João Victor Quintanilha
11/24/2024
CS 119

Gedanken Project: Agent-Based Model of the Water Consumption for Industries and Households of Karachi

Technical Design

The proposed system integrates an agent-based model (ABM) with large language models (LLMs) to simulate water consumption dynamics in Karachi. By leveraging advanced big data techniques, scalable infrastructure, and real-time data integration, the system provides policymakers with actionable insights into resource distribution and policy impact. This technical design outlines the datasets, computational infrastructure, and end-user tools critical to the solution.

Datasets and Data Features

The simulation incorporates a combination of static and dynamic datasets to model water distribution effectively.

Static Data: It includes historical water consumption records segmented by sector (household, industrial, and informal) from the Karachi Water & Sewerage Board (KWSB). Supplementary data such as demographic distributions, income levels, and industrial outputs further refine the model's initial conditions (Tabassum et al., 2016).

Dynamic Data streams, including weather patterns, economic indicators, and policy announcements, are continuously ingested to reflect real-time conditions. This dynamic input enables adaptive modeling and scenario testing, crucial for Karachi's rapidly changing environment (Deshmukh & Garg, 2023).

Textual Data, such as government reports and social media analyses, is processed by GPT-4 to refine agent archetypes. Archetypes categorize agents based on socio-economic, geographic, and behavioral parameters, ensuring diversity and stochasticity in simulation outcomes (Chopra et al., 2023).

These datasets are updated at varying frequencies: static datasets monthly, dynamic streams in near real-time, and textual data as new input becomes available.

Big Data Techniques

To model Karachi's 4 million households (assuming 5 people per house on average) and 6,000 industrial agents, the system employs multi-threading within a high-performance computing (HPC) environment powered by Apache Spark and Hadoop. Multi-threading enables concurrent processing of agent interactions, significantly reducing computation time while maintaining agent heterogeneity (Goldsby & Pancerella, 2013).

Implementation of Multi-Threading: Spark's task scheduler partitions agent data across nodes, assigning threads to process subsets of agents in parallel. This structure, combined with Hadoop's MapReduce framework, efficiently preprocesses, partitions, and aggregates large-scale simulation data. This setup ensures that even computationally intensive tasks like recording interactions or updating agent states are handled seamlessly.

Storage Optimization: With the assumption that each agent interaction generates approximately 1 KB of data over 100 timesteps, the system requires around 410 GB per simulation iteration. Hadoop Distributed File System (HDFS) manages this data, offering redundancy and scalability, while it also serves as a secondary storage layer for historical and aggregated results.

Processing Needs: Each HPC node features multi-core processors with at least 16 GB of RAM, supporting up to 10,000 agents per node (Google Cloud, n.d.). The system scales across 400 nodes, leveraging Spark's in-memory caching to minimize latency and accelerate iterative simulations (Goldsby & Pancerella, 2013).

Integration of LLMs

GPT-4 enhances the ABM by generating archetypes that encapsulate agent behaviors, making simulations both realistic and scalable. Archetypes derived from structured and unstructured data serve as baseline templates for agents, which are further diversified through stochastic adjustments during the simulation (Chopra et al., 2024). For instance, industrial agents may vary water consumption in response to fluctuating production demands, modeled probabilistically by GPT-4.

By querying LLMs at the archetype level rather than for individual agents, the system achieves computational efficiency while preserving behavioral complexity. The MESA framework, extended with AgentTorch, encodes GPT-4 outputs as parameter distributions, introducing randomness and adaptability into agent interactions.

Technological Components

As observed in Attachment 1, the system employs robust tools to ensure functionality and user accessibility:

1. **Agent-Based Modeling**: MESA, enhanced with AgentTorch, handles large-scale simulations, seamlessly integrating LLM outputs.
2. **Data Processing**: Apache Spark processes real-time data updates, while SparkSQL structures raw outputs into relational formats for visualization.
3. **Visualization**: Tableau serves as the user interface, transforming simulation results into interactive dashboards. SparkSQL directly connects to

Tableau, enabling dynamic queries and intuitive graphics, such as geospatial heatmaps and time-series analyses (Salesforce Tableau, n.d.).

End-Product and Usage

The final product is an interactive ABM accessible via a Tableau dashboard. Policymakers can set simulation parameters, such as household-to-industrial water allocation ratios, and analyze outcomes through geospatial maps, trend graphs, and scarcity indices. The dashboard allows iterative refinements, enabling users to adjust policies dynamically and rerun scenarios.

This model is particularly impactful in Karachi, where obtaining direct data on water distribution can be life-threatening due to the influence of the tanker mafia, which controls a significant portion of the city's water supply and operates with impunity (Dawn, 2023). By relying on the ABM to simulate and predict water dynamics, policymakers can base their decisions on reliable, modeled data without exposing themselves to physical risks. This represents a critical innovation, transforming how water management policies are developed and implemented in Karachi.

Similar applications have validated the efficacy of ABMs in water management. For example, a study in Raleigh, North Carolina, demonstrated how targeted pricing and restriction policies could effectively reduce demand during droughts (Mashhadi et al., 2017). The Karachi model adapts these principles to a more complex socio-political environment, ensuring scalability and relevance to other water-scarce regions.

References:

Karachi Water & Sewerage Board. (n.d.). Water management. *Karachi Water & Sewerage Board.* Retrieved October 31, 2024, from https://www.kwsb.gos.pk/water-management/

Tabassum, R., Arsalan, M. H., & Imam, N. (2016). *Estimation of water demand for commercial units in Karachi City. FAST-NU Research Journal*, 2(1), 21-26.

