23/10/21

ADBMS
TT-1

JUNAID. GIRKAR
G00U4190057
TE COMPS A-4

Q1

ANS PROPERTIES OF B-TREE :-

- A B-tree is defined by the term minimum degree 't'. The value of t depends upon disk block size
- Time complexity to search, insert and delete is o(log n)
- Insertion of a node happens only at leaf node
- All nodes may contain at most 2*t - 1 keys

PROPERTIES OF B+ TREE :

- All leaves are at the same level
- Each node can contain a maximum of m-1 keys and a minimum of m/2 - 1 keys
- It is a balanced tree in which every path from the root to a leaf is of the same length.
- The root node has atleast two children.

JAGirkar

G0000190057

Q2

ANS Amazon Prime Video uses Amazon Kinesis Data Streams (KDS) which is a massivly scalable and durable real-time data streaming service. KDS can continuosly capture gigabytes of data per second from hundreds and thousands of sources such as website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events. The data collected is available in milliseconds to enable real-time analysis use cases such as real-time dashboards, real time anomaly detection, dynamic pricing and more.

Characteristics of a Media Database :

- Affinity to structured data which results in an efficient overall system.
- Efficient media timeline modelling which helps to serve various types of media timeline data ranging from periodic sample oriented ones to event based ones is a fundamental trait of a media database.

* Here in this example, the release of the movie 'SARDAR UDHAM' was an event based-date controlled release on the 16th of October 2021.

- Spatio-temporal query ability where a media database makes it easy to check if a contiguous sequence of video frames contains text in a specific spatial region. Such a query could come in handy for detecting collisions between text present in video and subtitles

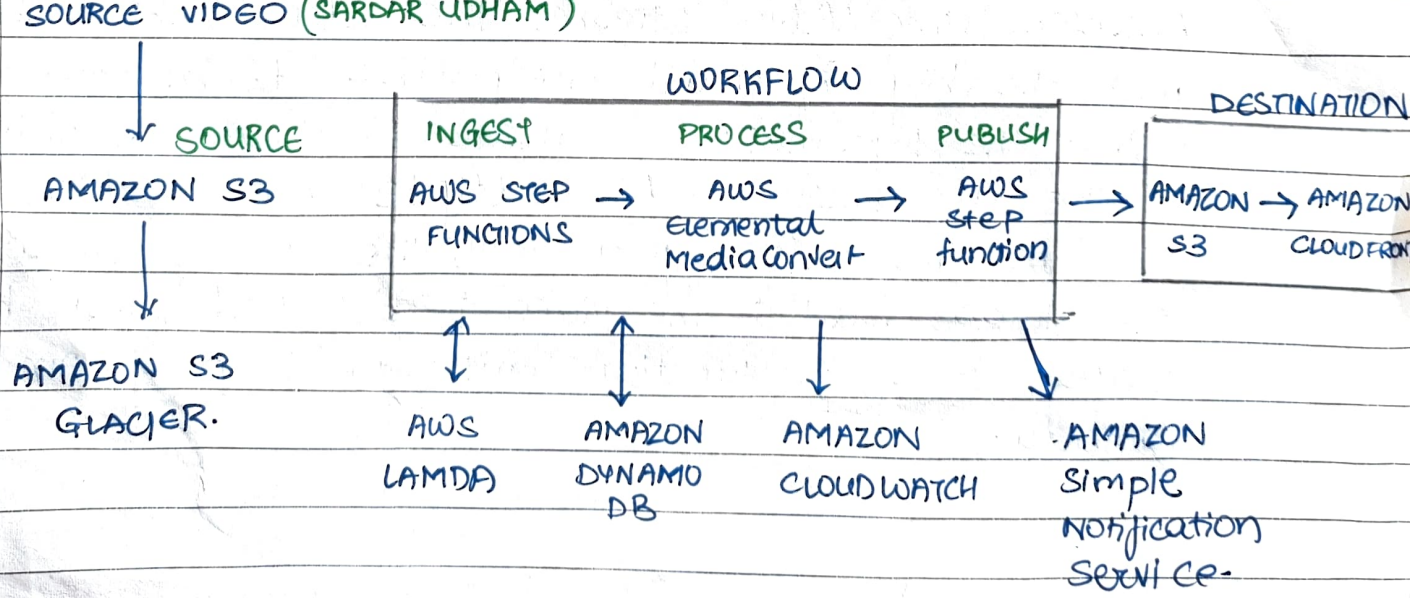
JAGUKAY

6000119005?

