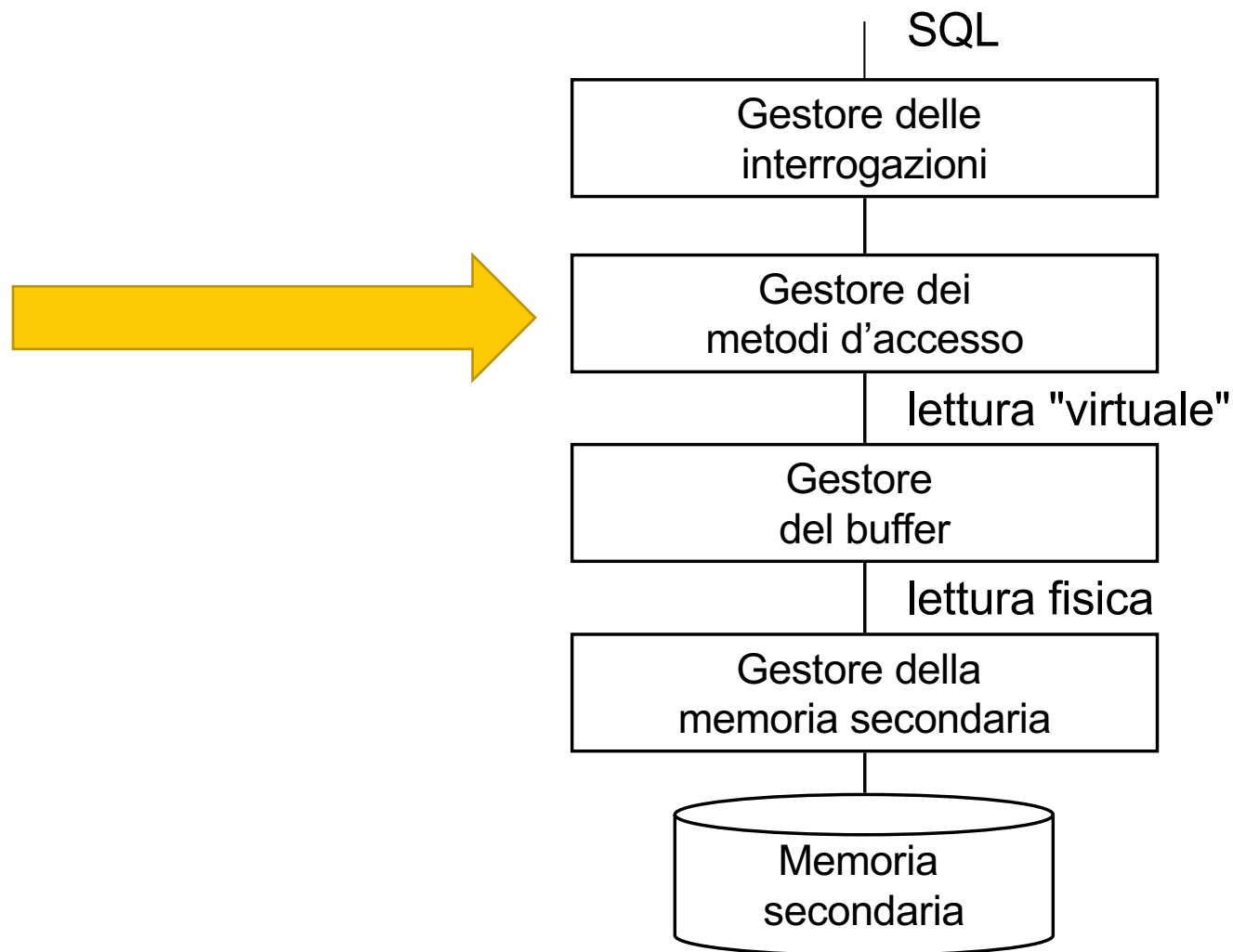


Gestore degli accessi e delle interrogazioni



Strutture ausiliari d'accesso - INDICI

- Struttura ausiliaria per l'accesso (efficiente) ai record di un file sulla base dei valori di un campo detto chiave
- **INDICE**: struttura d'accesso definita su campo chiave/campo di indicizzazione dove (per ogni) valore del campo chiave si memorizza il puntatore al blocco che memorizza i record con quel valore del campo
- IDEA: **ordinare l'indice** in base al valore del campo chiave (indici ordinati) per facilitare la ricerca
 - Analogo all'indice analitico di un libro: lista di coppie (termine, pagina), ordinata alfabeticamente sui termini
- **Ordine sul campo chiave/indicizzazione permette ricerca binaria sul file indice**

Indici ordinati

- Un indice \mathcal{I} per un file f è un altro file con record/voce definiti su due campi:
 - Campo chiave/indicizzazione
 - Puntatore al record (numero blocco)

I record/voci dell'indice sono ordinati secondo i valori del campo chiave

In base all'organizzazione dei file dati, possiamo avere diversi tipi di indici ordinati:

- Indici creati su campi usati per l'ordinamento fisico del file
 - Indice primario
 - Indice di clustering
- Indici creati su file non ordinati o creati su campi non usati per ordinamento del file
 - Indici secondari

INDICE PRIMARIO

- Si applica nel caso di file dati ordinato
- L'indice primario è un indice specificato sul campo chiave utilizzato nel file dati per ordinare i record
 - Il file dati è ordinato su campo K
 - Indice primario è creato su campo K
- Record/voce dell'indice primario:
 - Campo chiave (chiave primaria): stesso tipo del campo di ordinamento del file dati
 - Puntatore al blocco che memorizza il record con quel valore di campo

E' necessario un record/voce per ogni valore del campo chiave?

