International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Scalable Geospatial Data Platforms for Utility Asset Monitoring and Risk Mitigation

Aditya Laljibhai Patel

Submitted: 03/05/2021 **Revised**: 17/05/2021 **Accepted**: 28/05/2021

Abstract: For organisations that need to keep an eye on linear and networked utility infrastructure like electricity transmission lines, pipelines and water distribution systems, this deluge of data—derived from sensors—has posed a challenge and an opportunity. This paper presents a synthesis of the literature of scalable geospatial data platforms, unmanned aerial vehicle (UAV) and Light Detection and Ranging (LiDAR) inspection tools, and spatial statistical risk assessment techniques, and proposes an integrated system for monitoring and risk mitigation of utility assets. Evidence from the distributed computing research community with spatial data shows that using specialized data structures can cut the latency of a range query on a dataset of tens of millions of objects from about two hundred seconds to two seconds compared to traditional distributed processing engines. Comparative database benchmarking shows that document-oriented, non-relational stores can be 3 to 6 times faster on point and compound queries than relational spatial extensions, and that relational systems can be a similar speed on radius-based queries. Nearly ninety-three percent of defects are detected and nearly ninety-two percent are correctly identified by deep-learning inspection systems configured with UAV imagery, while the average error for defect-analysis is less than ten centimetres using the LiDAR-based clearance anomaly detection. The National-scale Flood Exposure characterization of critical infrastructure shows sectoral variation in flood exposure ranging from 2.7% to 27.1% of critical infrastructure within the 100-year floodplain, highlighting significant spatial disparities in exposure. The synthesis also demonstrates that cloud-based parallel processing reduces simulation runs from 12 days to two hours. The results are combined into a multi-layered ingest, distributed storage, spatial indexing, analytics, and risk-prioritisation architecture. The utility challenges, remaining interoperability, computational expense, and data governance issues are explored, as are implications for utility operators, regulators, and infrastructure planners.

Keywords: *geospatial data infrastructure, utility asset monitoring, risk assessment, LiDAR, unmanned aerial vehicles, convolutional neural network, spatial database, cloud computing, distributed processing, flood exposure, pipeline integrity, spatial indexing*

1. Introduction

Utility networks such as electricity transmission corridors, oil and gas pipelines, potable water distribution networks span thousands of kilometres and are constantly under threat from environmental, mechanical and third-party hazards (Bonvicini et al., 2018; Cevik & Topal, 2003). Traditional condition assessment was based on a manual inspection method performed at irregular intervals, which is labour intensive and not suitable for today's fast-changing and large-scale infrastructure networks (Chen et al., 2018). The convergence of three technological trends has started by 2020, scalable geospatial data platforms storing and querying billions of spatially referenced records (Eldawy & Mokbel, 2015; Rodrigues Zalipynis, 2018), UAV-borne and airborne LiDAR sensing and deep-learning image analysis for automated asset inspection (Siddiqui et al., 2018; Wang et al., 2019; Wu et al., 2018), and the refining of spatial statistical methods for prioritising mitigation resources based on quantified risk (Hou et al., 2020; Qiang, 2019).

In this paper, these three areas of research are combined into a comprehensive discussion of how scalable geospatial data platforms can help with

Employer: Logic Explorer LLC End Client: PG&E

Email Id / Password: Consult.AdityaPatel@gmail.com

utility asset monitoring and risk mitigation. Its goal is threefold: to describe, in terms of technical systems, the building blocks of a scalable geospatial platform; to measure, based on benchmarking evidence reported in the literature, the performance improvements that can be realized at each architectural level; and to explain how geospatial platforms can transform raw spatial data into actionable operational decisions, prioritized according to risk.

2. Scalable Geospatial Data Infrastructure

2.1 The Big-Data Character of Utility Geospatial Data

Today's utility monitoring produces data that is "big" with volume, velocity, veracity and variety, the value of which needs to be extracted by purpose-designed processing systems (Yang et al., 2017). Sources include coarse-resolution weather and hazard data, used for exposure modelling, as well as fine-resolution point cloud data from airborne or UAV-based laser scanners, and fine-resolution continuous data from mobile inspection units via Global Positioning System (GPS) telemetry (Goodchild, 2020). Spatial data is known to have multidimensional geometry and irregular access patterns, which makes it difficult to fit into conventional relational database management systems and general-purpose distributed computing engines, and spurred the development of spatially-aware extensions to existing big data frameworks (Eldawy & Mokbel, 2015).

2.2 Distributed Spatial Processing Frameworks

Purpose-built spatial indexing can deliver orders of magnitude performance gains, as seen in the case of SpatialHadoop, which adds a spatial component to the language, storage, MapReduce, and operations layers of the Apache Hadoop ecosystem (Eldawy & Mokbel, 2015). The framework is implemented by adapting the Grid, R-tree and R+-tree index structures into a two-level spatial index and adding the SpatialFileSplitter and SpatialRecordReader components to it, which allows for efficient range queries, k-nearest-neighbour queries and spatial joins directly in the MapReduce paradigm. Figure 1 and Table 1 illustrate a range query on seventy million spatial objects on a twenty-node cluster that takes about two hundred seconds with traditional Hadoop, but only two seconds with SpatialHadoop, a two order-of-magnitude improvement that is due to index-aware partitioning, not hardware.

