Memorie Di Massa (Gabbox ft Chiavax)

Caratteristiche Generali

è la componente più lenta, voluminosa e variegata dell'elaboratore. Devono accedere facilmente e in modo uniforme. deve ottimizzare I/O, perché tante volte è questo che rallenta.

Il ***File System*** dal punto di vista logico si divide in tre parti:

- L'interfaccia al programmatore
- Le strutture dati interne e gli algoritmi di allocazione
- Il livello più basso è legato alla memorizzazione su memoria secondaria

Il File System si preoccupa di fornire un'interfaccia logica comune a dispositivi di memoria di massa che hanno caratteristiche molto differenti. In particolare ci sono due tipi di memoria di massa:

- Dischi fissi, ovvero dischi magnetici
- Dispositivi a stato solido, SSD o Solid State Drive

Dischi Fissi

Ancora utilizzati nonostante l'arrivo delle SSD, sono costituiti da piatti con diametro di 1.8 e 3.5 pollici rivestiti di materiale magnetico (Ossido di ferro) originariamente in alluminio.
Ora la maggior parte i dischi sono su vetro per diversi motivi

- hanno una superficie più uniforme che comporta ad una maggiore affidabilità
- sono più rigidi e più resistenti
- permette di ridurre la distanza tra testina dalla superficie

I dischi sono composti da un insieme di piatti ed ogni disco è diviso in tracce, a sua volta divise in settori. Infine l'insieme di tracce alla stessa distanza dal centro presenti su tutti i dischi o piatti è detto cilindro.

Tutte le informazioni sono organizzate in tracce e sono disposte sulla superficie.
Le tracce sono numerate da 0 a partire dal bordo del disco verso l'interno.
Il settore, detto l'unità minima di trasferimento, in origine aveva una dimensione di 512 Byte ma dal 2010 è diventata 4 Kb.
Quando L'OS formatta il disco organizza in unità di allocazione dette Cluster.

Tutto assieme cilindro tracce e settori formano la geometria del hard disk.
L'attuatore permette di avere collegate tutte le testine che servono per scrivere o leggere su disco.

Struttura dei dischi

Il disco è un vettore monodimensionale di blocchi logici ed è mappato sequenzialmente sui settori del disco:

- Dalla traccia 0 del cilindro più esterno
- Lungo la traccia
- Verso cilindri interni.

La lettura e scrittura avviene tramite alla testina (head) e la punta permette di fare la scrittura e la lettura:

- **Scrittura**: la corrente tramite la bobina produce un campo magnetico e le particelle aciculari dell'ossido che si orientano in base al campo magnetico e tramite questo processo chimico memorizzano le informazioni (0 e 1).
- **Lettura**: Il campo magnetico muovendosi su disco in base alla testina induce corrente nella bobina e muovendosi interpreta le informazioni.
- **Le fasi della lettura e scrittura**:
    - Posizionamento della testina sulla traccia.
    - Attesa del passaggio del settore di interesse
    - Lettura o scrittura del dato
- Le testine planano su disco per le alte velocità che deve sopportare il disco e si tengono a una distanza di 10 ^ -4 mm

Settore

Il settore rappresenta l'unità più piccola che può essere letta, e per accedervi bisogna specificare **la superficie del disco, la traccia e il settore stesso**. I settori sono raggruppati in cluster in base logica, la dimensione del cluster è 2kb a 32kb a seconda dell'OS, un file occupa almeno un cluster.

Dischi Magnetici

I dischi ruotano ad una velocità tra i 60 e i 250 rps
***Seek time:* il tempo che impiega la testina a posizionarsi su settore.**
***Latency time:* Il tempo che ci mette il settore desiderato a posizionarsi sulla testina e dipende dalla velocità di rotazione**
***Transfer time:* il tempo di lettura e scrittura cioè al settore di passare sotto la testina.**
***Velocità di trasferimento* è la velocità con cui i dati fluiscono dall'unità disco alla RAM**
***Tempo di accesso* == Seek Time+Latency Time+Transfer Time**

Il **crollo della testina** è quando la testina tocca il disco, essi sono separati da pochi e micron e al contatto il disco può danneggiarsi.

Ci possono essere dei blocchi logici che possono essere danneggiati,questi possono caratterizzare di fatto la mappatura dei blocchi stessi, quindi il fatto di riuscire poi ad utilizzare tutte le parti del disco dipende anche da questi.Inoltre restringendosi il disco offre spazio sempre minore. Tutte le info più importanti sono slavate delle aree più esterne. la capacità va da i 500 Gb ai 20 Tb (ora anche fino ai 30 Tb)
La mappatura dei blocchi logici dipende se c'è un disco danneggiato oppure dalla dimensione del disco.

Cose da sapere a memoria perché si:

- velocità di trasferimento (teorica) : 6 Gb/s, l'effettiva velocità è 1Gb/s
- Seek time dai 3 ai 12 msec
- Tempo di latenza è calcolato in base alla velocità di rotazione del disco 1 / (rpm/60) es
- la latenza media è di ½ giro. (esempio)
- tempo di accesso medio: **seek time medio + tempo di latenza media** (più veloci: 3 msec+2 msc), più lenti (9 msc + 5.55 msc = 14.55 msc)
- Tempo medio di i/o = tempo medio di accesso+quantità di dati da trasferire rispetto alla velocità di trasferimento + overhead
- tempo di trasferimento: **quantità dati / velocità di trasferimento**

Dischi a Stato Solido:

Utilizzano una tecnologia elettronica per memorizzare i dati.
Fanno parte di questa categoria le memorie flash-based , DRAM-based ed altre memorie di stato solido.
Non contiene alcuna parte in movimento
Utilizza circuiti elettronici integrati per memorizzare i dati in forma permanente

Gli **SSD** utilizzano un'interfaccia che permette di andare a gestire le info come negli hard disk perché sono compatibili con gli hard disk, quindi sostituiscono gli hard disk con l'SSD

Caratteristiche SSD:

- Sono più resistenti a sollecitazioni fisiche,
- sono silenziosi
- tempo di accesso a latenza notevolmente inferiore
- Costa molto di più di un HDD, ma sta diminuendo rispetto il passato.
- Seek time di 0.1ms

Attualmente la maggior parte degli SSD usa delle memorie con tecnologia MLC NAND flash-based. (non volatile)

Per applicazioni che richiedono più velocità vengono utilizzati anche SSD basati su RAM. Tali dispositivi tengono le info salvate tramite delle batterie tampone che tengono alimentata la RAM per un certo tempo dopo lo spegnimento dell'alimentazione elettrica per mantenere lo stato dei dati.

