

# The Noblis Data Lakehouse

Delivering faster, better insights and accelerating AI/ML growth



## Challenge

Organizations need data management—ingestion, transformation, storage and serving—to fully leverage their data. Separate from relational database management systems, which are optimized for online transaction processing (such as inserting, updating or deleting single rows), data platforms are optimized for analysis. As the complexity of data management has grown, organizations have adopted a variety of different data platform architectures to make sure they are getting the full value of the data they collect, but these systems are often fraught with issues that include reliability, data staleness, high Total Cost of Ownership, and limited support for AI and machine learning (ML). With the ever-growing volume of unstructured data (Figure 1) and use of ML, the need for an optimal solution is greater than ever.

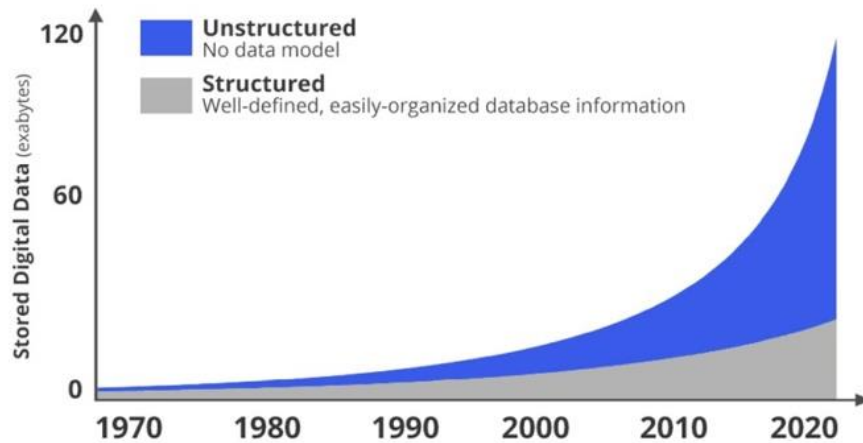


Figure 1: Growth comparison of structured and unstructured data over time.

## Our Solution

As a component of Noblis' AI Stack, the Noblis Data Lakehouse has a scalable data infrastructure that provides replication, consistency and reliability; an efficient and scalable engine for data processing; and support for faster analytic queries and the ML development lifecycle. We have developed this capability by bringing together a combination of open-source and custom-developed software solutions and through development of ingestion pipelines for multiple Noblis-owned and publicly available data sources (Figure 2).

### How it Works

The data is stored in the Delta Table format using a trusted, industry standard object storage solution. Delta Table is a secure Parquet file format that is an encoded layer over data, providing guarantees that help maintain data integrity. Data is then processed using a highly scalable engine. End-users can access the data using multiple user-interfaces which are commonly used by organizations. MLFlow tracks ML experimentation so that model training can be optimized and trained models can be easily found for different use cases. Moreover, Noblis has developed streamlined processing for mitigating bias in ML training data.