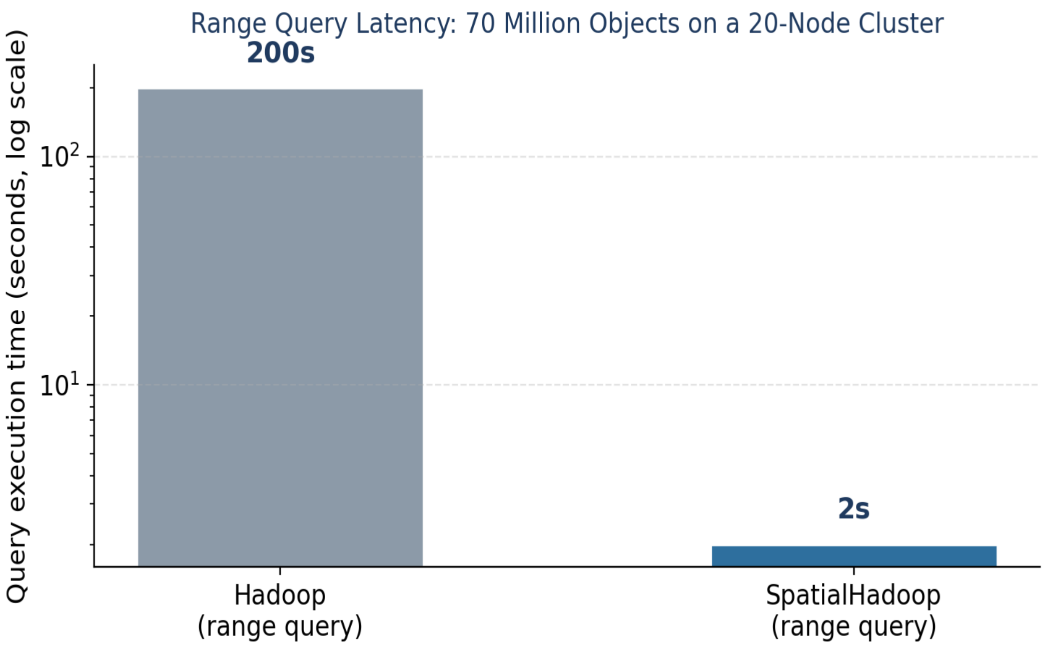

Figure 1. Range-Query Latency for 70 Million Spatial Objects on a 20-Node Cluster, Comparing Conventional Hadoop and SpatialHadoop.

Note. Data are derived from Eldawy and Mokbel (2015), which also reports validation against 60 GB and 300 GB real-world datasets sourced from TIGER/Line files and OpenStreetMap.

2.3 Relational, Non-Relational, and Array-Based Storage Models

In addition to distributed batch-processing frameworks, the selection of relational spatial extensions versus non-relational document stores can have a significant impact on the latency of the queries needed for a workload of interactive asset monitoring. A controlled benchmarking study of PostGIS (PostgreSQL spatial extension) and the document-oriented MongoDB database revealed that MongoDB was ~3 times faster than PostGIS on point in point query performance, and ~6 times faster on compound query performance (nearest-neighbour and intersection predicates) with an increasing search radius, while PostGIS remained ~3 times faster than MongoDB on radius-based queries (Bartoszewski et al., 2019). The results shown in Figure 2 suggest that there is no single storage paradigm which is clearly superior for all spatial operation types, and hence, when designing a platform, choose the storage-engine type of the platform according to the prevalent query pattern of the intended application.

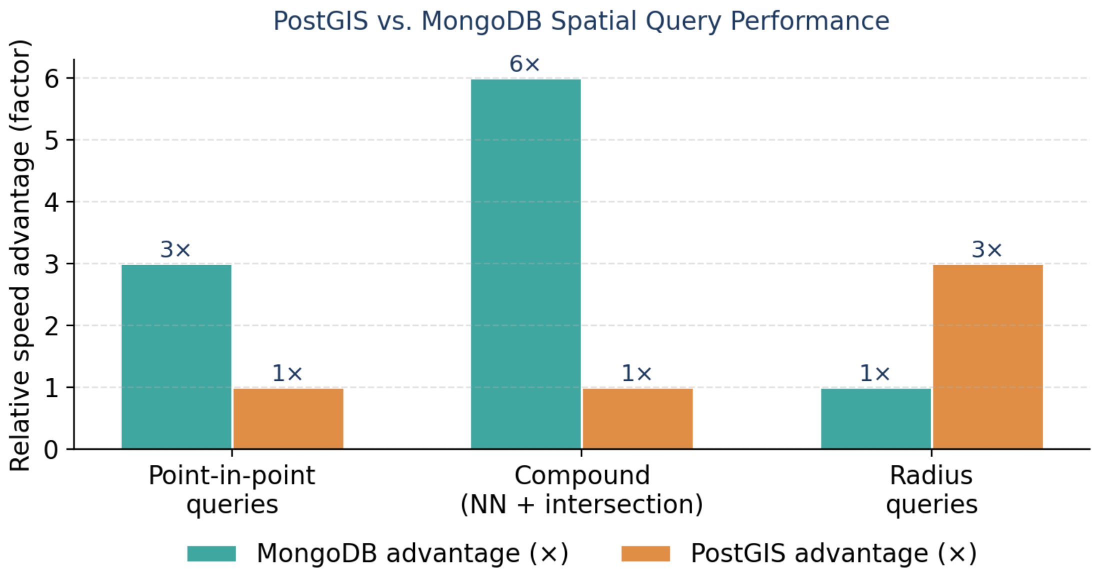

Figure 2. Relative Query-Speed Advantage of MongoDB Versus PostGIS Across Three Spatial Query Types.

Note. Advantage factors are derived from benchmarking results reported by Bartoszewski et al. (2019).

There are complementary requirements for which alternative architectures work. ChronosDB, a distributed, file-based geospatial array database management system, is designed for multidimensional raster and array data, like satellite imagery time series, separating the logical array operations from underlying distributed file system, thus offering horizontal scalability without any special storage formats (Rodrigues Zalipynis, 2018). Spatial capability is extended to linked-data environments via semantic-web-oriented platforms: Benchmarking of Resource Description Framework (RDF) stores with spatial predicates shows it is possible to integrate geospatial query into next-generation spatial data infrastructures with emphasis on interoperability over raw throughput (Huang et al., 2019). These architectures are complementary layers on top of an extensible platform; on one side, the batch frameworks for historical analyses, on the other side document or relational stores for operational queries, array databases for raster time series and semantic stores for data exchange, a taxonomy summarised in Table 1.