INDICE PRIMARIO

Data file

(Primary
key field)

Name	Ssn	Birth_date	Job	Salary	Sex
Aaron, Ed					
Abbot, Diane					
⋮					
Acosta, Marc					

Adams, John					
Adams, Robin					
⋮					
Akers, Jan					

Alexander, Ed					
Alfred, Bob					
⋮					
Allen, Sam					

Allen, Troy					
Anders, Keith					
⋮					
Anderson, Rob					

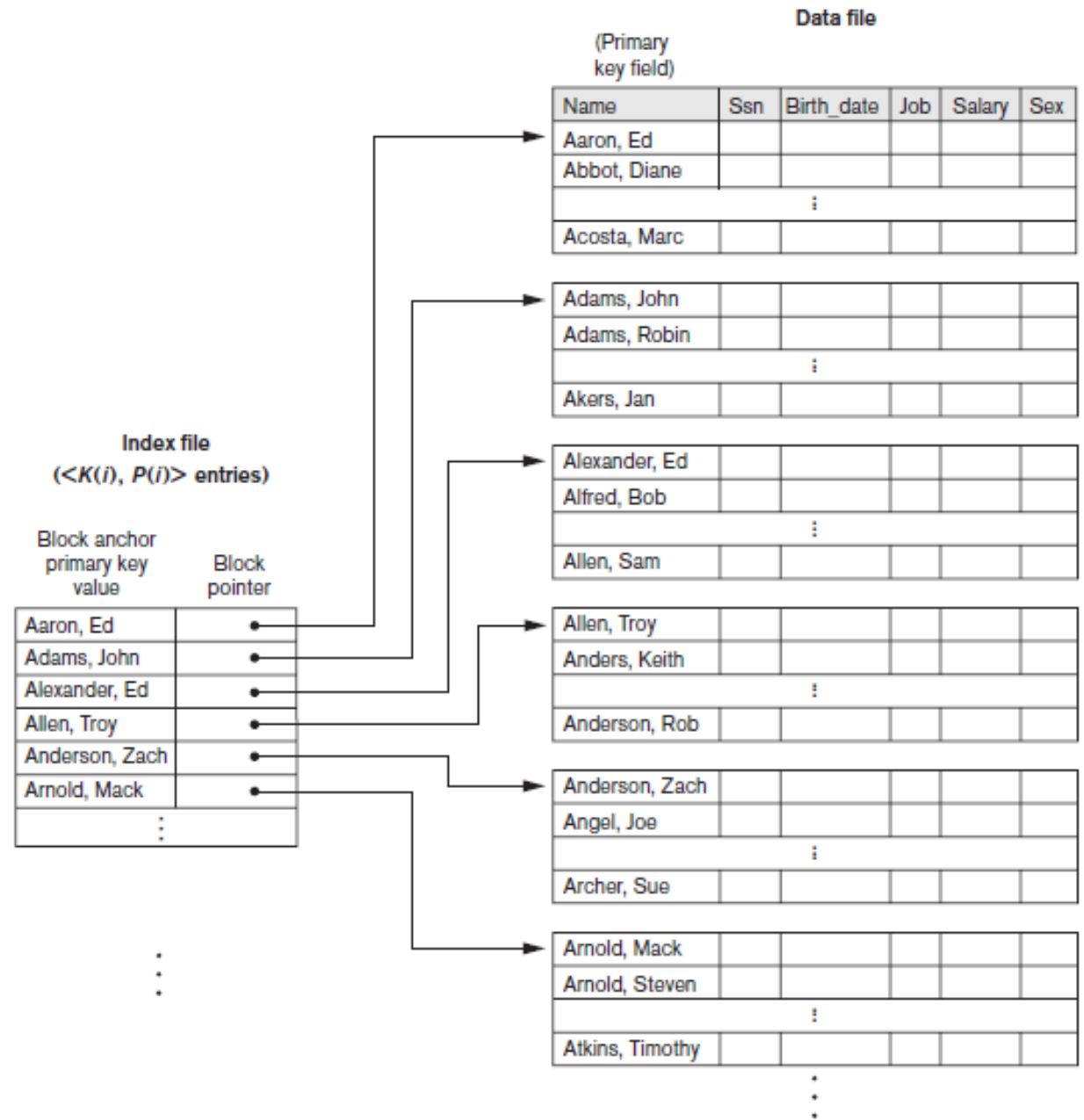
Anderson, Zach					
Angel, Joe					
⋮					
Archer, Sue					

Arnold, Mack					
Arnold, Steven					
⋮					
Atkins, Timothy					

⋮

INDICE PRIMARIO

- Sfrutto l'ordinamento del file dati per avere una voce che punta al primo record di ogni blocco (o ultimo)
- Indice primario ha un numero di record/voci pari al numero di blocchi del file dati
- Indice primario è SPARSO
 - Contiene voci solo per alcuni valori del campo chiave



INDICE PRIMARIO

- Costo d'accesso?
- Un file indice per un indice primario occupa meno blocchi rispetto al file dei dati:
 - ha meno voci/record
 - ogni voce/record è più piccolo
- La ricerca binaria su un indice primario è più efficiente che una ricerca binaria su file ordinato

Esercizio: Costo ricerca in termini di numero accessi

- Si supponga un file ordinato su campo ID con $r=300.000$ record memorizzati su disco con blocco $B=4096$ di lunghezza. I record hanno lunghezza fissa $R=100$ byte e sono memorizzati in modo unspanned.
- Si assuma un indice primario su campo ID dove:
 - Dimensione ID è 9 byte
 - Dimensione puntatore è 6 byte

Mediamente, quanti accessi sono richiesti per una ricerca sul campo ID?

Esercizio: Costo ricerca in termini di numero accessi

- Si supponga un file ordinato su campo ID con $r=300.000$ record memorizzati su disco con blocco $B=4096$ di lunghezza. I record hanno lunghezza fissa $R=100$ byte e sono memorizzati in modo unspanned.
- Si assuma un indice primario su campo ID dove:
 - Dimensione ID è 9 byte
 - Dimensione puntatore è 6 byte

Mediamente, quanti accessi sono richiesti per una ricerca sul campo ID?

