AMCIS 2022 Proceedings

SIG DSA - Data Science and Analytics for Decision Support

Aug 10th, 12:00 AM

Spatial Processing in Cloud-Based Architectures

Katarzyna H. Tuszynska
Claremont Graduate University, ktuszynska@esri.com

Spatial Processing in Cloud-Based Architectures

Emergent Research Forum (ERF)

Katarzyna H. Tuszynska
Claremont Graduate University
Katarzyna.Tuszynska@cgu.edu

Abstract

With spatial data, traditionally associated with on-premises GIS solutions, takes part in organization-wide digital transformation architectures, the resulting cloud-native data pipelines will become a reference architecture as workflows begin to extend further out to the edge, fog and individual units as distributed computing environments. This work explores the current landscape of that progression, affected industries and most impacted markets.

Keywords

GIS, Spatial, Cloud Architecture, Spatial Big Data

Introduction

Spatial data processing, in the form of Geographic Information System (GIS) solutions, has been a successful, niche submarket for over 50 years. As enterprise architectures migrate to the cloud, analytical processes for spatial data are gradually being integrated within holistic, general data pipelines. The early industry adoptions of this trend are a foreshadowing of larger market adoption patterns in the future. This paper provides an overview of the changing data market, the services underlying this change, a current snapshot of the architecture, and the current “players” in the market.

Changing Data Market

Processing of spatial data has been traditionally served by on-premises, solutions most successfully presented as platforms or a grouping of related software products that could persist the context, or the specific characteristics of spatial data between applications. Solutions provided by Esri, Hexagon or Pitney Bowes gave users a choice of client and server solutions that processed the many formats or representations of spatial data such as vector, raster, image or JSON along with its spatial reference information. However, as volume and velocity of data gathered by an organization grew, initiating the trend of migrating to cloud-based deployments, so did the amount of spatial data.

Enterprises found themselves initially looking for “lift and shift” migrations, but subsequently found that the artificial isolation of their data processes, previously necessitated by the limitations of on-premises deployments, did not have to bind their architectural choices in the cloud. Potentially introducing spatial processing into every stage of their organizational data pipeline, offering the benefit of machine-based spatial decisions closer to the event and opening up the availability of spatial data to the broader Business Intelligence (BI) analyst group inside of their organization, rather than delegating it to the few GIS analysts concentrated in a single department.

The challenge is the availability of spatially cognizant, cloud-native, persistence and analytical environments as well as the domain expertise necessary to construct an architecture of a data pipeline that will re-envision the approach to processing spatial data in the cloud architectural context.

Spatially Cognizant, Cloud-Native Data Products

The primary cloud services predicating an architecture that includes spatial data are services that persist or store spatial data and analytical environments that allow the data to be processed.

Persistence

Spatial data stored in the cloud can be persisted in a number of services depending on the original format of the data and its complexity. Data Lakes refer to object stores that persist files, such as CSVs, images, or more recently Parquet files. Relational databases have traditionally served the role of operational data stores, persisting data in spatial types, offering a set of spatial relationship functions and the ability to transform between spatial references. More recently cloud-native data warehouses have begun to support spatial data and are blurring the lines between analytical environments, by separating persistence from analytics and including a variety of connectors to external processing environments such as Apache Spark or machine learning.

Analytical Environments

Spatial analytics can vary based on the representation and the ultimate purpose, for example: Imagery and raster data processes numeric information as matrices, vector represents data as distinct features modeled as some extent or zoom level. Lidar data is represented as meshes of points usually indicating measurements between entities. As the volume of data increases solutions need to be available at ever growing computational capacities, offering a comparable level of functionality. At present most spatial processing occurs in python, R, relational databases and by proprietary server products. However, Apache Spark and machine learning environments are growing in their ability to process spatial data. The new environments create opportunities for re-envisioning established workflows, however the continuity of a spatially capable data pipeline which includes these environments is still emerging.