Table 1. Comparative Overview of Scalable Geospatial Data Platform Technologies Synthesised from References Up to 2020.

Technology	Primary Data Model	Reported Strength	Reference
SpatialHadoop	Distributed grid / R-tree indexed vector	~100× range-query speed-up vs. plain Hadoop	Eldawy & Mokbel (2015)
MongoDB (spatial)	Document-oriented, non-relational	3–6× faster on point/compound queries	Bartoszewski et al. (2019)
PostGIS / PostgreSQL	Relational, spatial extension	~3× faster on radius queries	Bartoszewski et al. (2019)
ChronosDB	Distributed file-based array DBMS	Scalable multidimensional raster/array storage	Rodrigues Zalipynis (2018)
Spatially enabled RDF stores	Semantic / linked-data graph	Interoperable next-generation SDI querying	Huang et al. (2019)
Cloud computing platforms	Elastic, on-demand compute/storage	Model runtime reduced from 12 days to 2 h	Yang et al. (2017)

Note. SDI = spatial data infrastructure. Reported strengths are drawn directly from the cited benchmarking or architectural studies and are not independently re-tested.

2.4 Cloud-Enabled Elastic Scalability

The technologies outlined above scale on demand thanks to cloud computing's elastic compute and storage substrate. Cloud computing, in its turn, helps to convert the four building blocks of big data – volume, velocity, veracity, and variety – into its fifth attribute: value, achieved by on-demand provisioning and rapid elasticity (Yang et al., 2017). An illustrative example is the modelling strategy used in the study, which is adaptive, loosely coupled, and uses lower-resolution models to identify sub-domains that require higher-resolution modelling, and then runs these higher-resolution models in parallel on cloud-provisioned clusters, limited to those sub-domains; this approach reduced a twelve-day simulation to about two hours (see Figure 3, in Yang et al., 2017). Similarly, comparing elasticity allows the ability to schedule complex computational operations, such as large-area LiDAR classification and network-wide risk simulation, to be run on demand instead of continually at peak capacity, and offers cost-efficiency implications as discussed in Section 5.

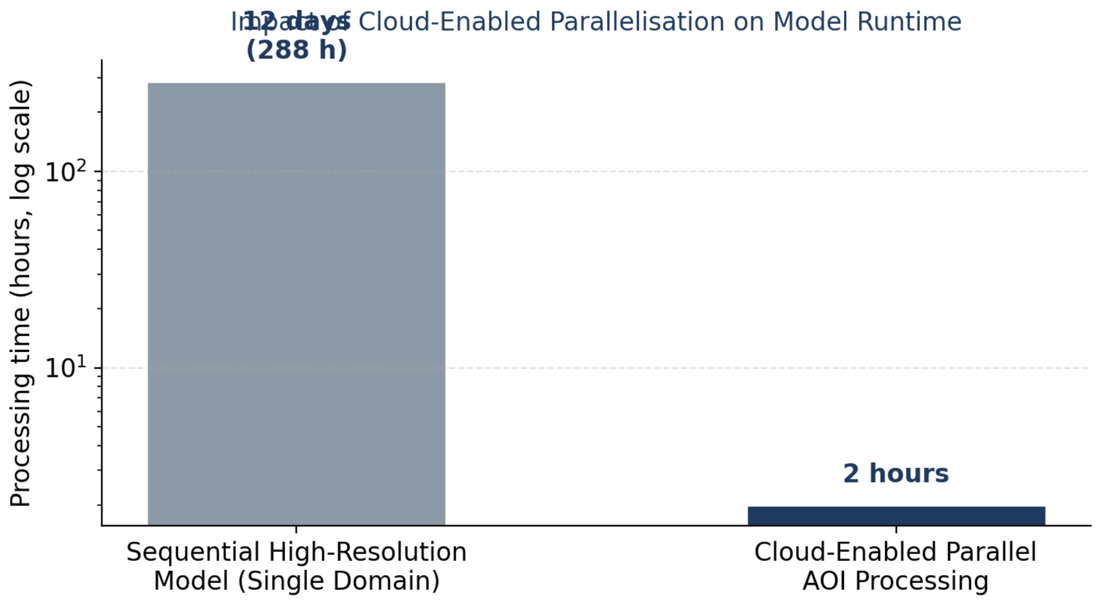

Figure 3. Impact of Cloud-Enabled Parallel Sub-Domain Processing on Simulation Runtime.

Note. Data are derived from Yang et al. (2017); values are converted to hours for direct comparison (12 days = 288 hours).

3. UAV- and LiDAR-Based Asset Extraction and Condition Monitoring

3.1 Point-Cloud Extraction of Transmission Infrastructure

The LiDAR sensors used are either airborne or UAV borne, and produce dense three-dimensional point clouds from which conductors, pylons, insulators and vegetation are automatically classified. For airborne LiDAR data acquired over urban areas, supervised classification methods using geometric and contextual features can successfully classify power-line points as power-line points and classify building returns as building returns with high accuracy (Wang et al., 2017); for airborne LiDAR data acquired over power-line corridors, voxels can be used to extract points as volumetric primitives before applying line-fitting algorithms, which can significantly reduce computational time and enhance the accuracy of the power-line points extracted (Yang & Kang, 2018). Extension of the extraction to accessory hardware, like clamps and spacers, is further achieved by semantic segmentation of the UAV images towards granular condition assessment (Wang et al., 2019). Three-dimensional reconstruction of substation LiDAR scans can also be used for automated as-built modelling of switchgear to minimize the use of manually surveyed drawings for asset registries (Wu et al., 2018).

