Overview of Distributed Database Indexing Techniques

R. Pavithra
Department of Computer science
KMCPGS
Puducherry, India

R. Kalaivani
Department of Computer Science
KMCPGS
Puducherry, India

***Abstract*— This paper studies the concept of indexing techniques in Distributed Database Management System (DDBMS). This paper starts with small introduction about Distributed Database, its indexing concepts and the types of indexing. The main objective of this paper to find the new concept in indexing process. Finally, this paper concludes with the new idea in indexing.**

Keywords*— *Distributed Database Architecture, Indexing Techniques, StarSparse Indexing.

I. INTRODUCTION

A. Distributed Database Management System

Distributed Database is a database in which all individual devices having separate Central Processing Unit (CPU) and connected with Centralized computer network equipment. This overall system controlled by Distributed Database Management System (DDBMS) also called Distributed Database System. It may be stored in multiple computers, located in the same physical location. Unlike parallel computing systems, in which the processors are closely coupled and constitute a single database system, a distributed database system consists of loosely coupled sites that share no physical components. If updating is made in one database should automatically replicated in other databases called consistency. Consistency plays the major role in Distributed Database. Distributed Database has two processes to involve that the distributed databases for update. They are replication and duplication. Both replication and duplication should have the data in all distributive systems.

B. Characteristics of Distributed Database

i) Each individual system has its own memory.
ii) Parallel processing is possible in the distributed database.
iii) It should always maintain consistency over the information stored in database.
iv) It should increase reliability and availability.
v) In distributed database Distribution transparency maintained.

C. Architecture of Distributed Database :

The below figure 1 explains the architecture of distributed database. It consists of individual nodes called systems, and each has separate database connected through computer network.

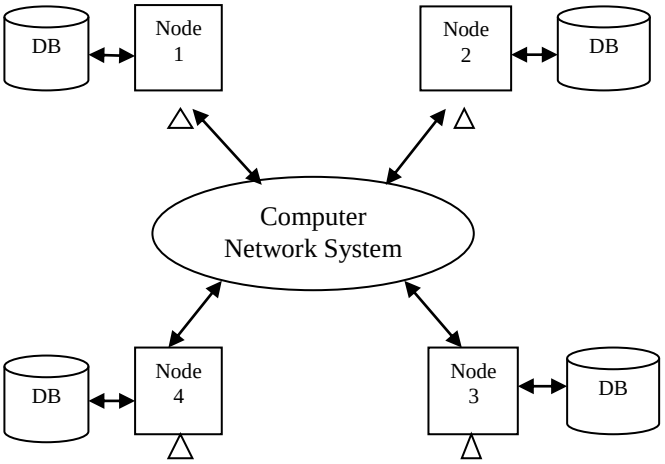


Fig 1: Distributed Database Architecture

II. INDEXING TECHNIQUES

A. Definition of Indexing

Indexing is a data structure technic to efficiently retrieve records from the database files based on some attributes on which the indexing has been done. Example, index page presents in a text book.

B. Three major categories of indexing

Primary index: Primary index is defined on an ordered data file. The data file is ordered on a key field. It is generally the primary key of the relation.

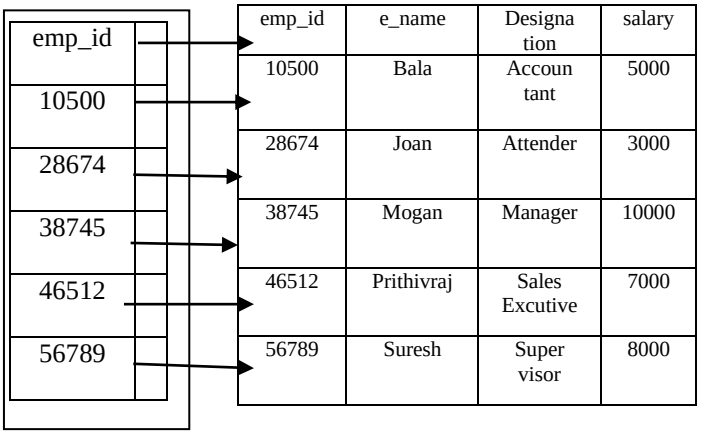
Secondary index: Secondary index may be generated from a field which is a candidate key and has a unique value in every record, or a non-key with duplicate values.