Gli ***SSD ibridi*** combinano caratteristiche SSD e HDD contenente un HDD ad alta capacità ed una cache SSD per migliorare le performance sui dati di frequente accesso.
Sono costituti da diversi chip di memoria NAND flash.
Altri tipi di SSD includono le unità USB e DRAM dotate di batterie di backup.

Nei telefoni le NVM montate sulla scheda madre, sono il dispositivo primario di archiviazione e sono più affidabili:

- sono più affidabili rispetto agli HDD
- hanno un costo elevato al Mb
- più veloci
- meno durature
- consumo minore di energia

Caratteristiche dei semiconduttori NAND:
La lettura scrittura avviene e per concetto di pagina, quindi un insieme di pagine caratterizzate da blocchi che possono avere dimensioni differenti, dove possono essere scritte o cancellate le varie info
si possono solo, mai effettivamente cancellate, sovrascrivere dati.

Si deteriora a ogni ciclo di cancellazione dopo tot cicli non può essere più utilizzato, quindi durata limitata nel tempo, quindi dopo un po' non si può più usare.
Si misura con DWPD , quanti cicli di lettura e scrittura esegue l'SSD al giorno. Su NAND da 1 Tb di classe 5 DWPD si possono scrivere 5 Tb al giorno senza errori per il periodo di garanzia.

NVM
Gli algoritmi di gestione ottimizzata non sono gestiti dal sistema operativo ma direttamente dall ssd il sistema operativo si limita a leggere e a scrivere.

Memoria flash: le info vengono registrate in un array di Floating GATE MOSFET una tipologia di transistor in grado di mantenere la carica elettrica per un tempo lungo.
Ogni transistor costituisce una cella di memoria che conosceva il valore di un bit.
Utilizzano un multilivello che permette di registrare il valore di più bit in un solo transistor.

Memoria NAND flash non hanno una sovrascrittura ma sono composti da pagine per tenere traccia dei dati ha una tabella della FTKL che mappa i blocchi logici validi.
implementa anche il **garbage collection** che ci serve a prendere spazio per cancellare dati non validi. un controllore SSD prima deve copiare tutti i dati validi nelle pagine vuote in un altro blocco eliminando tutti i dati da cancellare e tutti i dati che successivamente. solo a quel punto può iniziare a riscrivere in quel punto.
Normalmente si mantiene un overprovisioning (7-20% pagine libere) per garantire uno spazio fisico per il lavoro del garbage collection ogni cella ha durata di vita limitata, quindi l'obiettivo è quello di mantenere l'usura del succo in maniera uniforme per garantire un numero giusto di celle che possono sostituire quelle che sono invalide.
La ***garbage collection*** è una modalità di gestione di memoria automatica fatta dal sistema operativo o compilatore che libera spazio inutile.

MEMORIE NAND 3D INTEGRANO IN STRATI MULTIPLI DI CELLE E IMPILATI VERTICALMENTE INTERCONNESSE, OFFRE DI FATTO UNA MAGGIORE CAPACITà DI STORAGE. Le celle NAND non sono progettate per durare in eterno. Dispongono di un numero illimitato di cicli di scrittura tramite le funzioni di livellamento dell'usura gestiti da un controller flash.

Connessioni
Un disco può essere rimovibile quindi collegato o scollegato in base alle necessità ed è collegato tramite un bus di i/o con tecnologia di:

- ide-ata
- sata
- fc
- scsi
- usb

I bus standard sono lenti per gli SSD quindi si utilizzano i NVM e il trasferimento dei dati è eseguito da dei controllori cioè adattatori posti all'estremità del bus per verificare se il dato è corretto.
Le unità disco sono indirizzate come giganteschi vettori monodimensionali di blocchi logici, dove il blocco logico rappresenta la minima unità di trasferimenti. i blocchi logici sono creati all'atto della formattazione di basso livello

Il vettore viene mappato sequenzialmente nei settori del disco o sulle pagine di un blocco di NVM.

il settore 0 è il primo settore della prima traccia del cilindro più esterno
la corrispondenza prosegue ordinante lungo la prima traccia i settori danneggiati sono esclusi

MAPPATURA DEGLI INDIRIZZI
Per dischi rigidi per mappare un indirizzo usiamo **<cilindro,traccia,settore>** in indirizzi progressivi lineari LBA (Logical Block Address) , nella pratica è difficile una traduzione diretta per la presenza di vari vari fattori:

- settore danneggiati
- diverso numero di settori per traccia
- la gestione interna del controllore dell'operazione di mappatura.

Tuttavia gli algoritmi di gestione degli HDD tendono a presumere che gli indirizzi logici siano relativamente correlati agli indirizzi fisici, ovvero ad associare la crescita dell'indirizzo logico con la crescita dell'indirizzo fisico.

CLV Constant Linear Velocity densità dei bit per traccia uniforme
Tracce più lontane dal centro del disco sono più lunghe e contengono un maggior numero di settori la velocità di rotazione aumenta spostandosi verso l'interno quindi la quantità di dati che passano sotto le testina nell'unità di tempo

I CAV Constant Angular Velocity velocità di rotazione costante.

la densità dei bit decresce dalle tracce interne alle più esterne per mantenere costante la quantità di dati che passano sotto le testine nell'unità di tempo. Si usano dischi magnetici.

Scheduling del disco

Il sistema operativo è responsabile dell'uso efficiente dell'hardware per garantire tempi di accessi veloci dati il TEMPO DI ACCESO si divide in due componenti principali:

- Tempo di ricerca: è il tempo impiegato per spostare la testina sul cilindro che contiene il settore desiderato
- Latenza di rotazione: è il tempo necessario perché il disco ruoti fino a portare il settore desiderato sotto l a testina

Per migliorare le prestazioni din può intervenire solo su tempo si rientra e si tenta quindi di minimizzare.