3.2 Clearance Anomaly and Vegetation Encroachment Detection

Clearing vegetation from the right-of-way, providing a minimum safe distance between right-of-way vegetation and other vegetation or structures, is a key factor in determining outage and wildfire risk. In the automatic clearance anomaly detection method, the point clouds from UAV-borne LiDAR are preprocessed with an adaptive two-step filter to remove the terrain returns, then processed through a feature-map method to identify pylons and calculate clearance distances on the span-wise direction of the tunnel, which are compared with the clearance regulations for safety (Chen et al., 2018). It was found that the pylon-detection correctness and completeness were 96.50 percent and 94.80 percent respectively, the average clearance error was 0.08 metres and the maximum was 0.14 metres, which represents decimetre-level precision (Chen et al., 2018). As shown in Figure 4, automated LiDAR-based clearance analysis can also be near survey level accuracy, without the manpower, material, and time burden of manned aerial or ground-based inspection — especially in the mountains.

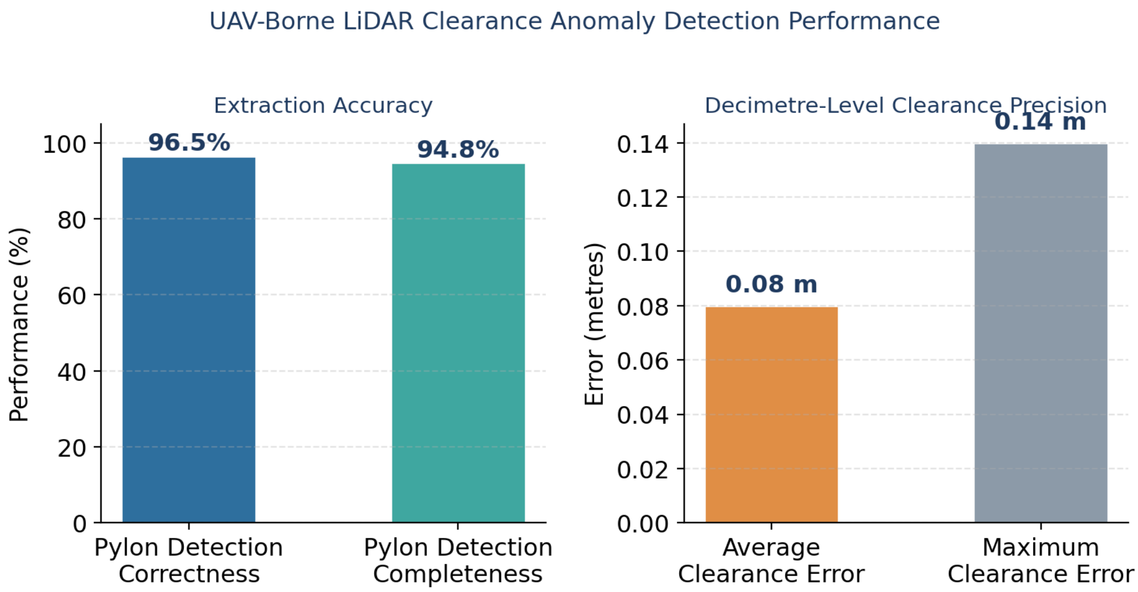

Figure 4. UAV-Borne LiDAR Clearance Anomaly Detection: Extraction Accuracy and Measurement Precision.

Note. Data are derived from Chen et al. (2018).

3.3 Convolutional Neural Network-Based Defect Detection

Along with point cloud analysis, convolutional neural network (CNN) architectures on UAV imagery provide automated detection of physical and electrical defects in powerline equipment. A dual-camera inspection system, consisting of a low resolution camera to locate insulators from long range shots, followed by a high resolution camera to detect the defects, achieved a detection recall of around ninety three percent and a precision of around ninety two percent for seventeen insulator types with cluttered backgrounds, and a dedicated defect analyser with rotation normalisation and ellipse detection achieved a defect-classification accuracy of up to ninety eight percent (Siddiqui et al., 2018). The figures in Figure 5 illustrate that deep-learning inspection can compete with manual visual inspection at a much lower cost, but this apparent accuracy may be higher for the majority of polymer insulators than for the minority of equipment, due to class imbalance, and highlight the importance of having a balanced training set.

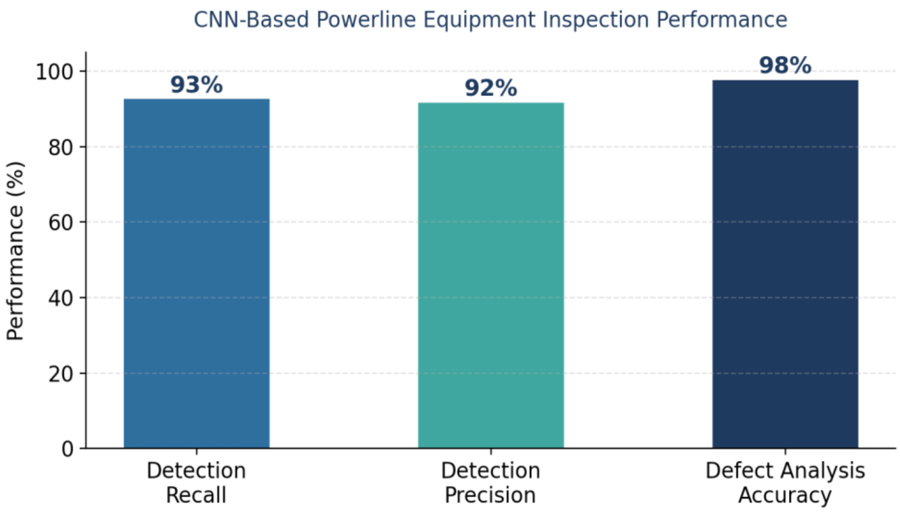

Figure 5. Detection and Defect-Classification Performance of a CNN-Based Powerline Inspection System.

Note. Data are derived from Siddiqui et al. (2018).

Table 2. Summary of UAV- and LiDAR-Based Asset Extraction and Inspection Studies.