Clustering index: Clustering index is defined on an ordered data file the data file is ordered on a non-key field.

C. Classificiation of Primary Index

i) **Bit Map Index:** Bit map index is a special kind of index that stores the bulk of its data as bit arrays and answers most queries by performing bit wise logical operation on this bit maps.
ii) **Reverse Index:** Reverse key index reverse the key value before entering it the index. Example: the value 47853 becomes 35874 in the index. Reversing the key value is particularly useful for indexing data such as sequence numbers where new key values monotonically increase.

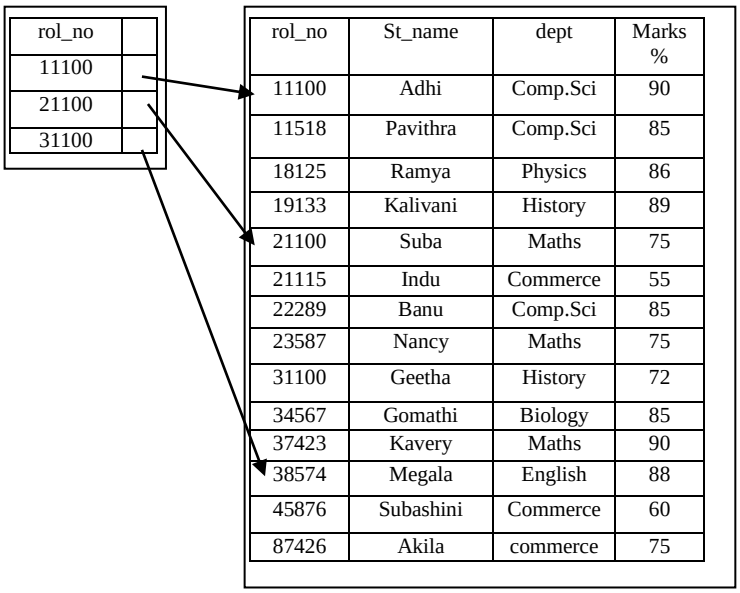
iii) **Dense Index:** A dense index in databases is a file with pairs of keys and pointers for every record in the data file. Every key in this file is associated with a particular pointer to a record in the sorted data file.
The below figure 2 explains the dense index, the employee details relations using emp_id as primary key.



emp_id	e_name	Designation	salary
10500	Bala	Accountant	5000
28674	Joan	Attender	3000
38745	Mogan	Manager	10000
46512	Prithivraj	Sales Excutive	7000
56789	Suresh	Supervisor	8000

Fig 2: employee details relation

iv) **Sparse index:** Sparse index in database is a file with pairs of keys and pointers for every block in the data file. Every key in the file is associated with a particular pointer to the block in the sorted data file. In clustered indices with the duplicate keys the sparse index points to the lowest search key in a search in each block. Below the figure 3 illustrate the sparse index search process.



rol_no	St_name	dept	Marks %
11100	Adhi	Comp.Sci	90
11518	Pavithra	Comp.Sci	85
18125	Ramya	Physics	86
19133	Kalivani	History	89
21100	Suba	Maths	75
21115	Indu	Commerce	55
22289	Banu	Comp.Sci	85
23587	Nancy	Maths	75
31100	Geetha	History	72
34567	Gomathi	Biology	85
37423	Kavery	Maths	90
38574	Megala	English	88
45876	Subashini	Commerce	60
87426	Akila	commerce	75

Figure 3: Sparse Index Search Process

The above diagram student details relation use the sparse index to search the records within particular range for example the range rol_no 11100 to 21100 and final range between 31100. It does not search the records above the range. For example rol_no 45876 and 87426. Each number range is considered a Block so each Block as its records only. The below figure 4 is illustrate blocks of sparse index.