L'AMPIEZZA DI BANDA DEL DISCO è il numero totale di byte trasferiti / il tempo trascorso tra la prima richiesta e il completamento dell'ultimo trasferimento quando un processo deve effettuare un operazione I/O effettua una chiamata al sistema operativo
la richiesta contiene:

- Specifica del tipo di operazione
- indirizzo su disco relativamente al quale effettuare il trasferimento
- indirizzo nella memoria relativamente al quale effettuare il trasferimento numero di byte da trasferire

Una richiesta di accesso al disco può venire soddisfatta se l'unità disco e controller sono disponibili, altrimenti viene aggiunta a una coda.
Il scheduling del disco ora è integrato dai controllori dei sds ipso sportivi di archiviazione che traducono direttamente gli indirizzi di blocco

In passato l'os era responsabile della gestione delle code, ovvero la schedulazione delle unità a disco

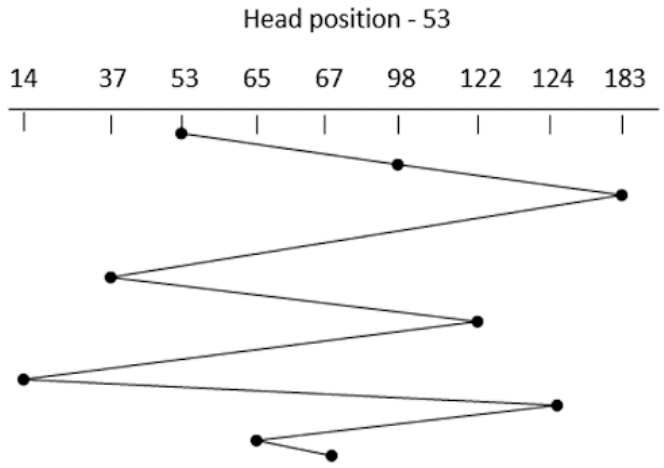
Scheduling del disco ora integrato ai controllore traducono
dei dispositivi di archiviazione che traducono direttamente gli indirizzi di blocco logico e gestione le code

Tuttavia lo scheduling deve mantenersi equo e tempestivo e garantire il raggruppamento di accessi che appaiono in sequenza, poiché si ottengono prestazioni migliori con I/O sequenziali

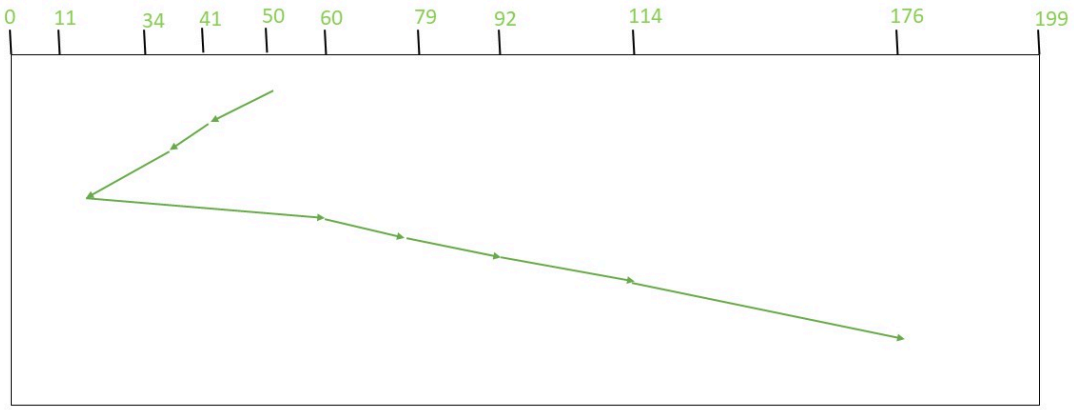
ESEMPIO

Gli algoritmi di scheduling vengono testati sulla coda di richieste per i cilindri (0 - 199)
98,183,37,122,14,124,65,67
La testina dell'unità a disco è posizionata sul cilindro 53

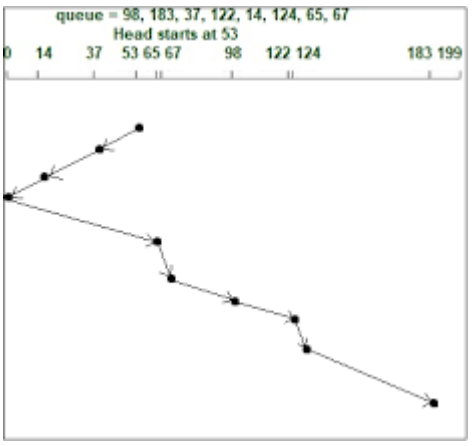
Algoritmo FCFS (first come first served) è un algoritmo che è considerato equo , si produce un movimento totale della testina pari a 640 cilindri. Si producono grandi oscillazioni



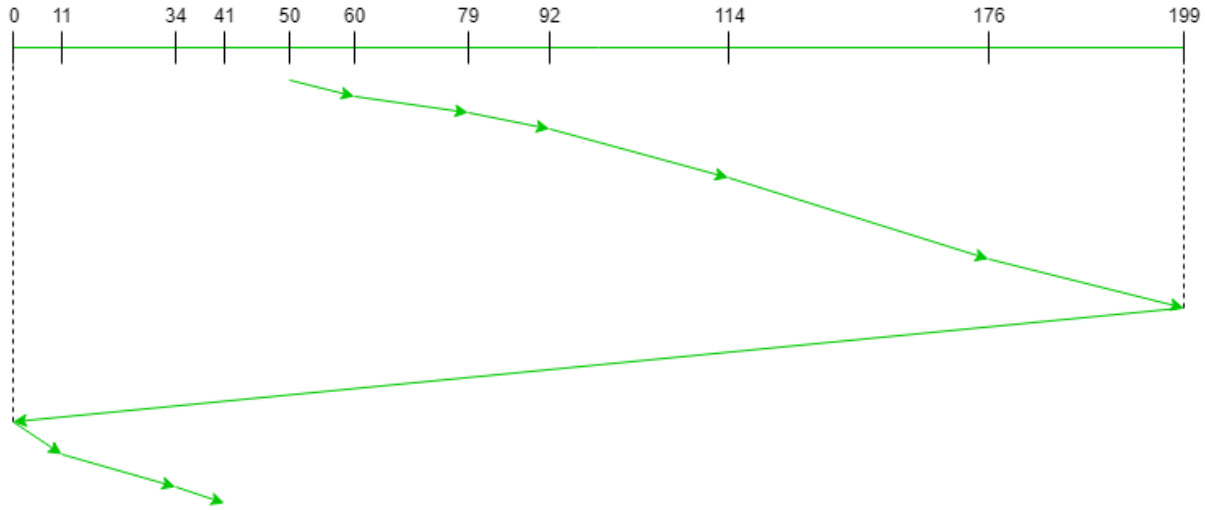
SSTF (shortest seek time first) opta per servire tutte le richiesta ma mando a minimizzare i movimenti radiali della testina, non è la scelta ottima e si può avere starvation e difficoltà nel servire le tracce periferiche si ha un movimento totale di 236 cilindri In pratica cerca di fare il minimo spostamento possibile tra le tracce.