- Fattore di blocco per il file dati $\lfloor 4096/100 \rfloor = 40$
- Numero blocchi per memorizzare i record: $\lceil 300.000/40 \rceil = 7500$
- Fattore di blocco per il file indice: $\lfloor 4096/15 \rfloor = 273$
- Nel file indice, una voce per ogni blocco di dati, quindi
 - Numero blocchi per memorizzare indice: $\lceil 7500/273 \rceil = 28$
- Numero accesso ai blocchi file indice per ricerca binaria $\lceil \log_2 28 \rceil = 5$
- N. Accessi per ricercare record $5+1$

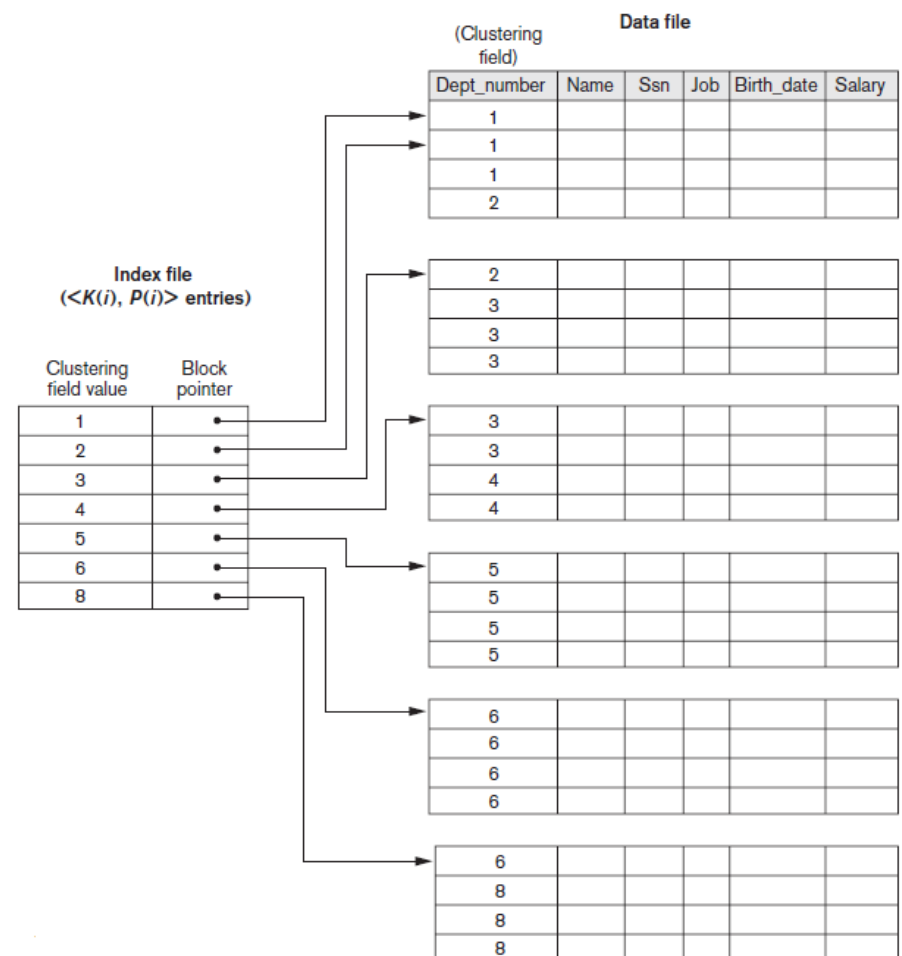
INDICE di CLUSTERING

- Si applica nel caso di file dati ordinato su un campo che non è UNIQUE/chiave primaria
 - Campo di clustering o campo di raggruppamento
 - File clustered
- Utile per velocizzare ricerca in caso di tanti record con lo stesso valore di raggruppamento

(Clustering field)		Data file				
Dept_number	Name	Ssn	Job	Birth_date	Salary	
1						
1						
1						
2						
2						
3						
3						
3						
4						
4						
5						
5						
5						
5						
6						
6						
6						
6						
6						
8						
8						
8						
8						

