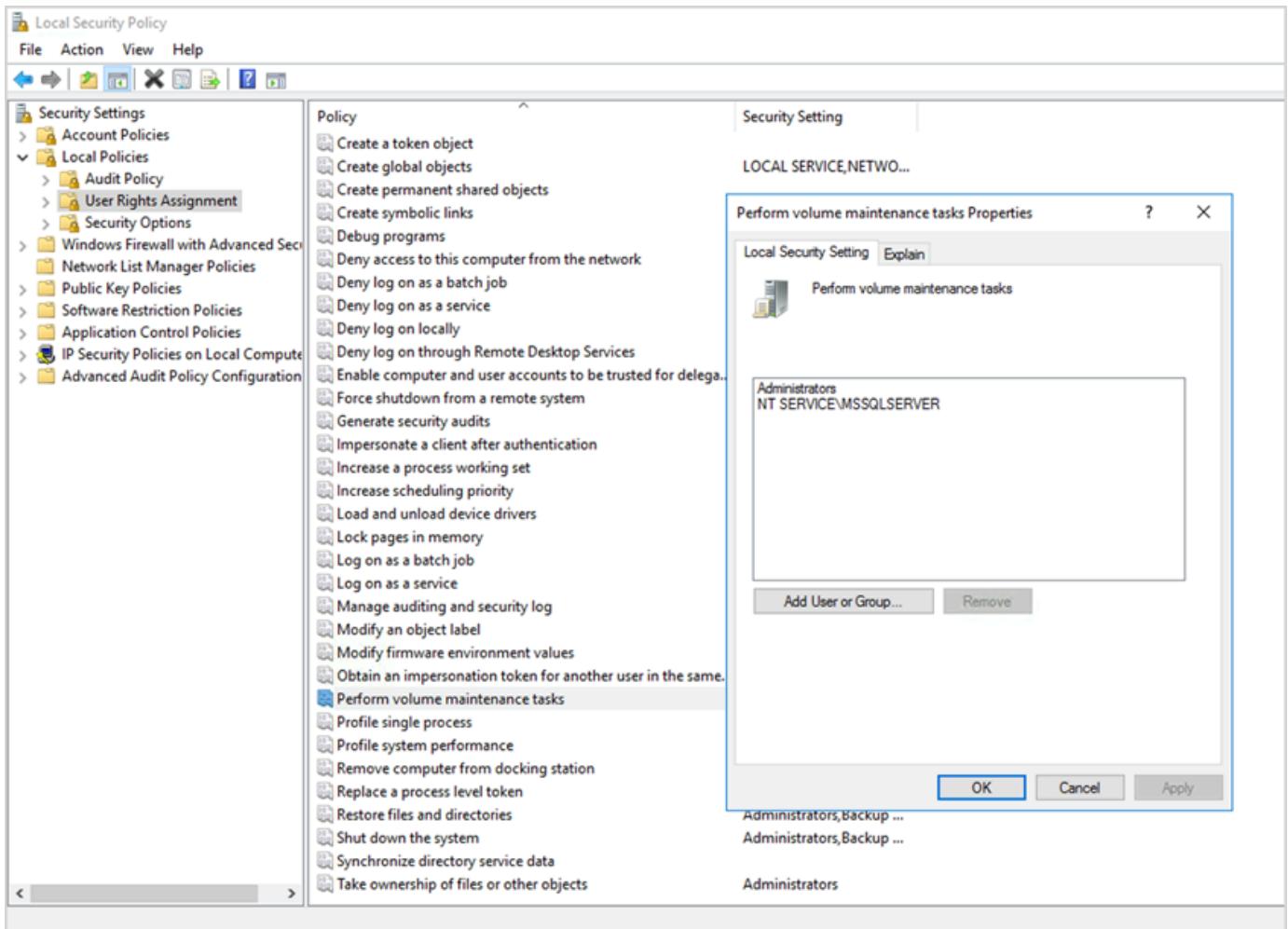
Questa impostazione ignora le operazioni di azzeramento per i file di dati. Cioè, i file di dati non vengono riempiti con valori zero ($0x0$) quando vengono inizializzati, il che può richiedere molto tempo. Il contenuto corrente sul disco viene sovrascritto solo quando vengono scritti nuovi dati sul disco.

Note

I file di registro non traggono vantaggio dall'inizializzazione istantanea dei file.

Per abilitare l'inizializzazione istantanea dei file:

1. Nella schermata Start, esegui `secpol.msc` per aprire la console Local Security Policy.
2. Scegli Criteri locali, Assegnazione diritti utente, Esegui attività di manutenzione del volume e aggiungi l'account del servizio SQL Server, come mostrato nella schermata seguente.



3. Riavviare l'istanza di SQL Server per rendere effettive le modifiche.

Per ulteriori informazioni sull'inizializzazione istantanea dei file, vedere la [documentazione di SQL Server](#) sul sito Web di Microsoft.

Nota sulla sicurezza

Quando si utilizza l'inizializzazione istantanea dei file, il disco viene sovrascritto solo quando vengono scritti nuovi dati nei file, quindi è possibile leggere il contenuto eliminato.

Mentre l'unità è collegata all'istanza, l'elenco di controllo degli accessi discrezionali (DACL) sul file riduce il rischio di divulgazione delle informazioni, poiché consente l'accesso solo all'account del servizio SQL Server e all'amministratore locale. Tuttavia, quando il file viene

scollegato, è possibile accedervi. Se la divulgazione del contenuto eliminato è un problema, è necessario disabilitare l'inizializzazione istantanea dei file per l'istanza di SQL Server.

Bloccare le pagine in memoria

Abilita l'opzione Blocca pagine in memoria per l'account di avvio di SQL Server per assicurarti che il sistema operativo non riduca il set di lavoro di SQL Server.

Per verificare se questa opzione è abilitata, usa la seguente query SQL:

```
SELECT sql_memory_model, sql_memory_model_desc  
FROM sys.dm_os_sys_info;
```

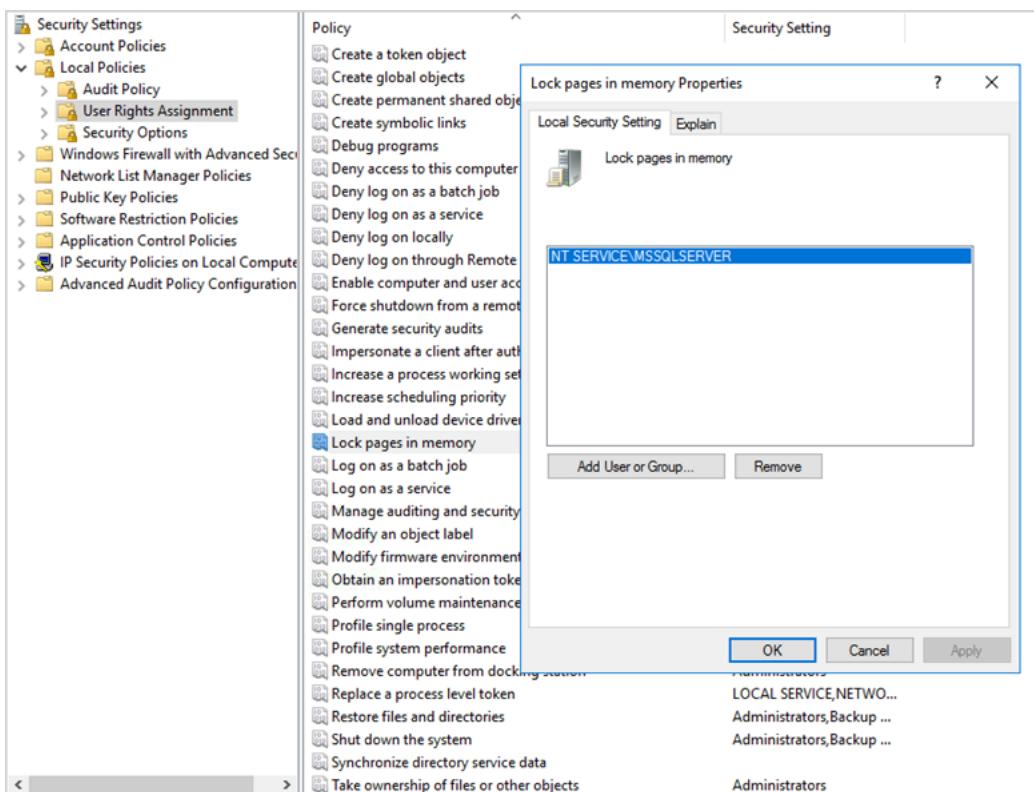
Output:

sql_memory_model	sql_memory_model_desc
1	CONVENTIONAL

"CONVENTIONAL" means it's not enabled.

Per abilitare l'opzione Blocca le pagine in memoria:

1. Nella schermata Start, esegui secpol.msc per aprire la console Local Security Policy.
2. Scegli Criteri locali, Assegnazione diritti utente, Blocca pagine in memoria e aggiungi l'account del servizio SQL Server, come mostrato nella schermata seguente.



3. Riavvia l'istanza di SQL Server per rendere effettive le modifiche.
4. Utilizza la seguente query SQL per confermare che l'opzione Blocca pagine in memoria è abilitata:

```
SELECT sql_memory_model, sql_memory_model_desc
FROM sys.dm_os_sys_info;
```

Output:

sql_memory_model	sql_memory_model_desc
2	LOCK_PAGES

"LOCK_PAGES" means it's enabled.

Per ulteriori informazioni sul modello di memoria di SQL Server, vedere `sql_memory_model` e `sql_memory_model_desc` nella documentazione [sys.dm_os_sys_info](#) sul sito Web di Microsoft.

Disattiva l'offload TCP e le impostazioni RSS

Se riscontri problemi di connettività casuali, come errori a livello di trasporto o errori di trasmissione di pacchetti durante l'esecuzione di carichi di lavoro SQL, potresti voler disabilitare l'offload TCP e le impostazioni RSS.

- L'offloading TCP (funzionalità TCP Chimney Offload) trasferisce l'elaborazione dei pacchetti TCP/IP dal processore alla scheda di rete, per liberare la CPU da dedicare ad altre attività.
- Il Receive-side scaling (RSS) aiuta a distribuire l'elaborazione del traffico di rete in entrata su sistemi multiprocessore. Bilancia il carico in modo efficiente l'elaborazione della rete tra i CPUs.

Per verificare le impostazioni correnti, al prompt dei comandi, esegui il netsh comando:

```
$ netsh int tcp show global
```

Ecco un esempio di output del comando. In questo esempio, Receive-Side Scaling State e Chimney Offload State sono entrambi disabilitati.

```
C:\Users\Administrator>netsh int tcp show global
Querying active state...

TCP Global Parameters
-----
Receive-Side Scaling State      : disabled
Chimney Offload State          : disabled
NetDMA State                   : disabled
Direct Cache Access (DCA)      : disabled
Receive Window Auto-Tuning Level: normal
Add-On Congestion Control Provider: none
ECN Capability                 : enabled
RFC 1323 Timestamps            : disabled
Initial RTO                     : 3000
Receive Segment Coalescing State: enabled
Non Sack Rtt Resiliency        : disabled
Max SYN Retransmissions       : 2
TCP Fast Open                  : disabled
```

Per ottenere informazioni di task offload su una connessione specifica, al prompt dei comandi, esegui:

```
netstat -t
```

e controlla il valore della colonna dello stato di offload.

Per disabilitare l'offloading TCP e l'RSS per Windows Server 2008 e 2012, esegui questi comandi al prompt dei comandi:

```
netsh int ip set global taskoffload=disabled  
netsh int tcp set global chimney=disabled  
netsh int tcp set global rss=disabled  
netsh int tcp set global netdma=disabled
```

Per ulteriori informazioni su queste impostazioni, consulta:

- Funzionalità di [TCP Chimney Offload, Receive Side Scaling e Network Direct Memory Access e introduzione a Receive Side Scaling sul sito Web Microsoft](#)
- [TCP Offloading nella documentazione di Amazon EC2](#)
- [Risoluzione dei problemi relativi ai driver fotovoltaici](#) nella documentazione di Amazon EC2

 **Important**

Non utilizzare IPsec Task Offload o TCP Chimney Offload. Secondo la [documentazione Microsoft](#), queste funzionalità di offload sono obsolete in Windows Server 2016 e potrebbero non essere supportate nelle versioni future. L'utilizzo di queste funzionalità potrebbe influire negativamente sulle prestazioni.

Determina i tuoi requisiti di IOPS e di throughput

Usa Windows Performance Monitor per ottenere informazioni su IOPS e velocità effettiva.

Per aprire Windows Performance Monitor, esegui lo perfmon al prompt dei comandi. I dati di IOPS e di throughput sono forniti dai seguenti contatori delle prestazioni:

- Disco = IOPS reads/sec + disk writes/sec
- Lettura su disco bytes/sec + disk write bytes/sec = velocità effettiva

Ti consigliamo di ottenere i dati IOPS e di throughput per i periodi di utilizzo di picco e anche per un ciclo di carico di lavoro tipico, per ottenere una buona stima dei tuoi requisiti. Assicurati che il tipo di istanza scelto per SQL Server supporti questi requisiti di I/O.

È importante che questa stima sia corretta. In caso contrario, è possibile che si verifichi un eccesso di approvvigionamento delle risorse, con conseguente sottoutilizzo delle risorse, oppure un approvvigionamento insufficiente delle risorse, con conseguenti gravi problemi di prestazioni.

Utilizzate lo striping per aggirare i limiti di IOPS e di throughput

Quando la tua applicazione SQL Server richiede più [IOPS e velocità effettiva massimi](#) disponibili su un volume EBS, prendi in considerazione la possibilità di eseguire lo striping del volume EBS per superare queste limitazioni.

I volumi di striping (RAID) consentono di soddisfare i requisiti di IOPS e di throughput e sono limitati dall'IOPS e dalla larghezza di banda massimi supportati da una particolare istanza. Per ulteriori informazioni sulle opzioni di striping, consulta la [configurazione RAID](#) nella EC2 documentazione di Amazon. Puoi anche utilizzare Storage Spaces su un server autonomo. Per ulteriori informazioni, consulta la [documentazione di Microsoft](#).

Escludere i file di SQL Server dal software antivirus

Quando configuri le impostazioni del software antivirus, assicurati di escludere i file e le directory di SQL Server dalla scansione antivirus. Per informazioni dettagliate e un elenco di file e directory da escludere, vedere [Come scegliere un software antivirus da eseguire su computer che eseguono SQL Server](#) sul sito Web di Microsoft.

Se non si escludono questi file di SQL Server, potrebbero essere danneggiati o messi in quarantena dal software antivirus quando SQL Server deve utilizzarli. La mancata esclusione di questi file può anche causare problemi di prestazioni.

Configurazione di SQL Server

Questa sezione fornisce le best practice per configurare i database SQL Server per ottimizzare le prestazioni, evitare gli errori più comuni e soddisfare i requisiti di sicurezza e disponibilità. Puoi implementare queste modifiche prima o dopo la migrazione dei database su Amazon EC2. Le seguenti sezioni forniscono suggerimenti e best practice per la configurazione.

Argomenti

- [Configura tempdb per ridurre i conflitti](#)
- [Imposta MAXDOP per prestazioni ottimali](#)
- [Modifica la soglia di costo del parallelismo](#)
- [Ottimizza per carichi di lavoro ad hoc](#)
- [Usa i trace flag per migliorare le prestazioni](#)
- [Installa le patch più recenti](#)
- [Limita la memoria massima del server per evitare la pressione della memoria](#)
- [Utilizza il massimo livello di compatibilità del database](#)
- [Controlla il numero di VLFs](#)
- [Controlla le impostazioni di crescita automatica del database](#)

Configura tempdb per ridurre i conflitti

Ti consigliamo di configurare tempdb con più file di dati di uguali dimensioni e con uguale fattore di crescita.

Su un server di database occupato che utilizza molto tempdb, potresti notare un forte blocco quando il server è sottoposto a un carico elevato. Potresti notare che le attività sono in attesa di attesa, risorse che puntano a pagine in tempdb. Queste pagine potrebbero essere [pagine Page Free Space \(PFS\) e Shared Global Allocation Map \(SGM\)](#) con il formato $2:x: x$ (ad esempio, 2:1:1 o 2:1:2).

