GE VERNOVA

WHITEPAPER

GridOS® DATA FABRIC
THE KEY TO UNLOCK GRID ORCHESTRATION

Discover, govern, and **utilize** energy data to orchestrate a reliable, resilient, and more sustainable grid.

GridOS® Orchestration Software

EXECUTIVE SUMMARY

With the energy transition, data has never been a more complex challenge for utilities. It's bigger, generates faster, and is more dispersed than ever before.

And yet, underneath all these complexities, data has also never held more opportunity for utilities.

The key is to harness it. More accessible energy data can help utilities embrace change to optimize resources, manage disruptions, and enable renewables at scale.

But for many utilities, that's easier said than done. Energy data often sits in IT, OT, and external silos, making it difficult for applications to leverage this energy data and for operators to access it.

What if you could improve your ability to **discover, govern,** and **utilize** data to make better decisions, faster?

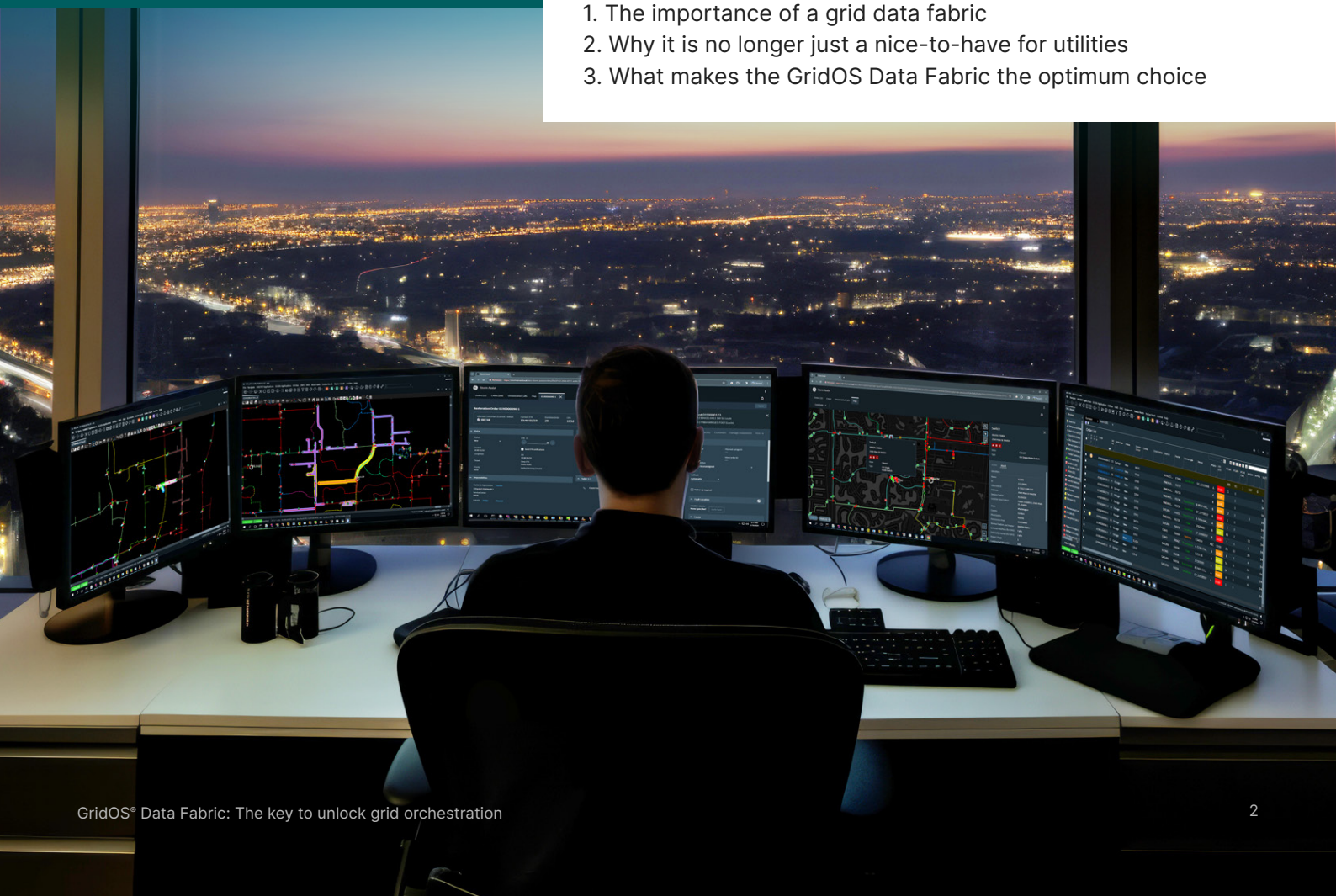
A solution exists to bridge siloed data, and it's called **GridOS® Data Fabric.**

GridOS Data Fabric is the grid data management layer of GE Vernova's GridOS platform, designed specifically for grid orchestration. GridOS Data Fabric helps utilities combine and contextualize internal and external data, pulling disparate information into a unified view capable of spanning transmission, distribution, and the edge.

This centralized view of decentralized data can give grid operators a more accurate look at the network, while access to shared data can power more effective artificial intelligence (AI)- and machine learning (ML)-driven applications for grid automation.

This whitepaper will explore:

1. The importance of a grid data fabric
2. Why it is no longer just a nice-to-have for utilities
3. What makes the GridOS Data Fabric the optimum choice



WHY THE GRID NEEDS A DATA FABRIC?