Deshmukh, R., & Garg, K. K. (2023). Sustainable groundwater management under climate change: A framework for India's agriculture. *Nature Sustainability.* https://doi.org/10.1038/s41893-023-01177-7

Chopra, A., Kumar, S., Raskar, R., & Giray-Kuru, N. (2023). Scalable multi-agent simulations for adaptive policies using large language models*. arXiv*. https://arxiv.org/abs/2303.08774

Goldsby, M. E., & Pancerella, C. M. (2013). Multithreaded agent-based simulation. *2013 Winter Simulations Conference (WSC)*, 1581–1591. https://doi.org/10.1109/WSC.2013.6721541

Google Cloud. (n.d.). CPU, RAM, and storage considerations. *Google Cloud VMware Engine*. Retrieved November 22, 2024, from https://cloud.google.com/kubernetes-engine/distributed-cloud/vmware/docs/how-to/cpu-ram-storage#:~:text=These%20are%20the%20default%20values,40%20GiB%20of%20storage

Salesforce Tableau. (n.d.). Example: Apache Spark SQL. *Tableau Documentation*. Retrieved November 22, 2024, from https://help.tableau.com/current/pro/desktop/en-us/examples_sparksql.htm

Dawn. (2023, January 24). Karachi's water crisis: Why the city always remains thirsty. *Dawn*. Retrieved October 31, 2024, from https://www.dawn.com/news/1733033

Mashhadi Ali, A., Shafiee, M. E., & Berglund, E. Z. (2017). Agent-based modeling to simulate the dynamics of urban water supply: Climate, population growth, and water shortages. *Sustainable Cities and Society*, *28*, 420–434. https://doi.org/10.1016/j.scs.2016.10.001

Attachments:

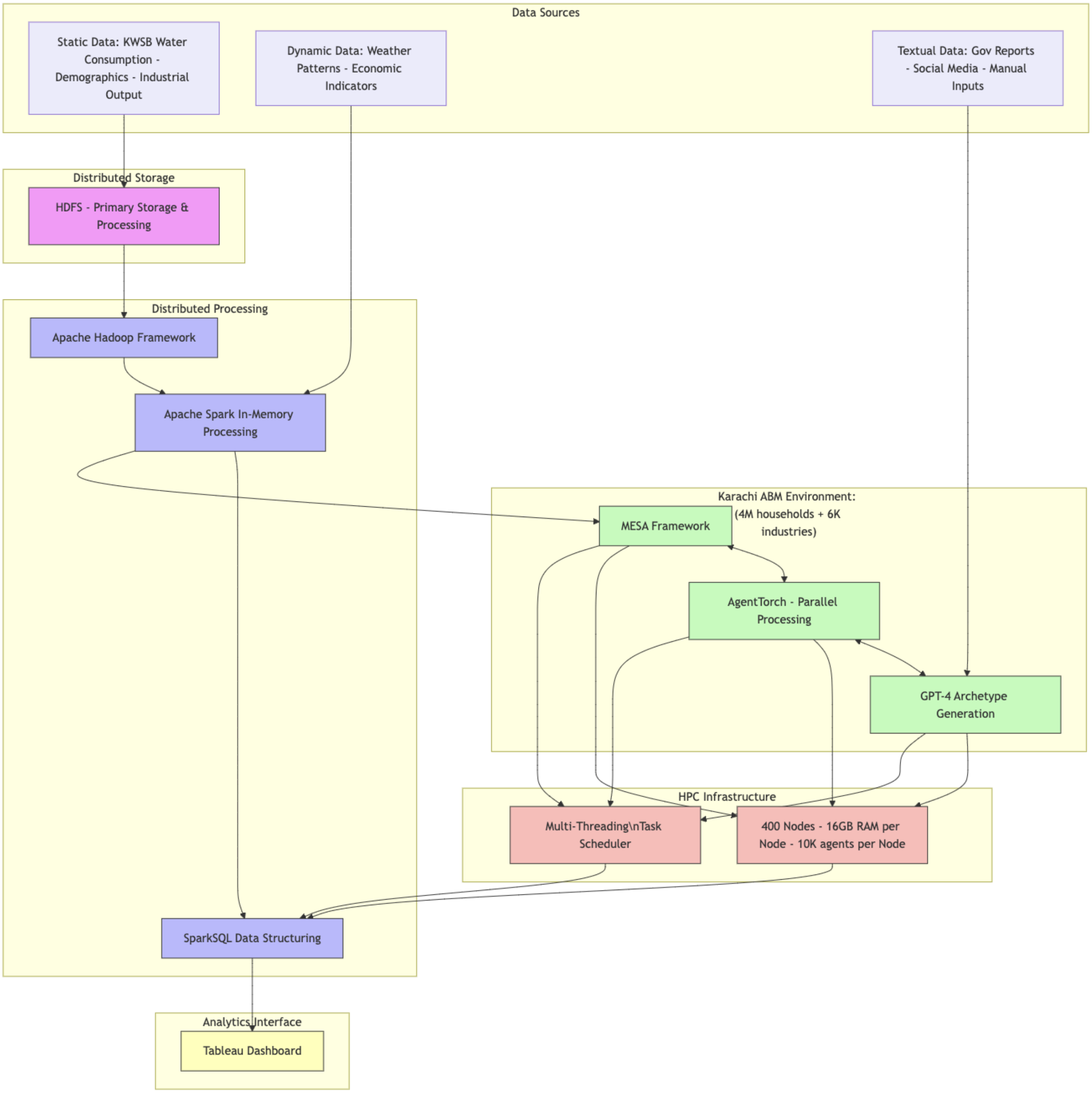

Figure 1: Technical Architecture Diagram of Karachi's ABM