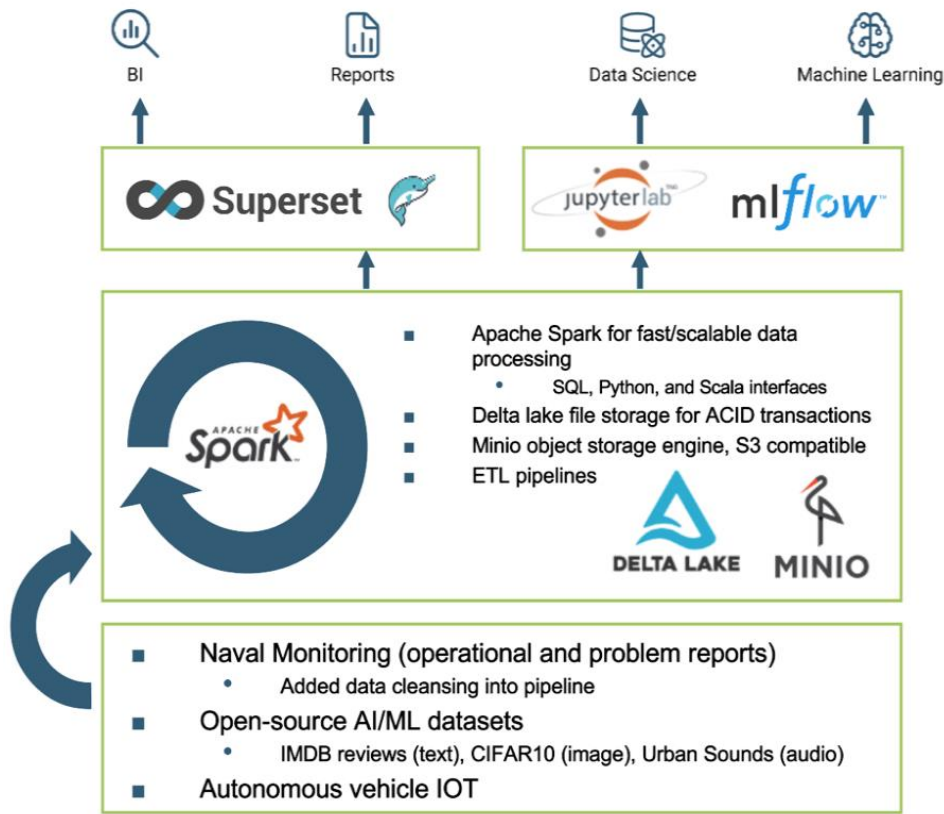


Figure 2: Data Lakehouse structure.

**Use Cases**

Federal government systems routinely need to store large amounts of structured and unstructured data every day and then access that data to glean meaningful insights. For example, law enforcement agencies routinely collect data from a variety of sources, including images and videos; smart cities collect, maintain and use data from sources such as water systems, electric grids and traffic control systems and some military systems, such as those used by naval fleets, present the unique challenge of relying on sporadic and low-bandwidth communications to gather large amounts of data. The Noblis Data Lakehouse can improve systems like these by providing a secure, easy-to-use and reliable platform to consolidate, store and access different types of data in a user-friendly interface to help agencies gain insights from their data more quickly and easily.

**In Conclusion**

The Noblis Data Lakehouse provides a single place for analysts to gather insights and for data scientists to develop ML algorithms. Compared to the data warehouse format, the Noblis Data Lakehouse allows organizations to better leverage unstructured data; compared to the data lake format, it provides faster analysis and more stringent data quality checks. As the amount of data gathered and analyzed by U.S. government systems grows, the need for a data platform like Noblis Data Lakehouse increases.

## Customer Success Story: Naval Fleet Readiness

Providing U.S. Navy vessels with spare parts is essential to maintaining their operational readiness. Meeting spare part demand can be challenging due to the large amount and siloed nature of ship data. To combat these challenges, Noblis created a dashboard for landing, platform, dock (LPD) class ship data. The Noblis Data Lakehouse allows ship maintenance experts to glean insights more quickly into this data by providing a centralized data store, ingestion and processing capabilities and the ability to build visualization dashboards such as the one shown below in figure 3.