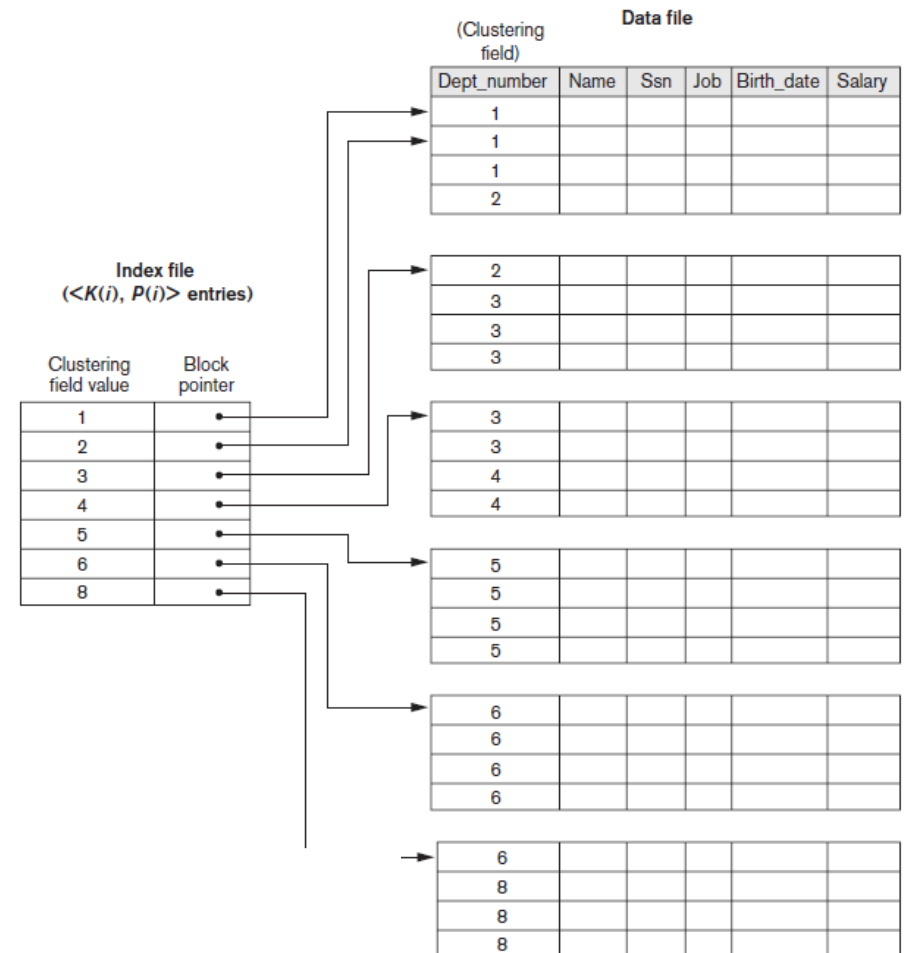
INDICE di CLUSTERING

- Si applica nel caso di file dati ordinato su un campo che non è UNIQUE/chiave primaria
 - Campo di clustering o campo di raggruppamento
 - File clustered
- Utile per velocizzare ricerca in caso di tanti record con lo stesso valore di raggruppamento
- Record nell'indice di clustering:
 - Valore del campo di raggruppamento
 - Puntatore al primo blocco che contiene il primo record con quel valore
- Numero voci nel file indice pari ai possibili valori di raggruppamento
- Indice SPARSO



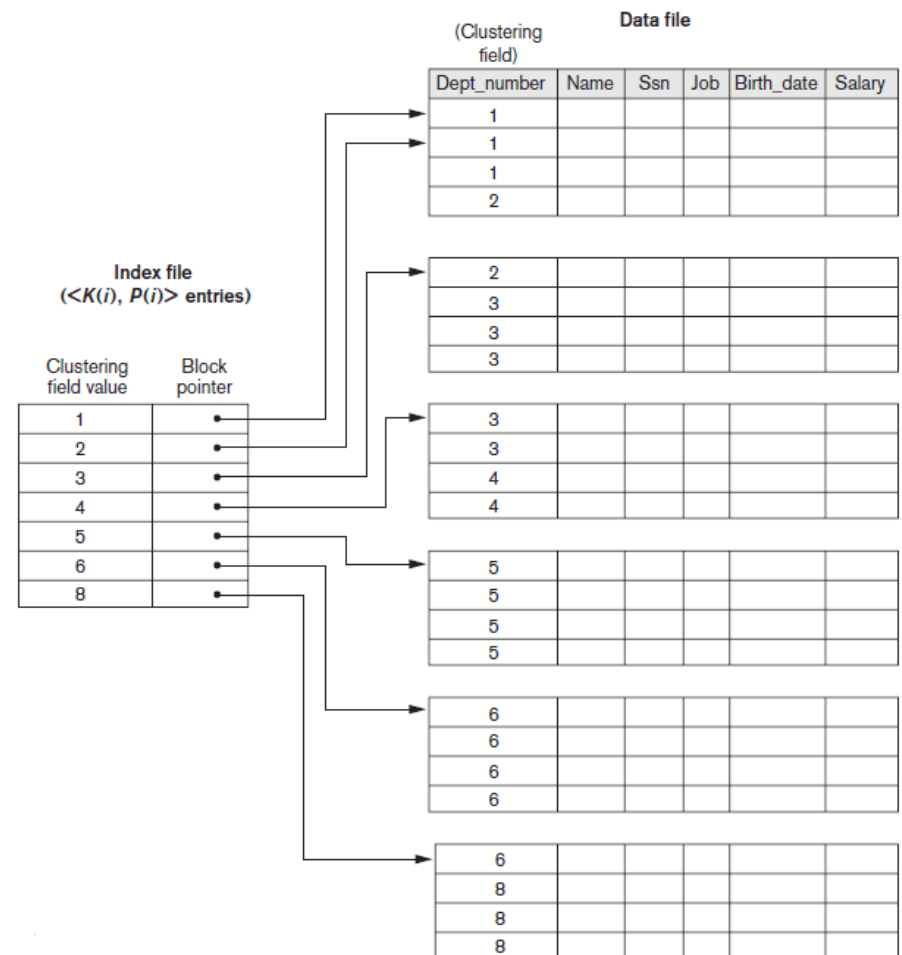
INDICE di CLUSTERING

- Si applica nel caso di file dati ordinato su un campo che non è UNIQUE/chiave primaria
 - Campo di clustering o campo di raggruppamento
 - File clustered
- Utile per velocizzare ricerca in caso di tanti record con lo stesso valore di raggruppamento
- Record nell'indice di clustering:
 - Valore del campo di raggruppamento
 - Puntatore al primo blocco che contiene il primo record con quel valore
- Numero voci nel file indice pari ai possibili valori di raggruppamento
- Indice SPARSO
- **Problema:** inserimento/cancellazione richiede riordino dei record