- Multi-tenacy - where it allows storing arbitrary data (db) provided it is structured and ~~ordered~~; data can also be associated to a particular time interval of the media resource, each tenant can then benefit from the efficient query-ability of our system.
  * Here in our example, this optimizes our query seach of the title "SARDAR UDHAM"

- Scalability. This means that the system has to adress issues related to availability and consistency under various load scenarios
  * In our example "SARDAR UDHAM" is currently trending at position number 1 in India and is therefore catering to a very high number of viewers and should scale up to that level so that the site doesn't crash.

ARCHITECTURE -



B0004190057

AMAZON S3 GLACIER is an online file storage web service that provides storage for data archiving and backup.

Amanon S3 is an object storage service through web service interface.

AWS elemental MediaConvert is a file-based video transcoding service with broadcast-grade features It allows you to easily create video-on-demand content for broadcast and multiscreen delivery at scale

AWS step function pushes requests to other services so that they can perform actions for your workflow, waits for the service task to complete, and then continues to the next step. An AWS step functions state machine can contain combinations of activity tasks and service tasks

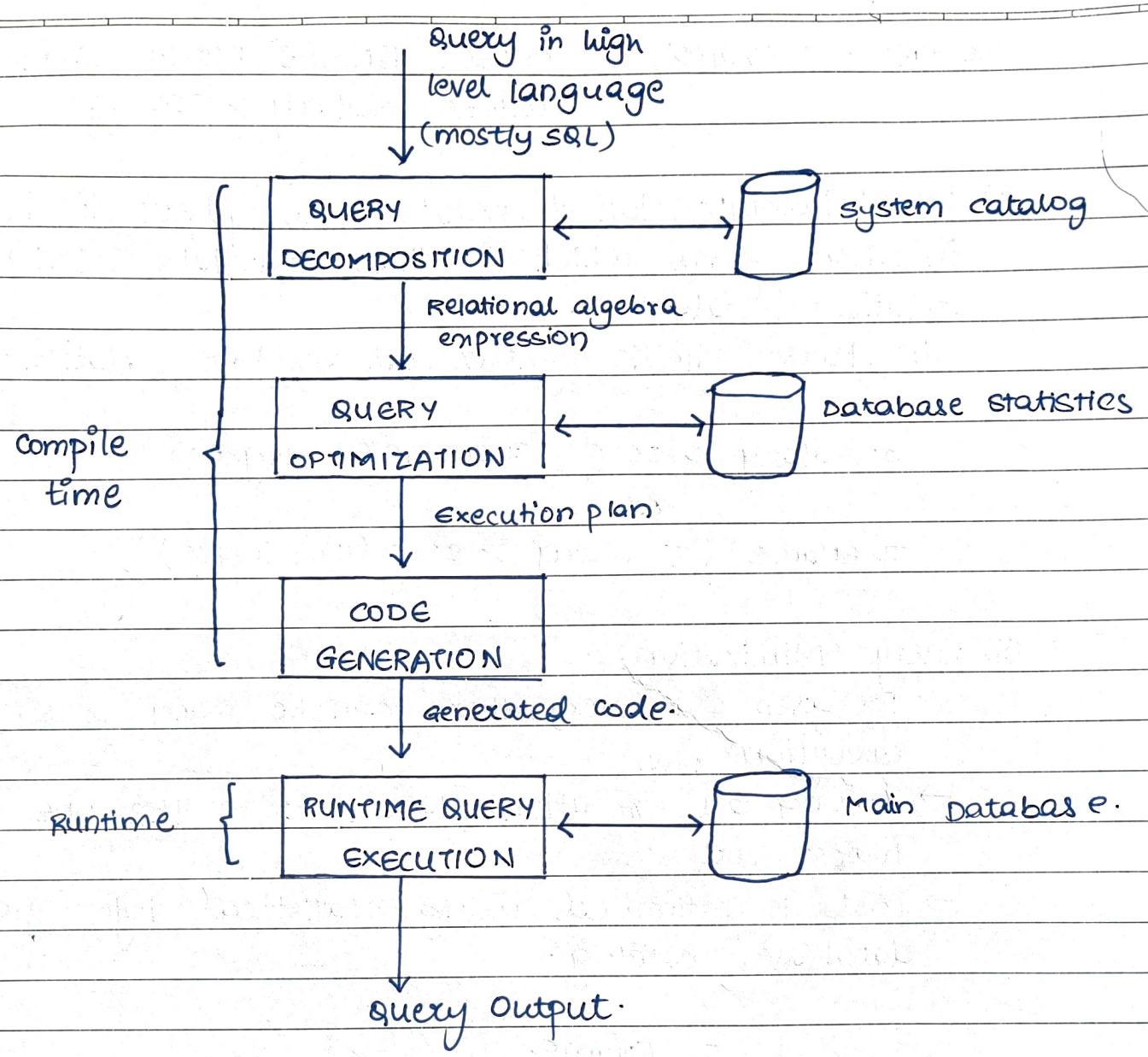
Amazon Prime Video streams the video ("SARDAR UDHAM") with the help of cloud computing technology using the DynamoDB database to show the videos to the user. It used multiple AWS service to stream the video, first with the use of amazon S3 the storage bucket the video is selected and has been scaled to step function with the help of Amazon lambda.