Study	Sensing Modality	Asset Focus	Key Reported Metric
Wang et al. (2017)	Airborne LiDAR	Urban power lines	Supervised classification separating conductors from vegetation/buildings
Yang & Kang (2018)	Airborne LiDAR	Transmission lines	Voxel-based extraction improving computational efficiency
Chen et al. (2018)	UAV-borne LiDAR	ROW clearance / vegetation	96.5% correctness; 94.8% completeness; 0.08 m mean error
Wang et al. (2019)	UAV imagery	Lines & accessory hardware	Semantic segmentation of accessories
Wu et al. (2018)	LiDAR point cloud	Substation structures	Automatic 3-D as-built reconstruction
Siddiqui et al. (2018)	UAV dual-camera imagery	Insulators / equipment defects	93% recall; 92% precision; 98% defect accuracy

Note. ROW = right-of-way.

4. Spatial Risk Assessment and Hazard Exposure Modelling

4.1 Geographic Information Systems in Hazard Prioritisation

Involving geographic information system (GIS)-based approaches in disaster risk prioritisation at the community scale has been done for a long time, by overlaying hazard, exposure and vulnerability layers to determine the location of disaster risk priority where mitigation investment is needed (Armenakis & Nirupama, 2013). GIS-based landslide susceptibility mapping along a problematic section of a natural gas pipeline showed that terrain, geological and hydrological factors could be integrated into a landslide susceptibility index that would be used to help guide realignment and reinforcement decisions (Cevik & Topal, 2003). The risk of environmental damage after large-scale onshore pipeline accidents has also been evaluated by spatially overlaying pipeline routes with ecologically sensitive receptors, allowing operators to prioritize pipeline corridor sections based on the potential environmental damage, not just on probability of failure (Bonvicini et al., 2018).

4.2 Spatial Statistical Prioritisation of Pipeline Inspection

If failure-event data are available from the past, a spatial statistical approach will enable inspection resources to be deployed using estimates of risk as opposed to fixed intervals. The spatial-statistic-based approach does not need to collect any extra field data to estimate risk uncertainty, and measures the consequence of risk in each inspected region as the total cost resulting from failure events in a region (Hou et al., 2020). The approach shows that, when budgets can be spent anywhere without spatial constraints, maintenance actions focus on a small number of critical zones, and limited budgets for inspections are allocated to those parts of the network that are disproportionately responsible for the overall network risk, thereby avoiding the human judgment which was found to be a source of bias in the standard approach to prioritisation (Hou et al., 2020).

4.3 Flood Exposure of Critical Infrastructure

Flooding is one of the most widespread threats to utilities. A one-hundred year flood exposure assessment for critical infrastructure in the U.S. revealed that the distribution of critical infrastructure by exposure to the one-hundred year floodplain is highly variable by sector (2.7 to 27.1 percent exposure across the nation), with Gulf Coast states, including Louisiana and Florida, having disproportionate exposure in most sectors (Qiang, 2019). There is considerable sub-national variability that is important for risk planning that can be hidden in the national figures, as illustrated in Figure 6. This complementary utilization of synthetic aperture radar (SAR) satellite data can capture and map flood-water propagation in near real-time to provide inundation delineation that can be combined with exposure assessment that is static in nature (Rahman & Thakur, 2018).

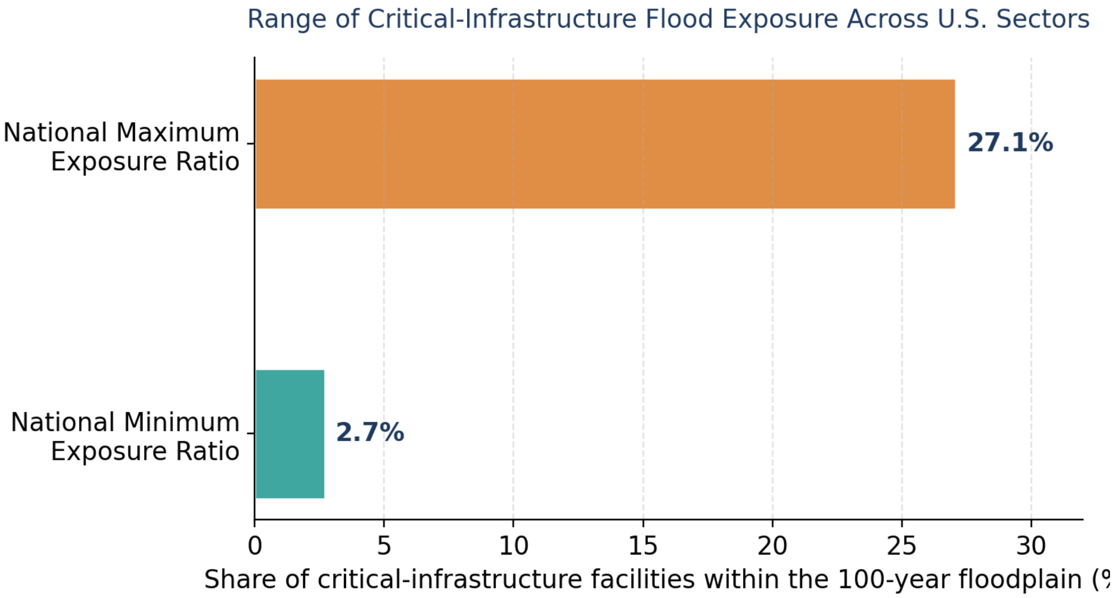

Figure 6. Range of Critical-Infrastructure Flood Exposure Ratios Across U.S. Sectors.

Note. Data are derived from Qiang (2019), based on assessment against the 100-year floodplain.