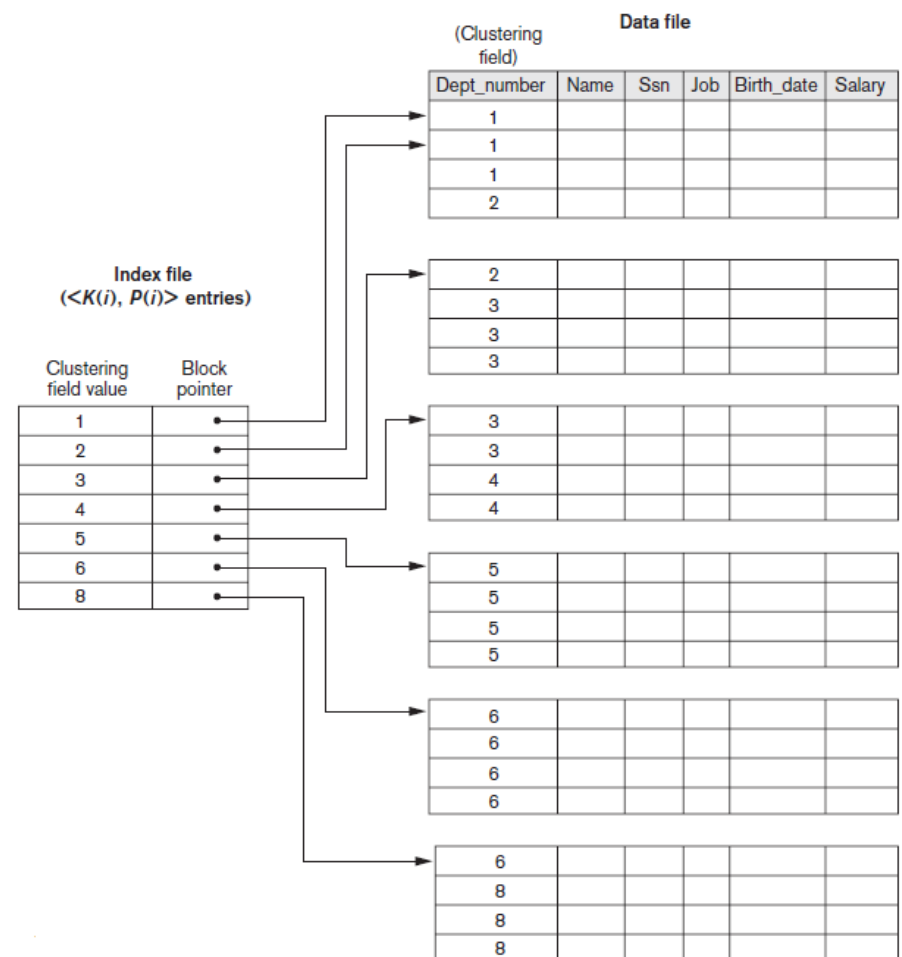
INDICE di CLUSTERING

- Si applica nel caso di file dati ordinato su un campo che non è UNIQUE/chiave primaria
 - Campo di clustering o campo di raggruppamento
 - File clustered
- Utile per velocizzare ricerca in caso di tanti record con lo stesso valore di raggruppamento
- Record nell'indice di clustering:
 - Valore del campo di raggruppamento
 - Puntatore al primo blocco che contiene il primo record con quel valore
- Numero voci nel file indice pari ai possibili valori di raggruppamento
- Indice SPARSO
- **Problema:** inserimento/cancellazione richiede riordino dei record
- Soluzione: un blocco differente o gruppi di blocchi per ogni valore di raggruppamento



