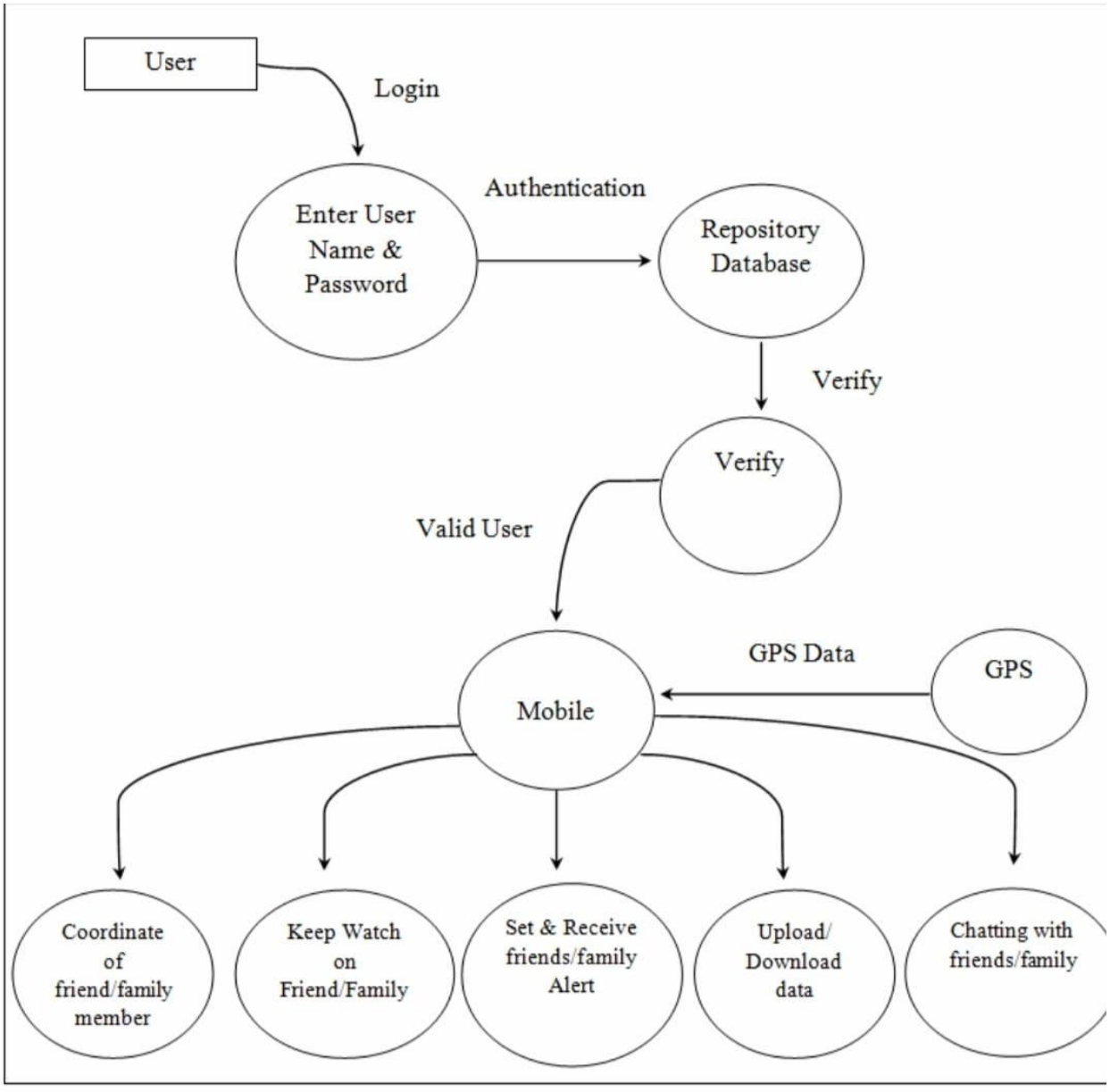
Group Task 2: Module 2

Data Essentials, Big Data, Processing, And Ethics

Big Data Processing mapping: groups select a real-world big data system (like Google Maps) and map out the entire data flow: data source, storage, processing, and output.



1. Data Source:

- In my diagram, the user and GPS are the primary sources.
- In Big Data, this is called Multi-source Ingestion.
- Velocity and Variety: It's not just one GPS coordinate; it's millions of "pings" per second. This includes structured data and unstructured data (chat logs).
- Event streams: Every time a user logs in or moves, an "event" is created. Big data systems use tools like Apache Kafka to catch these events so they don't get lost if the database is busy
- .Metadata: Beyond coordinates, the source provides device type, battery level, and network strength, which helps in "verifying" the user's environment.

2. Storage:

- The diagram shows a single repository, but Big Data usually splits this into a Polyglot storage strategy (using different databases for different jobs).
- NoSQL for speed: To handle "chatting" and "coordinates," systems use NoSQL databases because they can work with data incredibly fast without a rigid structure.
- Data lakes: For long-term history, data is moved to the Data lake (like Hadoop or Amazon S3). This is cheap, high-capacity storage.
- In-Memory caching: For the login and verify steps, systems use Redis. This keeps user sessions in the computer's RAM so authentication happens in milliseconds.

3. Processing (The Logic Layer):

- This is where the Verify and Set/Receive Alert bubbles happen. In Big Data, processing is split into two "speeds":
- Stream Processing (Real-time): This is critical for your Alert bubble. The system must process GPS data as it arrives to see if a family member has entered a Danger Zone.
- Batch processing (Historical): This analyzes data in large chunks. For example, calculating the "average time" a friend spends at a specific coordinate over a week to provide better "keep watch" insights.
- Validation engines: The verify step isn't just checking a password; it's checking for anomalies. (Ex: Is this user logging in from a new country suddenly?)

4. Output:

- The bottom row of your diagram represents the consumption layer.
- Visualization: Converting raw coordinates (x, y) into a visual map interface for the "coordinate of friend" feature.
- Push notification: This is a specialized output. When a friend/family alert is triggered, the system sends an asynchronous signal to a mobile push service.

- APIs (Application Programming Interface): The mobile bubble acts as the gateway. It requests specific slices of the data from the Big Data pool so the phone doesn't get overwhelmed with too much information.

Mapping a diagram to a Big Data process transforms a simple system into a high-scale, real-time ecosystem. By upgrading from a single database to distributed storage and replacing basic logic with stream processing, the system moves from simply sharing a location to providing instant, predictive intelligence. The goal of this Big Data architecture is to ensure that whether you have ten users or 10 million, the "alerts" are immediate, the chat is seamless, and the GPS history is stored safely.

Technologies Used

- Route optimization algorithms
- Traffic alerts sent to users
- Cloud computing platforms
- Distributed storage systems
- Real-time data streaming tools
- Machine learning algorithms
- Big Data analytics frameworks

Application of Big Data in Traffic Management

- Smart traffic signal control
- Accident detection and response
- Route optimization
- Urban traffic planning
- Reduced fuel consumption
- Lower air pollution
- Improved road safety

Challenges of Big Data in Traffic Systems

- Handling massive data volume
- Real-time processing difficulty
- Data privacy issues