4.4 GIS Support for Water Distribution Network Management

In addition to hazard exposure, GIS platforms can also be used for routine operations of utility networks to seamlessly integrate asset inventories, network topology, and customer records into a single spatial context. Use of GIS in a city water distribution system showed that spatial query and network analysis can be applied to a water distribution network on a single platform, preventing the fragmentation of various water asset systems, water hydraulic systems and customer systems (Kadhim et al., 2021). This mirrors the wider idea that geospatial infrastructure – standards and interoperable data models – is foundational organisational infrastructure like physical utility networks (Goodchild, 2020).

Table 3. Summary of Spatial Risk Assessment and Hazard Exposure Studies.

Study	Hazard / Application	Method	Key Reported Finding
Armenakis & Nirupama (2013)	Community disaster risk	GIS overlay prioritisation	Layered hazard-exposure-vulnerability ranking
Cevik & Topal (2003)	Landslide susceptibility (pipeline)	GIS susceptibility mapping	Terrain/geology-based route risk index
Bonvicini et al. (2018)	Environmental damage (pipeline accidents)	Spatial overlay with sensitive receptors	Corridor segment ranking by consequence
Hou et al. (2020)	Pipeline network integrity	Spatial statistics on failure events	Risk concentrated in few critical zones

Study	Hazard / Application	Method	Key Reported Finding
Qiang (2019)	Flood exposure (critical infrastructure)	Hotspot analysis (Getis-Ord Gi*)	Exposure ratio range: 2.7%–27.1% by sector
Rahman & Thakur (2018)	Flood-water propagation	SAR satellite mapping	Near-real-time inundation delineation
Kadhim et al. (2021)	Water distribution management	Integrated GIS network analysis	Unified asset/hydraulic/customer platform

5. An Integrated Architecture for Utility Asset Monitoring

5.1 Layered Platform Design

Combining the technologies and methods addressed in Section 2 to 4, a scalable, geospatial platform for utility asset monitoring is envisioned as the layered architecture shown in Figure 7. Heterogeneous sensing streams, UAV and LiDAR sensing, field GPS records, weather and hazard feeds, and volunteered geographic information are accepted as input to a distributed storage layer, based on a scalable file system or object store (Eldawy & Mokbel, 2015; Goodchild, 2020). A spatial indexing layer based on grid or R-tree structures is used as an intermediary between the stored data and a query engine that supports batch and spatial SQL query processing (Eldawy & Mokbel, 2015). The image and point cloud data is processed with deep learning to extract asset geometry and defect data, which is sent to a risk assessment layer that employs spatial statistical and hazard-exposure methods to prioritize inspection and mitigation, and then finally sent to an operations dashboard that allows the risk-based inspection scheduling.The imagery and point-cloud data is processed with deep-learning to extract asset geometry and defect data, which is fed into a risk assessment layer that uses spatial statistical and hazard-exposure methods to prioritize inspection and mitigation, and ultimately into an operations dashboard for risk-based inspection scheduling.

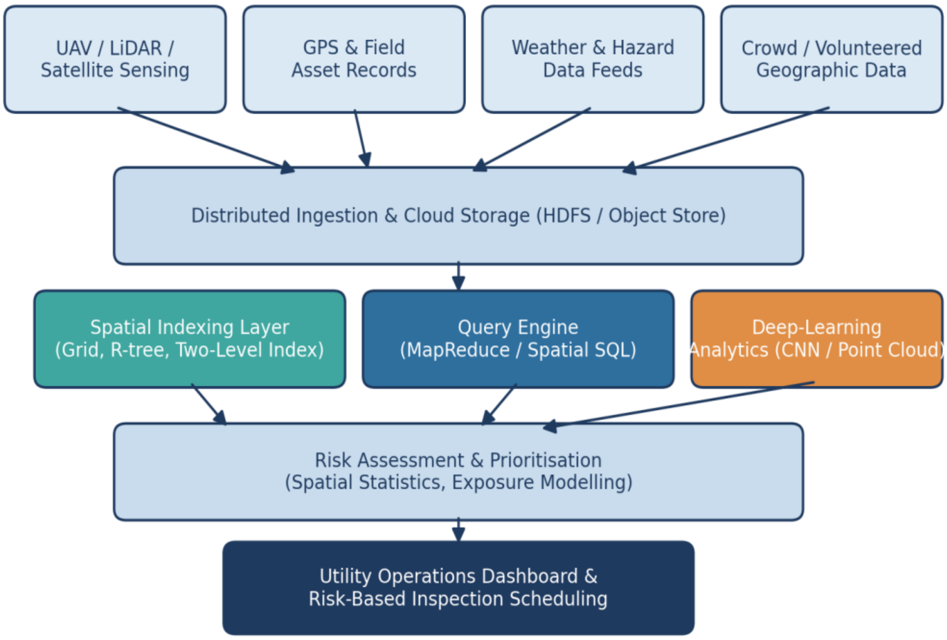

Figure 7. Integrated Layered Architecture for Scalable Geospatial Utility Asset Monitoring.

Note. Architecture synthesised from Eldawy and Mokbel (2015), Yang et al. (2017), Siddiqui et al. (2018), Hou et al. (2020), and Goodchild (2020).

5.2 Comparative and Scalability Assessment

All the main quantitative indicators reported in the literature examined are summarised in Table 4, which covers distributed query processing, database comparison, inspection accuracy and hazard exposure. Several patterns emerge. The biggest relative improvement, by orders of magnitude, comes from faster software-level indexing that is aware of the spatial properties of the data (Eldawy & Mokbel, 2015). Second, storage-engine performance is load dependent: There is no single storage engine that is better at handling all types of queries (point, compound, and radius) (Bartoszewski et al., 2019), so it is important to test against representative utility query patterns. Third, the accuracy of sensing and analytics, both clearance error and detection recall and precision, have reached a level in 2020 that can meet the requirements of replacing manual inspection in many situations rather than supplementing the work of manual inspection (Chen et al., 2018; Siddiqui et al., 2018). Fourth, there is significant variation in regional and sectoral hazard exposure, which would suggest that hazard exposure-based resource allocation would have the maximum mitigation value per dollar. (Qiang, 2019)

Table 4. Consolidated Quantitative Benchmarks Reported Across Synthesised References.