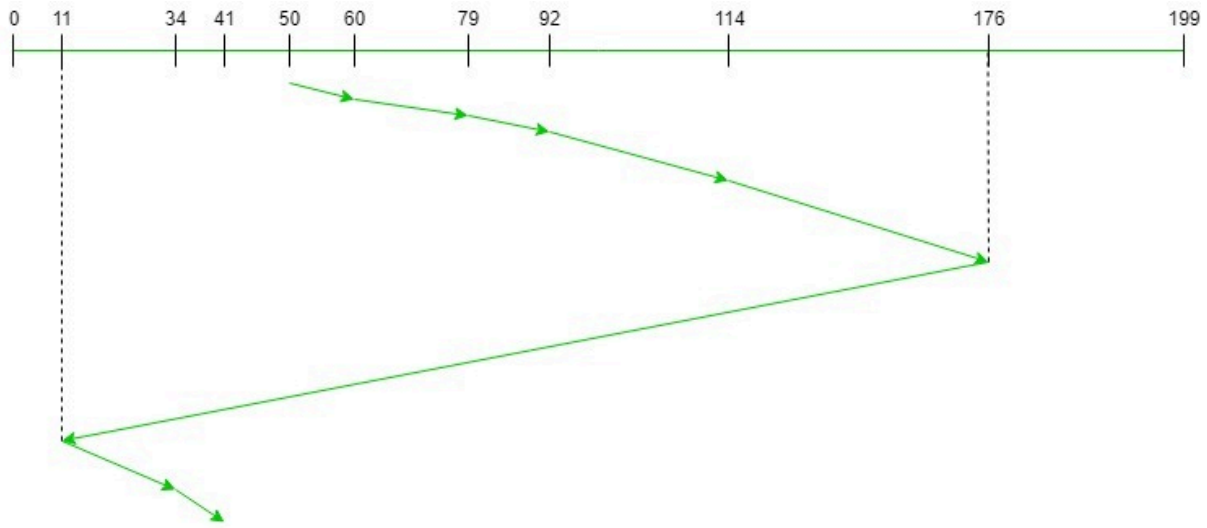
SCAN si muove da un estremo all'altro del disco servendo sequenzialmente le richieste giunte da un setrakian inverte si ha un movimento di 236 cilindri. è chiamato algoritmo dell'ascensore. se gli accessi sono distribuiti uniformemente quando la testina inverte il proprio movimento la maggior parte di richieste si ha all'esterno opposto del . avranno anche tempi di attesa maggiori.



C-SCAN garantisce un tempo di attesa uniforme rispetto allo scan
La testina si muove da un estremo all'altro del disco servendo sequenzialmente le richieste. Quando raggiunge l'ultimo cilindro non va a toccare il bordo ma direttamente torna indietro, senza servire richieste durante il viaggio di ritorno.
Considera i cilindri come organizzati secondo una lista circolare con l'ultimo cilindro adiacente al primo.
Si ha un movimento completo di 383 cilindri



C-look è un miglioramento del c scan il braccio non si sposta fino alla fine della richiesta estrema ma fino al settore indicato come ultimo e poi inverte direttamente la direzione



La scelta di un algoritmo di scheduling

SSTF molto comune e semplice da implementare e abbastanza efficiente
SCAN e C-scan migliore per i sistemi con un grande carico di i/o le performance dipendono dalle richieste.
Le performance dipendono dal numero di tipi e di richieste
le richieste ai dischi dipendono da come vengono allocati i file, ossia da come è implementato il file system.
L'algoritmo di scheduling dei dischi dovrebbe essere un modulo separato dal resto del kernel, facilmente rimpiazzabile se necessario
Sia **sstf** che **look** sono buone scelte come algoritmi di default

Tempo di lettura = N blocchi * (n cilindri* tempo di ricerca di spostamento cilindro + latenza di rotazione + tempo di trasferimento dati per blocco).

t = nb* (nc*tr+ lr + ttdb)

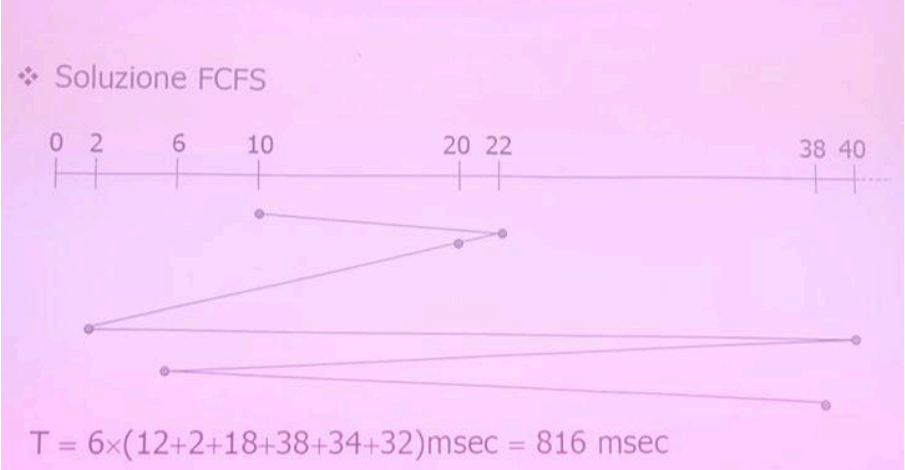
Esempio 1
Un disco ha c cilindri
6ms spostamento tra un cilindro e l'altro, latenza media 10 ms e tempo di trasferimento 0.25 ms.
Se abbiamo 20 blocchi e 13 cilindri?

T = 20 * (13*6+10+0.25) ms = 20 * (78+10+0.25) = 1765 ms

Se abbiamo 100 blocchi e 2 cilindri?