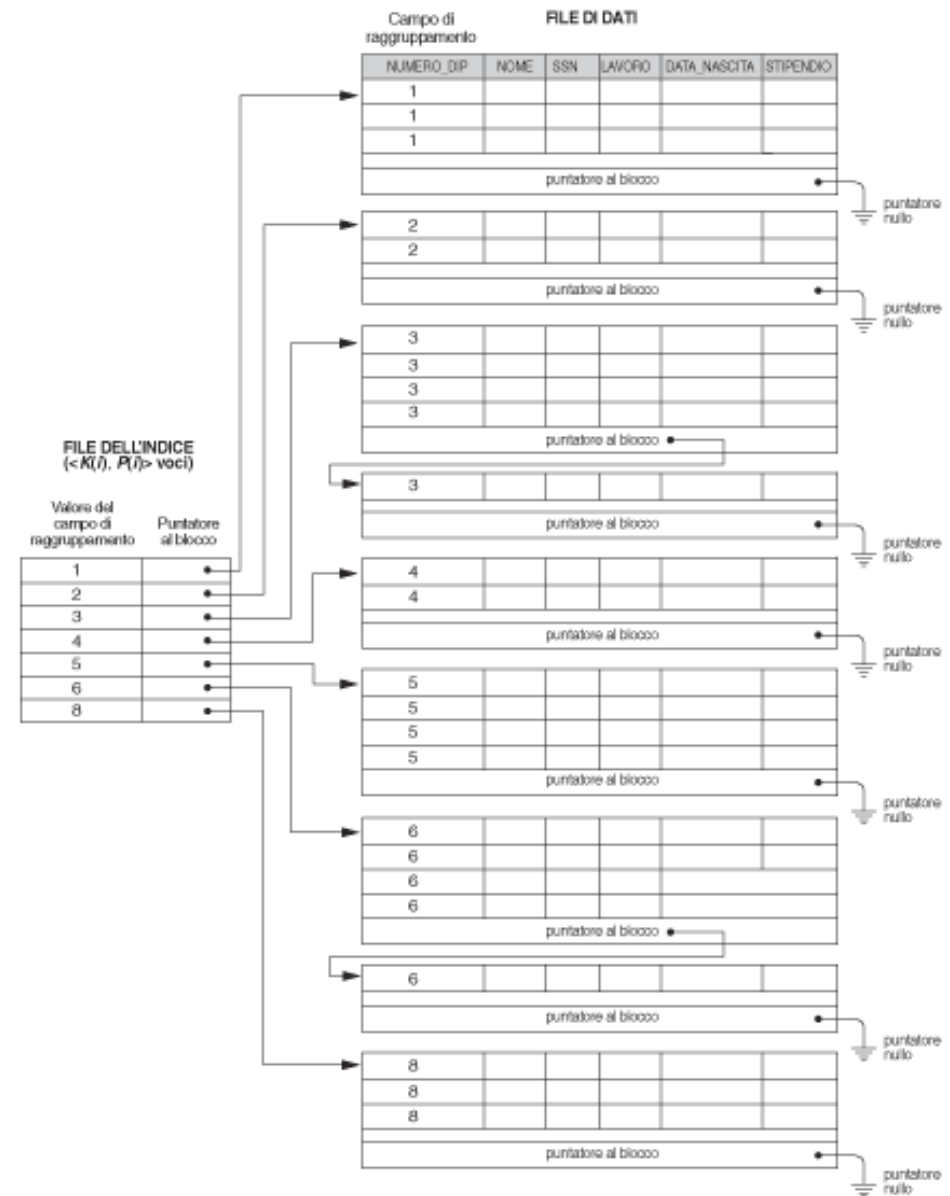
INDICE di CLUSTERING

- Si applica nel caso di file dati ordinato su un campo che non è UNIQUE/chiave primaria
 - Campo di clustering o campo di raggruppamento
 - File clustered
- Utile per velocizzare ricerca in caso di tanti record con lo stesso valore di raggruppamento
- Record nell'indice di clustering:
 - Valore del campo di raggruppamento
 - Puntatore al primo blocco che contiene il primo record con quel valore
- Numero voci nel file indice pari ai possibili valori di raggruppamento
- Indice SPARSO
- **Problema:** inserimento/cancellazione richiede riordino dei record
- Soluzione: un blocco differente o gruppi di blocchi per ogni valore di raggruppamento



INDICE di CLUSTERING

- Si applica nel caso di file dati ordinato su un campo che non è UNIQUE/chiave primaria
 - Campo di clustering o campo di raggruppamento
 - File clustered
- Utile per velocizzare ricerca in caso di tanti record con lo stesso valore di raggruppamento
- Record nell'indice di clustering:
 - Valore del campo di raggruppamento
 - Puntatore al primo blocco che contiene il primo record con quel valore
- Numero voci nel file indice pari ai possibili valori di raggruppamento
- Indice SPARSO
- **Problema:** inserimento/cancellazione richiede riordino dei record
- Soluzione: un blocco differente o gruppi di blocchi per ogni valore di raggruppamento



INDICE SECONDARIO

- L'indice secondario fornisce una struttura d'accesso indipendentemente dalla struttura fisica
 - Il file di dati può essere ordinato, non ordinato, hash
- E' possibile realizzare per il medesimo file dati diversi indici secondari
- L'indice secondario si può creare su un campo del record che:
 - è chiave primaria/UNIQUE
 - non è chiave primaria/UNIQUE

INDICE SECONDARIO su chiave/UNIQUE

- I record nel file NON sono fisicamente ordinati rispetto ai valori del campo dell'indice secondario,
 - non è possibile riferire solo al blocco che contiene il primo record

Data file

Indexing field
(secondary
key field)

