Abhisth Chatterji

500084948

R2142201733

B-3 6th Semester

BTech CSE (CCVT) Non-Hons

Week 1: Idea Framework and Flowchart

Cloud Application Development

Submitted to: Saurabh Shanu Sir

National DNA Database

A DNA database, also known as **a DNA databank,** is a collection of DNA profiles that can be used for **genetic disease study, criminal justice applications of genetic fingerprinting, and genetic genealogy.**

Use Case:

In many cases where a dead body is in a bad condition and is not identified the **case gets closed in 72 hours**. If we have a national DNA database then we can identify the dead body by matching his DNA with his existing DNA profile collected and stored in the database earlier. **Once the victim is identified then it will be easy to get hints about the accused.**

Advantages of a DNA database:

1. It can add another line of support to the evidence.

An individual's DNA can be used to prove their presence at the site of a crime when there are no eyewitnesses. Researchers can check acquired samples against historical data using a DNA database to see if any matches exist. The case against a suspect gets significantly stronger if there is a match. Because an accurate match can offer such high degrees of assurance, the existence of a DNA database aids in the prevention of crime.

2. Rates of crime decrease are possible.

DNA databases can aid in reducing crime in areas where repeat offenders exhibit criminal behaviours.

3. Studies on genetics can be made using the information.

There are already some genetic research projects that employ DNA databases. When there are near matches to previously recorded DNA profiles, they are also used to aid in locating genetic relations. Although this information might be sensitive, particularly if a law enforcement emphasis is placed on it, it can also be used to advance our understanding of what it means to be a person. We can fix

mistakes that these fundamental components of life may make for some people if we have a better understanding of our DNA.

Possible Outcomes:

We can use this database for research and medical needs in addition to criminal cases. **US, UK, Canada, Malaysia, South Africa and Australia, and around 70 more countries have DNA data banks.** This aids in the investigation of criminal cases and the quickening of justice. India will take part in it as well.

Challenges in National DNA Database:

1. Data breaches are possible.

A DNA database does not have to be accessible to the general public for its contents to be stolen. The versatility of hackers in gaining access to data when they want it has already been demonstrated numerous times. More than 123 million people's personal information was made available due to the Experian data breach. In certain cases, retail data breaches revealed the personal information of 50–70 million consumers.

2. The information might be utilized against the people it represents.

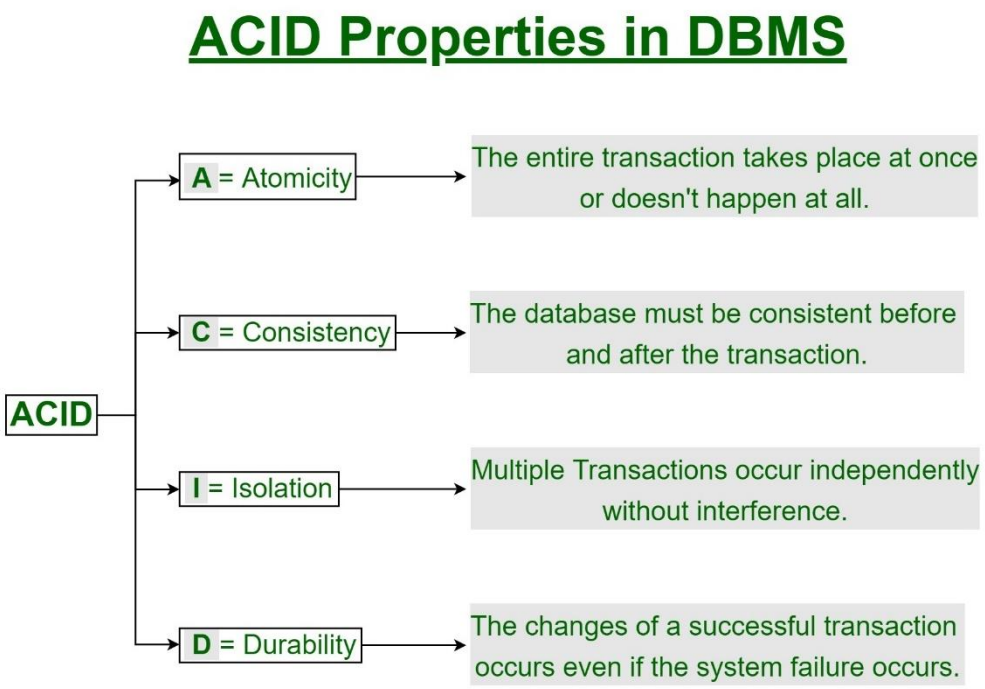
If a hacker gains access to a DNA database. They see that their neighbour who lives further up the street has a profile in the database. They go over to the neighbour’s house, dig through the garbage, and pull out a few things that probably have DNA samples in them. After gathering the evidence, they commit a crime and leave the scene. It would be challenging for these innocent neighbours to defend themselves because many people think DNA evidence is incontrovertible. When DNA databases are available, that may become a reality.