T = 100 * (2 * 6+10+0.25) ms = 100 * (12+10+0.25) = 2225 ms

Esempio 2
Al driver arrivano nell'ordine seguente richieste per i cilindri 10,22,20,2,40,6,38 Uno spostamento da una traccia all'altra ci impiega 6 ms. Ricava il tempo totale di lettura / scrittura del dato in base alla sequenza data



T = 6 * (12+2+18+38+34+32) = 816 ms

Scheduling su NVM

Le unità nvm dove non esistono delle parti mobili si utilizza di solito una politica FCFS, l'unica ottimizzazione possibile riguarda il servizio combinato di richieste relative a indirizzi logici adiacenti. tuttavia il vantaggio dei dispositivi nvm è meno sensibile in caso di accesso sequenziale contrariamente al tempo di ricerca sugli HHD quindi le prestazioni equivalenti o anche peggiori in caso di elevata usura del dispositivo. Problema dell'amplificazione di scrittura quando si attivano operazioni aggiuntive per la garbage collection. Non hanno parti meccaniche, basandosi infatti su sistemi di memoria solida. L'accesso è molto più veloce e permette un accesso ai dati simile a quello dei dischi tradizionali.
Ovviamente per questa categoria di dispositivi si pongono problematiche molto differenti:

- Gli algoritmi tradizionali di scheduling delle richieste non hanno più senso
- Insorgono problemi di write amplification con conseguente pericolo di usura precoce.
- I sistema operativi devono fornire del supporto esplicito per una corretta gestione degli SSD (supporto dell'istruzione TRIM)

Operazione TRIM

è una funzionalità di sottosistemi di memorie di massa dentro ad un elaboratore, che può essere utilizzato dall'OS per segnalare al dispositivo di memoria le pagine non più utilizzare che possono essere riallocate.
La funzione TRIM è una caratteristica tipica di tutti gli SSD ed è utilizzabile invocando l'omonimo comando attraverso l'interfaccia di controllo.
Generalmente nell'operazione di cancellazione eseguita da un sistema operativo i blocchi data vengono contrassegnati come non in uso. Il trim permette all 'OS di passare su questa info al controller dell' SSD che altrimenti non sarebbe in grado di sapere quali blocchi eliminare.

L'unità minima leggibile/scaricabile di un SSD è la pagina. Un disco nuovo permette di scrivere i nuovi dati a livello di singole pagine, Trattativa per riscrivere una pagina bisogna prima cancellare l'intero blocco, anche se nel blocco da cancellare ci sono altre pagine valide queste ultime vanno spostate in un altro blocco di prakla quindi di write amplification: (dati scritti su memoria flash/dati inviati dal SO)
sorge la necessità di un garbage collector per mantenere alte le performance
L'istruzione TRIM consente di informare il controller dell'unità SSD sulle pagine non più in uso da parte del file system

Rilevamento e correzione di errori
determina se si è verificato un problema: Il sistema può interrompere prima che l'errore si propaghi e la rivelazione viene eseguita frequentemente tramite il bit di parità. è una forma di CHECKSUM che utilizza l'aritmetica modulare per calcolare , archiviare e confrontare valori. Un altro metodo di rilevamento degli errori comune nelle reti è **il controllo di ridondanza ciclica** che utilizza una funzione hash per rivelare errori su più bit. Il codice di corrispondenza degli errori può anche correggere gli errori.

Memorie Terziarie:

- Pen Drive
- Memory Card
- Dischi Ottici

C'è sempre una testina, ma al posto del campo magnetico c'è un raggio laser che crea sottili scanalature su disco creando un'alternanza di zone scure o bianche, questo perché la luce riflette di meno sul buio e più su bianco. Quindi un rivelatore fotoelettrico misura la differenza di tale intensità e converte i segnali in una sequenza binaria.
La velocità dei dischi varia tra 60 e 250 giri al secondo, i parametri da considerare è la velocità di trasferimento e la velocità con cui i dati funzionano dal disco alla RAM ed è proporzionale alla velocità di rotazione.
Il tempo di posizionamento e il tempo necessario per spostare la testina in corrispondenza del cilindro, tutte le testine sono collegate assieme.
Influito dalla latenza di rotazione.
E' posizionata sospesa su un cuscinetto d'aria.
I dischi sono rimovibili perché i dischi si possono danneggiare e devono essere sostituiti.

GESTIONE DELLA MEMORIA SECONDARIA

Il SO si deve occupare anche di:

- Analizzare i disposti vuoti di memoria scendere mediante formattazione sia di tipo fisico che di tipo logico,
-
- Gestire i blocchi difettosi del disco
- Gestire in modo efficiente l'area di swap

FORMATTAZIONE

- Alla produzione il disco magnetico non è diviso in settori e dunque non può essere letto o scritto dal controller.
- L'operazione con cui il disco viene diviso in settori è detta formattazione fisica o formattazione di basso livello.
- L'area di dati è tipicamente di 512 byte. Intestazione e coda servono al controller per inserire informazioni di servizio. il codice di correzione Error-Correcting Code (ECC) è usato dal controller per rilevare malfunzionamenti sul settore, per capire se funziona o meno. Se il codice calcolato run-time del controller è diverso da quello memorizzato, allora si è verificato un errore su quel settore.
- Disco Vergine: Niente
- Formattazione Fisica:
    - Suddivisione in settori
    - Identificazione dei settori
    - Aggiunta si pazio di correzione

Formattazione Logica
File System
Lista spazio occupato o libero.

Formattazione di basso livello o fisica:

- Si suddivide il disco in settori, che possono essere letti e scritti dal controllore del disco
- dimensione standard pari a 4kb
- Nel caso di NVM devono essere iniziate le pagine e creata la tabella FTL(flash translation layer)
- In entrambi i casi la formazione di basso livello inserisce nel dispositivo una struttura speciale di dati per ogni blocco di memoria
- Dimensioni standard pari a 4kb
- Nel caso di nvm devono essere inizializzate le pagine create
- In entrambi i casi, la formattazione di basso livello inserisce nel dispositivo una speciale struttura dati per ogni blocco di memoria. .Intestazione e coda contengono info ad uso del controllore

Per poter impiegare un dispositivo al fine di memorizzare i nostri file il SO deve mantenere le proprie strutture dati sul disco (HDD/SSD)
Si partiziona il dispositivo in uno o più gruppi di cilindri/piadine ognuno dei quali rappresenta un disco logico.

Formattazione Logica o creazione di un file system Strutture dati del SO per la descrizione dello spazio libero/occupato e creazione di una directory iniziale vuota.

