Azure Modern Analytics Architecture for Synthetic Syndromic Surveillance

Arash Jalali, MPH, MSHI, Sean Huang, MD, Vineet Srivastava, Akshay Barapatre, Shree Parida, Sage Kim, PhD, & Karl Kochendorfer, MD, FAAFP, FAMIA, University of Illinois Hospital & Health Sciences System, Chicago, IL

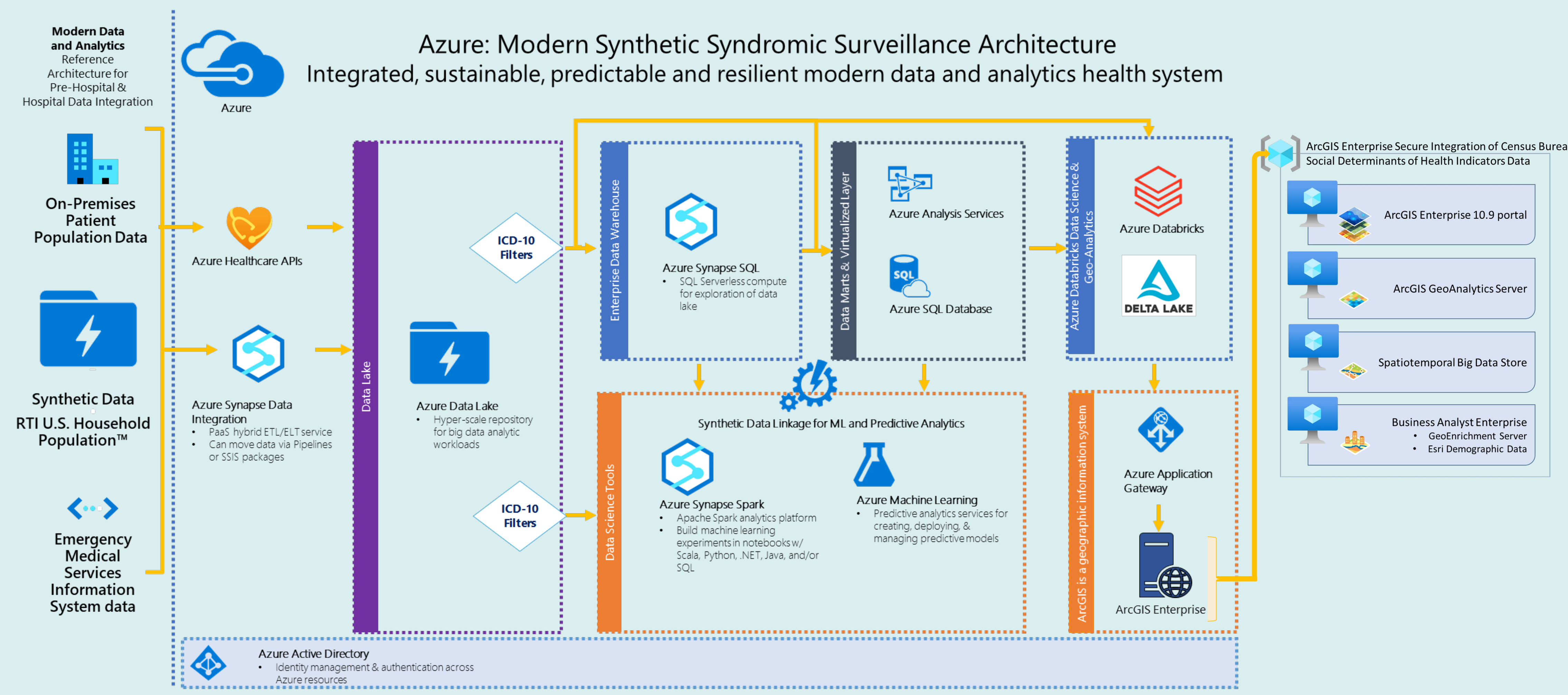
UI Health

UIC PUBLIC HEALTH
Health Policy and Administration

Background

The use of machine learning to build predictive models often requires the harmonization of disparate data sources. For instance, at our institution at the University of Illinois at Chicago, we have found a need to meld real-time geographic data with prehospital data and enriched geospatial information. This required a complex architecture for data analytics over cloud services which had not been published in the literature over the past decade. This poster aims to discuss the construction and design of synthetic syndromic surveillance cloud architecture for hospital data integration.