3. Information preservation practices may vary between countries.

2012 saw the deletion of more than 1.6 million fingerprint records in the United Kingdom. Additional 7.7 million DNA samples were destroyed, including 480,000 from children. Some nations prioritise preserving freedoms. Some people may not. There is no assurance that other nations will keep those records secure or obliterate them on demand if a DNA database is provided. A patchwork of database regulations created by different rules for data storage could put a person's genetic information in danger on a worldwide scale.

4. High availability and abiding ACID properties.

The Database should be available 24*7. The database should follow fundamental acid properties



Source: https://media.geeksforgeeks.org/wp-content/cdn-uploads/20191121102921/ACID-Properties.jpg

These limitations can be overcome by using Cloud Computing Services

I will be using **AWS** for developing National DNA Database. Reasons for choosing AWS over other cloud platforms are:

Free-tier Service Availability: It is advantageous for customers when a service is available that they can use for free for a set period to get accustomed to and better comprehend the service. Businesses can use AWS Free-tier to get a hands-on taste of the services without paying a penny. For an entire year with specific usage

restrictions, businesses can explore the services and assess whether their operations are acceptable for using AWS as the platform.

Model for On-Demand Pricing - Pay-as-you-go pricing from AWS also provides a discount on several services. Without any complicated licensing or long-term contracts, only pay for the specific services as they are required for the duration of the requirement.

Most Secure: Of all the cloud platforms, AWS is designed to be the most secure, dependable, and adaptable. The most sensitive data is protected by the core infrastructure. Services are provided to safeguard data, accounts, and workloads from illegal access, and the network activity and account behaviours of the cloud environment are continuously monitored.

Fast Pace Innovation - AWS has developed a new lineup of EC2 instances to ease client adoption to the Cloud in terms of pricing. Businesses may experiment and innovate swiftly and at a much faster rate when using the most recent technologies. AWS is continually developing new business-changing technology.

Widest Partner Network - AWS is the owner of the biggest and most vibrant community of partners and customers.

Data breaches issues can be handled with AWS

AWS continuously scans the landscape of privacy laws and regulations to spot changes and evaluate what solutions might be necessary for our customers to meet compliance requirements. Customer trust requires continual commitment. It makes an effort to keep us informed about the privacy and data security policies, procedures, and tools we've implemented. Among its commitments are:

Access: As a client, we are entirely in charge of setting up access to AWS services and resources as well as the content that you upload to the AWS services using your AWS account. To assist us in doing this successfully, we get a sophisticated collection of access, encryption, and logging tools (e.g., AWS Identity and Access Management, AWS Organizations and AWS CloudTrail).

Storage: We get to pick which AWS Region(s) your material is kept in. Our content can be duplicated and backed up in many AWS Regions.

Security: We get to decide how to protect our content. It provides us with the choice to maintain our encryption keys and equips us with industry-leading encryption tools to safeguard our content both in transit and at rest. These data protection attributes consist of:

There are data encryption features in more than 100 AWS services.

Customers can choose to have AWS handle their encryption keys or retain full control thanks to the flexible key management options provided by the AWS Key Management Service (KMS).

Customer content disclosure: Unless AWS is compelled to do so by law or court order, it will not divulge customers. If a governmental entity contacts AWS with a request for your custom content, it will make an effort to direct the request to us. Unless AWS is legally forbidden from doing so, if AWS required to reveal your custom content to a government entity, it will give you adequate notice of the demand to allow the customer to seek a protective order or another suitable remedy.

Security Assurance: To assist us in operating securely within AWS and to make the most of our security control environment, AWS has built a security assurance program that incorporates best practices for worldwide privacy and data protection. Multiple third-party independent assessments independently authenticate these security safeguards and management procedures.

Information preservation practices may vary between countries. This can be curbed by making a central system which can be accessed and handled by respective bodies so that database can abide by international standards

This can be achieved using cloud by using a public cloud. The public cloud has a multi-tenant design. With this architecture, one person can share resources while keeping their data private from other users. The availability of the network is necessary for quick data transmission. It can be utilized to put together a lot of data resources while still being affordable. We can access the cloud remotely from any device thanks to the public cloud. Because of this, the gadget frequently needs to conduct a little amount of calculation or even none at all. Public cloud architecture is dependent on the available structure.