Per migliorare la concorrenza di tempdb, è possibile aumentare il numero di file di dati in tempdb per massimizzare la larghezza di banda del disco e ridurre i conflitti nelle strutture di allocazione. Ecco alcune linee guida:

- Se il numero di processori logici è uguale o inferiore a 8: utilizzare lo stesso numero di file di dati e processori logici.

- Se il numero di processori logici è superiore a 8: utilizzare 8 file di dati.

Se il conflitto persiste, aumenta il numero di file di dati in multipli di 4 fino a quando il conflitto non viene risolto, fino a raggiungere il numero di processori logici presenti sul server. Ciò contribuirà a evitare il conflitto SGAM in tempdb. [Se utilizzi SQL Server 2014 o una versione precedente, devi anche abilitare trace flag 1118](#). Questo flag impone l'allocazione delle pagine su estensioni uniformi anziché su estensioni miste, il che riduce al minimo le scansioni sulla pagina SGAM e riduce i conflitti.

A partire da SQL Server 2016 (13.x), questo comportamento è controllato dalle opzioni e di AUTOGROW_SINGLE_FILE AUTOGROW_ALL_FILES ALTER DATABASE Per esempio:

```
alter database <database name> MODIFY FILEGROUP [PRIMARY] AUTOGROW_ALL_FILES
```

Per ulteriori informazioni sull'impostazione di queste opzioni, vedere la [documentazione di Microsoft SQL Server](#).

Imposta MAXDOP per prestazioni ottimali

Il grado massimo di parallelismo (MAXDOP) è un'opzione di configurazione del server per l'esecuzione di SQL Server su più server. CPUs Controlla il numero di processori utilizzati per eseguire una singola istruzione nell'esecuzione parallela del piano. Il valore predefinito è 0, che consente a SQL Server di utilizzare tutti i processori disponibili. Ciò può influire sulle prestazioni e non è ottimale per la maggior parte dei casi d'uso.

Utilizza le seguenti linee guida per configurare il valore MAXDOP per SQL Server.

Nodi NUMA	Processori logici	Valore MAXDOP
Singolo	≤ 8	4, 2 o numero di core (per uno o due core)
Singolo	> 8	8, 4 o 2
Più di uno	≤ 16	8, 4 o 2
Più di uno	> 16	16, 8, 4 o 2

Note

L'impostazione di MAXDOP su 2, 4 o 8 generalmente fornisce i migliori risultati nella maggior parte dei casi d'uso. Ti consigliamo di testare il tuo carico di lavoro e monitorare eventuali tipi di attesa legati al parallelismo, come CXPACKET

È possibile utilizzare la seguente query per raccogliere la configurazione NUMA corrente per SQL Server 2016 e versioni successive:

```
select @@SERVERNAME,
SERVERPROPERTY('ComputerNamePhysicalNetBIOS'),
cpu_count,
hyperthread_ratio,
softnuma_configuration,
softnuma_configuration_desc,
socket_count,
numa_node_count
from
sys.dm_os_sys_info
```

dove:

- **cpu_count** si riferisce al numero di elementi logici CPUs presenti nel sistema.
- **hyperthread_ratio** è il rapporto tra il numero di core esposti da un processore fisico.
- **softnuma_configuration** è 01, o2:
 - 0 (OFF): impostazione predefinita
 - 1 (automated): Soft-NUMA
 - 2 (manual): Soft-NUMA
- **softnuma_configuration_desc** è OFF, o: ON MANUAL
 - OFF indica che la funzione soft-NUMA è disattivata.
 - ON indica che SQL Server decide automaticamente le dimensioni dei nodi NUMA.
 - MANUAL indica che soft-NUMA è configurato manualmente.
- **socket_count** è il numero di socket del processore.
- **numa_node_count** è il numero di nodi NUMA disponibili nel sistema.

Per controllare il valore MAXDOP corrente, usa:

```
$ sp_configure 'max_degree_of_parallelism'
```

Per ulteriori informazioni su MAXDOP, vedere la documentazione di [Microsoft SQL Server](#).

Modificare la soglia di costo del parallelismo

La soglia di costo del parallelismo determina quali query sono candidate all'esecuzione parallela. Il valore predefinito di questa proprietà è 5, il che significa che l'ottimizzatore passa a un piano parallelo se il costo di un piano seriale è superiore a 5 (che si riferisce a un'unità di costo astratta, non al tempo stimato). Si consiglia di impostare questa proprietà su un numero più alto.

Il valore predefinito era appropriato quando i processori avevano prezzi elevati, la potenza di elaborazione era bassa e l'elaborazione delle query era più lenta di adesso. I processori odierni sono molto più veloci. Di conseguenza, le query relativamente più piccole (ad esempio, data una soglia di costo di 32) non trarranno grandi vantaggi dall'esecuzione parallela, soprattutto in considerazione del sovraccarico associato al coordinamento dell'esecuzione parallela.

Nella maggior parte dei casi, una soglia di costo del parallelismo impostata su 50 è un buon punto di partenza. Ecco un esempio di come configurare la soglia di costo del parallelismo:

```
USE sampledb;
GO
EXEC sp_configure 'show advanced options', 1 ;
GO
RECONFIGURE
GO
EXEC sp_configure 'cost threshold for parallelism', 50 ;
GO
RECONFIGURE
GO
```

Ottimizza per carichi di lavoro ad hoc

Abilita l'opzione di ottimizzazione per carichi di lavoro ad hoc per migliorare l'efficienza della cache del piano per i carichi di lavoro che contengono molti batch ad hoc monouso. Quando si esegue

inizialmente una query ad hoc, il motore di database memorizza nella cache uno stub del piano compilato anziché il piano di esecuzione completo, risparmiando così spazio nella cache del piano. Se si esegue nuovamente il batch ad hoc, il motore di database riconosce che il batch è stato eseguito in precedenza e sostituisce lo stub del piano compilato con il piano completo e compilato nella cache del piano.

Per verificare se questa opzione è abilitata, utilizzate la query:

```
$ sp_configure 'optimize for ad hoc workloads'
```

Per ulteriori informazioni sull'ottimizzazione per carichi di lavoro ad hoc, consulta la documentazione di [Microsoft SQL Server](#).

Usa i trace flag per migliorare le prestazioni

Valuta la possibilità di utilizzare i flag di traccia di SQL Server applicabili al tuo ambiente per migliorare le prestazioni. Per esempio:

- 4199: abilita le modifiche all'ottimizzatore di query (QO) rilasciate in SQL Server Cumulative Updates (CUs) e Service Pack (). SPs
- 8048: converte gli oggetti di memoria partizionati con NUMA in oggetti di memoria partizionati con CPU.
- 9024: converte un oggetto di memoria del pool di log globale in un oggetto di memoria partizionato con NUMA.

Gli esempi seguenti illustrano come attivare e disattivare i flag di traccia per SQL Server su Amazon. EC2 Se riscontri problemi quando abiliti il tracciamento, assicurati di disporre delle autorizzazioni appropriate per l'account.

Per attivare il trace flag 4199, esegui:

```
dbcc traceon (4199, -1);
```

Per controllare lo stato del flag di traccia, esegui:

```
dbcc tracestatus (4199);
```

Per disattivare il trace flag 4199, esegui:

```
dbcc traceoff (4199, -1);  
dbcc tracestatus (4199);
```

Per un elenco completo dei trace flag, consulta la [documentazione di Microsoft SQL Server](#).

Installa le patch più recenti

A partire da SQL Server 2017, Microsoft ha [interrotto il rilascio dei Service Pack \(SPs\)](#). Rilascia solo aggiornamenti cumulativi (CUs) e aggiornamenti critici (GDRs).

SPs include correzioni cruciali per SQL Server, quindi assicurati che sia stato installato l'SP più recente. Inoltre, se possibile, installa il pacchetto CU più recente.

Per informazioni sugli ultimi aggiornamenti di SQL Server, vedere [Aggiornamenti più recenti per Microsoft SQL Server](#) sul sito Web Microsoft.

Limita la memoria massima del server per evitare la pressione della memoria

Per motivi di prestazioni, SQL Server non rilascia memoria già allocata. All'avvio, SQL Server occupa lentamente la memoria specificata nell'opzione min_server_memory, quindi continua a crescere fino a raggiungere il valore specificato nell'opzione max_server_memory. (Per ulteriori informazioni su queste impostazioni, vedere Opzioni di [configurazione della memoria del server nella documentazione di SQL Server](#).)

La memoria di SQL Server ha due componenti: il pool di buffer e il pool non buffer (chiamato anche memory to leave o MTL). Il valore dell'opzione max_server_memory determina la dimensione del buffer pool di SQL Server, che comprende buffer cache, procedure cache, plan cache, buff structures e altre cache.

A partire da SQL Server 2012, gli account min_server_memory e max_server_memory per tutte le allocazioni di memoria per tutte le cache, tra cui,,, e. SQLGENERAL SQLBUFFERPOOL SQLQUERYCOMPILE SQLQUERYPLAN SQLQUERYEXEC SQLOPTIMIZER SQLCLR [Per un elenco completo degli addetti alla memoria in max_server_memory, vedere sys.dm_os_memory_clerks nella documentazione di Microsoft SQL Server](#).

Per verificare il valore attuale di max_server_memory, usa il comando:

```
$ sp_configure 'max_server_memory'
```

Ti consigliamo di limitare max_server_memory a un valore che non causi una pressione della memoria a livello di sistema. Non esiste una formula universale applicabile a tutti gli ambienti, ma abbiamo fornito alcune linee guida in questa sezione. max_server_memory è un'opzione dinamica, quindi può essere modificata in fase di esecuzione.

Come punto di partenza, puoi determinare max_server_memory come segue:

```
max_server_memory = total_RAM - (memory_for_the_OS + MTL)
```

dove:

- La memoria per il sistema operativo è di 1-4 GB.
- MTL (memory to leave) include la dimensione dello stack, che è di 2 MB su macchine a 64 bit per thread di lavoro e può essere calcolata come segue: $MTL = \text{stack_size} * \text{max_worker_threads}$

In alternativa, puoi usare:

```
max_server_memory = total_RAM - (1 GB for the OS  
+ memory_basis_amount_of_RAM_on_the_server)
```

dove la quantità di memoria RAM di base è determinata come segue:

- Se la RAM sul server è compresa tra 4 GB e 16 GB, lascia 1 GB per 4 GB di RAM. Ad esempio, per un server con 16 GB, lascia 4 GB.
- Se la RAM sul server è superiore a 16 GB, lascia 1 GB per 4 GB di RAM fino a 16 GB e 1 GB per 8 GB di RAM oltre 16 GB.

Ad esempio, se un server dispone di 256 GB di RAM, il calcolo sarà:

- 1 GB per il sistema operativo
- Fino a 16 GB di RAM: $16/4 = 4$ GB
- RAM rimanente superiore a 16 GB: $(256-16) / 8 = 30$

- RAM totale da lasciare: $1 + 4 + 30 = 35$ GB
- memoria massima del server: $256 - 35 = 221$ GB

Dopo la configurazione iniziale, monitora la memoria che puoi liberare per una durata tipica del carico di lavoro per determinare se è necessario aumentare o diminuire la memoria allocata a SQL Server.

Note

Windows segnala la notifica di risorse di memoria insufficienti a 96 MB, quindi è necessario un buffer, ma è possibile impostare i Mbyte disponibili in modo che siano superiori a 1 GB su server più grandi con 256 GB o più RAM.

Per ulteriori informazioni, vedere la [Memory Management Architecture Guide](#) nella documentazione di Microsoft SQL Server.

Utilizza il massimo livello di compatibilità del database

Assicurati di utilizzare l'attuale livello di compatibilità del database per sfruttare i miglioramenti più recenti di SQL Server. Questo è importante da verificare perché quando ripristini un database da una versione precedente a una versione superiore, manterrà il livello di compatibilità della versione precedente. Alcuni dei miglioramenti più recenti del database sono efficaci solo quando si imposta la compatibilità del database al livello più recente disponibile per la versione del motore installata.

Per verificare l'attuale compatibilità del database, usa:

```
$ select name, compatibility_level from sys.databases
```

Per ulteriori informazioni sui livelli di compatibilità dei database, vedere la [documentazione di Microsoft SQL Server](#).

Controlla il numero di VLFs

Pre-alloca la dimensione massima dei dati e dei file di registro. Per prestazioni migliori, controllate il numero di file di registro virtuali (VLFs) preallocando lo spazio e correggendo le impostazioni di crescita automatica (crescita automatica) per i file di registro.

In genere, un fattore di crescita automatica di 8 GB funziona bene nella maggior parte degli ambienti di produzione. Prendi in considerazione la possibilità di espandere i file di registro delle transazioni in blocchi da 8 GB. Un numero maggiore di file VLFs può prolungare i tempi di backup e ripristino del database e può causare problemi di prestazioni con qualsiasi operazione (ad esempio la replica) che richieda l'analisi dei file di registro.

[Per ulteriori informazioni sull'algoritmo di creazione e crescita del VLF, consultate il SQLskills blog.](#)

Controlla le impostazioni di crescita automatica del database

Qualsiasi transazione che richieda la crescita dei dati o del file di registro include il tempo impiegato dall'operazione di crescita del file. Il file cresce in base alla dimensione di incremento definita dall'opzione FILEGROWTH. È possibile cercare gli eventi di crescita dei file nelle tracce del profiler di SQL Server. Se la crescita dei file richiede molto tempo, è possibile che si verifichino tipi di attesa come ASYNC_IO_COMPLETION, ad esempio, quando l'elaborazione dei dati è molto lenta. Tali tipi di attesa non solo influiscono sulle prestazioni, ma possono anche causare timeout delle transazioni. Se tale transazione blocca le risorse ricercate da altre transazioni, il timeout comporterebbe gravi problemi di blocco del server.

Per questo motivo, ti consigliamo di configurare le impostazioni di crescita automatica con molta attenzione. Tieni inoltre presente che:

- La crescita dei file è una delle operazioni più costose in SQL Server.
- La crescita automatica frequente in piccoli blocchi può portare alla frammentazione del disco.
- [La crescita automatica frequente dei file di registro genera un numero elevato di file di registro virtuali \(VLFs\) e influisce sulle prestazioni, come illustrato nella sezione precedente.](#)

Tutti questi motivi potrebbero comportare un rallentamento dell'avvio del database e un aumento dei tempi di backup e ripristino.

Idealmente, è consigliabile pre-ampliare i file in modo proattivo, sulla base di un monitoraggio regolare. Scegliete con attenzione se impostare la crescita automatica come percentuale o come valore statico (in MB). In genere, impostare la crescita automatica su un ottavo della dimensione del file è un buon punto di partenza, ma potrebbe non essere la scelta giusta. (Ad esempio, questa percentuale sarebbe troppo alta se il file di dati ha dimensioni diverse TBs).

