**Name:  Hassan Ashraf Hassan Al Sadaawy**

# *Assignment 1 Solution*

# *Thndr*

## Data Sources

Thndr, as a digital investment platform, collects data from three main sources: information you provide directly, data collected automatically through your use of their services, and data obtained from third parties.

## *Main Sources of Data for Thndr*

→ **Information You Provide Directly:** This includes data you enter when using the app for account creation or transactions. Examples include identity data (name, date of birth, government ID), contact data (email, address), financial data (bank details, income), and profile data (username, preferences).

→ **Automatically Collected Data:** This data is gathered as you interact with the platform using tracking technologies. It includes usage and device data (how you access services, IP address, device identifiers) and location data for fraud protection and service improvement.

→ **Data from Other Sources and Third Parties:** Thndr may combine the data they collect with information from external sources. This can include data from business and marketing partners, third-party services linked to your account (like Google or Facebook), and publicly available data. They also obtain market data and official records from entities such as Misr for Central Clearing, Depository and Registry (MCDR) and EGID, the official provider of Egyptian Stock Market data.

---

# *Storage layer used*

financial technology  companies usually use a combination of specialized storage types, most commonly a **data lake** for raw data and a **data warehouse** for processed data. Data may be stored in Egypt or other countries where Thndr operates, subject to data protection laws.

**Thndr stores its data in two main layers:**

→ **Data Lake:** Used to store raw market data, user activity logs, and unprocessed events. It supports large volumes of structured and unstructured data.

→ **Data Warehouse:** Stores cleaned and structured data such as trading history, user profiles, and financial metrics. Used for analytics, reporting, and dashboards.

---

## *Processing Layer*

In the data processing layer of an architecture, data is cleaned, transformed, and prepared for use through a structured process, which can be implemented via **batch processing** or **real-time processing**.

→ **Batch Processing:** Used for daily portfolio calculations, end-of-day summaries, cleaning market data, and preparing financial reports.

→ **Real-Time Processing:** Handles live market prices, instant trade execution, and user in-app events. Streaming systems process data as it arrives to ensure low-latency updates.

---

## *Serving Layer*

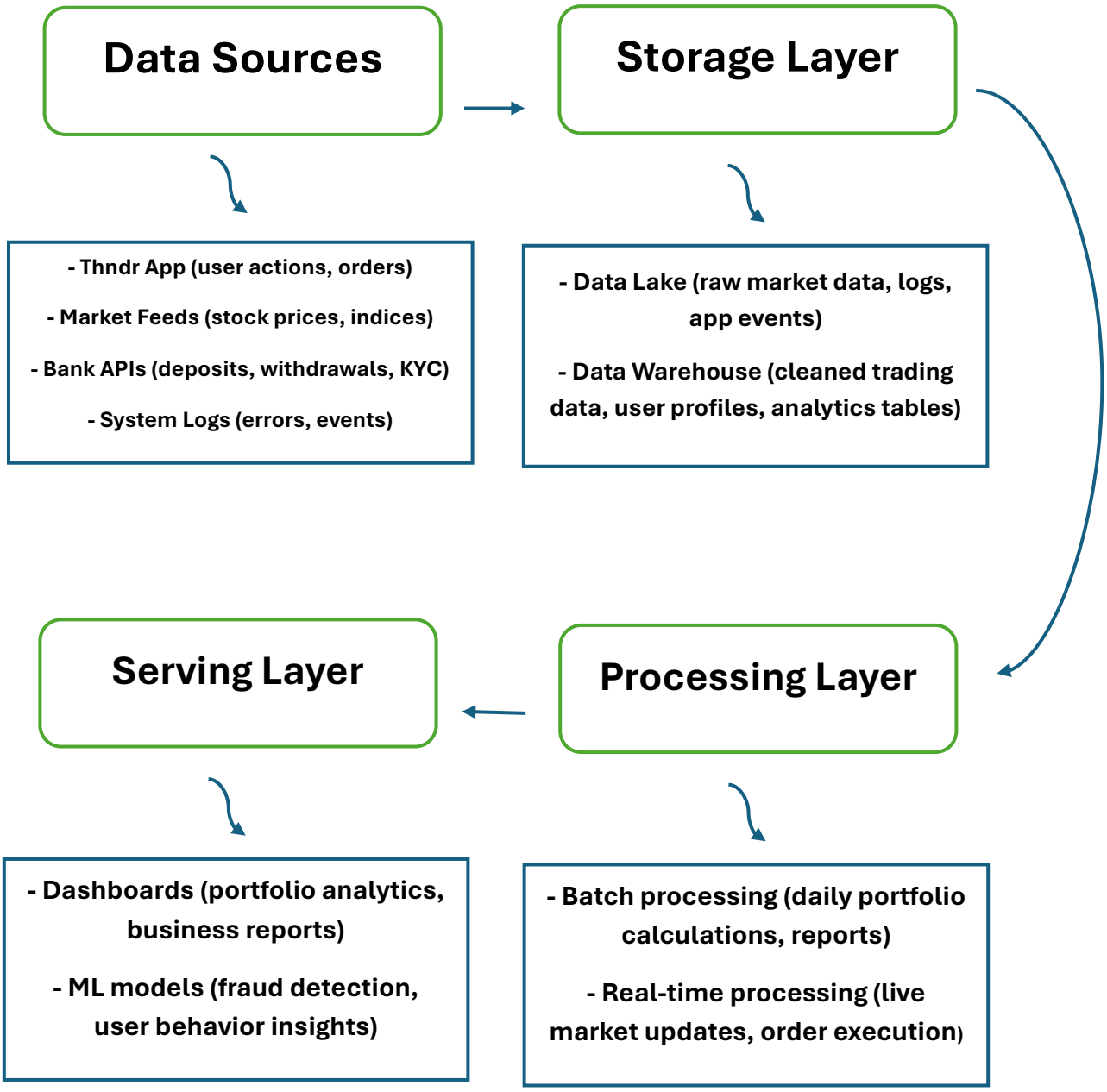
Financial services firms use processed data from their serving layer to drive business value by enabling data-informed decision-

making, improving operational efficiency, and enhancing customer experience. This data is primarily consumed through two main avenues:

➔ **Dashboards:** Used for portfolio analytics, user activity tracking, market insights, and internal performance monitoring.

➔ **Machine Learning Models:** Support fraud detection, user behavior analysis, and personalized investment recommendations.

# Thndr Abstracted Data Pipeline Flow



**Sources:**

Thndr Privacy Policy

ThndrBlog

Deloitte