Layer	Metric	Reported Value	Reference
Distributed query processing	Range-query latency (70M objects, 20 nodes)	200 s (Hadoop) → 2 s (SpatialHadoop)	Eldawy & Mokbel (2015)
Database benchmarking	Point-in-point query advantage	MongoDB 3× faster than PostGIS	Bartoszewski et al. (2019)
Database benchmarking	Compound query advantage	MongoDB 6× faster than PostGIS	Bartoszewski et al. (2019)
Database benchmarking	Radius query advantage	PostGIS ~3× faster than MongoDB	Bartoszewski et al. (2019)
LiDAR extraction	Pylon correctness / completeness	96.5% / 94.8%	Chen et al. (2018)
LiDAR extraction	Clearance measurement error	0.08 m average; 0.14 m maximum	Chen et al. (2018)
CNN inspection	Detection recall / precision	93% / 92%	Siddiqui et al. (2018)
CNN inspection	Defect-analysis accuracy	Up to 98%	Siddiqui et al. (2018)
Hazard exposure	CI flood exposure ratio range	2.7%–27.1% (by sector)	Qiang (2019)
Cloud elasticity	Simulation runtime reduction	12 days → ~2 hours	Yang et al. (2017)

Note. CI = critical infrastructure. Values are reported as published in the cited sources and reflect distinct experimental settings; they are not directly comparable across rows.

5.3 Cost and Resource Efficiency Considerations

The synthesised references do not report any standardised monetary cost values for direct cost-benefit comparison between studies, but a few proxy indicators of resource efficiency are common across the different studies. The manpower, material and time savings of automated UAV-LiDAR clearance detection over manned aerial or ground-based inspection are reported, particularly in difficult terrain (Chen et al., 2018). Risk-based pipeline inspection prioritisation focuses on maintenance investments on a small number of high-risk areas, signifying more efficient use of a pipeline maintenance budget (Hou et al., 2020). Cloud elasticity takes fixed investment in peak capacity infrastructure and transforms it into so-called on-demand investment, as seen in the reduction of a 12-day simulation to ~2 hours using parallel sub-domain processing (Yang et al., 2017 The efficiency dimensions are summarised in Table 5.

Table 5. Resource-Efficiency Dimensions Associated with Platform Components.

Architectural Layer	Efficiency Dimension	Reported Effect	Reference
Sensing / inspection	Labour and time substitution	Reduced manpower, materials, and time vs. manned/ground inspection	Chen et al. (2018)
Risk prioritisation	Budget concentration	Maintenance focused on minority of high-risk zones	Hou et al. (2020)
Compute infrastructure	Elastic provisioning	Runtime reduced from 12 days to ~2 hours	Yang et al. (2017)
Storage engine selection	Workload-matched querying	Up to 6× speed differential depending on query type	Bartoszewski et al. (2019)
Data integration	Fragmentation reduction	Unified asset, hydraulic, and customer data platform	Kadhim et al. (2021)

6. Discussion: Challenges, Governance, and Future Directions

6.1 Interoperability and Standardisation

One thing that stood out throughout the synthesised literature is the challenge of balancing performance optimisation for the platform and cross-organisational interoperability. The use of specialised distributed frameworks and array databases provides significant performance improvements, at least partly due to the data models and indexing schemes used for the data, but it can become difficult to exchange data between operators, regulators and emergency management agencies without the use of standardised distributed interfaces (Eldawy & Mokbel, 2015; Rodrigues Zalipynis, 2018). Spatial enabled semantic-web technologies such as RDF stores are expected to be a partial solution, emphasizing interoperability, a key characteristic of next-generation spatial data infrastructure (SDI) (Huang et al., 2019; Goodchild, 2020).

6.2 Data Quality, Model Bias, and Verification

The vulnerabilities of deep-learning inspection systems are passed on from training data. A classic example of reported class imbalance—the ratio of the number of majority equipment types to training samples—is that of headline recall and precision values that underrepresent the lower performance of rarer but more serious defect categories (Siddiqui et al., 2018). The combination of automatic and manual field surveys (as used in LiDAR clearance validation) continues to be a valuable precautionary measure in the move from experimental to operational deployment (Chen et al., 2018).

6.3 Regulatory and Ethical Considerations

Pipeline risk assessment methods have regulatory implications in that they are used to evaluate risk to pipeline infrastructure related to floodplain management and infrastructure siting after major disaster events (Qiang, 2019). High exposure levels in certain areas and industries create equity concerns in terms of mitigation investments, as uniform national-level mitigation policy may have disproportionate impact on certain communities. Likewise, data governance issues that arise when fine-grained UAV imagery and GPS telemetry are integrated across corridors that span both private and public property can only be covered indirectly by the technical literature summarised here (Goodchild, 2020).

6.4 Future Directions

In the field of the synthesised evidence, various pathways can be identified. The gap between Batch and Interactive query processing is likely to be small with continued development of two-level and learned spatial indexes (Eldawy & Mokbel, 2015). Joint hazard-exposure and asset-condition queries will be supported by a unified platform based on the integration of array-based storage for raster time series with vector-oriented asset databases (Rodrigues Zalipynis, 2018). Spatial infrastructure, especially expansion of semantic spatial infrastructure, is poised to enable enhanced interoperability for disaster response across agencies (Huang et al., 2019; Goodchild, 2020). Finally, to scale up spatial statistical risk methodology, which has been validated so far for pipelines and flood exposure, to other hazard types is a logical next step (Hou et al., 2020).