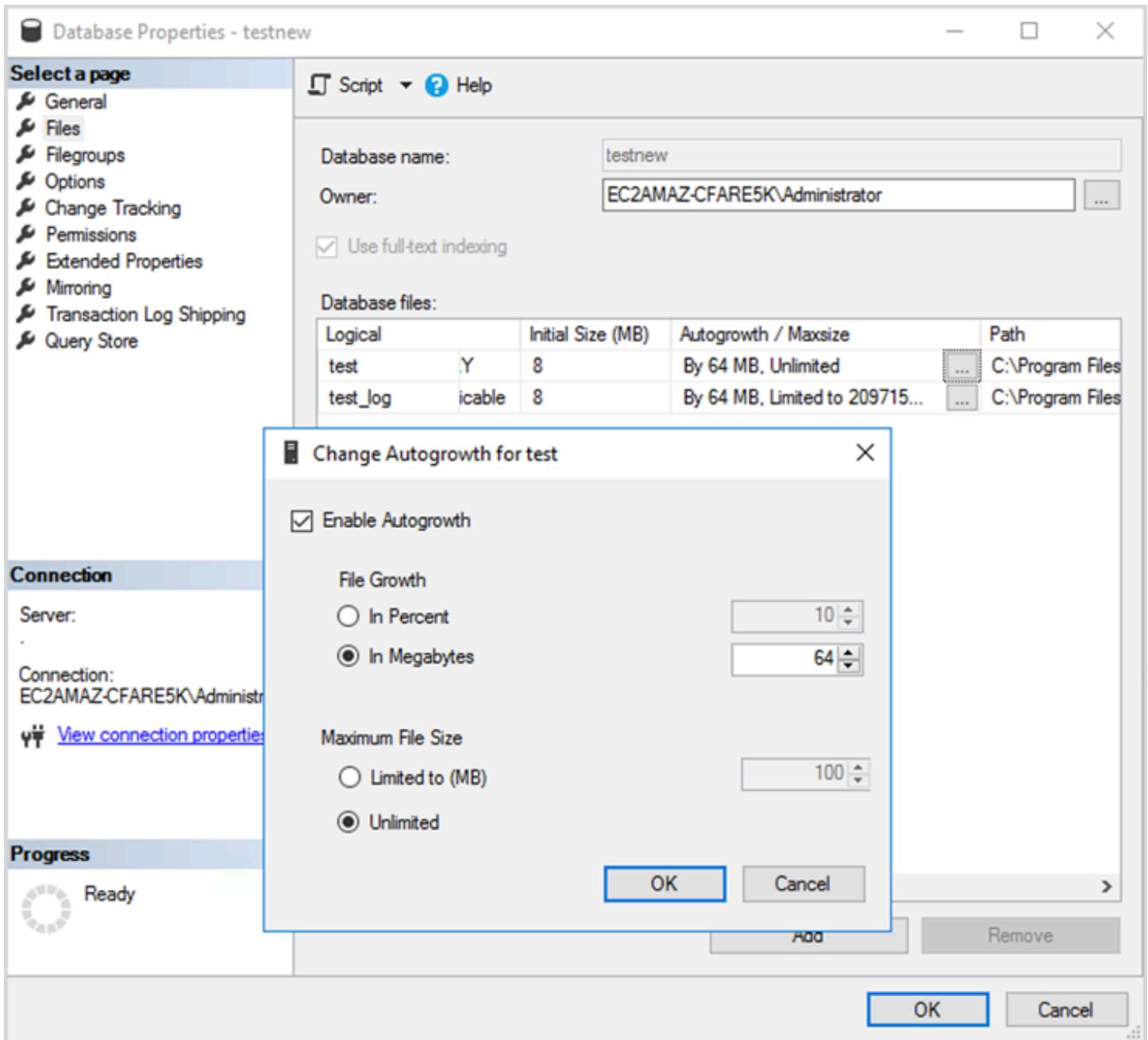
Nella maggior parte dei casi, un valore di crescita automatica di 1024 MB è ideale per i file di dati presenti nella maggior parte dei database di grandi dimensioni. Per i file di registro, 512 MB sono un

buon punto di partenza. Per le misure di emergenza, consigliamo vivamente di impostare il valore di crescita automatica, ma di ingrandire i file manualmente per alcuni mesi in base alle tendenze passate.

 Note

L'impostazione della crescita automatica dovrebbe essere una misura di emergenza, quindi dovresti impostarla dopo aver preallocato lo spazio di archiviazione su un file.

È possibile modificare le impostazioni di crescita automatica utilizzando SQL Server Management Studio (SSMS) o Transact-SQL. La seguente illustrazione della schermata mostra le impostazioni di crescita automatica in SSMS.



Quando utilizzate l'opzione FILEGROWTH per dati e file di registro, scegliete con attenzione se impostarla come percentuale o come valore statico (in MB). L'impostazione di una percentuale comporta una crescita sempre maggiore dei file, quindi potresti preferire utilizzare una dimensione statica per un migliore controllo del rapporto di crescita.

- Nelle versioni precedenti a SQL Server 2022 (16.x), i log delle transazioni non possono utilizzare l'inizializzazione istantanea dei file, pertanto i tempi di crescita prolungati dei log sono particolarmente importanti.

- A partire da SQL Server 2022 (16.x, tutte le edizioni), l'inizializzazione istantanea dei file può favorire gli eventi di crescita dei log delle transazioni fino a 64 MB. L'incremento predefinito delle dimensioni di crescita automatica per i nuovi database è di 64 MB. Gli eventi di crescita automatica del file di registro delle transazioni di dimensioni superiori a 64 MB non possono trarre vantaggio dall'inizializzazione istantanea dei file.

Configurazione dei gruppi di disponibilità Always On

Se utilizzi librerie client native per SQL Server versione 2012 e successive e librerie .NET Framework 4.5, puoi utilizzare il `MultiSubnetFailover` parametro per modificare il comportamento di connessione. Si consiglia di impostare questo parametro su `TRUE`. Ciò consentirà un failover più rapido con i gruppi di disponibilità Always On.

Note

Se disponi di applicazioni legacy che non possono utilizzare il `MultiSubnetFailover` parametro, puoi posizionare un Network Load Balancer davanti alle istanze di SQL Server. Il sistema di bilanciamento utilizza un controllo dello stato che determina quale database SQL Server è attivo e invia il traffico all'istanza che attualmente ospita quel database. Il load balancer si estende su una o più zone di disponibilità. È possibile utilizzare una porta dedicata come 59999 per il controllo dello stato di salute e quindi modificare il parametro del gruppo di cluster per rispondere a quella porta. Ciò consente di ridurre il tempo di failover di SQL Server a circa un minuto senza utilizzare il `MultiSubnetFailover` parametro. Per istruzioni dettagliate, consulta il post sul blog [Ridurre i tempi di failover per SQL Server su EC2 un'istanza Amazon utilizzando Network Load Balancer](#).

Due impostazioni influiscono sul modo in cui il listener del gruppo di disponibilità viene registrato con DNS: `RegisterAllProviders` IP e TTL. `HostRecord`

Imposta `RegisterAllProviders` IP su true quando usi i gruppi di disponibilità Always On

Ti consigliamo di impostare `RegisterAllProvidersIP` su 1 (true). Quando il listener del gruppo di disponibilità viene creato con `RegisterAllProvidersIP` impostato su 1, tutti gli indirizzi IP di quel listener vengono registrati nel DNS. Quando `RegisterAllProvidersIP` è impostato su 0 (false), viene registrato un solo IP attivo.

In caso di failover, quando la replica primaria passa da una sottorete all'altra, il vecchio indirizzo IP non viene registrato e il nuovo indirizzo IP viene registrato. Il DNS viene aggiornato con il nuovo IP quando il listener del gruppo di disponibilità è online. Tuttavia, i sistemi client non risolveranno il nome del listener nel nuovo indirizzo IP fino alla scadenza della voce attualmente memorizzata nella cache.

Imposta HostRecord TTL su 60 o meno quando usi i gruppi di disponibilità Always On

L'impostazione HostRecordTTL controlla il Time to Live (TTL) per le voci DNS memorizzate nella cache. Il valore predefinito è 1200 secondi. Ti consigliamo di modificare il HostRecordTTL su un'impostazione molto più bassa (60 secondi o meno). Ciò fa sì che il valore memorizzato nella cache scada prima, quindi in caso di failover, i sistemi client possono risolvere il nuovo IP più rapidamente.

Disattiva il fallback automatico per il gruppo di cluster Always On

Verifica che il fallback automatico sia disabilitato per i gruppi di disponibilità Always On in Windows Cluster Manager.

Configurazione dei backup

Come discusso nella sezione [Ottimizzazione del layout del disco o della distribuzione dei file](#), consigliamo di inviare i backup nativi di SQL Server su un'unità separata. Valuta anche la possibilità di scattare un'istantanea pianificata del volume EBS in cui risiedono i file di backup.

Miglioramento dell'ottimizzazione del database

Questa sezione fornisce le migliori pratiche per migliorare le prestazioni quando si lavora con l'ottimizzatore di query di SQL Server. Descrive in che modo la ricostruzione degli indici e l'aggiornamento regolare delle statistiche delle tabelle possono aiutare a ottimizzare i piani di esecuzione. Le sezioni seguenti forniscono suggerimenti e procedure consigliate per la configurazione.

Argomenti

- [Ricostruisci gli indici](#)
- [Aggiorna le statistiche](#)

Ricostruisci gli indici

Affinché l'ottimizzatore di query generi i migliori piani di query possibili e utilizzi gli indici giusti, gli indici non devono essere frammentati. Gli indici si frammentano nel tempo in base alla frequenza di aggiornamento, inserimento o eliminazione. Assicurati che le tabelle vengano reindividuate regolarmente. La frequenza di ricostruzione dipende dalla velocità con cui il database gestisce le operazioni DML (Data Manipulation Language).

Un buon punto di partenza sarebbe ricostruire gli indici frammentati per più del 30% e riorganizzare gli indici frammentati per meno del 30%. Il valore del 30% funziona nella maggior parte dei casi d'uso, ma se i piani di query persistono a causa di indici non utilizzati, potrebbe essere necessario rivedere questa percentuale.

Usa una query come la seguente per verificare la frammentazione:

```
SELECT OBJECT_NAME(OBJECT_ID), index_id, index_type_desc, index_level,
       avg_fragmentation_in_percent, avg_page_space_used_in_percent, page_count
  FROM sys.dm_db_index_physical_stats
 (DB_ID(N'<your_database>'), NULL, NULL, NULL, 'SAMPLED')
 ORDER BY avg_fragmentation_in_percent DESC
```

Ti consigliamo di creare un processo di manutenzione per ricostruire gli indici su base regolare.

Aggiorna le statistiche

Come per gli indici frammentati, se l'ottimizzatore non dispone di up-to-date informazioni sulla distribuzione dei valori chiave (statistiche) delle colonne della tabella, non può generare piani di esecuzione ottimali. Si consiglia di aggiornare regolarmente le statistiche per tutte le tabelle. La frequenza degli aggiornamenti dipende dalla velocità con cui il database gestisce le operazioni DML, ma in genere viene eseguita due volte a settimana durante le ore non di punta. Tuttavia, evita di aggiornare le statistiche nei giorni in cui ricostruisci gli indici. Per ulteriori informazioni sull'aggiornamento delle statistiche, vedere la [documentazione di Microsoft SQL Server](#).

Per l'ottimizzazione del database, si consiglia di utilizzare uno script di gestione degli indici e delle statistiche. Per un esempio, vedi lo [script di gestione degli indici e delle statistiche di SQL Server](#) disponibile nel sito Web di SQL Server Maintenance Solution.

Ottimizzazione delle distribuzioni di SQL Server su Amazon EC2 con Launch Wizard AWS

AWS Launch Wizard è il metodo principale per le distribuzioni di SQL Server a istanza singola e ad alta disponibilità (HA) su Amazon EC2. Le implementazioni di Launch Wizard si basano sul [AWS Well-Architected Framework](#) e sono ottimizzate per sicurezza, affidabilità, efficienza delle prestazioni e risparmio sui costi.

Launch Wizard semplifica le distribuzioni di SQL Server e semplifica anche la configurazione di SQL Server. Le sue funzionalità includono:

- Selezione automatica AWS delle risorse: Launch Wizard può consigliare il tipo di istanza ottimale in base alla CPU virtuale (vCPU), alla memoria e ai requisiti di rete. Può anche consigliare il tipo di volume in base all'unità di archiviazione e al throughput.
- Monitoraggio con un clic: Launch Wizard si integra [con CloudWatch Amazon Application Insights](#) per configurare il monitoraggio delle distribuzioni di SQL Server HA su AWS. Quando selezioni questa opzione, Application Insights imposta automaticamente le metriche, i log e gli allarmi pertinenti e inizia a monitorare i carichi di lavoro appena distribuiti. CloudWatch
- Gruppi di risorse delle applicazioni per una facile individuabilità: Launch Wizard crea un gruppo di risorse per AWS tutte le risorse create per l'applicazione SQL Server. È possibile gestire, applicare patch e mantenere le applicazioni SQL Server dalla AWS Systems Manager console.

Launch Wizard fornisce modelli di codice riutilizzabili AWS CloudFormation. Questi modelli possono fungere da base per le successive distribuzioni delle applicazioni. Per ulteriori informazioni, consulta la [panoramica](#) e la guida per l'[utente](#) di AWS Launch Wizard.

Passaggi successivi

Questa guida ha illustrato alcune delle migliori pratiche per configurare ed eseguire carichi di lavoro di Microsoft SQL Server su Amazon EC2. Seguire queste linee guida nelle fasi di pianificazione e implementazione del processo di migrazione ti aiuterà a stabilire un server stabile nell'ambiente di produzione.

Per ulteriori informazioni su queste attività di configurazione, consultate i collegamenti forniti in ogni sezione e visitate le pagine Web elencate nella sezione [Risorse aggiuntive](#).

Risorse aggiuntive

Strategie, guide e modelli correlati

- [Strategia di migrazione per database relazionali](#)
- Sezione Patterns del [sito Web AWS Prescriptive Guidance](#) (puoi utilizzare i filtri disponibili in quella pagina per visualizzare i modelli in base alla strategia di migrazione, ad esempio re-architect, rehost, relocate o replatform)

AWS resources

- [AWS documentazione](#)
- [AWS riferimento generale](#)
- [AWS glossario](#)

AWS servizi

- [Amazon EBS](#)
- [Amazon EC2](#)

Altre risorse

- [Come spostare Microsoft SQL Server tempdb su dischi su instance/ephemeral Amazon EC2](#)
- [Stripe Windows Ephemeral Disks al momento del lancio](#)
- [Di quanta memoria necessita effettivamente il mio SQL Server?](#)
- [SQL Server Wait Statistics \(o per favore dimmi dove fa male...\)](#)
- [Timeout di connessione nel gruppo di disponibilità multi-subnet](#)
- [Pianifica la cache e l'ottimizzazione per carichi di lavoro ad hoc](#)
- [RAM, memoria virtuale, file di pagina e gestione della memoria in Windows](#)
- [Come determinare la dimensione appropriata del file di pagina per le versioni a 64 bit di Windows](#)

Cronologia dei documenti

La tabella seguente descrive le modifiche significative apportate a questa guida. Per ricevere notifiche sugli aggiornamenti futuri, puoi abbonarti a un [feed RSS](#).

Modifica	Descrizione	Data
Informazioni corrette	Sono state corrette le informazioni sulla soglia di costo della proprietà di parallelismo . Il suo valore è misurato in unità di costo, non in termini di tempo.	4 dicembre 2023
Linee guida aggiornate	Sono state aggiornate le sezioni sull'impostazione della dimensione dell'unità di allocazione NTFS, sul blocco delle pagine in memoria, sull'utilizzo delle funzionalità di offload delle attività, sull'uso dello striping, sulla modifica della soglia di costo del parallelismo, sull'uso dei flag di traccia e sull'utilizzo delle impostazioni di crescita automatica del database.	8 agosto 2023
Sono state aggiunte linee guida	Sono state aggiunte informazioni sull'utilizzo di Network Load Balancer per applicazioni legacy che non possono utilizzare il MultiSubnetFailover rparametro.	11 novembre 2022
Codice fisso	È stato corretto il PowerShell codice nella sezione	27 giugno 2022

sull'[inizializzazione](#)
[dell'instance store.](#)

[È stata aggiunta una nuova sezione](#)

Sono state aggiunte informazioni su [AWS Launch Wizard for SQL Server.](#)

18 agosto 2021

[Pubblicazione iniziale](#)

—

21 luglio 2020

AWS Glossario delle linee guida prescrittive

I seguenti sono termini di uso comune nelle strategie, nelle guide e nei modelli forniti da AWS Prescriptive Guidance. Per suggerire voci, utilizza il link Fornisci feedback alla fine del glossario.