The following list of the public cloud's primary advantages:

- It enables businesses to lower their maintenance and investment expenditures.
- With scalability, user requests may be addressed with ease.
- fewer waste of resources
- dependable

These days, public cloud service providers offer better security solutions. To automate security operations and detect any irregularities or abnormalities in the system, a dedicated workforce is employed. When making user data accessible to cloud tenants, strong regulations are used to protect it. The public cloud can be used in a hybrid environment to obtain authorization for higher security levels.

We can use IAM services can help us decide the roles and responsibilities of professionals and individuals.

High Availability and fault tolerance

The type of workload and which cloud service from the provider is used to deploy it will determine how administrators execute AWS high availability and fault tolerance.

AWS occasionally makes high availability available by default. For instance, the orchestrators in each service automatically try to restart on healthy nodes to preserve availability whether you use Elastic Container Service or Elastic Kubernetes Service to deploy containerized apps.

Some AWS products do not come standard with high availability. Depending on the storage tier you select, data in an Amazon S3 storage bucket could be kept by default in just one availability zone. Data will become unavailable if an S3 storage service interruption impacts that availability zone.

Some AWS services provide some level of fault tolerance by default, similar to high availability. Data is automatically replicated across several availability zones by some S3 storage levels. Data will continue to be accessible via the other availability zones without any delays or information loss if one availability zone is disrupted.

Use additional availability zones or additional regions, which cover several availability zones, if a workload is not fault-tolerant by default. Administrators can replicate workloads across physical data centres in the same AWS region using different availability zones. The utilization of numerous regions distributes workloads across many geographic locations, including various U.S. states, various European nations, and more. Configuring deployment across various availability zones and regions to improve fault tolerance might be challenging depending on the workload. Data replication across several zones or regions is simple if the workload solely consists of data.

Maintaining distinct copies of the same workload running within the same availability zone or region is an alternate method for fault tolerance in AWS. The risk of a disruption to the entire location is not avoided by this strategy. However, if one instance fails due to an issue with the virtual host infrastructure, this strategy does maintain the workload operational. To provide EC2 fault tolerance, for instance, Elastic IP addresses can instantly reroute a workload from one EC2 server to another.

Preserving Acid Properties of Database

A group of characteristics (atomicity, consistency, isolation, and durability) known as "ACID transactions" guarantees data integrity in database transactions. By guaranteeing read consistency for queries on the data lake, ACID transactions allow multiple users to concurrently and reliably add and delete Amazon S3 items in an atomic manner while isolating any existing queries. The Athena SQL data manipulation language now has the single-table capability for insert, delete, update, and time travel operations thanks to Athena ACID transactions (DML). You can reliably alter Amazon S3 data at the row level using Athena ACID transactions, together with numerous concurrent users. Athena transactions take care of locking semantics and coordination naturally; they don't need a special record-locking solution.

Working of the National DNA Database System

A database will be created consisting of information on every individual who is a citizen of India. The primary Key of the database will be **DNA Index** to create a **combined DNA index number system. CODIS is a national DNA database bank that enables state and local crime laboratories to retain and analyze DNA profiles from evidence found at crime scenes and from criminals who have been found guilty.** Storing other essential information in the database can also fasten up the process of knowing more about a person and start making connections so that we can find the accused. Aadhar number will be the foreign key so that we can also easily discover the victim's families and also get their DNA profiles .

Sample table Database Structure:

DNA Profile - PK	Aadhar Number - FK	Required information
Xxx1	xxxx-xxxx-xxx1	Depends
Xxx2	xxxx-xxxx-xxx2	Depends
Xxx3	xxxx-xxxx-xxx3	Depends
Xxx4	xxxx-xxxx-xxx4	depends