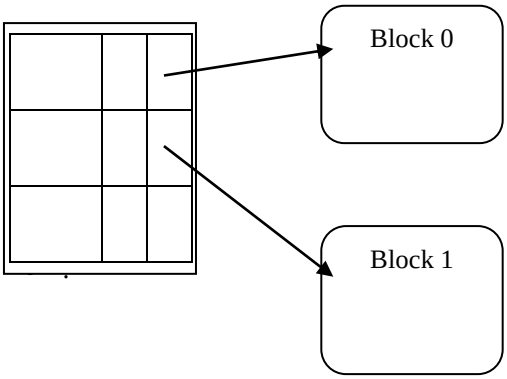


Fig 4: Sparse index blocks

v) **Multilevel Index:** Multilevel index first breaks the entire index into two again the first patitions is devided into two.If the database is grows automatically the number of indices also grows. Multilevel indexing processed like linear search process. The below figure 5 illustrate the structure of multilevel indexing.

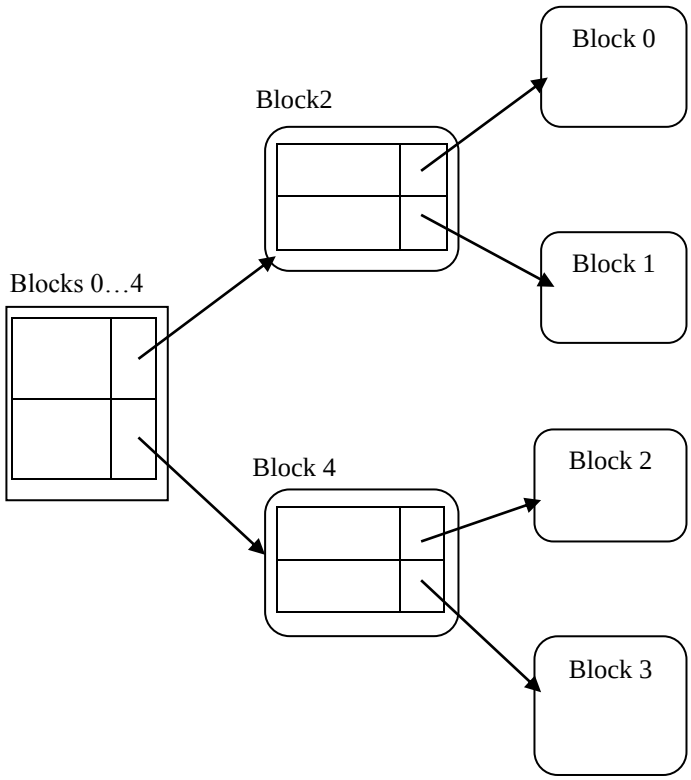


Fig 5: Multilevel Indexing structure

D. Disadvantages of sparse and multilevel indexing

1. The major disadvantage of multilevel indexing is it requires more storage space than other techniques.
2. When the database is large the number of indices required also increase.
3. In sparse indexing it checks the blocks in sequential order when the number is out of its range finding of record is becomes difficult process.
4. The time spent to search the particular range also become waste.

So, to overcome the disadvantage of sparse indexing this paper produce the new concept called StarSparse indexing. Star Sparce indexing process also include sparse Index with some added features.