Research Design

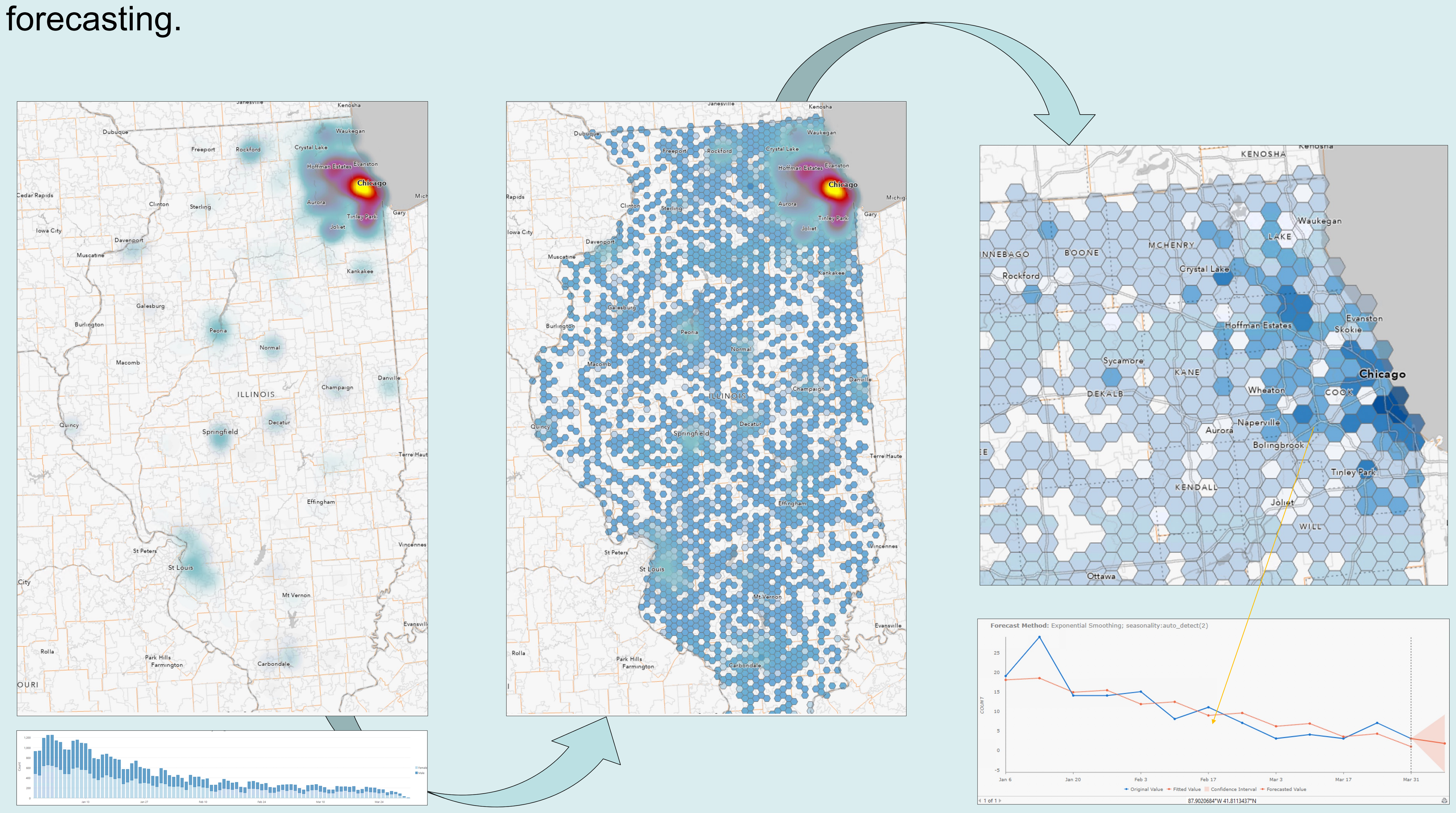


Methods

The open-source cloud computing data analytics architecture that we used is shown in our Research Design section. Data can be obtained from various sources, including patients from an electronic health record. We used Illinois patient data acquired from the COMPdata Informatics and statistically matched the patient records with household population records from the RTI U.S. Synthetic Household Population™ database. Afterwards, the data is stored in a hyper-scale repository Azure Data Lake. After filtering with ICD-10 codes, the data is ingested into an Azure Synapse SQL Server. Azure pipelines help to clean and filter the data inside the Azure SQL data warehouse. Data analysis can be performed using Azure Synapse Spark as an analytics platform and Azure Machine Learning for its analytic services. We can move the data into ArcGIS Enterprise for geospatial analytics. In addition, predictive models can be created, deployed, and managed as part of a synthetic syndromic surveillance system.

Results

The following figures display the system's ability to use forecasting tools to predict hotspots. In the maps below, we see various shades of blue throughout Chicago, with darker blue hues indicating higher forecasting of COVID-19. Graphs accompany each picture to demonstrate a predictive model. The blue lines follow actual cases while the orange line indicate our forecasting based on geospatial data. Presented are example of exponential smoothing forecasting.



Conclusion

The presence of health data, synthetic data, and the lack of interoperability between these forms for research necessitates designing a new system to perform modern data analytics efficiently. Such experiments often require the use of a complex data warehouse as well as a cloud computing environment. Here we have described an open-source architecture that we have developed to successfully build our unique synthetic syndromic surveillance systems.

Future Plans

We plan to improve our team-based data science processes through frequent data acquisition from the Illinois Department of Public Health and continued synthetic linkage to COMPdata Informatics. This will allow us to provide up-to-date spatial analysis for improved community health advocacy. We have also decided to focus more on feature engineering, model training, and model evaluation.

Acknowledgements

The analyses described in this poster presentation were conducted with data collected from the Illinois Health and Hospital Association and were analyzed using artificial intelligence tools accessed through Microsoft Azure. The research is supported by the Microsoft US EDU Customer Success Unit, via a Microsoft AI for Health Grant, and via a Community Health Advocacy Grant.

References

1. Maciejewski R, Hafen R, Rudolph S, Tebbetts G, Cleveland WS, Grannis SJ, et al. Generating Synthetic Syndromic-Surveillance Data for Evaluating Visual-Analytics Techniques. CG-M. 2009;29(3):18-28.

2. The Azure Secure Enclave for Research [Internet].; 2022 [updated September 7,; cited Septemeber 2022]. Available from: https://github.com/microsoft/Azure-Secure-Enclave-for-Research.

3. Data Lakehouse & Synapse [Internet].; 2020 [updated September 10,; cited July 2021]. Available from: https://www.jamesserra.com/archive/2020/09/data-lakehouse-synapse/.