JAGadal

G0004190057.

Q3

ANS



QUERY PROCESSING IN DBMS.

Steps for Query Processing

(a) Query Decomposition
(b) Query optimization
(c) Code generation
(d) Runtime query execution

we take the example of SELECT EName FROM Employee WHERE Salary > 5000 ;

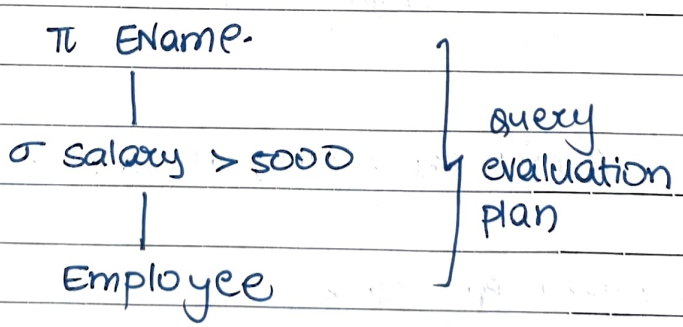
(i) Query Decomposition translates the query to its internal form which is than translated into relational algebra.
The Parser checks syntax and verifier relation.

σ Salary > 5000 (π EName (Employee))

or

π EName (σ Salary > 5000 (Employee))

(ii) Query optimization:
→ For given SQL query, there can be many possible execution plan.
→ Among all execution plans, choose the one with lowest cost
→ Cost is estimated using statistical info from database catelog.



iii) Code generation : The query evaluation indem takes a query evaluation plan, executes the plan and returns the generated code to the query.
iv) The generated code is executed in runtime environment

JAGaska

7

6000U190057

Q4

ANS The cost estimation of a query evaluation plan is calculated in terms of various resources that include :-

- Number of disk accesses
- Execution time taken by the CPU to execute a query
- Communication costs in a distributed or parallel database system.

To estimate the cost of a query evaluation plan, we use the number of blocks transferred from the disk, and the number of disk seeks. suppose the disk has an average block access time of t_s seconds and takes an average of t_T seconds to transfer x data blocks. The block access time is the sum of disk seeks time and rotational latency. If it peyorms 's' seeks then the time taken will be $b * t_T + s * t_s$ seconds.

The response time, i.e, the time required to execute the plan, could be used for estimating the cost of the query evaluation plan but due to the following reasons, it becomes difficult to calculate the response time without actually executing the query evaluation plan:

- when the query begins its execution, the response time becomes dependent on the contents stored in the buffer. But this information is difficult to retrieve when the query is in optimized mode, or it is not available also
- when a system with multiple disks is present, the response time depends on an interrogation that in

Sundaram

JAGIKAY

"what way accesses are distributed among the disks?". It is difficult to estimate without having detailed knowledge of the data layout present over the disk.

- consequently, instead of minimizing the response time for a query evaluation plan, the optimizers finds it better to reduce the total resource consumption of the query plan. Thus to estimate the cost of a query evaluation plan, it is good to minimize the resources used for accessing the disk or the extra resources

B0004190057

Q5

ANS A2 algorithm is a search algorithm in primary B+ tree indexes, searching by using equality on keys.

