Atlas on Cloud: Utilizing modern cloud infrastructure for hosting OMOP tools.

Siriraj Informatics and Data Innovation Center

Presenter: Natpatchara Pongjirapat

Email: Natpatchara.pon@mahidol.ac.th

Background

Cloud technology has grown significantly in recent years. In particular, Atlas has shown significant benefits from cloud computing. Various reports illustrate the advantages of utilizing the cloud for deploying Atlas. However, most of these articles do not adequately capture the processes and challenges one might encounter. We aim to address these gaps by sharing our process and the problems we faced.

Method

We started by researching the design, requirements, and configuration for Atlas and OHDSI WebAPI. Broadsea provides an excellent document, which we used as a model for setting up services using the provided images. Azure was chosen as our cloud platform for implementation primarily due to its availability within our institution and the support provided by Microsoft's templates.

Azure offers several services for hosting containerized applications. Among these options, we chose Azure Container Apps because it allows us to leverage the benefits of containerized applications without managing a Kubernetes cluster. Azure PostgreSQL was used for hosting our metadata database, while Azure SQL Database is used for our CDM data. Additionally, Microsoft Entra service is set up for Single Sign-On (SSO) authentication.

Table 1: Overview of Atlas deployment of Siriraj hospital on cloud

Azure service	Name	Description
Containerized service		
Azure Container app	Atlas container	Containerized service for hosting main Atlas Web application utilizing docker technology.
Azure container app	OHDSI WebAPI container	Containerized service for hosting undelying OHDSI WebAPI service utilizing docker technology.
Database		
Azure database for PostgresSQL flexible server	WebAPI postgres database	Fully managed PostgresSQL database for OHDSI WebAPI metadata including security and project configuration.
Azure database for PostgresSQL flexible server	OMOP(synthea1k) postgres database	Fully managed PostgresSQL database for Demo patient data adhering to OMOP CDM standard (Synthea10k was used in the framework.)
Azure SQL server (not include with template)	OMOP de-id MSSQL database	Fully managed MSSQL database for Demo patient data adhering to OMOP CDM standard based on characteristic found at Siriraj hospital (not included in the template.)
Storage		
Azure storage account and file share service	file shares	File storage service for runtime configuration data of Application and database.
Others		
Azure key vault	key vault	Key vault for stroing Secret and key used to initiated connection and security between service.
Azure log analytics workspace	log analytics workspace	Realtime log and workspace for monitoring and troubleshoot runtime error.