	9				
	5				
	13				
	8				

	6				
	15				
	3				
	17				

	21				
	11				
	16				
	2				

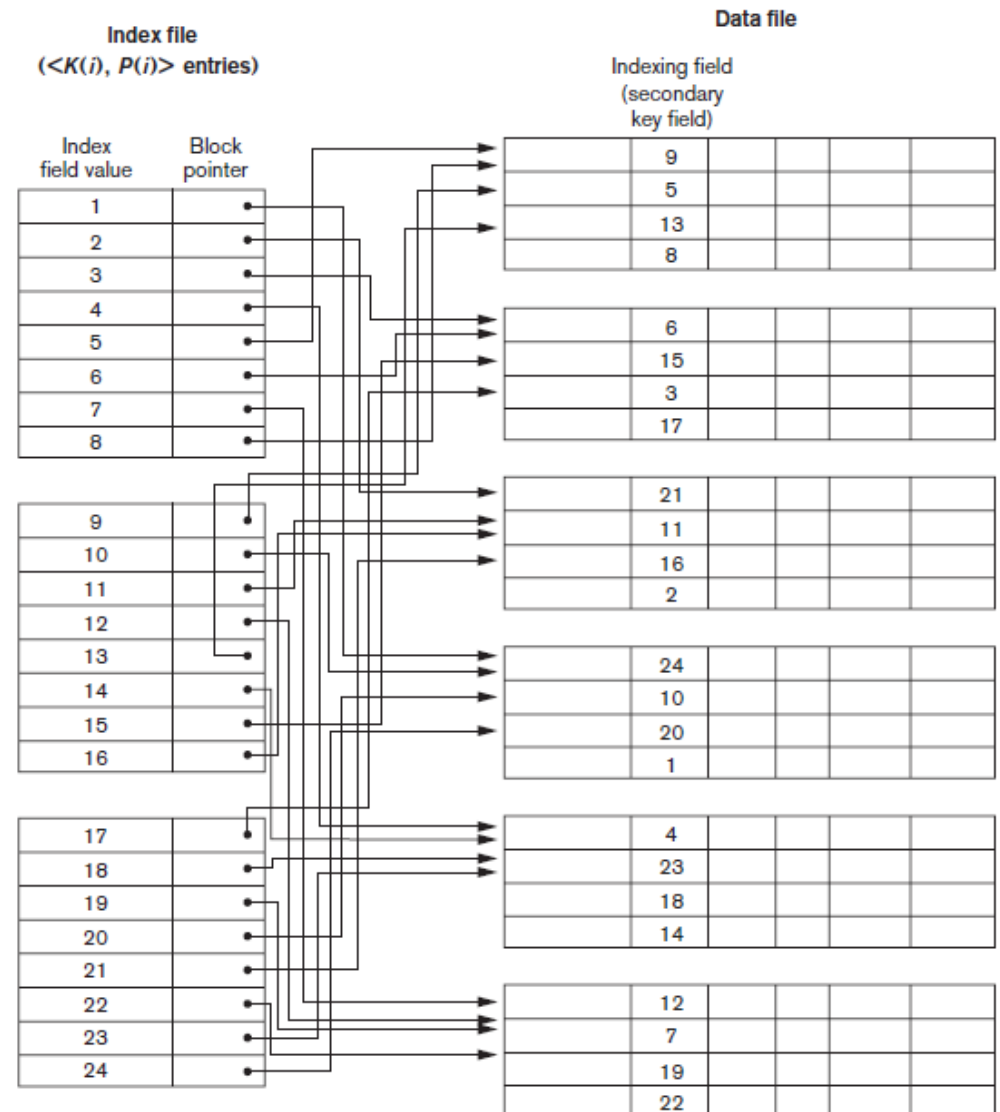
	24				
	10				
	20				
	1				

	4				
	23				
	18				
	14				

	12				
	7				
	19				
	22				

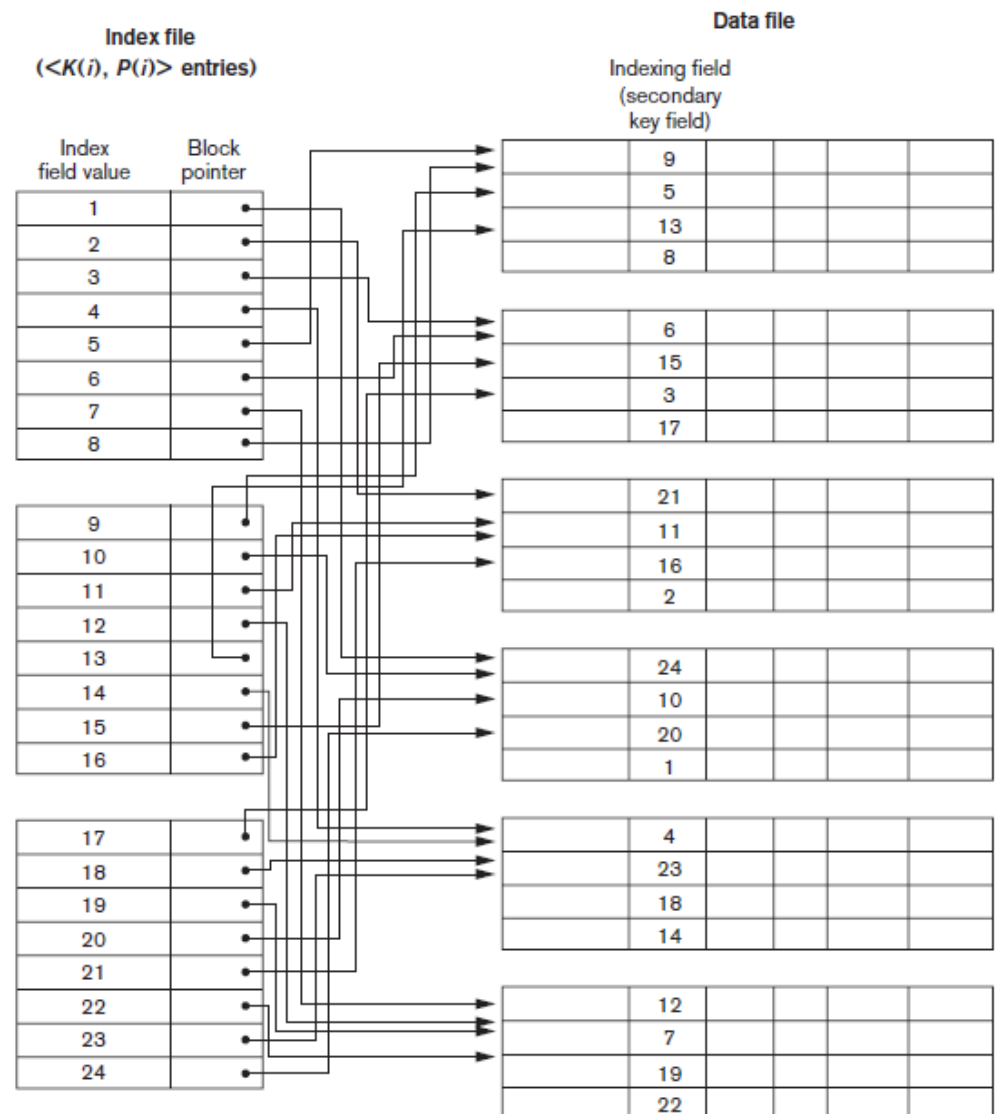
INDICE SECONDARIO su chiave/UNIQUE

- I record nel file NON sono fisicamente ordinati rispetto ai valori del campo dell'indice secondario,
 - non è possibile riferire solo al blocco che contiene il primo record
- Nel file indice serve una voce per ogni possibile valore del campo dell'indice secondario
 - Indice DENSO



INDICE SECONDARIO su chiave/UNIQUE

- I record nel file NON sono fisicamente ordinati rispetto ai valori del campo dell'indice secondario,
 - non è possibile riferire solo al blocco che contiene il primo record
- Nel file indice serve una voce per ogni possibile valore del campo dell'indice secondario
 - Indice DENSO
- Un indice secondario ha bisogno di più spazio di memorizzazione e richiede un maggior tempo di ricerca rispetto all'indice primario



Esercizio: Costo ricerca in termini di numero accessi

- Si supponga un file con $r=300.000$ record memorizzati su disco con blocco $B=4096$ di lunghezza. I record hanno lunghezza fissa $R=100$ byte e sono memorizzati in modo unspanned.
- Si assuma un indice secondario su campo V dove:
 - Dimensione V è 9 byte
 - Dimensione puntatore è 6 byte

Mediamente, quanti accessi sono richiesti per una ricerca sul campo V ?