7. Conclusion

This paper is a synthesis for scalable geospatial data platforms and their utility for utility asset monitoring and risk mitigation. The evidence shows that when it comes to query latency, spatially aware distributed processing frameworks can deliver about two orders of magnitude lower latency compared with traditional big-data engines – a workload-dependent difference that depends on the specific type of queries being processed – and that combining UAV- and LiDAR-based inspection with deep-learning analytics can achieve, at best, the same reliability as human inspection, while simultaneously lowering labour and time costs; and that, when it comes to managing finite resources for hazard mitigation, spatially aware distributed processing frameworks can allow operators to focus those resources on portions of the pipeline that contribute disproportionately to overall risk. A layered architecture that brings together these elements, from ingestion to storage, indexing, analytics, and risk prioritisation, ensures a unified approach to how the spatial data is used to inform decisions. Despite the ongoing issues with interoperability, model verification and governance, technical progress is not enough and continued focus on standardisation and equitable resource access will be needed to achieve the maximum mitigation capacity on these platforms.

References

[1] Armenakis, C., & Nirupama, N. (2013). Prioritization of disaster risk in a community using GIS. Natural Hazards, 66(1), 15–29. https://doi.org/10.1007/s11069-012-0167-8

[2] Bartoszewski, D., Piórkowski, A., & Lupa, M. (2019). The comparison of processing efficiency of spatial data for PostGIS and MongoDB databases. In S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, & D. Kostrzewa (Eds.), Beyond databases, architectures and structures. Paving the road to smart data processing and analysis (pp. 291–302). Springer. https://doi.org/10.1007/978-3-030-19093-4_22

[3] Bonvicini, S., Antonioni, G., & Cozzani, V. (2018). Assessment of the risk related to environmental damage following major accidents in onshore pipelines. Journal of Loss Prevention in the Process Industries, 56, 505–516. https://doi.org/10.1016/j.jlp.2018.11.005

[4] Cevik, E., & Topal, T. (2003). GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). Environmental Geology, 44(8), 949–962. https://doi.org/10.1007/s00254-003-0838-6

[5] Chen, C., Yang, B., Song, S., Peng, X., & Huang, R. (2018). Automatic clearance anomaly detection for transmission line corridors utilizing UAV-borne LiDAR data. Remote Sensing, 10(4), 613. https://doi.org/10.3390/rs10040613

[6] Eldawy, A., & Mokbel, M. F. (2015). SpatialHadoop: A MapReduce framework for spatial data. In 2015 IEEE 31st International Conference on Data Engineering (pp. 1352–1363). IEEE. https://doi.org/10.1109/ICDE.2015.7113382

[7] Goodchild, M. F. (2020). Building geospatial infrastructure. Geo-spatial Information Science, 23(1), 1–9.

https://doi.org/10.1080/10095020.2019.1698274

[8] Hou, P., Yi, X., & Dong, H. (2020). A spatial statistic based risk assessment approach to prioritize the pipeline inspection of the pipeline network. Energies, 13(3), 685. https://doi.org/10.3390/en13030685

[9] Huang, W., Raza, S. A., Mirzov, O., & Harrie, L. (2019). Assessment and benchmarking of spatially enabled RDF stores for the next generation of spatial data infrastructure. ISPRS International Journal of Geo-Information, 8(7), 310. https://doi.org/10.3390/ijgi8070310

[10] Kadhim, N. R., Abdulrazzaq, K. A., & Hashim, A. (2021). The management of water distribution network using GIS application case study: AL-Karada area. Journal of Physics: Conference Series, 1895, 012038. https://doi.org/10.1088/1742-6596/1895/1/012038

[11] Qiang, Y. (2019). Flood exposure of critical infrastructures in the United States. International Journal of Disaster Risk Reduction, 39, 101240. https://doi.org/10.1016/j.ijdrr.2019.101240

[12] Rahman, M. R., & Thakur, P. K. (2018). Detecting, mapping and analysing of flood water propagation using synthetic aperture radar (SAR) satellite data and GIS: A case study from the Kendrapara District of Orissa State of India. The Egyptian Journal of Remote Sensing and Space Sciences, 21(S1), S37–S41. https://doi.org/10.1016/j.ejrs.2017.10.002

[13] Rodriges Zalipynis, R. A. (2018). ChronosDB: Distributed, file based, geospatial array DBMS. Proceedings of the VLDB Endowment, 11(10), 1247–1261. https://doi.org/10.14778/3231751.3231754

[14] Siddiqui, Z. A., Park, U., Lee, S., Jung, N., Choi, M., Lim, C., & Seo, J. (2018). Robust powerline equipment inspection system based on a convolutional neural network. Sensors, 18(11), 3837. https://doi.org/10.3390/s18113837

[15] Wang, L., Chen, Z., Hua, D., & Zheng, Z. (2019). Semantic segmentation of transmission lines and their accessories based on UAV-taken images. IEEE Access, 7, 80829–80839. https://doi.org/10.1109/ACCESS.2019.2923024

[16] Wang, Y., Chen, Q., Liu, L., Zheng, D., Li, C., & Li, K. (2017). Supervised classification of power lines from airborne LiDAR data in urban areas. Remote Sensing, 9(8), 771. https://doi.org/10.3390/rs9080771

[17] Wu, Q., Yang, H., Wei, M., Remil, O., Wang, B., & Wang, J. (2018). Automatic 3D reconstruction of electrical substation scene from LiDAR point cloud. ISPRS Journal of Photogrammetry and Remote Sensing, 143, 57–71. https://doi.org/10.1016/j.isprsjprs.2018.04.024

[18] Yang, C., Yu, M., Hu, F., Jiang, Y., & Li, Y. (2017). Utilizing cloud computing to address big geospatial data challenges. Computers, Environment and Urban Systems, 61, 120–128. https://doi.org/10.1016/j.compenvurbsys.2016.10.010

[19] Yang, J., & Kang, Z. (2018). Voxel-based extraction of transmission lines from airborne LiDAR point cloud data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(11), 3892–3904. https://doi.org/10.1109/JSTARS.2018.2869542