Assuming that the bo blocks of a relation are stored contiguosly. The cost of this operation is determined by the expression : $(hi + 1) \times (t_T + t_S)$

where hi denotes the height of index.

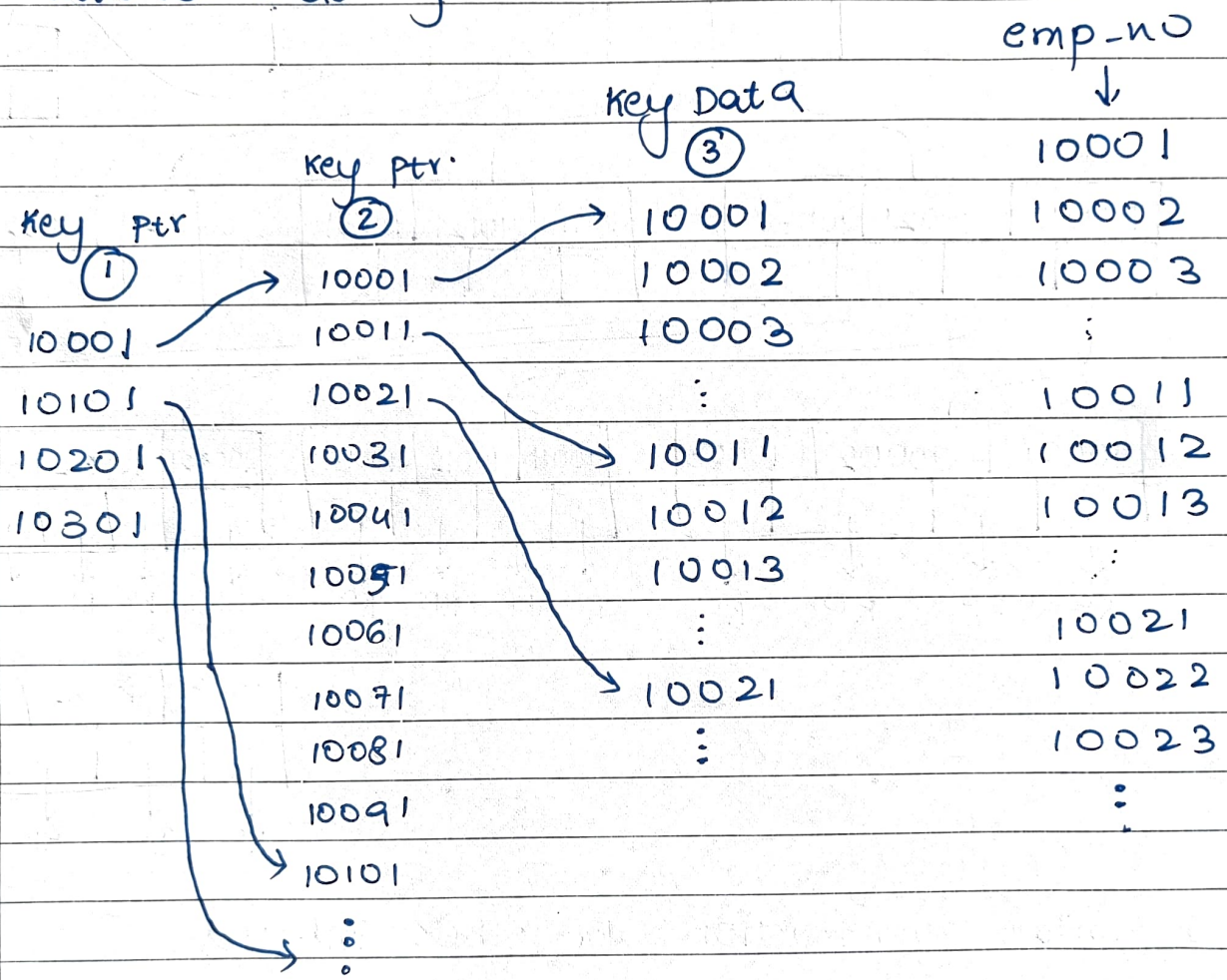
Index lookup traverses the height of the tree plus one I/O to fetch the record ; each of these I/O operations require a seek and a block transfer.