Per migliorare le prestazioni, la maggior parte dei file system accorpa i blocchi in gruppi, detti Cluster: I/O su disco fatto per blocchi, I/O via file system fatto per cluster, File e metadati vicini su HDD per diminuire i movimenti della testina.
 Al momento dell'accensione il computer mette in funzione un programma iniziare bootstrap molto semplice che inizializza il sistema e avvia il SO

In informatica, il ***boot*** (o ***bootstrap***, o più raramente ***booting***) o l'**avvio del computer** è, in generale, l'insieme dei processi che vengono eseguiti da un computer dall'accensione fino al completo caricamento in memoria primaria del *kernel* del sistema operativo a partire dalla memoria secondaria.

Il bootstrap può essere memorizzato:

Su una ROM sola lettura ma è difficile sostituire il codice per migliorarlo, Su un disco: in questo caso viene utilizzata un ‘area del disco detta area di boot.
In realtà l’approccio è misto: Nella rom è collocato un piccolissimo programma che avvia il boot.
Il programma di avvio avviene in modo completo nell'area di boot del disco di sistema, quando viene letta l’area di boot il SO non è ancora stato caricato, non ha i driver e non esiste il file system.
Nel settore di boot quindi non esiste una strutturazione logica dei file. Quindi l'accesso nell’area di boot non è una normale lettura di un file bensì la lettura di alcuni settori fisici in cui è contenuta direttamente l’immagine da caricare in memoria.

blocco di boot:
Un disco di avviamento o disco di sistema ha una partizione di boot. Il primo blocco logico del dispositivo è il blocco di avvio o detto boot block.
La partizione di boot contiene il SO, altre partizioni possono contenere altri SO, altri file system o essere partizioni raw. Viene montata all’avvio del sistema
Altre partizioni possono essere montate automaticamente o manualmente
Al momento del montaggio di ogni partizione, si verifica la coerenza del file system (controllando la correttezza dei metadato). Si aggiorna la tabella di montaggio.

gestione:
Nel boot block sono contenute le info necessarie per l’inizializzazione del sistema. In windows si chiama MBR (Master Boot Record) esecuzione del codice del bootstrap loader contenuto nel firmware Lettura/esecuzione del codice contenuto nel MBR che contiene una tabella delle partizioni.
caricamento della partizione di boot , del kernel e dei sottosistemi del SO

I dischi magnetici sono strutturalmente proni a malfunzionamenti, perché costituiti da parti mobili con basse tolleranze. Si impiega l'accantonamento dei settori come modalità di gestione dei blocchi difettosi:
Durante la formattazione fisica si mantiene un gruppo di settori di riserva non visibili al SO. Il controllore “è istruito” per sostituire, dal punto di vista logico, un settore difettoso con uno dei settori di riserva inutilizzati. Anche i dispositivi NVM possono contenere pagine difettose, che vengono logicamente sostituite o con pagine accantonate o appartenenti alla riserva costituita o appartenenti alla riserva costituita dell'over-provisioning.

area dello swap:
l'area di swap è parte di disco usata dal gestore della memoria come estensione della memoria principale. Può essere ricavata dal file system normale o in una partizione separata. Gestione dell'area di :

swap: 4.3BSD: alloca lo spazio appena parte il processo per i segmenti text e data. Per lo stack, lo spazio viene allocato man mano che cresce.

Solaris 2: si colloca una pagina dello stack solo quando di deve fare un NPN page-out , non alla ricreazione della pagina virtuale.

gestione:
La memoria virtuale impiega lo spazio su disco come un'estensione della memoria centrale. Pratica attualmente meno comune, grazie all'incremento nella capacità delle memorie
L'obiettivo principale nella progettazione e realizzazione dell'area di swap è di fornire la migliore produttività per il sistema di memoria virtuale

Lo spazio di swap può essere ricavato all'interno del normale file system o più comunemente, si può trovare in una partizione separata del disco.

Protezione raw:
adotta algoritmi ottimizzati rispetto alla velocità di accesso piuttosto che all'occupazione di spazio

L'area di swap, in linux, è utilizzata solo per la memoria anonima, ovvero per dati che non corrispondono a file.
Linux permette di una o più aree di avvicendamento, sia in file che in una partizione raw.
Un 'area dia avvicendano è formata da una serie di moduli detti di slot delle pagine

Ogni area dispone di una mappa di avvicendamento, un array di contatori interi, ciascuno dei quali corrisponde ad uno slot dell'area.

Se un contatore vale 0, la pagina che gli corrisponde è disponibile, valori maggiori di 0 indicano che lo slot è pccipatro da un a delle pagine avvicendate. Il valore del contatore indica il numero di collegamenti alla pagina, se vale 3 la pagina da parte dello spazio degli indirizzi virtuali di tre processi distinti.

Affidabilità e performance dei dischi:

Cause:
- Aumenta sempre più la differenza di velocità tra applicazione e dischi.
- Le memorie cache non sempre sono efficaci

Soluzione:
- Suddividere il carico tra più dischi che cooperano per offrire l'immagine di un disco unitario virtuale più efficiente.

RAID

RAID in info è un sinonimo di archiviazione
Originariamente il termine faceva riferimento a una serie ridondante di isole poco costosi. In seguito, l'acronimo è stato aggiornato e oggi la definizione fa riferimento a Redundant Array Of Independent Disks, ovvero una serie ridondante di dischi indipendenti.

è un metodo molto comune per proteggere i dati delle nostre applicazioni sia un'unità disco fisso che su storage allo stato solido.

-L'idea originale è di combinare una serie di dischi a vasto i costi in modo da ottimizzare il sistema in termini di capacità, affidabilità e velocità rispetto a un disco di ultima generazione.
Ne esistono diversi tipi che bilanciano il livello di protezione, in base al loro prezzo

i dischi più economici sono gli:
IDE (acronimo) da cercare
ATA (acronimo)
SATA (acronimo)
i dischi più costosi sono gli SCSI (acronimo)

Raggruppando le singole unità fisiche in modo da formare un set di dischi, quindi il raid rappresenta tutte queste unità fisiche come un disco logico su server. è chiamato numero di unità logica LUN Logical Unit Number

I dati vengono partizionati in sezione di uguale lunghezza e trascritti su dischi differenti utilizzando un algoritmo per la distribuzione.

Quando si richiede una lettura di dimensione superiore all'unità di sezionamento,questa tecnologia distribuisce il carico di lavoro su più dischi in parallelo, aumentando così le prestazioni.

La ridondanza viene gestita dal controller
Diversi livelli a seconda del tipi di ridondanza:

- 0 livello: striping: i dati vengono affettati e parallelizzati. Altissima perfomance, non c'è ridondanza.