Numeri

7 R

Sette strategie di migrazione comuni per trasferire le applicazioni sul cloud. Queste strategie si basano sulle 5 R identificate da Gartner nel 2011 e sono le seguenti:

- Rifattorizzare/riprogettare: trasferisci un'applicazione e modifica la sua architettura sfruttando appieno le funzionalità native del cloud per migliorare l'agilità, le prestazioni e la scalabilità. Ciò comporta in genere la portabilità del sistema operativo e del database. Esempio: migra il tuo database Oracle locale all'edizione compatibile con Amazon Aurora PostgreSQL.
- Ridefinire la piattaforma (lift and reshape): trasferisci un'applicazione nel cloud e introduci un certo livello di ottimizzazione per sfruttare le funzionalità del cloud. Esempio: migra il tuo database Oracle locale ad Amazon Relational Database Service (Amazon RDS) per Oracle in Cloud AWS
- Riacquistare (drop and shop): passa a un prodotto diverso, in genere effettuando la transizione da una licenza tradizionale a un modello SaaS. Esempio: migra il tuo sistema di gestione delle relazioni con i clienti (CRM) su Salesforce.com.
- Eseguire il rehosting (lift and shift): trasferisci un'applicazione sul cloud senza apportare modifiche per sfruttare le funzionalità del cloud. Esempio: migra il database Oracle locale su Oracle su un'istanza in EC2 Cloud AWS
- Trasferire (eseguire il rehosting a livello hypervisor): trasferisci l'infrastruttura sul cloud senza acquistare nuovo hardware, riscrivere le applicazioni o modificare le operazioni esistenti. Si esegue la migrazione dei server da una piattaforma locale a un servizio cloud per la stessa piattaforma. Esempio: migra un'applicazione suMicrosoft Hyper-V. AWS
- Riesaminare (mantenere): mantieni le applicazioni nell'ambiente di origine. Queste potrebbero includere applicazioni che richiedono una rifattorizzazione significativa che desideri rimandare a un momento successivo e applicazioni legacy che desideri mantenere, perché non vi è alcuna giustificazione aziendale per effettuarne la migrazione.
- Ritirare: disattiva o rimuovi le applicazioni che non sono più necessarie nell'ambiente di origine.

A

ABAC

Vedi controllo degli accessi [basato sugli attributi](#).

servizi astratti

Vedi [servizi gestiti](#).

ACIDO

Vedi [atomicità, consistenza, isolamento, durata](#).

migrazione attiva-attiva

Un metodo di migrazione del database in cui i database di origine e di destinazione vengono mantenuti sincronizzati (utilizzando uno strumento di replica bidirezionale o operazioni di doppia scrittura) ed entrambi i database gestiscono le transazioni provenienti dalle applicazioni di connessione durante la migrazione. Questo metodo supporta la migrazione in piccoli batch controllati anziché richiedere una conversione una tantum. È più flessibile ma richiede più lavoro rispetto alla migrazione [attiva-passiva](#).

migrazione attiva-passiva

Un metodo di migrazione del database in cui i database di origine e di destinazione vengono mantenuti sincronizzati, ma solo il database di origine gestisce le transazioni provenienti dalle applicazioni di connessione mentre i dati vengono replicati nel database di destinazione. Il database di destinazione non accetta alcuna transazione durante la migrazione.

funzione di aggregazione

Una funzione SQL che opera su un gruppo di righe e calcola un singolo valore restituito per il gruppo. Esempi di funzioni aggregate includono SUM e. MAX

Intelligenza artificiale

Vedi [intelligenza artificiale](#).

AIOps

Guarda le [operazioni di intelligenza artificiale](#).

anonimizzazione

Il processo di eliminazione permanente delle informazioni personali in un set di dati.

L'anonymizzazione può aiutare a proteggere la privacy personale. I dati anonimi non sono più considerati dati personali.

anti-modello

Una soluzione utilizzata di frequente per un problema ricorrente in cui la soluzione è controproducente, inefficace o meno efficace di un'alternativa.

controllo delle applicazioni

Un approccio alla sicurezza che consente l'uso solo di applicazioni approvate per proteggere un sistema dal malware.

portfolio di applicazioni

Una raccolta di informazioni dettagliate su ogni applicazione utilizzata da un'organizzazione, compresi i costi di creazione e manutenzione dell'applicazione e il relativo valore aziendale.

Queste informazioni sono fondamentali per [il processo di scoperta e analisi del portfolio](#) e aiutano a identificare e ad assegnare la priorità alle applicazioni da migrare, modernizzare e ottimizzare.

intelligenza artificiale (IA)

Il campo dell'informatica dedicato all'uso delle tecnologie informatiche per svolgere funzioni cognitive tipicamente associate agli esseri umani, come l'apprendimento, la risoluzione di problemi e il riconoscimento di schemi. Per ulteriori informazioni, consulta la sezione [Che cos'è l'intelligenza artificiale?](#)

operazioni di intelligenza artificiale (AIOps)

Il processo di utilizzo delle tecniche di machine learning per risolvere problemi operativi, ridurre gli incidenti operativi e l'intervento umano e aumentare la qualità del servizio. Per ulteriori informazioni su come AIOps viene utilizzato nella strategia di AWS migrazione, consulta la [guida all'integrazione delle operazioni](#).

crittografia asimmetrica

Un algoritmo di crittografia che utilizza una coppia di chiavi, una chiave pubblica per la crittografia e una chiave privata per la decripttografia. Puoi condividere la chiave pubblica perché non viene utilizzata per la decripttografia, ma l'accesso alla chiave privata deve essere altamente limitato.

atomicità, consistenza, isolamento, durabilità (ACID)

Un insieme di proprietà del software che garantiscono la validità dei dati e l'affidabilità operativa di un database, anche in caso di errori, interruzioni di corrente o altri problemi.

Controllo degli accessi basato su attributi (ABAC)

La pratica di creare autorizzazioni dettagliate basate su attributi utente, come reparto, ruolo professionale e nome del team. Per ulteriori informazioni, consulta [ABAC AWS](#) nella documentazione AWS Identity and Access Management (IAM).

fonte di dati autorevole

Una posizione in cui è archiviata la versione principale dei dati, considerata la fonte di informazioni più affidabile. È possibile copiare i dati dalla fonte di dati autorevole in altre posizioni allo scopo di elaborarli o modificarli, ad esempio anonimizzandoli, oscurandoli o pseudonimizzandoli.

Zona di disponibilità

Una posizione distinta all'interno di un edificio Regione AWS che è isolata dai guasti in altre zone di disponibilità e offre una connettività di rete economica e a bassa latenza verso altre zone di disponibilità nella stessa regione.

AWS Cloud Adoption Framework (CAF)AWS

Un framework di linee guida e best practice AWS per aiutare le organizzazioni a sviluppare un piano efficiente ed efficace per passare con successo al cloud. AWS CAF organizza le linee guida in sei aree di interesse chiamate prospettive: business, persone, governance, piattaforma, sicurezza e operazioni. Le prospettive relative ad azienda, persone e governance si concentrano sulle competenze e sui processi aziendali; le prospettive relative alla piattaforma, alla sicurezza e alle operazioni si concentrano sulle competenze e sui processi tecnici. Ad esempio, la prospettiva relativa alle persone si rivolge alle parti interessate che gestiscono le risorse umane (HR), le funzioni del personale e la gestione del personale. In questa prospettiva, AWS CAF fornisce linee guida per lo sviluppo delle persone, la formazione e le comunicazioni per aiutare a preparare l'organizzazione all'adozione del cloud di successo. Per ulteriori informazioni, consulta il [sito web di AWS CAF](#) e il [white paper AWS CAF](#).

AWS Workload Qualification Framework (WQF)AWS

Uno strumento che valuta i carichi di lavoro di migrazione dei database, consiglia strategie di migrazione e fornisce stime del lavoro. AWS WQF è incluso in (). AWS Schema Conversion Tool AWS SCT Analizza gli schemi di database e gli oggetti di codice, il codice dell'applicazione, le dipendenze e le caratteristiche delle prestazioni e fornisce report di valutazione.

B

bot difettoso

Un [bot](#) che ha lo scopo di interrompere o causare danni a individui o organizzazioni.

BCP

Vedi la [pianificazione della continuità operativa](#).

grafico comportamentale

Una vista unificata, interattiva dei comportamenti delle risorse e delle interazioni nel tempo. Puoi utilizzare un grafico comportamentale con Amazon Detective per esaminare tentativi di accesso non riusciti, chiamate API sospette e azioni simili. Per ulteriori informazioni, consulta [Dati in un grafico comportamentale](#) nella documentazione di Detective.

sistema big-endian

Un sistema che memorizza per primo il byte più importante. Vedi anche [endianness](#).

Classificazione binaria

Un processo che prevede un risultato binario (una delle due classi possibili). Ad esempio, il modello di machine learning potrebbe dover prevedere problemi come "Questa e-mail è spam o non è spam?" o "Questo prodotto è un libro o un'auto?"

filtro Bloom

Una struttura di dati probabilistica ed efficiente in termini di memoria che viene utilizzata per verificare se un elemento fa parte di un set.

distribuzioni blu/verdi

Una strategia di implementazione in cui si creano due ambienti separati ma identici. La versione corrente dell'applicazione viene eseguita in un ambiente (blu) e la nuova versione dell'applicazione nell'altro ambiente (verde). Questa strategia consente di ripristinare rapidamente il sistema con un impatto minimo.

bot

Un'applicazione software che esegue attività automatizzate su Internet e simula l'attività o l'interazione umana. Alcuni bot sono utili o utili, come i web crawler che indicizzano le informazioni su Internet. Alcuni altri bot, noti come bot dannosi, hanno lo scopo di disturbare o causare danni a individui o organizzazioni.

botnet

Reti di [bot](#) infettate da [malware](#) e controllate da un'unica parte, nota come bot herder o bot operator. Le botnet sono il meccanismo più noto per scalare i bot e il loro impatto.

ramo

Un'area contenuta di un repository di codice. Il primo ramo creato in un repository è il ramo principale. È possibile creare un nuovo ramo a partire da un ramo esistente e quindi sviluppare funzionalità o correggere bug al suo interno. Un ramo creato per sviluppare una funzionalità viene comunemente detto ramo di funzionalità. Quando la funzionalità è pronta per il rilascio, il ramo di funzionalità viene ricongiunto al ramo principale. Per ulteriori informazioni, consulta [Informazioni sulle filiali](#) (documentazione). GitHub

accesso break-glass

In circostanze eccezionali e tramite una procedura approvata, un mezzo rapido per consentire a un utente di accedere a un sito a Account AWS cui in genere non dispone delle autorizzazioni necessarie. Per ulteriori informazioni, vedere l'indicatore [Implementate break-glass procedures](#) nella guida Well-Architected AWS .

strategia brownfield

L'infrastruttura esistente nell'ambiente. Quando si adotta una strategia brownfield per un'architettura di sistema, si progetta l'architettura in base ai vincoli dei sistemi e dell'infrastruttura attuali. Per l'espansione dell'infrastruttura esistente, è possibile combinare strategie brownfield e [greenfield](#).

cache del buffer

L'area di memoria in cui sono archiviati i dati a cui si accede con maggiore frequenza.

capacità di business

Azioni intraprese da un'azienda per generare valore (ad esempio vendite, assistenza clienti o marketing). Le architetture dei microservizi e le decisioni di sviluppo possono essere guidate dalle capacità aziendali. Per ulteriori informazioni, consulta la sezione [Organizzazione in base alle funzionalità aziendali](#) del whitepaper [Esecuzione di microservizi containerizzati su AWS](#).

pianificazione della continuità operativa (BCP)

Un piano che affronta il potenziale impatto di un evento che comporta l'interruzione dell'attività, come una migrazione su larga scala, sulle operazioni e consente a un'azienda di riprendere rapidamente le operazioni.

C

CAF

Vedi [AWS Cloud Adoption Framework](#).

implementazione canaria

Il rilascio lento e incrementale di una versione agli utenti finali. Quando sei sicuro, distribuisci la nuova versione e sostituisci la versione corrente nella sua interezza.

CCoE

Vedi [Cloud Center of Excellence](#).

CDC

Vedi [Change Data Capture](#).

Change Data Capture (CDC)

Il processo di tracciamento delle modifiche a un'origine dati, ad esempio una tabella di database, e di registrazione dei metadati relativi alla modifica. È possibile utilizzare CDC per vari scopi, ad esempio il controllo o la replica delle modifiche in un sistema di destinazione per mantenere la sincronizzazione.

ingegneria del caos

Introduzione intenzionale di guasti o eventi dirompenti per testare la resilienza di un sistema. Puoi usare [AWS Fault Injection Service \(AWS FIS\)](#) per eseguire esperimenti che stressano i tuoi AWS carichi di lavoro e valutarne la risposta.

CI/CD

Vedi [integrazione continua e distribuzione continua](#).

classificazione

Un processo di categorizzazione che aiuta a generare previsioni. I modelli di ML per problemi di classificazione prevedono un valore discreto. I valori discreti sono sempre distinti l'uno dall'altro. Ad esempio, un modello potrebbe dover valutare se in un'immagine è presente o meno un'auto.

crittografia lato client

Crittografia dei dati a livello locale, prima che il destinatario li Servizio AWS riceva.

Centro di eccellenza cloud (CCoE)

Un team multidisciplinare che guida le iniziative di adozione del cloud in tutta l'organizzazione, tra cui lo sviluppo di best practice per il cloud, la mobilitazione delle risorse, la definizione delle tempistiche di migrazione e la guida dell'organizzazione attraverso trasformazioni su larga scala. Per ulteriori informazioni, consulta gli [CCoE post](#) sull' Cloud AWS Enterprise Strategy Blog.

cloud computing

La tecnologia cloud generalmente utilizzata per l'archiviazione remota di dati e la gestione dei dispositivi IoT. Il cloud computing è generalmente collegato alla tecnologia di [edge computing](#).

modello operativo cloud

In un'organizzazione IT, il modello operativo utilizzato per creare, maturare e ottimizzare uno o più ambienti cloud. Per ulteriori informazioni, consulta [Building your Cloud Operating Model](#).

fasi di adozione del cloud

Le quattro fasi che le organizzazioni in genere attraversano quando migrano verso Cloud AWS:

- Progetto: esecuzione di alcuni progetti relativi al cloud per scopi di dimostrazione e apprendimento
- Fondamento: effettuare investimenti fondamentali per scalare l'adozione del cloud (ad esempio, creazione di una landing zone, definizione di una CCo E, definizione di un modello operativo)
- Migrazione: migrazione di singole applicazioni
- Reinvenzione: ottimizzazione di prodotti e servizi e innovazione nel cloud

Queste fasi sono state definite da Stephen Orban nel post sul blog [The Journey Toward Cloud-First & the Stages of Adoption on the Enterprise Strategy](#). Cloud AWS [Per informazioni su come si relazionano alla strategia di AWS migrazione, consulta la guida alla preparazione alla migrazione](#).

CMDB

Vedi [database di gestione della configurazione](#).

repository di codice

Una posizione in cui il codice di origine e altri asset, come documentazione, esempi e script, vengono archiviati e aggiornati attraverso processi di controllo delle versioni. Gli archivi cloud più comuni includono GitHub o Bitbucket Cloud. Ogni versione del codice è denominata ramo. In una struttura a microservizi, ogni repository è dedicato a una singola funzionalità. Una singola pipeline CI/CD può utilizzare più repository.

cache fredda

Una cache del buffer vuota, non ben popolata o contenente dati obsoleti o irrilevanti. Ciò influisce sulle prestazioni perché l'istanza di database deve leggere dalla memoria o dal disco principale, il che richiede più tempo rispetto alla lettura dalla cache del buffer.

dati freddi

Dati a cui si accede raramente e che in genere sono storici. Quando si eseguono interrogazioni di questo tipo di dati, le interrogazioni lente sono in genere accettabili. Lo spostamento di questi dati su livelli o classi di storage meno costosi e con prestazioni inferiori può ridurre i costi.

visione artificiale (CV)

Un campo dell'[intelligenza artificiale](#) che utilizza l'apprendimento automatico per analizzare ed estrarre informazioni da formati visivi come immagini e video digitali. Ad esempio, Amazon SageMaker AI fornisce algoritmi di elaborazione delle immagini per CV.

deriva della configurazione

Per un carico di lavoro, una modifica della configurazione rispetto allo stato previsto. Potrebbe causare la non conformità del carico di lavoro e in genere è graduale e involontaria.

database di gestione della configurazione (CMDB)

Un repository che archivia e gestisce le informazioni su un database e il relativo ambiente IT, inclusi i componenti hardware e software e le relative configurazioni. In genere si utilizzano i dati di un CMDB nella fase di individuazione e analisi del portafoglio della migrazione.