Emerging Market

Traditional GIS provided the ability to manage real world entities. Those entities varied by industry, the computational capabilities of GIS environments focused on insights generated by a domain savvy, human analysts, at times long after the occurrence of the event. The data becoming a static snapshot in time. With progressively faster compute environments, possibly located closer to the occurrence itself, spatial processing can be at the core of near-real time systems.

Meaning that mobile entities (Adeosun, Olajide & Adeyanju, Ibrahim & Omotosho. I & Ibrahim, M. & Folaranmi, Oluwatosin. (2015) could now be processed within the time of their movement for management and shortly after for maintenance. At this time industries reliant on movement such as assisted or autonomous automobiles and trucking and logistics companies are at the vanguard of innovating cloud-native data pipeline architectures that rely on progressively more spatial data processing at every step of the pipeline.

These implementations could guide the emergence of architectures for other movement-based entities such as cars, trucks, drones, submersibles, subterranean and pipeline systems, and pave the way for asset heavy industries such as the oil and gas or the utility market and circling back to public organizations that could manage municipal and state assets with “digital twin” models in “smart city” platforms.

The cloud architectures, as new as they are, could themselves serve as a reference architecture for data pipelines that extend all the way back to the unit itself.

Autonomous Vehicles are an example of a mobile unit that processes significant amount of data on the unit, with some subset of the data being sent either directly to a group or “flock” (Myers 2019) of nearby vehicles or to the edge device, usually co-located with a cell tower. Where it can be relayed to other “flocks” (TE 2018) or aggregated as a batch of data to be sent further up the pipeline. Fog devices at times perform the same function as Edge devices, but in some architectures, they gather data from multiple edges, (Hardik 2017) potentially process data in near-real time to be shared back to units, or aggregate larger volumes of data and send it to the cloud. The cloud receives data batches at different frequencies and processes the data for its many uses within an organization. The Edge and Fog workflows are still in their infancy with significant potential to redefine the capabilities of near-real time information systems such as utility pipelines or electrical grids.

	Unit	**"Flock" of Units**	**Edge**	**Fog**	**Cloud**	**On Premises**
Physical Implemen-tation	Static & Mobile Assets Cars, cell phones	Static & Mobile Assets	Cell Tower LTEBase Station Gateway	Core Network Private Cloud Data Centers	AWS, Azure, GCP, Baidu, JBCloud, Huawei	Enterprises
Role of data	Decisions for unit Send out alarm Send out status	Exchange of info. "just in time" predictions Exchange of info.	Aggregation Exchange of information from unit to flock, send status, alarm Bounded by location and time	Aggregation Exchange of information from unit to flock, send status, alarm	Aggregate of status and alarm Enrichment inference Prediction	Organization specific insights
Analytics	Sensor Fusion, ML On-chip (GPU)	ML On-chip (GPU) Through a gateway	MQTT Gateway Local compute Messaging Data cashing	MQTT Gateway Precompute per location/time (event) Send status (batches)	PY, RDBMS, DW, Spark, ML, GPU DBs,	PY, R, Web, Mobile Apps Dashboards Data Information Products
Spatial	Sensors: camera, HD Maps, Lidar, radar Self-location Location to the route	Identity through space and time	Asset tracks Asset deduplication location as an identifier Spatial aggregation	Asset as an id Spatial aggregation	Aggregation Dimention-alization Exploration – OLTP Insights Prediction	Application of insights Maintenance Communi-cation Visualization (Maps, Apps)

Table 1. Elements in a holistic data pipeline, from the unit to the user.

Market Examples

The public cloud market is at the initial stages of integrating spatial processes into various stages of compute in a data pipeline. The industry use case leading the charge seems to be assisted and autonomous vehicle pipelines, arguably because a lot of spatial processing occurs on the vehicle, and a portion of the data is than conveyed to a cloud backend. Followed by other complex, mobility use cases such as Transpiration and Logistics which traditionally rely on proprietary solutions for different parts of their data pipeline, but recently this industry is beginning to coalesce and re-envision their processes in a single, holistic cloud implementation. There are two distinct adoption patterns, one visible states-side and the other in China.

