Designing a System for Data Collection

(i) Ten Important Questions to Ask Your Client

1. **What is the primary goal of the data collection system?**
    - To understand whether the focus is on customer behavior, sales trends, or operational efficiency.
2. **What types of data will be collected?**
    - Details such as structured (e.g., sales receipts) or unstructured data (e.g., camera footage).
3. **What is the expected volume of data?**
    - Helps determine storage and processing needs.
4. **How frequently will the data be updated or collected?**
    - To decide on real-time vs. batch processing.
5. **What are the key metrics or insights you want to derive from the data?**
    - To align system design with business objectives.
6. **What are the current data sources and their formats?**
    - Examples: cash register receipts, sensor logs, or video feeds.
7. **Are there any existing systems or databases that need integration?**
    - To ensure compatibility and avoid redundancy.
8. **What are the security and privacy requirements for the collected data?**
    - To comply with regulations like GDPR and ensure customer trust.
9. **Who will access the system, and what are their roles?**
    - To define user permissions and access control.
10. **What is your budget and timeline for implementing this system?**
    - To ensure feasibility within constraints.

(ii) Likely Data and File Formats

1. **Structured Data**:
    - *Formats*: CSV, Excel files, relational database tables.
    - *Examples*: Cash register receipts containing product ID, department, price, and timestamp.
2. **Unstructured Data**:
    - *Formats*: JSON, XML.
    - *Examples*: Sensor logs tracking customer movement in stores.

3. **Time-Series Data**:
    - *Formats*: Parquet (for efficient storage), CSV.
    - *Examples*: Timestamps from cameras or sensors monitoring foot traffic over time.
4. **Multimedia Data**:
    - *Formats*: MP4 (video), JPEG/PNG (images).
    - *Examples*: Camera footage for tracking customer movements.
5. **Metadata**:
    - *Formats*: JSON, YAML.
    - *Examples*: Descriptions of sensor locations, camera specifications, or timestamp annotations.

(iii) Suggested Database System and Justification

Recommended System: A hybrid database system combining relational databases (e.g., PostgreSQL) with NoSQL databases (e.g., MongoDB).

1. **Relational Database (PostgreSQL)**:
    - Ideal for structured data like sales receipts.
    - Supports complex queries for analyzing customer spending patterns.
    - Ensures consistency through ACID compliance.
2. **NoSQL Database (MongoDB)**:
    - Suitable for unstructured data such as sensor logs or JSON files.
    - Scales horizontally to handle high volumes of data from multiple stores.
    - Flexible schema accommodates diverse data formats without predefined structures.
3. **Justification**:
    - The combination of relational and NoSQL systems allows efficient handling of both structured and unstructured data.
    - Relational databases provide robust querying capabilities for financial analysis, while NoSQL databases excel in storing large-scale, flexible datasets like sensor outputs.