- 1 livello: mirroring o shadowing: vengono duplicati dischi in modo interno. Eccellente resistenza ai crash, basse performance in scrittura.

- 5 livello: Block interleaved parity: un disco a turno per gpo striper viene dedicato a contenere l'informazione di parità del resto della striper. Alta resistenza, discrete performance.

Caratteristiche Raid 0:

- In raid 0 i dati vengono suddivisi in blocchi ed ognuno di questi viene memorizzato in un disco dell'array a disposizione secondo lo schema.
- La performance dell/i/o aumenta notevolmente dato che il carico di lavoro viene suddiviso tra più ischi. Non c'è ridondanza

Il sezionamento del disco o data striping tratta un gruppo di dischi come un'unica unità di memorizzazione:

- ogni blocco di dati è suddivisa in sottoblocchi memorizzati su dischi distinti. Il tempo di trasferimento per rotazioni sincronizzate diminuisce proporzionalmente al numero di dischi che abbiamo in batteria.

gli schemi raid migliorano prestazione ed affidabilità del sistema memorizzando i dati ridondanti: RAID 1 conserva duplicati su ciascun disco

Caratteristiche Raid 1: 2 dischi

- I dati vengono duplicati su coopir di schi
- La performance della lettura dei dati raddoppia, mentre quella della scrittura rimane uguale al caso di un singolo disco.
- Nel caso di guasto ad un disco, si può utilizzare immediatamente l'arto
- Aspetto negativo: spreco di spazio.

Caratteristiche Raid 2: 7 dischi

- Questo livello era nato per dischi che non possedevano meccanismi propri per gestire gli errori di lettura/scrittura
- pera a livello di bit utilizzati poi i codici di Hamming(7,4): 4 bit di dati ognuno su uno dei 4 dischi e 3 bit di parità ognuno su un disco di parità
- Può correggere errori dovuti all'inversione di un singolo bit e rilevare errori dovuti all'inversione di due bit.

Caratteristiche Raid 3: 3 dischi

- I dati sono organizzati a livello di buit p byte: come ECC utilizza XOR
- Prestazioni molto elevate in lettura/scrittura
- Ottima tolleranza ai guasti
- In generale si può evadere una singola richiesta per volta.

Caratteristiche Raid 4: 3 dischi

- i dati suddivisi in blocchi: ogni blocco viene memorizzato su un disco nell'array seconde lo schema,
- Prestazioni molto elevate in lettura
- Ottima Tolleranza ai guasti
- Il disco di parità è un collo di bottiglia.

Caratteristiche Raid 5: 3 dischi:

- Il principio di funzionamento è quello del RAID 4, ma i blocchi di parità sono menati in modo distribuito sui dischi dell'array
- Prestazioni migliorate in scrittura (RANDOM)
- Sistema è complesso e costoso da realizzare

Caratteristiche Raid 6: 4 dischi:

- Il principio di funzionamento del 5, ma vengono utilizzati due tipi di controllo di errori
- Grande ridondanza e sicurezza dei dati
- In caso di guasto la ricostruzione dei dati è molto lenta a causa del doppio controllo ECC.

Connessione dei dispositivi di memoria

Connessione

- I calcolatori accedono alla memoria secondaria in 3 medif
    - Tramite un dispositivo collegato alla macchina
    - Tramite un dispositivo connesso alla rete
    - In cloud

- Alla memoria secondaria connessa alla macchina si accede dalle porte locali di I/O che sono collegate al bus.
    - Nei PC, con interfaccia SATA, due unità al più per ciascun bus di i/o.

Per accedere a un maggiore spazio di archiviazione utilizzo di porte e cavi USB, FireWire e Thunderbolt

FC è un architettura seriale ad alta velocità:
Può gestire uno spazio di indirizzi a 24 bit, che è alla base storage area network, nelle quali molti host sono connessi con altrettante unità di memorizzazione.
Per eseguire un'operazione di i/o si inserisce un comodo opportuno nell'adattatore, generalmente mediante porte i/o mappate in
memoria. L'adattatore invia il comando al controllore del disco, che gestisce l'hardware dell'unità,dell'unità, per portare a termine il compito richiesto. Il trasferimento dei dati avviene tra la superficie del disco e la cache incorporata nel controllore.
Il trasferimento verso l'host avviene a velocità elevata tramite **DMA** → Direct access memory : di un computer e quel meccanismo che permette ad altri sottosistemi, quali ad esempio le periferiche, di interferire direttamente alla memoria interna per scambiare dati in lettura e/o scrittura senza coinvolgere la CPU.

Un dispositivo di memoria secondaria connessa alla rete (**NAS**) è un sistema di memoria specializzato al quale si accede in modo attraverso la rete di trasmissione dati. I client accedono alla memoria connessa alla rete tramite un'interfaccia RPC, supportato da protocolli quelli NFS e CIFS. Le chiamate di procedura remota sono implementate per mezzo di TCP o UDP, sopra una rete IP, di solito la stessa rete che supporta tutto il traffico dei dati tra client e server.

Memoria secondaria in cloud:

Similmente al NAS, il cloud fornisce l'accesso allo storage tramite rete. A differenza del NAS, l'accesso al data center remoto avviene tramite Internet o Wan.
Il NAS si presenta come un file system mentre lo storage cloud è basato su API, con programmi che utilizzano le API per fornire l'accesso: Si impiegano le API a causa delle lunghe latenze e per i numerosi scenari di errore che sono comuni sulle WAN.
Esempi di cloud storage sono : Google Drive, Microsoft Onedrive, Dropbox, Amazon s3, ICloud.

SAN:
Reti private (che impiegano protocolli specifici per la memorizzazione) tra server e unità di memoria secondaria. Flessibilità : si possono connettere alla stessa SAN molti calcolatori e molti storage array.

Partizionamento del disco

Dopo aver esaminato la struttura del disco da un punto di vista fisico ora andremo ad analizzare l'organizzazione dei dati memorizzati all'interno di un disco, che logicamente è suddiviso in partizioni. Si definisce partizione di u n disco: una parte di disco separata da un punto di vista logico.
Il partizionamento di un disco può permettere di: Separare da punti di vista logico i file che costituiscono dai file dei programmi applicativi e dai dati degli utenti, memorizzando su partizioni separate per avere una migliore organizzazione dei dati sulla memoria. Avere la possibilità di installare su più SO sullo stesso computer. Avere più file system che possono essere usati sullo stesso sistema operativo. Prevede un'area di Backup. Per i PC portatili, definire un'area per memorizzare lo stato del sistema, quando è in modalità standby.