US Public Cloud Companies

AWS, Azure and Google Cloud Platform (GCP) are the predominant public cloud providers in United States. Each has differentiated by focusing on different product sets, from Infrastructure as a Service (IaaS) provided by AWS, attractive to implementations lead by IT or DevOps., focused on Platform as a Service (PaaS) offerings, implemented by enterprises that look for collaborative workflows. GCP has strong Software as a Service (SaaS) offerings that focuses on flexibility and variety for analytical processes. As companies re-envision their Digital Transformation journey, they chose cloud offerings based on the aspect most important to them, often migrating between offerings as priorities change.

Chinese Public Cloud Companies

The Chines cloud market differs from those in the western countries. Focusing on the pattern of use by the enterprise user, there is a heavier usage of private clouds spanning smaller organizations, with public cloud providers serving a more customer facing role and as a backend to large enterprises. This difference and the significant growth of the Chinese economy, prior to the Covid 19 pandemic, has led to more industry centric products. Connected architectures for "Smart Manufacturing" or "Smart Marketing" (Alibaba 2022) has led to product bundles where data and application context is carried through, essentially offering a pre-architected industry centric deployment. When examining solutions for the autonomous or assisted vehicle market the product groupings are articulated as a progression that the user does not have to architect, the user is led through a pre-determined pipeline. This may be seen as a guidance for future US based architectures for markets with many spatial data components.

Conclusion

The Digital Transformation journey for many companies has the potential of including spatial processing workflows throughout the architecture. These types of workflows are already being signaled by Autonomous and Assisted Vehicle implementations and by architectures deployed by Transpiration and Logistics companies, with the bulk of the focus on cloud-centric backends and emergent implementations in Fog and Edge environments. Once these industries have solidified on an implementation, these architectures can be used as a reference for a broader scope of implementations deployed by asset centric industries for static and mobile entities.

REFERENCES

Adeosun, O & Adeyanju, I & Omotosho, I. & Ibrahim, M. & Folaranmi, O. 2015. "Simulation of a Fault Tolerant Mobile Telecom Network. International Journal of Electronics Communication and Computer Engineering", 6. 2278-4209. (https://www.researchgate.net/publication/281557435_Simulation_of_a_Fault_Tolerant_Mobile_Telecom_Network, accessed on February 24, 2022)

Alibaba. 2022. "Industrial Brain" (https://www.alibabacloud.com/solutions/intelligence-brain/industrial, accessed February 24, 2022).

McKendrick J. 2021. "Digital Transformation Is Only the Beginning: A Company's Journey", (https://www.forbes.com/sites/joemckendrick/2021/09/12/digital-transformation-is-only-the-beginning-a-companys-journey/?sh=473540db2420 , accessed February 24, 2022).

Myers, A. 2019. "What can autonomous vehicles learn from the flight behavior of birds?" https://engineering.stanford.edu/magazine/article/what-can-autonomous-vehicles-learn-flight-behavior-birds, accessed February 24, 2022).

Shah H. 2017. "Edge computing and Fog computing for enterprise IoT", (https://www.simform.com/blog/iot-edge-fog-computing/, accessed February 24, 2022).

Tang J, Phillips D. 2020. "Field Notes: Building an Autonomous Driving and ADAS Data Lake on AWS", (https://aws.amazon.com/blogs/architecture/field-notes-building-an-autonomous-driving-and-adas-data-lake-on-aws/, accessed February 24, 2022).

TE, Connectivity. 2018. "6 Key Connectivity Requirements of Autonomous Driving", (https://spectrum.ieee.org/6-key-connectivity-requirements-of-autonomous-driving, accessed February 24, 2022).

Wikipedia, The Free Encyclopedia. 2022. "Geographic Information Systems" https://en.wikipedia.org/wiki/Geographic_information_system, accessed February 24, 2022).

Wikipedia, the free Encyclopedia. 2022. "Edge computing" (https://en.wikipedia.org/wiki/Edge_computing, accessed February 24, 2022).

Yifa, P. 2020 "What is a Lift and Shift cloud Migration?" NetApp (https://cloud.netapp.com/blog/what-is-a-lift-and-shift-cloud-migration , accessed February 24, 2022).