Pacchetto di conformità

Una raccolta di AWS Config regole e azioni correttive che puoi assemblare per personalizzare i controlli di conformità e sicurezza. È possibile distribuire un pacchetto di conformità come singola entità in una regione Account AWS and o all'interno di un'organizzazione utilizzando un modello YAML. Per ulteriori informazioni, consulta i [Conformance](#) Pack nella documentazione. AWS Config

integrazione e distribuzione continua (continuous integration and continuous delivery, CI/CD)

Il processo di automazione delle fasi di origine, compilazione, test, gestione temporanea e produzione del processo di rilascio del software. CI/CD viene comunemente descritto come una pipeline. CI/CD può aiutarvi ad automatizzare i processi, migliorare la produttività, migliorare la qualità del codice e velocizzare le consegne. Per ulteriori informazioni, consulta [Vantaggi](#)

della distribuzione continua. CD può anche significare continuous deployment (implementazione continua). Per ulteriori informazioni, consulta [Distribuzione continua e implementazione continua a confronto](#).

CV

Vedi [visione artificiale](#).

D

dati a riposo

Dati stazionari nella rete, ad esempio i dati archiviati.

classificazione dei dati

Un processo per identificare e classificare i dati nella rete in base alla loro criticità e sensibilità. È un componente fondamentale di qualsiasi strategia di gestione dei rischi di sicurezza informatica perché consente di determinare i controlli di protezione e conservazione appropriati per i dati. La classificazione dei dati è un componente del pilastro della sicurezza nel AWS Well-Architected Framework. Per ulteriori informazioni, consulta [Classificazione dei dati](#).

deriva dei dati

Una variazione significativa tra i dati di produzione e i dati utilizzati per addestrare un modello di machine learning o una modifica significativa dei dati di input nel tempo. La deriva dei dati può ridurre la qualità, l'accuratezza e l'equità complessive nelle previsioni dei modelli ML.

dati in transito

Dati che si spostano attivamente attraverso la rete, ad esempio tra le risorse di rete.

rete di dati

Un framework architettonico che fornisce la proprietà distribuita e decentralizzata dei dati con gestione e governance centralizzate.

riduzione al minimo dei dati

Il principio della raccolta e del trattamento dei soli dati strettamente necessari. Praticare la riduzione al minimo dei dati in the Cloud AWS può ridurre i rischi per la privacy, i costi e l'impronta di carbonio delle analisi.

perimetro dei dati

Una serie di barriere preventive nell' AWS ambiente che aiutano a garantire che solo le identità attendibili accedano alle risorse attendibili delle reti previste. Per ulteriori informazioni, consulta [Building a data perimeter](#) on. AWS

pre-elaborazione dei dati

Trasformare i dati grezzi in un formato che possa essere facilmente analizzato dal modello di ML. La pre-elaborazione dei dati può comportare la rimozione di determinate colonne o righe e l'eliminazione di valori mancanti, incoerenti o duplicati.

provenienza dei dati

Il processo di tracciamento dell'origine e della cronologia dei dati durante il loro ciclo di vita, ad esempio il modo in cui i dati sono stati generati, trasmessi e archiviati.

soggetto dei dati

Un individuo i cui dati vengono raccolti ed elaborati.

data warehouse

Un sistema di gestione dei dati che supporta la business intelligence, come l'analisi. I data warehouse contengono in genere grandi quantità di dati storici e vengono generalmente utilizzati per interrogazioni e analisi.

linguaggio di definizione del database (DDL)

Istruzioni o comandi per creare o modificare la struttura di tabelle e oggetti in un database.

linguaggio di manipolazione del database (DML)

Istruzioni o comandi per modificare (inserire, aggiornare ed eliminare) informazioni in un database.

DDL

Vedi linguaggio di [definizione del database](#).

deep ensemble

Combinare più modelli di deep learning per la previsione. È possibile utilizzare i deep ensemble per ottenere una previsione più accurata o per stimare l'incertezza nelle previsioni.

deep learning

Un sottocampo del ML che utilizza più livelli di reti neurali artificiali per identificare la mappatura tra i dati di input e le variabili target di interesse.

defense-in-depth

Un approccio alla sicurezza delle informazioni in cui una serie di meccanismi e controlli di sicurezza sono accuratamente stratificati su una rete di computer per proteggere la riservatezza, l'integrità e la disponibilità della rete e dei dati al suo interno. Quando si adotta questa strategia AWS, si aggiungono più controlli a diversi livelli della AWS Organizations struttura per proteggere le risorse. Ad esempio, un defense-in-depth approccio potrebbe combinare l'autenticazione a più fattori, la segmentazione della rete e la crittografia.

amministratore delegato

In AWS Organizations, un servizio compatibile può registrare un account AWS membro per amministrare gli account dell'organizzazione e gestire le autorizzazioni per quel servizio. Questo account è denominato amministratore delegato per quel servizio specifico. Per ulteriori informazioni e un elenco di servizi compatibili, consulta [Servizi che funzionano con AWS Organizations](#) nella documentazione di AWS Organizations .

implementazione

Il processo di creazione di un'applicazione, di nuove funzionalità o di correzioni di codice disponibili nell'ambiente di destinazione. L'implementazione prevede l'applicazione di modifiche in una base di codice, seguita dalla creazione e dall'esecuzione di tale base di codice negli ambienti applicativi.

Ambiente di sviluppo

[Vedi ambiente.](#)

controllo di rilevamento

Un controllo di sicurezza progettato per rilevare, registrare e avvisare dopo che si è verificato un evento. Questi controlli rappresentano una seconda linea di difesa e avvisano l'utente in caso di eventi di sicurezza che aggirano i controlli preventivi in vigore. Per ulteriori informazioni, consulta [Controlli di rilevamento](#) in Implementazione dei controlli di sicurezza in AWS.

mappatura del flusso di valore dello sviluppo (DVSM)

Un processo utilizzato per identificare e dare priorità ai vincoli che influiscono negativamente sulla velocità e sulla qualità nel ciclo di vita dello sviluppo del software. DVSM estende il processo di

mappatura del flusso di valore originariamente progettato per pratiche di produzione snella. Si concentra sulle fasi e sui team necessari per creare e trasferire valore attraverso il processo di sviluppo del software.

gemello digitale

Una rappresentazione virtuale di un sistema reale, ad esempio un edificio, una fabbrica, un'attrezzatura industriale o una linea di produzione. I gemelli digitali supportano la manutenzione predittiva, il monitoraggio remoto e l'ottimizzazione della produzione.

tabella delle dimensioni

In uno [schema a stella](#), una tabella più piccola che contiene gli attributi dei dati quantitativi in una tabella dei fatti. Gli attributi della tabella delle dimensioni sono in genere campi di testo o numeri discreti che si comportano come testo. Questi attributi vengono comunemente utilizzati per il vincolo delle query, il filtraggio e l'etichettatura dei set di risultati.

disastro

Un evento che impedisce a un carico di lavoro o a un sistema di raggiungere gli obiettivi aziendali nella sua sede principale di implementazione. Questi eventi possono essere disastri naturali, guasti tecnici o il risultato di azioni umane, come errori di configurazione involontari o attacchi di malware.

disaster recovery (DR)

La strategia e il processo utilizzati per ridurre al minimo i tempi di inattività e la perdita di dati causati da un [disastro](#). Per ulteriori informazioni, consulta [Disaster Recovery of Workloads su AWS: Recovery in the Cloud in the AWS Well-Architected Framework](#).

DML

Vedi linguaggio di manipolazione [del database](#).

progettazione basata sul dominio

Un approccio allo sviluppo di un sistema software complesso collegandone i componenti a domini in evoluzione, o obiettivi aziendali principali, perseguiti da ciascun componente. Questo concetto è stato introdotto da Eric Evans nel suo libro, *Domain-Driven Design: Tackling Complexity in the Heart of Software* (Boston: Addison-Wesley Professional, 2003). Per informazioni su come utilizzare la progettazione basata sul dominio con il modello del fico strangolatore (Strangler Fig), consulta la sezione [Modernizzazione incrementale dei servizi Web Microsoft ASP.NET \(ASMX\) legacy utilizzando container e il Gateway Amazon API](#).

DOTT.

Vedi [disaster recovery](#).

rilevamento della deriva

Tracciamento delle deviazioni da una configurazione di base. Ad esempio, è possibile AWS CloudFormation utilizzarlo per [rilevare deviazioni nelle risorse di sistema](#) oppure AWS Control Tower per [rilevare cambiamenti nella landing zone](#) che potrebbero influire sulla conformità ai requisiti di governance.

DVSM

Vedi la [mappatura del flusso di valore dello sviluppo](#).

E

EDA

Vedi [analisi esplorativa dei dati](#).

MODIFICA

Vedi [scambio elettronico di dati](#).

edge computing

La tecnologia che aumenta la potenza di calcolo per i dispositivi intelligenti all'edge di una rete IoT. Rispetto al [cloud computing](#), l'edge computing può ridurre la latenza di comunicazione e migliorare i tempi di risposta.

scambio elettronico di dati (EDI)

Lo scambio automatizzato di documenti aziendali tra organizzazioni. Per ulteriori informazioni, vedere [Cos'è lo scambio elettronico di dati](#).

crittografia

Un processo di elaborazione che trasforma i dati in chiaro, leggibili dall'uomo, in testo cifrato.

chiave crittografica

Una stringa crittografica di bit randomizzati generata da un algoritmo di crittografia. Le chiavi possono variare di lunghezza e ogni chiave è progettata per essere imprevedibile e univoca.

endianità

L'ordine in cui i byte vengono archiviati nella memoria del computer. I sistemi big-endian memorizzano per primo il byte più importante. I sistemi little-endian memorizzano per primo il byte meno importante.

endpoint

Vedi service endpoint.

servizio endpoint

Un servizio che puoi ospitare in un cloud privato virtuale (VPC) da condividere con altri utenti. Puoi creare un servizio endpoint con AWS PrivateLink e concedere autorizzazioni ad altri Account AWS o a AWS Identity and Access Management (IAM) principali. Questi account o principali possono connettersi al servizio endpoint in privato creando endpoint VPC di interfaccia. Per ulteriori informazioni, consulta [Creazione di un servizio endpoint](#) nella documentazione di Amazon Virtual Private Cloud (Amazon VPC).

pianificazione delle risorse aziendali (ERP)

Un sistema che automatizza e gestisce i processi aziendali chiave (come contabilità, [MES](#) e gestione dei progetti) per un'azienda.

crittografia envelope

Il processo di crittografia di una chiave di crittografia con un'altra chiave di crittografia. Per ulteriori informazioni, vedete [Envelope encryption](#) nella documentazione AWS Key Management Service (AWS KMS).

ambiente

Un'istanza di un'applicazione in esecuzione. Di seguito sono riportati i tipi di ambiente più comuni nel cloud computing:

- ambiente di sviluppo: un'istanza di un'applicazione in esecuzione disponibile solo per il team principale responsabile della manutenzione dell'applicazione. Gli ambienti di sviluppo vengono utilizzati per testare le modifiche prima di promuoverle negli ambienti superiori. Questo tipo di ambiente viene talvolta definito ambiente di test.
- ambienti inferiori: tutti gli ambienti di sviluppo di un'applicazione, ad esempio quelli utilizzati per le build e i test iniziali.

- ambiente di produzione: un'istanza di un'applicazione in esecuzione a cui gli utenti finali possono accedere. In una CI/CD pipeline, l'ambiente di produzione è l'ultimo ambiente di distribuzione.
- ambienti superiori: tutti gli ambienti a cui possono accedere utenti diversi dal team di sviluppo principale. Si può trattare di un ambiente di produzione, ambienti di preproduzione e ambienti per i test di accettazione da parte degli utenti.

epica

Nelle metodologie agili, categorie funzionali che aiutano a organizzare e dare priorità al lavoro. Le epiche forniscono una descrizione di alto livello dei requisiti e delle attività di implementazione. Ad esempio, le epopee della sicurezza AWS CAF includono la gestione delle identità e degli accessi, i controlli investigativi, la sicurezza dell'infrastruttura, la protezione dei dati e la risposta agli incidenti. Per ulteriori informazioni sulle epiche, consulta la strategia di migrazione AWS , consulta la [guida all'implementazione del programma](#).

ERP

Vedi [pianificazione delle risorse aziendali](#).

analisi esplorativa dei dati (EDA)

Il processo di analisi di un set di dati per comprenderne le caratteristiche principali. Si raccolgono o si aggregano dati e quindi si eseguono indagini iniziali per trovare modelli, rilevare anomalie e verificare ipotesi. L'EDA viene eseguita calcolando statistiche di riepilogo e creando visualizzazioni di dati.

F

tabella dei fatti

Il tavolo centrale con [schema a stella](#). Memorizza dati quantitativi sulle operazioni aziendali. In genere, una tabella dei fatti contiene due tipi di colonne: quelle che contengono misure e quelle che contengono una chiave esterna per una tabella di dimensioni.

fallire velocemente

Una filosofia che utilizza test frequenti e incrementali per ridurre il ciclo di vita dello sviluppo. È una parte fondamentale di un approccio agile.

limite di isolamento dei guasti

Nel Cloud AWS, un limite come una zona di disponibilità Regione AWS, un piano di controllo o un piano dati che limita l'effetto di un errore e aiuta a migliorare la resilienza dei carichi di lavoro. Per ulteriori informazioni, consulta [AWS Fault Isolation Boundaries](#).

ramo di funzionalità

Vedi [filiale](#).

caratteristiche

I dati di input che usi per fare una previsione. Ad esempio, in un contesto di produzione, le caratteristiche potrebbero essere immagini acquisite periodicamente dalla linea di produzione.

importanza delle caratteristiche

Quanto è importante una caratteristica per le previsioni di un modello. Di solito viene espresso come punteggio numerico che può essere calcolato con varie tecniche, come Shapley Additive Explanations (SHAP) e gradienti integrati. Per ulteriori informazioni, consulta [Interpretabilità del modello di machine learning con AWS](#).

trasformazione delle funzionalità

Per ottimizzare i dati per il processo di machine learning, incluso l'arricchimento dei dati con fonti aggiuntive, il dimensionamento dei valori o l'estrazione di più set di informazioni da un singolo campo di dati. Ciò consente al modello di ML di trarre vantaggio dai dati. Ad esempio, se suddividi la data "2021-05-27 00:15:37" in "2021", "maggio", "giovedì" e "15", puoi aiutare l'algoritmo di apprendimento ad apprendere modelli sfumati associati a diversi componenti dei dati.

prompt con pochi scatti

Fornire a un [LLM](#) un numero limitato di esempi che dimostrino l'attività e il risultato desiderato prima di chiedergli di eseguire un'attività simile. Questa tecnica è un'applicazione dell'apprendimento contestuale, in cui i modelli imparano da esempi (immagini) incorporati nei prompt. I prompt con pochi passaggi possono essere efficaci per attività che richiedono una formattazione, un ragionamento o una conoscenza del dominio specifici. [Vedi anche zero-shot prompting](#).

FGAC

Vedi il controllo [granulare degli accessi](#).

controllo granulare degli accessi (FGAC)

L'uso di più condizioni per consentire o rifiutare una richiesta di accesso.

migrazione flash-cut

Un metodo di migrazione del database che utilizza la replica continua dei dati tramite [l'acquisizione dei dati delle modifiche](#) per migrare i dati nel più breve tempo possibile, anziché utilizzare un approccio graduale. L'obiettivo è ridurre al minimo i tempi di inattività.

FM

[Vedi modello di base.](#)

modello di fondazione (FM)

Una grande rete neurale di deep learning che si è addestrata su enormi set di dati generalizzati e non etichettati. FMs sono in grado di svolgere un'ampia varietà di attività generali, come comprendere il linguaggio, generare testo e immagini e conversare in linguaggio naturale. Per ulteriori informazioni, consulta [Cosa sono i modelli Foundation](#).

G

AI generativa

Un sottoinsieme di modelli di [intelligenza artificiale](#) che sono stati addestrati su grandi quantità di dati e che possono utilizzare un semplice prompt di testo per creare nuovi contenuti e artefatti, come immagini, video, testo e audio. Per ulteriori informazioni, consulta [Cos'è l'IA generativa](#).

blocco geografico

[Vedi restrizioni geografiche.](#)

limitazioni geografiche (blocco geografico)

In Amazon CloudFront, un'opzione per impedire agli utenti di determinati paesi di accedere alle distribuzioni di contenuti. Puoi utilizzare un elenco consentito o un elenco di blocco per specificare i paesi approvati e vietati. Per ulteriori informazioni, consulta [Limitare la distribuzione geografica dei contenuti](#) nella CloudFront documentazione.