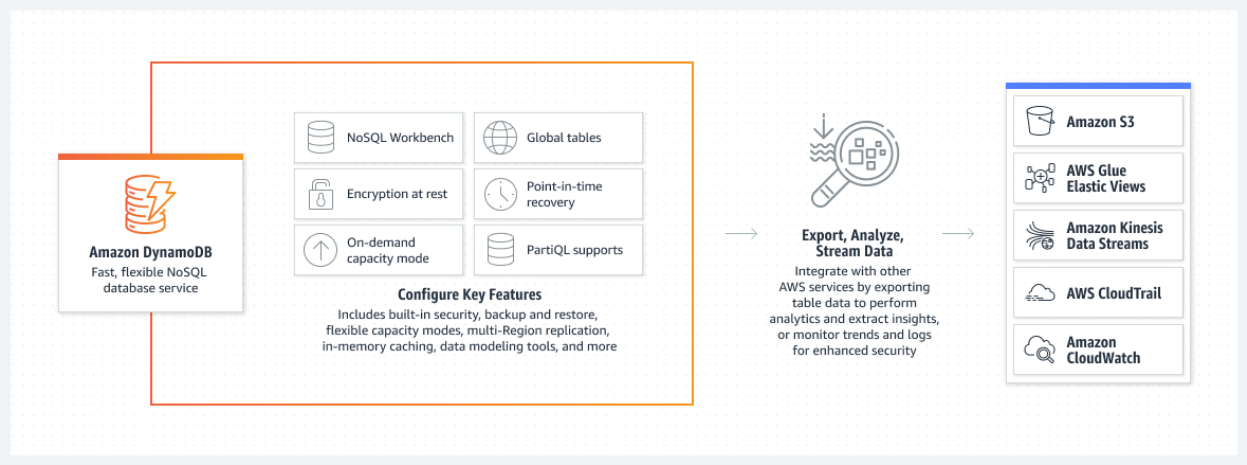
Using AWS we can authenticate and assign a role to different individual tasks so that there is an isolation that an individual who has only permission to view the data will not be able to edit the database and so on.

Aws services being used are :

IAM

Source: https://aws.amazon.com/iam/#

DynamoDB



Source: https://d1.awsstatic.com/product-page-diagram_Amazon-

S3

Source: https://d1.awsstatic.com/s3-pdp-redesign/product-page-diagram_Amazon-S3_HIW.cf4c2bd7aa02f1fe77be8aa120393993e08ac86d.png

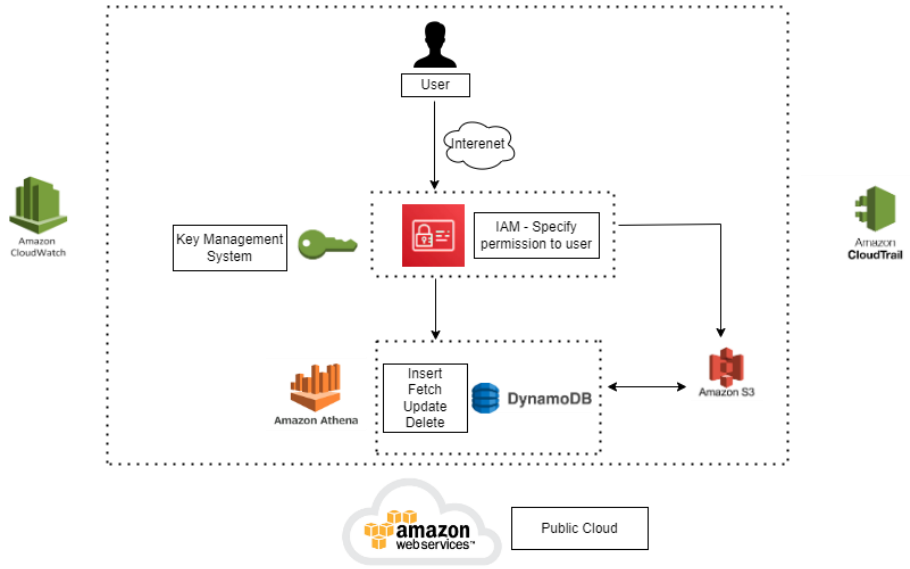
Both S3 and Dynamo Db will be used together.

Athena

Source: https://us-east-1.console.aws.amazon.com/athena/home?region=us-east-1#/landing-page

Apart from these, I will use KMS for making the application more secure it is an Amazon Web Services product called AWS Key Management Service (KMS) that enables administrators to create, remove, and manage the keys that encrypt data stored in AWS databases and products. Cloud watch and cloud trail for managing and monitoring resources and user activities.

Flow Chart



References:

1. https://www.nature.com/scitable/topicpage/forensics-dna-fingerprinting-and-codis-736/#:~:text=Databases%20of%20DNA%20Profiles&text=The%20unique%20profile%20of%20each,from%20past%20and%20future%20crimes.
2. https://docs.aws.amazon.com/
3. https://vittana.org/11-significant-dna-database-pros-and-cons#:~:text=It%20can%20provide%20another%20layer,determine%20if%20matches%20are%20present.