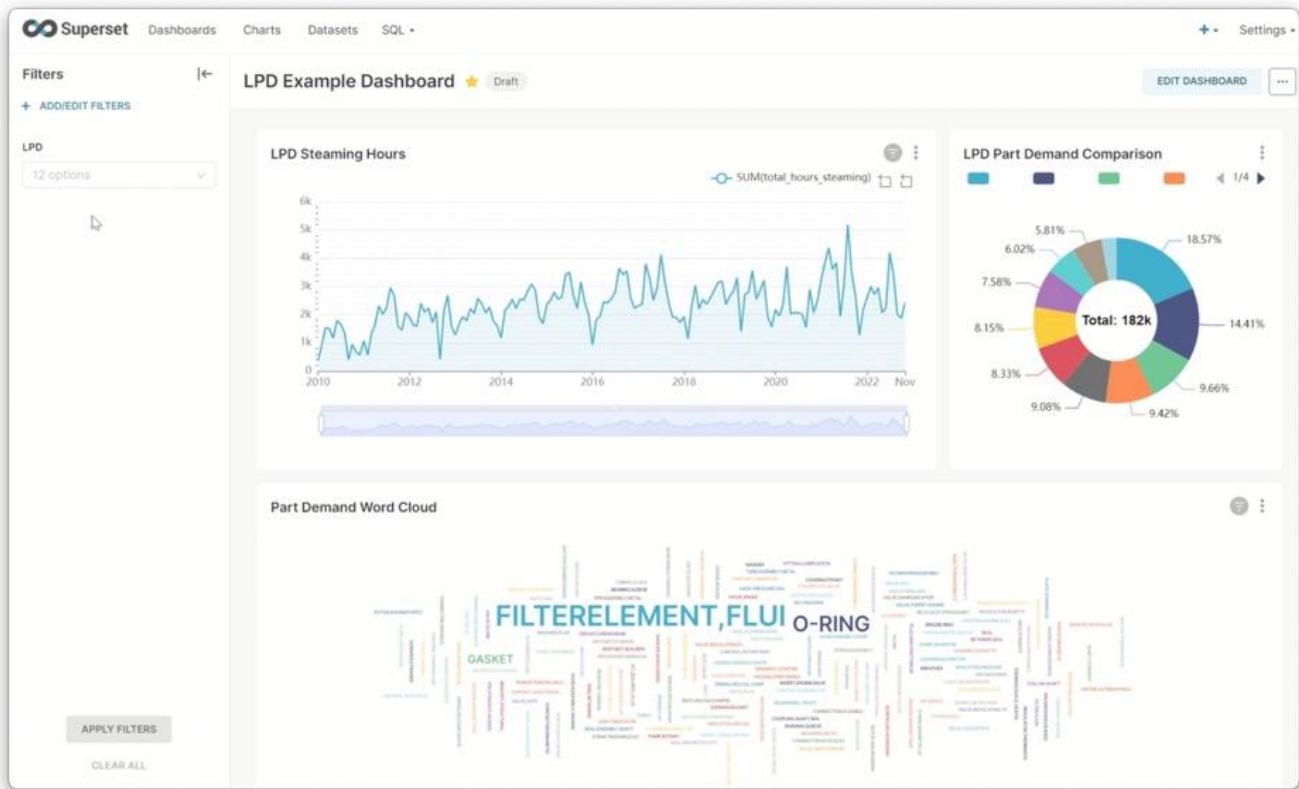


Figure 3: Demo dashboard built with Noblis Data Lakehouse.

## About Noblis

---

We exist to enrich lives and make our nation safer with our shared passion for excellence and innovation.

For more than 25 years, Noblis has been an innovator within the federal government, committed to solving the challenges of today and investing in the mission of tomorrow. As a nonprofit, Noblis works for the public good, bringing together the best possible capabilities, including science and technology expertise and solutions, in an environment of independence and objectivity to deliver enduring impact on federal missions.

## Working with Us

Government agencies can access Noblis through a variety of contracting mechanisms. We have several contracts in place and available to government agencies. We are also a GSA Schedule holder.

For a full list of vehicles, visit [noblis.org/contracting](https://noblis.org/contracting) or call us at 703.610.2000. Email us at [answers@noblis.org](mailto:answers@noblis.org).

## Technical Points of Contact

---

### **Nathan Spivy**

[Nathan.Spivy@noblis.org](mailto:Nathan.Spivy@noblis.org)

### **Matthew Kersting**

[Matthew.Kersting@noblis.org](mailto:Matthew.Kersting@noblis.org)

### **Nick Barlow, Ph.D.**

[Nicholas.Bartlow@noblis.org](mailto:Nicholas.Bartlow@noblis.org)