III. STARSPARSE INDEX

A. INTRODUCTION

In this star parse indexing, flag is used when the process of indexing started , if the flag is set to 1 or true that particular block only gets permission to further indexing process. All the blocks having the records between some ranges this range is considered as maximum value and minimum value. The search records are between the maximum and minimum value mean the indexing process successfully completed. In other words the flag set to 1.Oherwise the flags set to 0 mean the search records are not present that maximum and minimum value

B. STUCTURE OF STARSPARSE INDEX

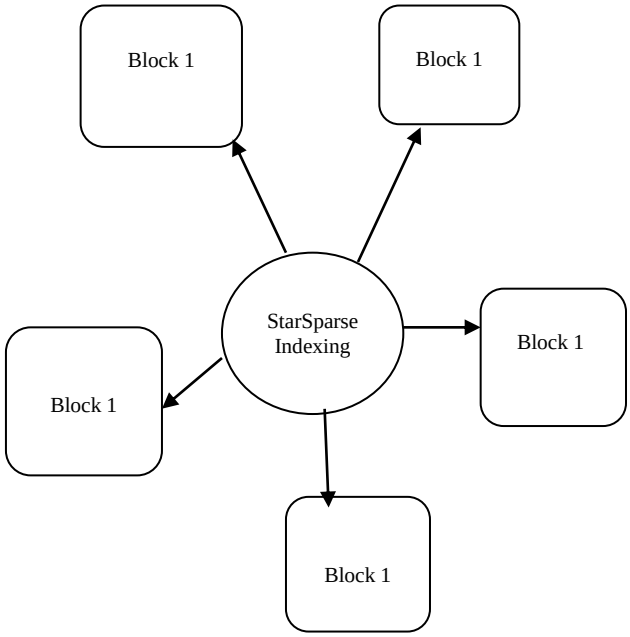


Fig 6: StarSparse index structure

C. ALGORITHM FOR STARSPARSE INDEX

The algorithm of StarSparse indexing consists of the variables such as search records , flag.

```
StarSparse( sr, b[],flag)

{
sr= searchrecord;
/ search record assigned by the user.

  b[n]= {block1,block2,block3…}
    /block are assigned to array b.n is assigned to
  number of
    / blocks.
  for(i=0;b[i]<=n;i++)
  {
   If (sr<max && sr>min) then
    print("block i consist the search record")
     flag=1;
  else
  print ("block i does not consist the search record")
  }
   return block;
  }
  End;
```

d. EXAMPLE OF STARSPARSE INDEX

This example contains the search record as number 28579,

there are 7 blocks and maximum minimum value as 10000 to

30000.

a) Steps followed in StarSparse Index

1) Assign the variable to sr=28579,max=30000 and min=10000,n=7.
2) Send the max ,min values to all the 7 blocks.
3) Each block check there header with max and min range.
4) If any block header match with the max and min value then, that particular block flag is set to 1.
5) Here the block 1 header has matched with the max and min values.
6) So, block1's flag is 1.
7) The further search process is take place in block 1 only not remaining blocks.

b) Flowchart of the above example:
The figure 7 illustrates the flowchart of above example.