Flusso di lavoro di GitFlow

Un approccio in cui gli ambienti inferiori e superiori utilizzano rami diversi in un repository di codice di origine. Il flusso di lavoro Gitflow è considerato obsoleto e il flusso di lavoro [basato su trunk è l'approccio moderno e preferito.](#)

immagine dorata

Un'istantanea di un sistema o di un software che viene utilizzata come modello per distribuire nuove istanze di quel sistema o software. Ad esempio, nella produzione, un'immagine dorata può essere utilizzata per fornire software su più dispositivi e contribuire a migliorare la velocità, la scalabilità e la produttività nelle operazioni di produzione dei dispositivi.

strategia greenfield

L'assenza di infrastrutture esistenti in un nuovo ambiente. Quando si adotta una strategia greenfield per un'architettura di sistema, è possibile selezionare tutte le nuove tecnologie senza il vincolo della compatibilità con l'infrastruttura esistente, nota anche come [brownfield](#). Per l'espansione dell'infrastruttura esistente, è possibile combinare strategie brownfield e greenfield.

guardrail

Una regola di alto livello che aiuta a governare le risorse, le politiche e la conformità tra le unità organizzative (). OUs I guardrail preventivi applicano le policy per garantire l'allineamento agli standard di conformità. Vengono implementati utilizzando le policy di controllo dei servizi e i limiti delle autorizzazioni IAM. I guardrail di rilevamento rilevano le violazioni delle policy e i problemi di conformità e generano avvisi per porvi rimedio. Sono implementati utilizzando Amazon AWS Config AWS Security Hub CSPM GuardDuty AWS Trusted Advisor, Amazon Inspector e controlli personalizzati AWS Lambda .

H

AH

Vedi [disponibilità elevata](#).

migrazione di database eterogenea

Migrazione del database di origine in un database di destinazione che utilizza un motore di database diverso (ad esempio, da Oracle ad Amazon Aurora). La migrazione eterogenea fa in

genere parte di uno sforzo di riprogettazione e la conversione dello schema può essere un'attività complessa. [AWS offre AWS SCT](#) che aiuta con le conversioni dello schema.

alta disponibilità (HA)

La capacità di un carico di lavoro di funzionare in modo continuo, senza intervento, in caso di sfide o disastri. I sistemi HA sono progettati per il failover automatico, fornire costantemente prestazioni di alta qualità e gestire carichi e guasti diversi con un impatto minimo sulle prestazioni.

modernizzazione storica

Un approccio utilizzato per modernizzare e aggiornare i sistemi di tecnologia operativa (OT) per soddisfare meglio le esigenze dell'industria manifatturiera. Uno storico è un tipo di database utilizzato per raccogliere e archiviare dati da varie fonti in una fabbrica.

dati di esclusione

[Una parte di dati storici etichettati che viene trattenuta da un set di dati utilizzata per addestrare un modello di apprendimento automatico.](#) È possibile utilizzare i dati di holdout per valutare le prestazioni del modello confrontando le previsioni del modello con i dati di holdout.

migrazione di database omogenea

Migrazione del database di origine in un database di destinazione che condivide lo stesso motore di database (ad esempio, da Microsoft SQL Server ad Amazon RDS per SQL Server). La migrazione omogenea fa in genere parte di un'operazione di rehosting o ridefinizione della piattaforma. Per migrare lo schema è possibile utilizzare le utilità native del database.

dati caldi

Dati a cui si accede frequentemente, come dati in tempo reale o dati di traduzione recenti. Questi dati richiedono in genere un livello o una classe di storage ad alte prestazioni per fornire risposte rapide alle query.

hotfix

Una soluzione urgente per un problema critico in un ambiente di produzione. A causa della sua urgenza, un hotfix viene in genere creato al di fuori del tipico DevOps flusso di lavoro di rilascio.

periodo di hypercare

Subito dopo la conversione, il periodo di tempo in cui un team di migrazione gestisce e monitora le applicazioni migrate nel cloud per risolvere eventuali problemi. In genere, questo periodo dura

da 1 a 4 giorni. Al termine del periodo di hypercare, il team addetto alla migrazione in genere trasferisce la responsabilità delle applicazioni al team addetto alle operazioni cloud.

|

IaC

Considera [l'infrastruttura come codice](#).

Policy basata su identità

Una policy associata a uno o più principi IAM che definisce le relative autorizzazioni all'interno dell'Cloud AWS ambiente.

applicazione inattiva

Un'applicazione che prevede un uso di CPU e memoria medio compreso tra il 5% e il 20% in un periodo di 90 giorni. In un progetto di migrazione, è normale ritirare queste applicazioni o mantenerle on-premise.

IIoT

Vedi [Industrial Internet of Things](#).

infrastruttura immutabile

Un modello che implementa una nuova infrastruttura per i carichi di lavoro di produzione anziché aggiornare, applicare patch o modificare l'infrastruttura esistente. [Le infrastrutture immutabili sono intrinsecamente più coerenti, affidabili e prevedibili delle infrastrutture mutabili](#). Per ulteriori informazioni, consulta la best practice [Deploy using immutable infrastructure in Well-Architected AWS Framework](#).

VPC in ingresso (ingress)

In un'architettura AWS multi-account, un VPC che accetta, ispeziona e indirizza le connessioni di rete dall'esterno di un'applicazione. La [AWS Security Reference Architecture](#) consiglia di configurare l'account di rete con funzionalità in entrata, in uscita e di ispezione VPCs per proteggere l'interfaccia bidirezionale tra l'applicazione e la rete Internet in generale.

migrazione incrementale

Una strategia di conversione in cui si esegue la migrazione dell'applicazione in piccole parti anziché eseguire una conversione singola e completa. Ad esempio, inizialmente potresti spostare

solo alcuni microservizi o utenti nel nuovo sistema. Dopo aver verificato che tutto funzioni correttamente, puoi spostare in modo incrementale microservizi o utenti aggiuntivi fino alla disattivazione del sistema legacy. Questa strategia riduce i rischi associati alle migrazioni di grandi dimensioni.

Industria 4.0

Un termine introdotto da [Klaus Schwab](#) nel 2016 per riferirsi alla modernizzazione dei processi di produzione attraverso progressi in termini di connettività, dati in tempo reale, automazione, analisi e AI/ML.

infrastruttura

Tutte le risorse e gli asset contenuti nell'ambiente di un'applicazione.

infrastruttura come codice (IaC)

Il processo di provisioning e gestione dell'infrastruttura di un'applicazione tramite un insieme di file di configurazione. Il processo IaC è progettato per aiutarti a centralizzare la gestione dell'infrastruttura, a standardizzare le risorse e a dimensionare rapidamente, in modo che i nuovi ambienti siano ripetibili, affidabili e coerenti.

IIoInternet delle cose industriale (T)

L'uso di sensori e dispositivi connessi a Internet nei settori industriali, come quello manifatturiero, energetico, automobilistico, sanitario, delle scienze della vita e dell'agricoltura. Per ulteriori informazioni, vedere [Creazione di una strategia di trasformazione digitale per l'Internet of Things \(IIoT\) industriale.](#)

VPC di ispezione

In un'architettura AWS multi-account, un VPC centralizzato che gestisce le ispezioni del traffico di rete tra VPCs (nello stesso o in modo diverso Regioni AWS), Internet e le reti locali. La [AWS Security Reference Architecture](#) consiglia di configurare l'account di rete con informazioni in entrata, in uscita e di ispezione VPCs per proteggere l'interfaccia bidirezionale tra l'applicazione e Internet in generale.

Internet of Things (IoT)

La rete di oggetti fisici connessi con sensori o processori incorporati che comunicano con altri dispositivi e sistemi tramite Internet o una rete di comunicazione locale. Per ulteriori informazioni, consulta [Cos'è l'IoT?](#)

interpretabilità

Una caratteristica di un modello di machine learning che descrive il grado in cui un essere umano è in grado di comprendere in che modo le previsioni del modello dipendono dai suoi input. Per ulteriori informazioni, vedere Interpretabilità del modello di [machine learning](#) con AWS.

IoT

Vedi [Internet of Things](#).

libreria di informazioni IT (ITIL)

Una serie di best practice per offrire servizi IT e allinearli ai requisiti aziendali. ITIL fornisce le basi per ITSM.

gestione dei servizi IT (ITSM)

Attività associate alla progettazione, implementazione, gestione e supporto dei servizi IT per un'organizzazione. Per informazioni sull'integrazione delle operazioni cloud con gli strumenti ITSM, consulta la [guida all'integrazione delle operazioni](#).

ITIL

Vedi la [libreria di informazioni IT](#).

ITSM

Vedi [Gestione dei servizi IT](#).

L

controllo degli accessi basato su etichette (LBAC)

Un'implementazione del controllo di accesso obbligatorio (MAC) in cui agli utenti e ai dati stessi viene assegnato esplicitamente un valore di etichetta di sicurezza. L'intersezione tra l'etichetta di sicurezza utente e l'etichetta di sicurezza dei dati determina quali righe e colonne possono essere visualizzate dall'utente.

zona di destinazione

Una landing zone è un AWS ambiente multi-account ben progettato, scalabile e sicuro. Questo è un punto di partenza dal quale le organizzazioni possono avviare e distribuire rapidamente carichi di lavoro e applicazioni con fiducia nel loro ambiente di sicurezza e infrastruttura. Per ulteriori

informazioni sulle zone di destinazione, consulta la sezione [Configurazione di un ambiente AWS multi-account sicuro e scalabile](#).

modello linguistico di grandi dimensioni (LLM)

Un modello di [intelligenza artificiale](#) di deep learning preaddestrato su una grande quantità di dati.

Un LLM può svolgere più attività, come rispondere a domande, riepilogare documenti, tradurre testo in altre lingue e completare frasi. [Per ulteriori informazioni, consulta Cosa sono. LLMs](#)

migrazione su larga scala

Una migrazione di 300 o più server.

BIANCO

Vedi controllo degli accessi [basato su etichette](#).

Privilegio minimo

La best practice di sicurezza per la concessione delle autorizzazioni minime richieste per eseguire un'attività. Per ulteriori informazioni, consulta [Applicazione delle autorizzazioni del privilegio minimo](#) nella documentazione di IAM.

eseguire il rehosting (lift and shift)

Vedi [Z](#) R.

sistema little-endian

Un sistema che memorizza per primo il byte meno importante. Vedi anche [endianità](#).

LLM

Vedi [modello linguistico di grandi dimensioni](#).

ambienti inferiori

Vedi [ambiente](#).

M

machine learning (ML)

Un tipo di intelligenza artificiale che utilizza algoritmi e tecniche per il riconoscimento e l'apprendimento di schemi. Il machine learning analizza e apprende dai dati registrati, come i dati

dell'Internet delle cose (IoT), per generare un modello statistico basato su modelli. Per ulteriori informazioni, consulta la sezione [Machine learning](#).

ramo principale

Vedi [filiale](#).

malware

Software progettato per compromettere la sicurezza o la privacy del computer. Il malware potrebbe interrompere i sistemi informatici, divulgare informazioni sensibili o ottenere accessi non autorizzati. Esempi di malware includono virus, worm, ransomware, trojan horse, spyware e keylogger.

servizi gestiti

Servizi AWS per cui AWS gestisce il livello di infrastruttura, il sistema operativo e le piattaforme e si accede agli endpoint per archiviare e recuperare i dati. Amazon Simple Storage Service (Amazon S3) Simple Storage Service (Amazon S3) e Amazon DynamoDB sono esempi di servizi gestiti. Questi sono noti anche come servizi astratti.

sistema di esecuzione della produzione (MES)

Un sistema software per tracciare, monitorare, documentare e controllare i processi di produzione che convertono le materie prime in prodotti finiti in officina.

MAP

Vedi [Migration Acceleration Program](#).

meccanismo

Un processo completo in cui si crea uno strumento, si promuove l'adozione dello strumento e quindi si esaminano i risultati per apportare le modifiche. Un meccanismo è un ciclo che si rafforza e si migliora man mano che funziona. Per ulteriori informazioni, consulta [Creazione di meccanismi nel AWS Well-Architected Framework](#).

account membro

Tutti gli account Account AWS diversi dall'account di gestione che fanno parte di un'organizzazione in AWS Organizations Un account può essere membro di una sola organizzazione alla volta.

MEH

Vedi [sistema di esecuzione della produzione](#).

Message Queuing Telemetry Transport (MQTT)

[Un protocollo di comunicazione machine-to-machine \(M2M\) leggero, basato sul modello di pubblicazione/sottoscrizione, per dispositivi IoT con risorse limitate.](#)

microservizio

Un servizio piccolo e indipendente che comunica tramite canali ben definiti ed è in genere di proprietà di piccoli team autonomi. APIs Ad esempio, un sistema assicurativo potrebbe includere microservizi che si riferiscono a funzionalità aziendali, come vendite o marketing, o sottodomini, come acquisti, reclami o analisi. I vantaggi dei microservizi includono agilità, dimensionamento flessibile, facilità di implementazione, codice riutilizzabile e resilienza. Per ulteriori informazioni, consulta [Integrazione dei microservizi utilizzando servizi serverless](#). AWS

architettura di microservizi

Un approccio alla creazione di un'applicazione con componenti indipendenti che eseguono ogni processo applicativo come microservizio. Questi microservizi comunicano attraverso un'interfaccia ben definita utilizzando sistemi leggeri. APIs Ogni microservizio in questa architettura può essere aggiornato, distribuito e dimensionato per soddisfare la richiesta di funzioni specifiche di un'applicazione. Per ulteriori informazioni, vedere [Implementazione dei microservizi](#) su AWS

Programma di accelerazione della migrazione (MAP)

Un AWS programma che fornisce consulenza, supporto, formazione e servizi per aiutare le organizzazioni a costruire una solida base operativa per il passaggio al cloud e per contribuire a compensare il costo iniziale delle migrazioni. MAP include una metodologia di migrazione per eseguire le migrazioni precedenti in modo metodico e un set di strumenti per automatizzare e accelerare gli scenari di migrazione comuni.

migrazione su larga scala

Il processo di trasferimento della maggior parte del portfolio di applicazioni sul cloud avviene a ondate, con più applicazioni trasferite a una velocità maggiore in ogni ondata. Questa fase utilizza le migliori pratiche e le lezioni apprese nelle fasi precedenti per implementare una fabbrica di migrazione di team, strumenti e processi per semplificare la migrazione dei carichi di lavoro attraverso l'automazione e la distribuzione agile. Questa è la terza fase della [strategia di migrazione AWS](#).

fabbrica di migrazione

Team interfunzionali che semplificano la migrazione dei carichi di lavoro attraverso approcci automatizzati e agili. I team di Migration Factory in genere includono addetti alle operazioni,

analisti e proprietari aziendali, ingegneri addetti alla migrazione, sviluppatori e DevOps professionisti che lavorano nell'ambito degli sprint. Tra il 20% e il 50% di un portfolio di applicazioni aziendali è costituito da schemi ripetuti che possono essere ottimizzati con un approccio di fabbrica. Per ulteriori informazioni, consulta la [discussione sulle fabbriche di migrazione](#) e la [Guida alla fabbrica di migrazione al cloud](#) in questo set di contenuti.

metadati di migrazione

Le informazioni sull'applicazione e sul server necessarie per completare la migrazione. Ogni modello di migrazione richiede un set diverso di metadati di migrazione. Esempi di metadati di migrazione includono la sottorete, il gruppo di sicurezza e l'account di destinazione. AWS modello di migrazione

Un'attività di migrazione ripetibile che descrive in dettaglio la strategia di migrazione, la destinazione della migrazione e l'applicazione o il servizio di migrazione utilizzati. Esempio: riorganizza la migrazione su Amazon EC2 con AWS Application Migration Service.