- Fattore di blocco per il file dati $\lfloor 4096/100 \rfloor = 40$
- Numero blocchi per memorizzare i record: $\lceil 300.000/40 \rceil = 7500$
- Fattore di blocco per il file indice: $\lfloor 4096/15 \rfloor = 273$
- Nell'indice secondario, una voce per record, quindi
 - Numero di blocchi per memorizzare indice: $\lceil 300.000/273 \rceil = 1099$
- Numero accesso ai blocchi file indice per ricerca binaria $\lceil \log_2 1099 \rceil = 11$
- N. Accessi per ricercare record $11+1$

INDICE SECONDARIO

su campo NON chiave

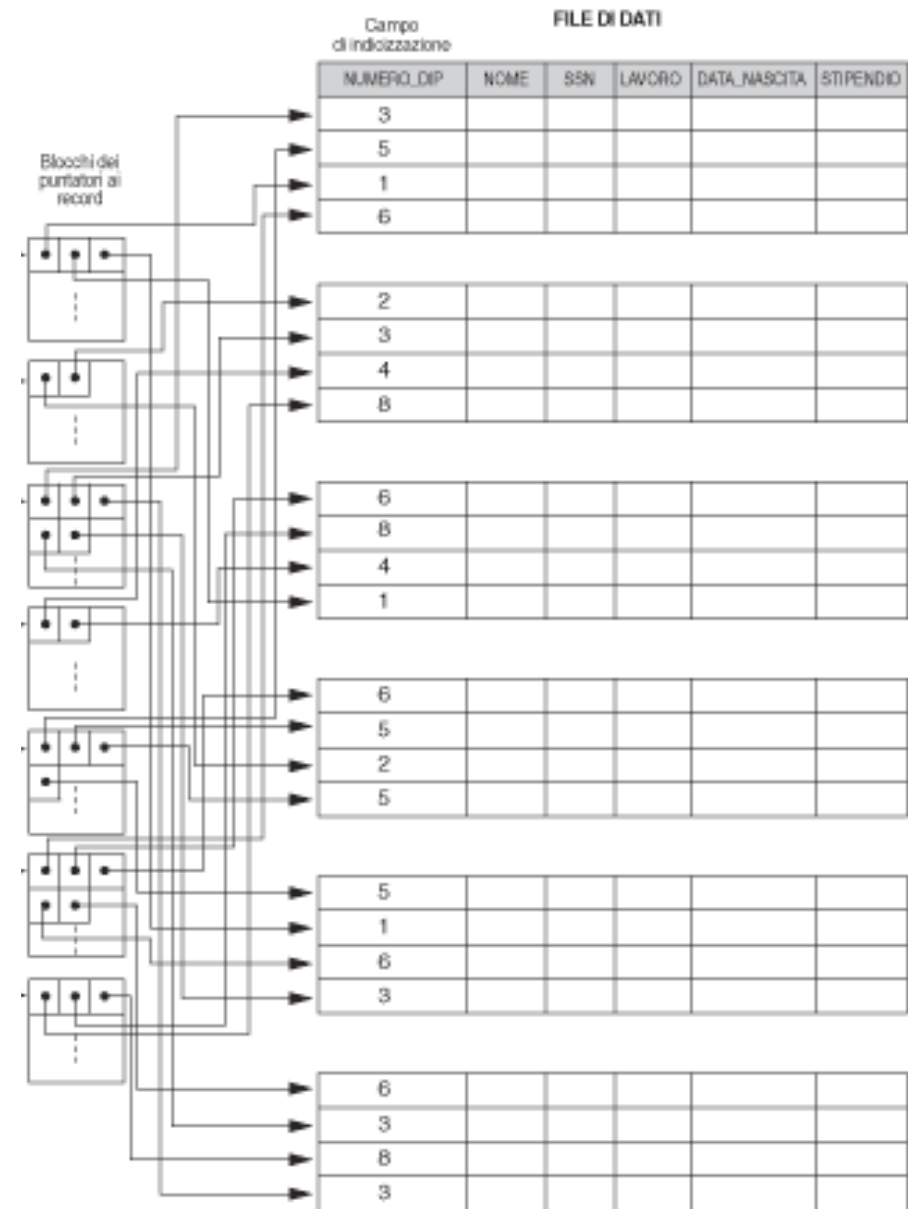
- Numerosi record del file possono avere lo stesso valore del campo indicizzazione
- Varie soluzioni per file indice
 - Inserire più voci con stesso valore del campo di indicizzazione
 - Usare record di lunghezza variabile per ospitare tutti i puntatori ai record

Campo di indicizzazione	FILE DI DATI				
NUMERO_DIP	NOME	SSN	LAVORO	DATA_NASCITA	STIPENDIO
3					
5					
1					
6					
2					
3					
4					
8					
6					
8					
4					
1					
6					
5					
2					
5					
5					
1					
6					
3					
6					
3					
8					
3					

INDICE SECONDARIO

su campo NON chiave

- Numerosi record del file possono avere lo stesso valore del campo indicizzazione
- Varie soluzioni per file indice
 - Inserire più voci con stesso valore del campo di indicizzazione
 - Usare record di lunghezza variabile per ospitare tutti i puntatori ai record
 - Creare un ulteriore livello per gestire i puntatori multipli. Una voce per valore con puntatore che punta al blocco dei puntatori



INDICE SECONDARIO su campo NON chiave

- Numerosi record del file possono avere lo stesso valore del campo indicizzazione
- Varie soluzioni per file indice
 - Inserire più voci con stesso valore del campo di indicizzazione
 - Usare record di lunghezza variabile per ospitare tutti i puntatori ai record
 - Creare un ulteriore livello per gestire i puntatori multipli. Una voce per valore con puntatore che punta al blocco dei puntatori