L'ambiente BIOS

Analizziamo ora l'organizzazione dei dati in ambiente **BIOS** (Basic Input Output System). è il programma che viene eseguito nella fase di accensione di un PC, permettendo la fase di avvio e tra le altre cose, di prelevare dalla memoria di massa il SO e trasferirlo all'interno della memoria centrale.

In ambiente BIOS, il primo blocco del disco con coordinate CHS, chiamato MBR ha una funzione FONDAMENTALE, perché fornisce alla CPU le istruzioni per eseguire il caricamento del SO e permette al SO stesso di usare correttamente il supporto di memoria. All'interno del MBR sono memorizzate alcune linee di codice in linguaggio macchina che permettono di leggere alla Tabella delle Partizioni alla ricerca della partizione attiva, è definita come quella porzione di disco che nel suo primo settore conteneva il Il Master Boot Code, che permette di caricare in memoria centrale del SO, cioè il Kernel.

ogni memoria di massa può contenere fino a quattro partizioni di cui tre primarie, con la caratteristica di essere bootable, cioè avviabili. Questo fa sì che ci sia la possibilità di caricare più sistemi operativi. Grazie al bootloader che esegue la fase di bootstrap di uno dei

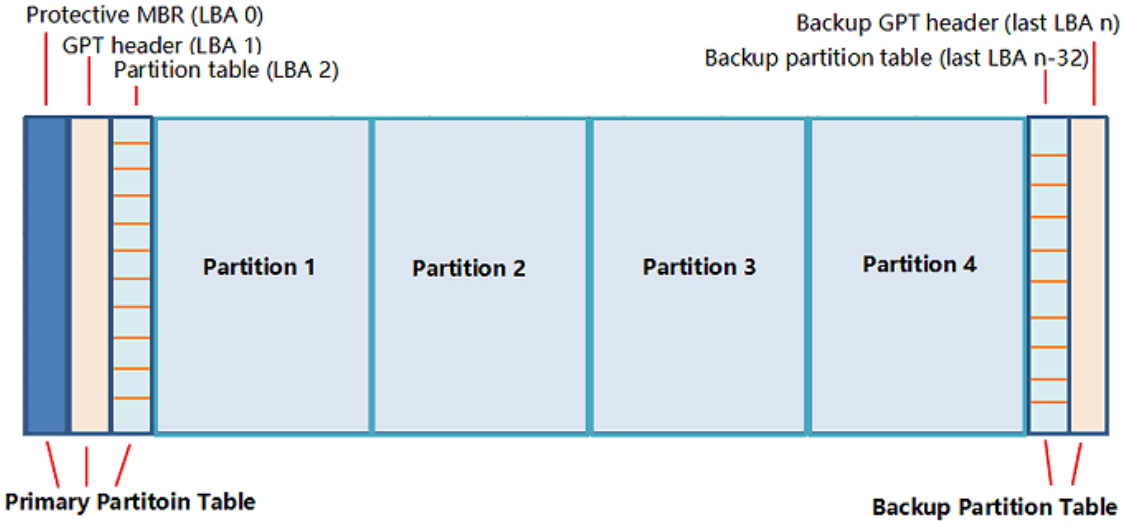
SO scelti dall'utente. La partizione estesa ha invece il compito di contenere altre partizioni logiche che non possono essere rese avviabili. A ogni partizione viene assegnata una lettera alfabetica maiuscola progressiva a partire dalla C. I dischi supportati possono avere una dimensione massima di 2 TB

Al termine del MBR, è memorizzata la Partition Table, che contiene l'elenco delle partizioni su disco. Ogni record della tabella contiene le seguenti info: La terna CHS del primo e ultimo blocco della partizione, il valore del LBA del primo blocco, La dimensione totale della partizione e se la partizione è avviabile oppure no.

Organizzazione logica dei dati su disco
AMBIENTE UEFI

Analizziamo ora l'organizzazione dei dati in ambiente UEFI è un nuovo firmware che è andato a inglobare in maniera progressiva, a partire dal 2010, il BIOS, aggiungendo però ulteriori funzionalità e superando alcune sue limitazioni. Il GPT è il nuovo standard per la definizione della tabella delle partizioni su unità di memoria a stato solido o disco fisso. Il GPT utilizza nuove potenzialità offerte dall'ambiente UEFI. Per motivi di protezione e compatibilità l'unità inizia con un riferimento MBR, cui segue il GPT stesso con la tabella delle partizioni. Il GPT utilizza l'indirizzamento LBA anziché l'indirizzamento di tipo CHS utilizzato dal MBR.

Un disco GPT è partizionato in: Primary Partition table, contiene il Protective MBR (LBA 0) ,l'intestazione della tabella di partizione GPT (LBA 1) e la tabella delle partizioni (LBA 2-33).
Data Partition, è la posizione fisica in cui il disco GPT memorizza i dati e i file.
Backup Partition Table, è un'area in cui il disco GPT conserva le info di backup per l'intestazione GPT e la tabella delle partizioni.



GPT Disk Layout

LBA 0

La partizione GPT PREVEDE UNA FUNZIONE DI RETROCOMPATIBILITÀ CHIAMATA protective MBR.
Ai SO che non sopportano GPT viene mostrata un unica partizione primaria che copre l'intero disco, in modo tale che software di partizionamenti obsoleti che non interpretano erroneamente i dischi GPT come non iniziallitatti oppure provvisti di partizioni. Questa partizione è in sola lettura, infatti per accedere ad un disco GPT il SO deve supportare esplicitamente.

LBA 2-33

sono dei record che contengono: I primi 16 bytes contengo il Type Partition GUID (acronimo), i successivi 16 bytes contengono il Unique Partition , 8 byte per LBA dell'inizio della partizione considerata, Il nome della partizione e agli attributi eventuali.

CANE QUA HA FATTO LA FOTO (serve foto caratteristiche tra MBR e GPT [CANE])

	MBR	GPT
Maximum Partition Capacity	2TB	9.4ZB (1 ZB is 1 billion terabytes)
Maximum Partition Number	4 primary partitions(or 3 primary + an infinite number of logical partitions)	128 primary partitions
Firmware Interface Support	BIOS	UEFI
Operating System Support	Windows 7 and older systems like Windows 95/98, Windows XP 32-bit, Windows 2000, Windows 2003 32-bit	Later systems like Windows 11, Windows 10 64-bit, Windows 8/8.1 64-bit