Valutazione del portfolio di migrazione (MPA)

Uno strumento online che fornisce informazioni per la convalida del business case per la migrazione a Cloud AWS MPA offre una valutazione dettagliata del portfolio (dimensionamento corretto dei server, prezzi, confronto del TCO, analisi dei costi di migrazione) e pianificazione della migrazione (analisi e raccolta dei dati delle applicazioni, raggruppamento delle applicazioni, prioritizzazione delle migrazioni e pianificazione delle ondate). [Lo strumento MPA](#) (richiede l'accesso) è disponibile gratuitamente per tutti i AWS consulenti e i consulenti dei partner APN.

valutazione della preparazione alla migrazione (MRA)

Il processo di acquisizione di informazioni sullo stato di preparazione al cloud di un'organizzazione, l'identificazione dei punti di forza e di debolezza e la creazione di un piano d'azione per colmare le lacune identificate, utilizzando il CAF. AWS Per ulteriori informazioni, consulta la [guida di preparazione alla migrazione](#). MRA è la prima fase della [strategia di migrazione AWS](#).

strategia di migrazione

L'approccio utilizzato per migrare un carico di lavoro verso Cloud AWS Per ulteriori informazioni, consulta la voce [7 R](#) in questo glossario e consulta [Mobilita la tua organizzazione per accelerare le migrazioni su larga scala](#).

ML

[Vedi machine learning.](#)

modernizzazione

Trasformazione di un'applicazione obsoleta (legacy o monolitica) e della relativa infrastruttura in un sistema agile, elastico e altamente disponibile nel cloud per ridurre i costi, aumentare l'efficienza e sfruttare le innovazioni. Per ulteriori informazioni, vedere [Strategia per la modernizzazione delle applicazioni in Cloud AWS](#)

valutazione della preparazione alla modernizzazione

Una valutazione che aiuta a determinare la preparazione alla modernizzazione delle applicazioni di un'organizzazione, identifica vantaggi, rischi e dipendenze e determina in che misura l'organizzazione può supportare lo stato futuro di tali applicazioni. Il risultato della valutazione è uno schema dell'architettura di destinazione, una tabella di marcia che descrive in dettaglio le fasi di sviluppo e le tappe fondamentali del processo di modernizzazione e un piano d'azione per colmare le lacune identificate. Per ulteriori informazioni, vedere [Valutazione della preparazione alla modernizzazione per le applicazioni in Cloud AWS](#)

applicazioni monolitiche (monoliti)

Applicazioni eseguite come un unico servizio con processi strettamente collegati. Le applicazioni monolitiche presentano diversi inconvenienti. Se una funzionalità dell'applicazione registra un picco di domanda, l'intera architettura deve essere dimensionata. L'aggiunta o il miglioramento delle funzionalità di un'applicazione monolitica diventa inoltre più complessa man mano che la base di codice cresce. Per risolvere questi problemi, puoi utilizzare un'architettura di microservizi. Per ulteriori informazioni, consulta la sezione [Scomposizione dei monoliti in microservizi](#).

MAPPA

Vedi [Migration Portfolio Assessment](#).

MQTT

Vedi [Message Queuing Telemetry Transport](#).

classificazione multiclasse

Un processo che aiuta a generare previsioni per più classi (prevedendo uno o più di due risultati). Ad esempio, un modello di machine learning potrebbe chiedere "Questo prodotto è un libro, un'auto o un telefono?" oppure "Quale categoria di prodotti è più interessante per questo cliente?"

infrastruttura mutable

Un modello che aggiorna e modifica l'infrastruttura esistente per i carichi di lavoro di produzione.

Per migliorare la coerenza, l'affidabilità e la prevedibilità, il AWS Well-Architected Framework consiglia l'uso di un'infrastruttura [immutable](#) come best practice.

O

OAC

Vedi [Origin Access Control](#).

QUERCIA

Vedi [Origin Access Identity](#).

OCM

Vedi [gestione delle modifiche organizzative](#).

migrazione offline

Un metodo di migrazione in cui il carico di lavoro di origine viene eliminato durante il processo di migrazione. Questo metodo prevede tempi di inattività prolungati e viene in genere utilizzato per carichi di lavoro piccoli e non critici.

OI

Vedi [l'integrazione delle operazioni](#).

OLA

Vedi accordo a [livello operativo](#).

migrazione online

Un metodo di migrazione in cui il carico di lavoro di origine viene copiato sul sistema di destinazione senza essere messo offline. Le applicazioni connesse al carico di lavoro possono continuare a funzionare durante la migrazione. Questo metodo comporta tempi di inattività pari a zero o comunque minimi e viene in genere utilizzato per carichi di lavoro di produzione critici.

OPC-UA

Vedi [Open Process Communications - Unified Architecture](#).

Comunicazioni a processo aperto - Architettura unificata (OPC-UA)

Un protocollo di comunicazione machine-to-machine (M2M) per l'automazione industriale.

OPC-UA fornisce uno standard di interoperabilità con schemi di crittografia, autenticazione e autorizzazione dei dati.

accordo a livello operativo (OLA)

Un accordo che chiarisce quali sono gli impegni reciproci tra i gruppi IT funzionali, a supporto di un accordo sul livello di servizio (SLA).

revisione della prontezza operativa (ORR)

Un elenco di domande e best practice associate che aiutano a comprendere, valutare, prevenire o ridurre la portata degli incidenti e dei possibili guasti. Per ulteriori informazioni, vedere [Operational Readiness Reviews \(ORR\)](#) nel Well-Architected AWS Framework.

tecnologia operativa (OT)

Sistemi hardware e software che interagiscono con l'ambiente fisico per controllare le operazioni, le apparecchiature e le infrastrutture industriali. Nella produzione, l'integrazione di sistemi OT e di tecnologia dell'informazione (IT) è un obiettivo chiave per le trasformazioni [dell'Industria 4.0](#).

integrazione delle operazioni (OI)

Il processo di modernizzazione delle operazioni nel cloud, che prevede la pianificazione, l'automazione e l'integrazione della disponibilità. Per ulteriori informazioni, consulta la [guida all'integrazione delle operazioni](#).

trail organizzativo

Un percorso creato da noi AWS CloudTrail che registra tutti gli eventi di un'organizzazione per tutti Account AWS . AWS Organizations Questo percorso viene creato in ogni Account AWS che fa parte dell'organizzazione e tiene traccia dell'attività in ogni account. Per ulteriori informazioni, consulta [Creazione di un percorso per un'organizzazione](#) nella CloudTrail documentazione.

gestione del cambiamento organizzativo (OCM)

Un framework per la gestione di trasformazioni aziendali importanti e che comportano l'interruzione delle attività dal punto di vista delle persone, della cultura e della leadership. OCM aiuta le organizzazioni a prepararsi e passare a nuovi sistemi e strategie accelerando l'adozione del cambiamento, affrontando i problemi di transizione e promuovendo cambiamenti culturali e organizzativi. Nella strategia di AWS migrazione, questo framework si chiama accelerazione delle

persone, a causa della velocità di cambiamento richiesta nei progetti di adozione del cloud. Per ulteriori informazioni, consultare la [Guida OCM](#).

controllo dell'accesso all'origine (OAC)

In CloudFront, un'opzione avanzata per limitare l'accesso per proteggere i contenuti di Amazon Simple Storage Service (Amazon S3). OAC supporta tutti i bucket S3 in generale Regioni AWS, la crittografia lato server con AWS KMS (SSE-KMS) e le richieste dinamiche e dirette al bucket S3.

PUT DELETE

identità di accesso origine (OAI)

Nel CloudFront, un'opzione per limitare l'accesso per proteggere i tuoi contenuti Amazon S3. Quando usi OAI, CloudFront crea un principale con cui Amazon S3 può autenticarsi. I principali autenticati possono accedere ai contenuti in un bucket S3 solo tramite una distribuzione specifica. CloudFront Vedi anche [OAC](#), che fornisce un controllo degli accessi più granulare e avanzato.

ORR

[Vedi la revisione della prontezza operativa.](#)

- NON

[Vedi la tecnologia operativa.](#)

VPC in uscita (egress)

In un'architettura AWS multi-account, un VPC che gestisce le connessioni di rete avviate dall'interno di un'applicazione. La [AWS Security Reference Architecture](#) consiglia di configurare l'account di rete con funzionalità in entrata, in uscita e di ispezione VPCs per proteggere l'interfaccia bidirezionale tra l'applicazione e Internet in generale.

P

limite delle autorizzazioni

Una policy di gestione IAM collegata ai principali IAM per impostare le autorizzazioni massime che l'utente o il ruolo possono avere. Per ulteriori informazioni, consulta [Limiti delle autorizzazioni](#) nella documentazione di IAM.

informazioni di identificazione personale (PII)

Informazioni che, se visualizzate direttamente o abbinate ad altri dati correlati, possono essere utilizzate per dedurre ragionevolmente l'identità di un individuo. Esempi di informazioni personali includono nomi, indirizzi e informazioni di contatto.

Informazioni che consentono l'identificazione personale degli utenti

Visualizza le [informazioni di identificazione personale](#).

playbook

Una serie di passaggi predefiniti che raccolgono il lavoro associato alle migrazioni, come l'erogazione delle funzioni operative principali nel cloud. Un playbook può assumere la forma di script, runbook automatici o un riepilogo dei processi o dei passaggi necessari per gestire un ambiente modernizzato.

PLC

Vedi [controllore logico programmabile](#).

PLM

Vedi la gestione [del ciclo di vita del prodotto](#).

policy

[Un oggetto in grado di definire le autorizzazioni \(vedi politica basata sull'identità\), specificare le condizioni di accesso \(vedi politica basata sulle risorse\) o definire le autorizzazioni massime per tutti gli account di un'organizzazione in \(vedi politica di controllo dei servizi\). AWS Organizations](#)

persistenza poliglotta

Scelta indipendente della tecnologia di archiviazione di dati di un microservizio in base ai modelli di accesso ai dati e ad altri requisiti. Se i microservizi utilizzano la stessa tecnologia di archiviazione di dati, possono incontrare problemi di implementazione o registrare prestazioni scadenti. I microservizi vengono implementati più facilmente e ottengono prestazioni e scalabilità migliori se utilizzano l'archivio dati più adatto alle loro esigenze. Per ulteriori informazioni, consulta la sezione [Abilitazione della persistenza dei dati nei microservizi](#).

valutazione del portfolio

Un processo di scoperta, analisi e definizione delle priorità del portfolio di applicazioni per pianificare la migrazione. Per ulteriori informazioni, consulta la pagina [Valutazione della preparazione alla migrazione](#).

predicate

Una condizione di interrogazione che restituisce o, in genere, si trova in una clausola `true`. `false` WHERE

predicato pushdown

Una tecnica di ottimizzazione delle query del database che filtra i dati della query prima del trasferimento. Ciò riduce la quantità di dati che devono essere recuperati ed elaborati dal database relazionale e migliora le prestazioni delle query.

controllo preventivo

Un controllo di sicurezza progettato per impedire il verificarsi di un evento. Questi controlli sono la prima linea di difesa per impedire accessi non autorizzati o modifiche indesiderate alla rete. Per ulteriori informazioni, consulta [Controlli preventivi](#) in Implementazione dei controlli di sicurezza in AWS.

principale

Un'entità in AWS grado di eseguire azioni e accedere alle risorse. Questa entità è in genere un utente root per un Account AWS ruolo IAM o un utente. Per ulteriori informazioni, consulta Principali in [Termini e concetti dei ruoli](#) nella documentazione di IAM.

privacy fin dalla progettazione

Un approccio di ingegneria dei sistemi che tiene conto della privacy durante l'intero processo di sviluppo.

zone ospitate private

Un contenitore che contiene informazioni su come desideri che Amazon Route 53 risponda alle query DNS per un dominio e i relativi sottodomini all'interno di uno o più VPCs. Per ulteriori informazioni, consulta [Utilizzo delle zone ospitate private](#) nella documentazione di Route 53.

controllo proattivo

Un [controllo di sicurezza](#) progettato per impedire l'implementazione di risorse non conformi. Questi controlli analizzano le risorse prima del loro provisioning. Se la risorsa non è conforme al controllo, non viene fornita. Per ulteriori informazioni, consulta la [guida di riferimento sui controlli](#) nella AWS Control Tower documentazione e consulta Controlli [proattivi in Implementazione dei controlli](#) di sicurezza su AWS.

gestione del ciclo di vita del prodotto (PLM)

La gestione dei dati e dei processi di un prodotto durante l'intero ciclo di vita, dalla progettazione, sviluppo e lancio, attraverso la crescita e la maturità, fino al declino e alla rimozione.

Ambiente di produzione

[Vedi ambiente.](#)

controllore logico programmabile (PLC)

Nella produzione, un computer altamente affidabile e adattabile che monitora le macchine e automatizza i processi di produzione.

concatenamento rapido

Utilizzo dell'output di un prompt [LLM](#) come input per il prompt successivo per generare risposte migliori. Questa tecnica viene utilizzata per suddividere un'attività complessa in sottoattività o per perfezionare o espandere iterativamente una risposta preliminare. Aiuta a migliorare l'accuratezza e la pertinenza delle risposte di un modello e consente risultati più granulari e personalizzati.

pseudonimizzazione

Il processo di sostituzione degli identificatori personali in un set di dati con valori segnaposto. La pseudonimizzazione può aiutare a proteggere la privacy personale. I dati pseudonimizzati sono ancora considerati dati personali.

publish/subscribe (pub/sub)

Un modello che consente comunicazioni asincrone tra microservizi per migliorare la scalabilità e la reattività. Ad esempio, in un [MES](#) basato su microservizi, un microservizio può pubblicare messaggi di eventi su un canale a cui altri microservizi possono abbonarsi. Il sistema può aggiungere nuovi microservizi senza modificare il servizio di pubblicazione.

Q

Piano di query

Una serie di passaggi, come le istruzioni, utilizzati per accedere ai dati in un sistema di database relazionale SQL.

regressione del piano di query

Quando un ottimizzatore del servizio di database sceglie un piano non ottimale rispetto a prima di una determinata modifica all'ambiente di database. Questo può essere causato da modifiche a statistiche, vincoli, impostazioni dell'ambiente, associazioni dei parametri di query e aggiornamenti al motore di database.

R

Matrice RACI

Vedi [responsabile, responsabile, consultato, informato \(RACI\)](#).

STRACCIO

Vedi [Retrieval Augmented Generation](#).

ransomware

Un software dannoso progettato per bloccare l'accesso a un sistema informatico o ai dati fino a quando non viene effettuato un pagamento.

Matrice RASCI

Vedi [responsabile, responsabile, consultato, informato \(RACI\)](#).

RCAC

Vedi controllo dell'[accesso a righe e colonne](#).

replica di lettura

Una copia di un database utilizzata per scopi di sola lettura. È possibile indirizzare le query alla replica di lettura per ridurre il carico sul database principale.

riprogettare

Vedi [7 Rs.](#)

obiettivo del punto di ripristino (RPO)

Il periodo di tempo massimo accettabile dall'ultimo punto di ripristino dei dati. Questo determina ciò che si considera una perdita di dati accettabile tra l'ultimo punto di ripristino e l'interruzione del servizio.

obiettivo del tempo di ripristino (RTO)

Il ritardo massimo accettabile tra l'interruzione del servizio e il ripristino del servizio.

rifattorizzare

Vedi [7 R.](#)

Regione

Una raccolta di AWS risorse in un'area geografica. Ciascuna Regione AWS è isolata e indipendente dalle altre per fornire tolleranza agli errori, stabilità e resilienza. Per ulteriori informazioni, consulta [Specificare cosa può usare Regioni AWS il tuo account](#).

regressione

Una tecnica di ML che prevede un valore numerico. Ad esempio, per risolvere il problema "A che prezzo verrà venduta questa casa?" un modello di ML potrebbe utilizzare un modello di regressione lineare per prevedere il prezzo di vendita di una casa sulla base di dati noti sulla casa (ad esempio, la metratura).

riospitare

Vedi [7 R.](#)

rilascio

In un processo di implementazione, l'atto di promuovere modifiche a un ambiente di produzione.

trasferisco

Vedi [7 Rs.](#)

ripiattaforma

Vedi [7 Rs.](#)

riacquisto

Vedi [7 Rs.](#)

resilienza

La capacità di un'applicazione di resistere o ripristinare le interruzioni. [L'elevata disponibilità e il disaster recovery](#) sono considerazioni comuni quando si pianifica la resilienza in Cloud AWS. Per ulteriori informazioni, vedere [Cloud AWS Resilience](#).

policy basata su risorse

Una policy associata a una risorsa, ad esempio un bucket Amazon S3, un endpoint o una chiave di crittografia. Questo tipo di policy specifica a quali principali è consentito l'accesso, le azioni supportate e qualsiasi altra condizione che deve essere soddisfatta.

matrice di assegnazione di responsabilità (RACI)

Una matrice che definisce i ruoli e le responsabilità di tutte le parti coinvolte nelle attività di migrazione e nelle operazioni cloud. Il nome della matrice deriva dai tipi di responsabilità definiti nella matrice: responsabile (R), responsabile (A), consultato (C) e informato (I). Il tipo di supporto (S) è facoltativo. Se includi il supporto, la matrice viene chiamata matrice RASCI e, se lo escludi, viene chiamata matrice RACI.

controllo reattivo

Un controllo di sicurezza progettato per favorire la correzione di eventi avversi o deviazioni dalla baseline di sicurezza. Per ulteriori informazioni, consulta [Controlli reattivi](#) in Implementazione dei controlli di sicurezza in AWS.

retain

Vedi [7 R.](#)

andare in pensione

Vedi [7 Rs.](#)

Retrieval Augmented Generation (RAG)

Una tecnologia di [intelligenza artificiale generativa](#) in cui un [LLM](#) fa riferimento a una fonte di dati autorevole esterna alle sue fonti di dati di formazione prima di generare una risposta. Ad esempio, un modello RAG potrebbe eseguire una ricerca semantica nella knowledge base o nei dati personalizzati di un'organizzazione. Per ulteriori informazioni, consulta [Cos'è il RAG](#).

rotazione

Processo di aggiornamento periodico di un [segreto](#) per rendere più difficile l'accesso alle credenziali da parte di un utente malintenzionato.

controllo dell'accesso a righe e colonne (RCAC)

L'uso di espressioni SQL di base e flessibili con regole di accesso definite. RCAC è costituito da autorizzazioni di riga e maschere di colonna.