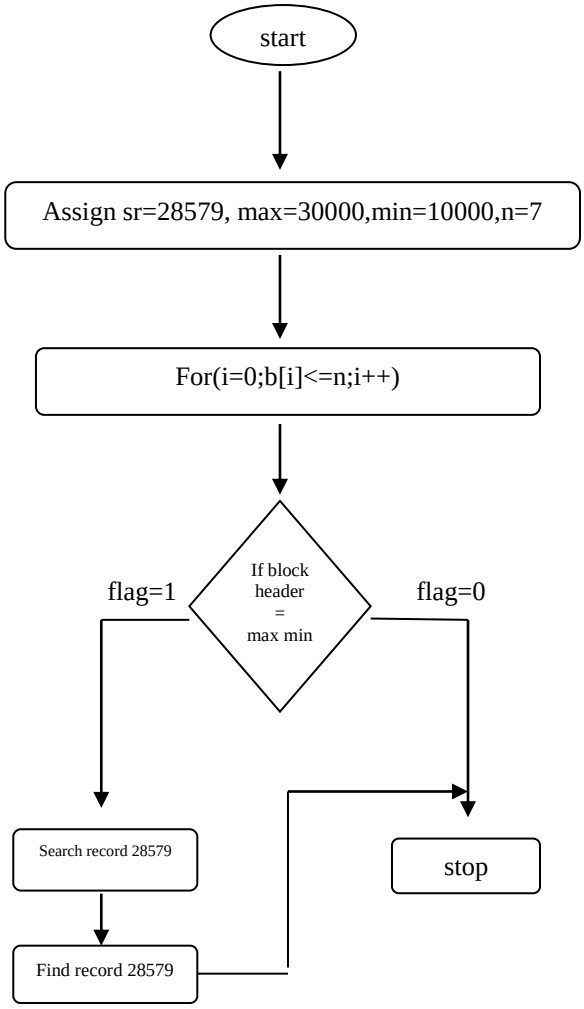


Fig 7: flow chart of example

E . ADVANTAGES OF STARSPARSE INDEXING

i) The traditional sparse indexing checks all the blocks and all the records but the StarSparse index check only one block.
ii) The time required for searching process is less than sparse index.
iii) The best case search time required the number records present in i th block.
iv) It requires same data structure used in sparse indexing.

CONCLUSION

This paper tried to give a new concept about indexing**.** StarSparse indexing is a useful concept for indexing process using this finding a particular record easily. It contains only few additional features of sparse indexing. So, it can be easy to implement. It effectively reduces the amount of time requires to overall indexing process.

REFERENCES:

[1] Cyrus Shahabi, Latifur Khan, Dennis Mcleod. A probe based techniques to optimize join queries in distributed internet bases, knowledge and information Systems.2000, 2.
[2] Syam Menon. Allocating fragments in distributed databases[J]. IEEE transactions on parallel and distributed Systems, 2005 16: 5577-585.
[3] TSAIPSM CHINALP, Optimizing queries with foreign function in a distributed environment, IEEE trans on knowledge and Data Eng. 2002, 14(2):809-824.
[4] LEEE Chiang, CHIH Chi-Sheng, CHEN yaw-huei. Optimizing large join queries using a graph –based approach, IEEE Trans on Knowledge and Data Eng.2001, 13(20:298-315
[5] S. Pramanik, D. Vineyard. Optimization Join Queries in Distribute Database. IEEE TSE. 2004.14(9):19:1326
[6] Peng- Yeng Yin, Shiuh_Sheng Yu, Pei-Pei Wang and Yi-Te Wang. A hybrid particle swarm optimization algorithm for optimal task assignment in distributed systems[J].Computer Standards & Interfaces, Volume 28, issue 4, April 2006,444-450.
[7] D. Kossamann, K. Stocker. Iterative Dynamic Programming: A New Class of Query Optimization Algorithms [J]. ACM Transactions on Database Systems, 2000, 25(1):43-82.
[8] Prof. (Dr.) Anand K. Tripathi and Mrs. Monica Tripathi: "A Frame: Why Information System Fails", Proceeding of the National Conference on "Managing complexity in Information World," publication Bhilai Institute of Technology, Durg,Bhilai(CG) dated 7-8 Nov- 2008. Edition-1,vlo-1,pf 112-115.
[9] Prof. (Dr) Anadn K.Tripathi and Mrs. Monica Tripathi AICTE Sponsored, National Conference on "Next generation Computing & iInformation systems" at Model Institute of Engineering & Technilogy, Jammu Feb.14-15,2009.edition2,vol-2,pg.724-731
[10] [Balt82] Balter. R.,Beard, P., and Decitre. P., "why Control of the concurrency Level in Distributed Systems is More Fundamental than Deadlock Management," Proc . 1 st ACM SIGACT-SIGOPS Symp.on Principles of Dist.Comp., Aug 1982.
[11] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol-1., Oxford: Clarendom, 1892, pp.68-73.
[12] I. S Jacobs and C.P Bean, "Fine Particuls , thin films and exchange anisotroy,"in Magnetism vol. III G.T Rado and H. Suhl, Eds. New York: Academic 1963,pp 271-350
[13] K. Elissa, "Title of Paper if Known," unpublished.