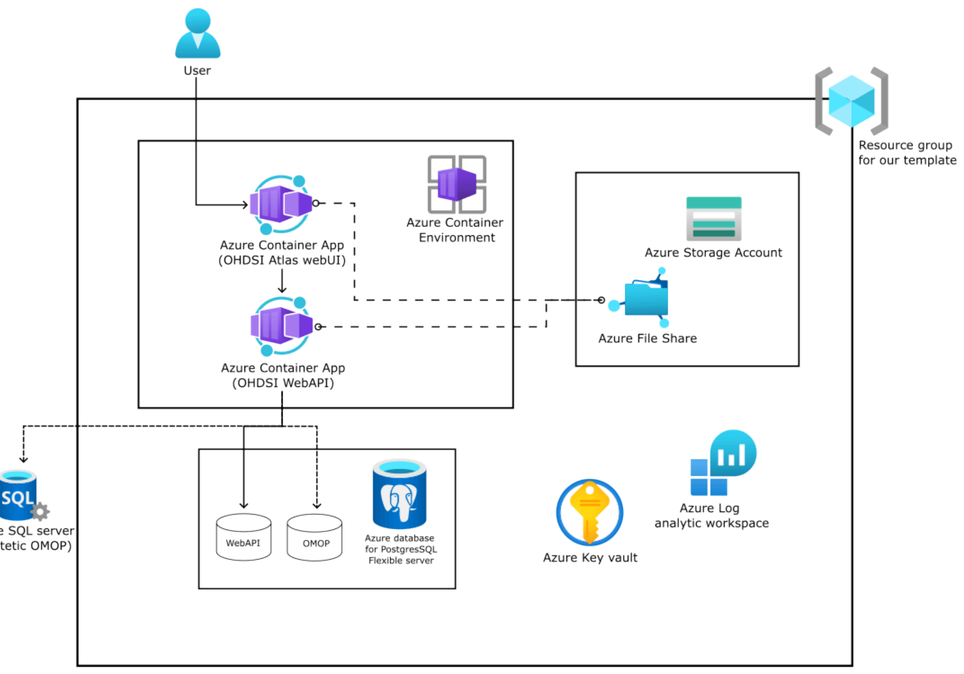

Figure 1: Framework for Atlas deployment used in our institution

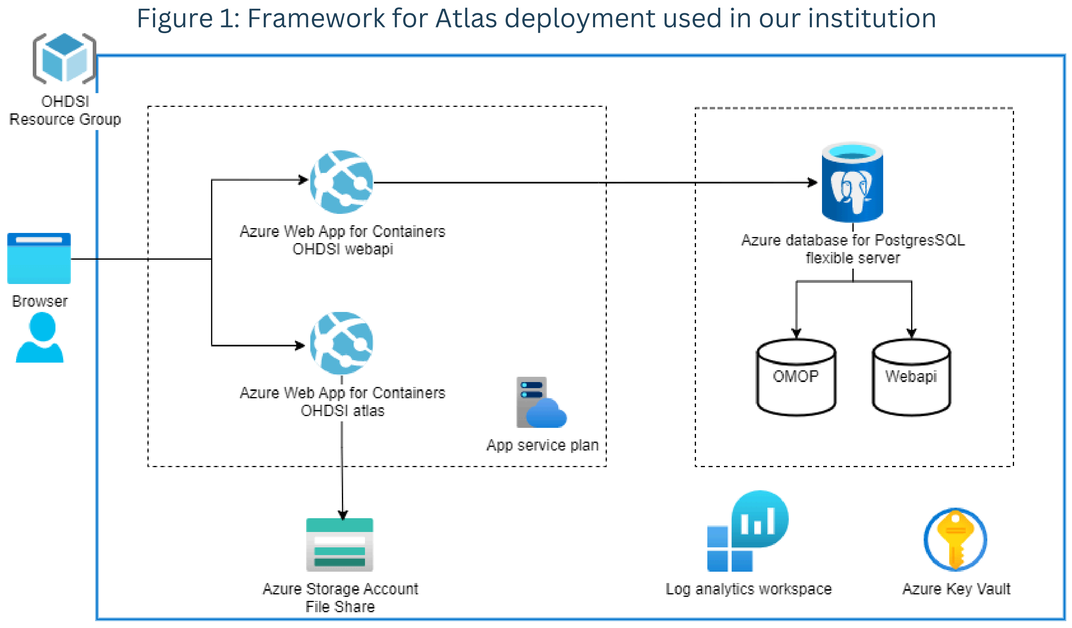

Figure 2: Framework provide by Microsoft for deploying Atlas on Azure

Figures 1 and 2 compare the framework for our deployment on the Azure platform with the one provided by Microsoft.

After the system was designed, we developed a Bicep code to automate the deployment process, enabling us to reproduce and redeploy our setup as needed. Our code, along with detailed documentation about our design, is publicly accessible on our GitHub.

Finally, synthesized clinical data, adhering to the OMOP CDM format, was created and uploaded to Azure SQL Database. Atlas was then configured to access both the Synthea10k synthetic data hosted on Azure PostgreSQL and our synthesized data on the SQL server.

Result

The application was hosted on our Azure platform from July to September 2024. A workshop was conducted on 3-4 September 2024. The application was accessed by users within Mahidol University. A total of 117 jobs were created throughout the period, with the majority occurring during the workshop. Several challenges were encountered during the process, some of which are outlined below.

Firstly, concerns were raised about the data privacy of our deployment. Many cloud service providers have data centers located outside Thailand, making data transfer to foreign countries unavoidable. One potential solution discussed by our team was the use of a hybrid network, where the database is hosted locally while the application is hosted in the cloud with secure access to the database.

All code used will be released in the future in our github.
To visit scan the qr code or go to https://github.com/sidataplus

Another topic of discussion is the cost of hosting such services. While cloud-managed services offer a pay-as-you-go model, costs can rise significantly as activity increases. Our estimates suggest that the cost of our Atlas deployment could range from approximately $20 per month to $200 per month or more if the application becomes active. This cost variability is a concern for our institution.

Finally, careful configuration is essential to fully leverage cloud capabilities. During the workshop, authentication through Microsoft Entra failed. This issue was not encountered during pre-workshop tests, likely due to the lower demand.

While migrating Atlas to a cloud platform may come with numerous challenges, we believe in the benefits gained from utilizing cloud capabilities. Thus, the institution should make an effort to support such development.

Natpatchara Pongjirapat. Krittaphas Chaisutyakorn. and Natthawut Adulyanukosol