RPO

Vedi l'obiettivo del punto [di ripristino](#).

RTO

Vedi l'[obiettivo del tempo di ripristino](#).

runbook

Un insieme di procedure manuali o automatizzate necessarie per eseguire un'attività specifica. In genere sono progettati per semplificare operazioni o procedure ripetitive con tassi di errore elevati.

S

SAML 2.0

Uno standard aperto utilizzato da molti provider di identità (IdPs). Questa funzionalità abilita il single sign-on (SSO) federato, in modo che gli utenti possano accedere Console di gestione AWS o chiamare le operazioni AWS API senza che tu debba creare un utente in IAM per tutti i membri dell'organizzazione. Per ulteriori informazioni sulla federazione basata su SAML 2.0, consulta [Informazioni sulla federazione basata su SAML 2.0](#) nella documentazione di IAM.

SCADA

Vedi [controllo di supervisione e acquisizione dati](#).

SCP

Vedi la [politica di controllo del servizio](#).

Secret

In Gestione dei segreti AWS, informazioni riservate o riservate, come una password o le credenziali utente, archiviate in forma crittografata. È costituito dal valore segreto e dai relativi metadati. Il valore segreto può essere binario, una stringa singola o più stringhe. Per ulteriori informazioni, consulta [Cosa c'è in un segreto di Secrets Manager?](#) nella documentazione di Secrets Manager.

sicurezza fin dalla progettazione

Un approccio di ingegneria dei sistemi che tiene conto della sicurezza durante l'intero processo di sviluppo.

controllo di sicurezza

Un guardrail tecnico o amministrativo che impedisce, rileva o riduce la capacità di un autore di minacce di sfruttare una vulnerabilità di sicurezza. [Esistono quattro tipi principali di controlli di sicurezza: preventivi, investigativi, reattivi e proattivi.](#)

rafforzamento della sicurezza

Il processo di riduzione della superficie di attacco per renderla più resistente agli attacchi. Può includere azioni come la rimozione di risorse che non sono più necessarie, l'implementazione di best practice di sicurezza che prevedono la concessione del privilegio minimo o la disattivazione di funzionalità non necessarie nei file di configurazione.

sistema di gestione delle informazioni e degli eventi di sicurezza (SIEM)

Strumenti e servizi che combinano sistemi di gestione delle informazioni di sicurezza (SIM) e sistemi di gestione degli eventi di sicurezza (SEM). Un sistema SIEM raccoglie, monitora e analizza i dati da server, reti, dispositivi e altre fonti per rilevare minacce e violazioni della sicurezza e generare avvisi.

automazione della risposta alla sicurezza

Un'azione predefinita e programmata progettata per rispondere o porre rimedio automaticamente a un evento di sicurezza. Queste automazioni fungono da controlli di sicurezza [investigativi](#) o [reattivi](#) che aiutano a implementare le migliori pratiche di sicurezza. AWS Esempi di azioni di risposta automatizzate includono la modifica di un gruppo di sicurezza VPC, l'applicazione di patch a un'istanza EC2 Amazon o la rotazione delle credenziali.

Crittografia lato server

Crittografia dei dati a destinazione, da parte di chi li riceve. Servizio AWS

Policy di controllo dei servizi (SCP)

Una politica che fornisce il controllo centralizzato sulle autorizzazioni per tutti gli account di un'organizzazione in. AWS Organizations SCPs definire barriere o fissare limiti alle azioni che un amministratore può delegare a utenti o ruoli. È possibile utilizzarli SCPs come elenchi consentiti o elenchi di rifiuto, per specificare quali servizi o azioni sono consentiti o proibiti. Per ulteriori informazioni, consulta [le politiche di controllo del servizio](#) nella AWS Organizations documentazione.

endpoint del servizio

L'URL del punto di ingresso per un Servizio AWS. Puoi utilizzare l'endpoint per connetterti a livello di programmazione al servizio di destinazione. Per ulteriori informazioni, consulta [Endpoint del Servizio AWS](#) nei Riferimenti generali di AWS.

accordo sul livello di servizio (SLA)

Un accordo che chiarisce ciò che un team IT promette di offrire ai propri clienti, ad esempio l'operatività e le prestazioni del servizio.

indicatore del livello di servizio (SLI)

Misurazione di un aspetto prestazionale di un servizio, ad esempio il tasso di errore, la disponibilità o la velocità effettiva.

obiettivo a livello di servizio (SLO)

[Una metrica target che rappresenta lo stato di un servizio, misurato da un indicatore del livello di servizio.](#)

Modello di responsabilità condivisa

Un modello che descrive la responsabilità condivisa AWS per la sicurezza e la conformità del cloud. AWS è responsabile della sicurezza del cloud, mentre tu sei responsabile della sicurezza nel cloud. Per ulteriori informazioni, consulta [Modello di responsabilità condivisa](#).

SIEM

Vedi il [sistema di gestione delle informazioni e degli eventi sulla sicurezza](#).

punto di errore singolo (SPOF)

Un guasto in un singolo componente critico di un'applicazione che può disturbare il sistema.

SLAM

Vedi il contratto sul [livello di servizio](#).

SLI

Vedi l'indicatore del [livello di servizio](#).

LENTA

Vedi obiettivo del [livello di servizio](#).

split-and-seed modello

Un modello per dimensionare e accelerare i progetti di modernizzazione. Man mano che vengono definite nuove funzionalità e versioni dei prodotti, il team principale si divide per creare nuovi team di prodotto. Questo aiuta a dimensionare le capacità e i servizi dell'organizzazione, migliora la produttività degli sviluppatori e supporta una rapida innovazione. Per ulteriori informazioni, vedere [Approccio graduale alla modernizzazione delle applicazioni in Cloud AWS](#)

SPOF

Vedi [punto di errore singolo](#).

schema a stella

Una struttura organizzativa di database che utilizza un'unica tabella dei fatti di grandi dimensioni per archiviare i dati transazionali o misurati e utilizza una o più tabelle dimensionali più piccole per memorizzare gli attributi dei dati. Questa struttura è progettata per l'uso in un [data warehouse](#) o per scopi di business intelligence.

modello del fico strangolatore

Un approccio alla modernizzazione dei sistemi monolitici mediante la riscrittura e la sostituzione incrementali delle funzionalità del sistema fino alla disattivazione del sistema legacy. Questo modello utilizza l'analogia di una pianta di fico che cresce fino a diventare un albero robusto e alla fine annienta e sostituisce il suo ospite. Il modello è stato [introdotto da Martin Fowler](#) come metodo per gestire il rischio durante la riscrittura di sistemi monolitici. Per un esempio di come applicare questo modello, consulta [Modernizzazione incrementale dei servizi Web legacy di Microsoft ASP.NET \(ASMX\) mediante container e Gateway Amazon API](#).

sottorete

Un intervallo di indirizzi IP nel VPC. Una sottorete deve risiedere in una singola zona di disponibilità.

controllo di supervisione e acquisizione dati (SCADA)

Nella produzione, un sistema che utilizza hardware e software per monitorare gli asset fisici e le operazioni di produzione.

crittografia simmetrica

Un algoritmo di crittografia che utilizza la stessa chiave per crittografare e decriptografare i dati.

test sintetici

Test di un sistema in modo da simulare le interazioni degli utenti per rilevare potenziali problemi o monitorare le prestazioni. Puoi usare [Amazon CloudWatch Synthetics](#) per creare questi test.

prompt di sistema

Una tecnica per fornire contesto, istruzioni o linee guida a un [LLM](#) per indirizzarne il comportamento. I prompt di sistema aiutano a impostare il contesto e stabilire regole per le interazioni con gli utenti.

T

tags

Copie chiave-valore che fungono da metadati per l'organizzazione delle risorse. AWS Con i tag è possibile a gestire, identificare, organizzare, cercare e filtrare le risorse. Per ulteriori informazioni, consulta [Tagging delle risorse AWS](#).

variabile di destinazione

Il valore che stai cercando di prevedere nel machine learning supervisionato. Questo è indicato anche come variabile di risultato. Ad esempio, in un ambiente di produzione la variabile di destinazione potrebbe essere un difetto del prodotto.

elenco di attività

Uno strumento che viene utilizzato per tenere traccia dei progressi tramite un runbook. Un elenco di attività contiene una panoramica del runbook e un elenco di attività generali da completare. Per ogni attività generale, include la quantità stimata di tempo richiesta, il proprietario e lo stato di avanzamento.

Ambiente di test

[Vedi ambiente.](#)

training

Fornire dati da cui trarre ispirazione dal modello di machine learning. I dati di training devono contenere la risposta corretta. L'algoritmo di apprendimento trova nei dati di addestramento i pattern che mappano gli attributi dei dati di input al target (la risposta che si desidera prevedere). Produce un modello di ML che acquisisce questi modelli. Puoi quindi utilizzare il modello di ML per creare previsioni su nuovi dati di cui non si conosce il target.

Transit Gateway

Un hub di transito di rete che puoi utilizzare per interconnettere le tue reti VPCs e quelle locali.

Per ulteriori informazioni, consulta [Cos'è un gateway di transito](#) nella AWS Transit Gateway documentazione.

flusso di lavoro basato su trunk

Un approccio in cui gli sviluppatori creano e testano le funzionalità localmente in un ramo di funzionalità e quindi uniscono tali modifiche al ramo principale. Il ramo principale viene quindi integrato negli ambienti di sviluppo, preproduzione e produzione, in sequenza.

Accesso attendibile

Concessione delle autorizzazioni a un servizio specificato dall'utente per eseguire attività all'interno dell'organizzazione AWS Organizations e nei suoi account per conto dell'utente.

Il servizio attendibile crea un ruolo collegato al servizio in ogni account, quando tale ruolo è necessario, per eseguire attività di gestione per conto dell'utente. Per ulteriori informazioni, consulta [Utilizzo AWS Organizations con altri AWS servizi](#) nella AWS Organizations documentazione.

regolazione

Modificare alcuni aspetti del processo di training per migliorare la precisione del modello di ML. Ad esempio, puoi addestrare il modello di ML generando un set di etichette, aggiungendo etichette e quindi ripetendo questi passaggi più volte con impostazioni diverse per ottimizzare il modello.

team da due pizze

Una piccola DevOps squadra che puoi sfamare con due pizze. Un team composto da due persone garantisce la migliore opportunità possibile di collaborazione nello sviluppo del software.

U

incertezza

Un concetto che si riferisce a informazioni imprecise, incomplete o sconosciute che possono minare l'affidabilità dei modelli di machine learning predittivi. Esistono due tipi di incertezza: l'incertezza epistemica, che è causata da dati limitati e incompleti, mentre l'incertezza aleatoria è causata dal rumore e dalla casualità insiti nei dati. Per ulteriori informazioni, consulta la guida [Quantificazione dell'incertezza nei sistemi di deep learning](#).

compiti indifferenziati

Conosciuto anche come sollevamento di carichi pesanti, è un lavoro necessario per creare e far funzionare un'applicazione, ma che non apporta valore diretto all'utente finale né offre vantaggi competitivi. Esempi di attività indifferenziate includono l'approvvigionamento, la manutenzione e la pianificazione della capacità.

ambienti superiori

[Vedi ambiente.](#)

V

vacuum

Un'operazione di manutenzione del database che prevede la pulizia dopo aggiornamenti incrementali per recuperare lo spazio di archiviazione e migliorare le prestazioni.

controllo delle versioni

Processi e strumenti che tengono traccia delle modifiche, ad esempio le modifiche al codice di origine in un repository.

Peering VPC

Una connessione tra due VPCs che consente di indirizzare il traffico utilizzando indirizzi IP privati. Per ulteriori informazioni, consulta [Che cos'è il peering VPC?](#) nella documentazione di Amazon VPC.

vulnerabilità

Un difetto software o hardware che compromette la sicurezza del sistema.

W

cache calda

Una cache del buffer che contiene dati correnti e pertinenti a cui si accede frequentemente. L'istanza di database può leggere dalla cache del buffer, il che richiede meno tempo rispetto alla lettura dalla memoria dal disco principale.

dati caldi

Dati a cui si accede raramente. Quando si eseguono interrogazioni di questo tipo di dati, in genere sono accettabili query moderatamente lente.

funzione finestra

Una funzione SQL che esegue un calcolo su un gruppo di righe che si riferiscono in qualche modo al record corrente. Le funzioni della finestra sono utili per l'elaborazione di attività, come il calcolo di una media mobile o l'accesso al valore delle righe in base alla posizione relativa della riga corrente.

Carico di lavoro

Una raccolta di risorse e codice che fornisce valore aziendale, ad esempio un'applicazione rivolta ai clienti o un processo back-end.

flusso di lavoro

Gruppi funzionali in un progetto di migrazione responsabili di una serie specifica di attività. Ogni flusso di lavoro è indipendente ma supporta gli altri flussi di lavoro del progetto. Ad esempio, il flusso di lavoro del portfolio è responsabile della definizione delle priorità delle applicazioni, della pianificazione delle ondate e della raccolta dei metadati di migrazione. Il flusso di lavoro del portfolio fornisce queste risorse al flusso di lavoro di migrazione, che quindi migra i server e le applicazioni.

VERME

Vedi [scrivere una volta, leggere molti](#).

WQF

Vedi [AWS Workload Qualification Framework](#).

scrivi una volta, leggi molte (WORM)

Un modello di storage che scrive i dati una sola volta e ne impedisce l'eliminazione o la modifica. Gli utenti autorizzati possono leggere i dati tutte le volte che è necessario, ma non possono modificarli. Questa infrastruttura di archiviazione dei dati è considerata [immutabile](#).

Z

exploit zero-day

[Un attacco, in genere malware, che sfrutta una vulnerabilità zero-day.](#)

vulnerabilità zero-day

Un difetto o una vulnerabilità assoluta in un sistema di produzione. Gli autori delle minacce possono utilizzare questo tipo di vulnerabilità per attaccare il sistema. Gli sviluppatori vengono spesso a conoscenza della vulnerabilità causata dall'attacco.

prompt zero-shot

Fornire a un [LLM](#) le istruzioni per eseguire un'attività ma non esempi (immagini) che possano aiutarla. Il LLM deve utilizzare le sue conoscenze pre-addestrate per gestire l'attività. L'efficacia del prompt zero-shot dipende dalla complessità dell'attività e dalla qualità del prompt. [Vedi anche few-shot prompting.](#)

applicazione zombie

Un'applicazione che prevede un utilizzo CPU e memoria inferiore al 5%. In un progetto di migrazione, è normale ritirare queste applicazioni.

Le traduzioni sono generate tramite traduzione automatica. In caso di conflitto tra il contenuto di una traduzione e la versione originale in Inglese, quest'ultima prevarrà.