JAGIKAS

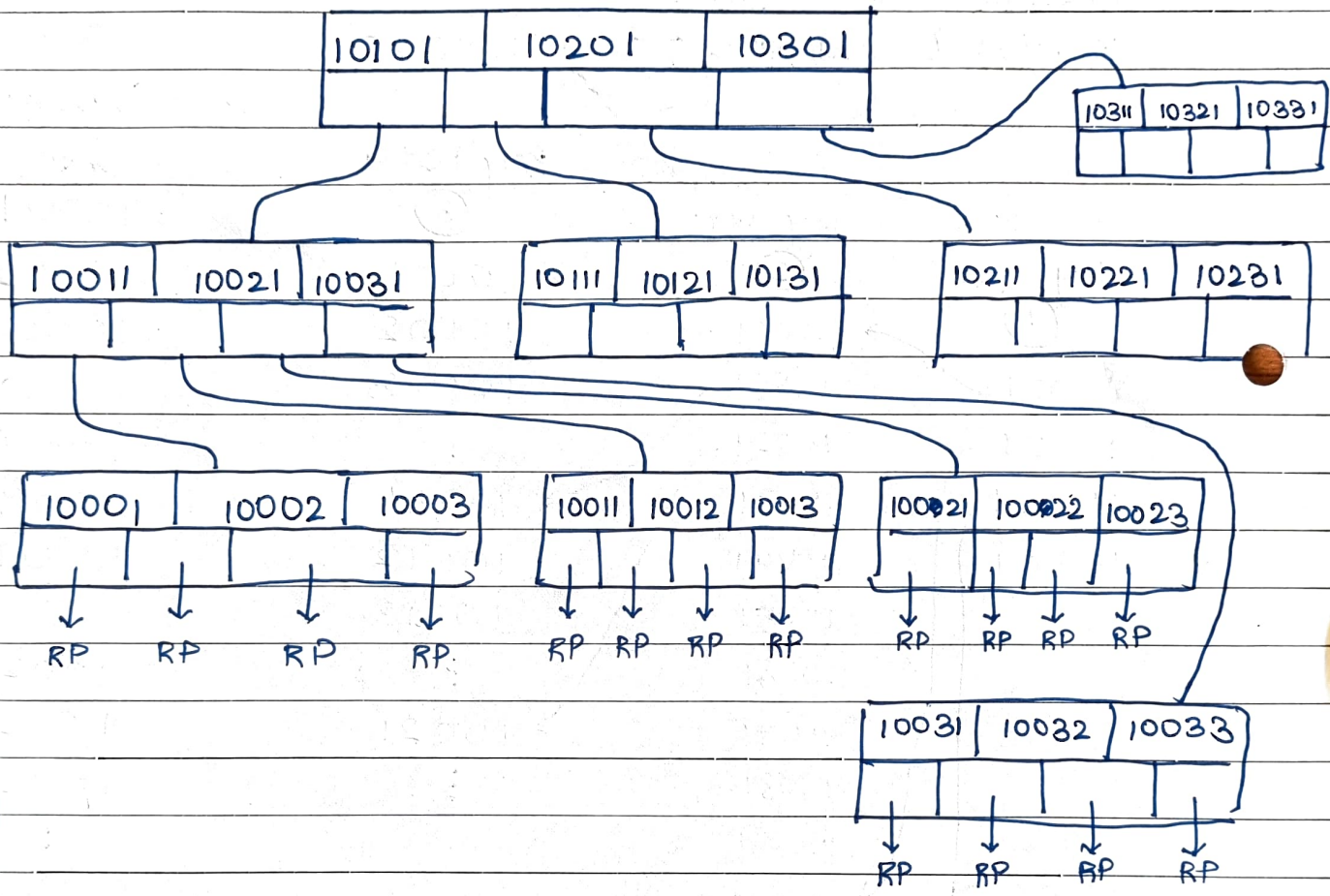
Q6 Database used for Experiment 2 : Employee database

Indexing on emp-no

Multilevel Indexing — consider :-



Converting to B+ tree :



Similar construction for other indexes

EXAMPLE : Select * from employee where emp_no = 10021

The first index table will show that the emp_no lies in the range 10001 to 10100 and hence will point to location of block containing 10001 in 2nd index