Grid data challenges

Grid data has never presented more challenges to the utilities that produce, own, and leverage it. Based on our domain observations, the following characteristics of modern grid data tend to cause the biggest headaches for grid operators:

1 Bigger.

Grids are now generating more data than ever before. There are two factors contributing to this exponential increase in grid data:

- **IoT-driven digitization.** Grids have added a large number of internet-connected sensors to their networks in recent years, each one recording a huge volume and variety of data.
- **Distributed energy resource (DER) growth.** Millions of DERs are currently connected to the grid and many thousands more are added every year. Each DER adds to the mountain of data utilities need to manage the grid.

2 Faster.

Data generation has increased in terms of frequency. Previously, grid sensors generated data over longer intervals. Today, grid data is generated in near-real time, and the insights need to be leveraged just as quickly. This is especially important when facing modern grid challenges and violations like backfeeds, faults, and voltage issues. In order to avoid any negative impacts, utilities need to identify violations the moment they appear in the data, and take corrective action just as quickly.

3 More volatile.

Grid data is becoming more volatile, especially with the growth of modern assets like DERs. DERs are not as tightly connected to the control room as SCADA connected assets, resulting in volatile data streams.

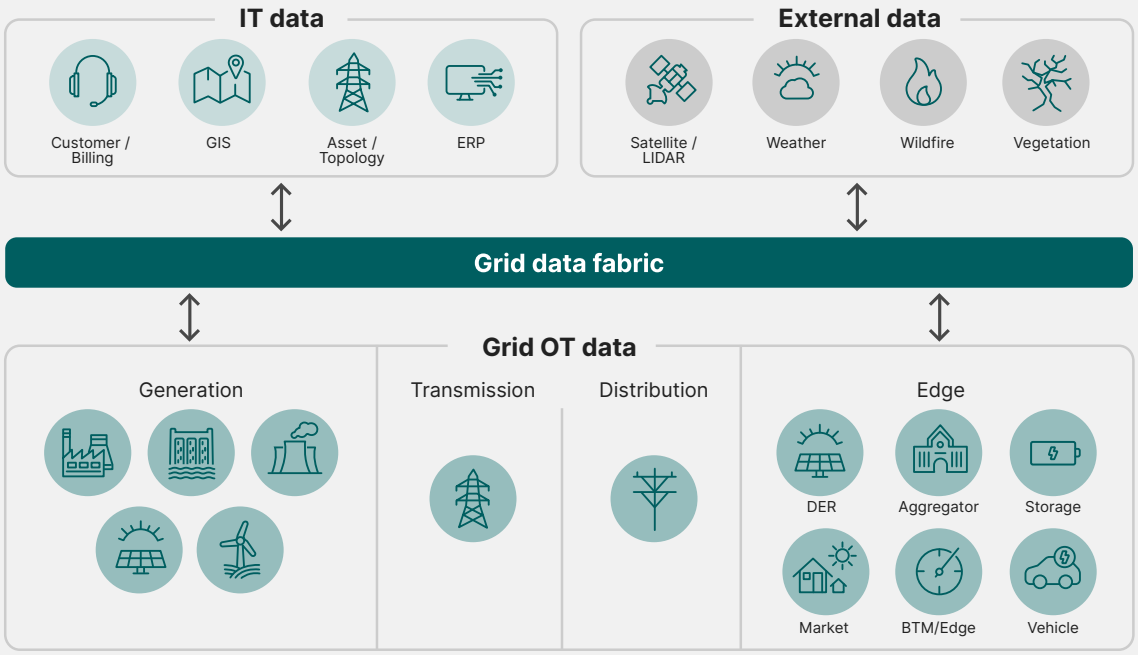
4 Siloed.

Perhaps the single biggest challenge associated with grid data is the fact that much of it rests in silos. Most grid data is split between three main silos:

- OT data
- IT data
- External data

But those are not the *only* data silos utilities must contend with. For example, grid OT data is often siloed further across generation, transmission, distribution, and the grid edge. Similarly, silos are a major cause of "dark data," or data that is unutilized and/or unknown. Many utilities are sitting on a wealth of information without even realizing it's there – simply because data silos obscure visibility to it. Silos make it difficult for utilities to access the information they need to deploy advanced applications and unlock intelligent use cases – both of which are crucial for modern grid management.

Distributed grid data necessitates a grid data fabric



5 Distributed.

Grid data is what data scientists call "distributed" in nature. In other words, grid data is generated and stored all over utility organizations, with various business units, applications, and functions owning their own unique data sets. Data may also come from outside the organization, as in the case of weather and wildfire data.

During the 2010s, many utilities attempted to bridge silos and resolve the problem by "centralizing" their grid data in data warehouses or data lakes. Many of these attempts struggled – simply because grid data is distributed by nature, and not easily centralized. The side effect of these centralization attempts was that data owners lost control of their data, and quality and accuracy suffered as a result. Critical grid processes like grid orchestration, AI- and ML-powered technologies, and advanced grid control applications need high-quality data to work effectively. Thus, most grid data remains distributed – and by extension, difficult to centralize.

6 Heterogenous.

There are numerous grid assets of many different types, each one generating its own data. As a result, grid data can come in a staggering array of different formats and data types. This extensive variability makes grid data incredibly heterogenous and difficult to work with for operators and applications alike.

But despite the challenges of identifying, collecting, and leveraging energy data, there's no questioning its crucial importance to the modern utility. Within this data lies the potential for utilities to unlock data-driven decision making, grid automation, and intelligent use cases.

This means that utilities absolutely must find a solution that enables them to access and utilize any type of data, exactly when they need it.

The ideal solution must be able to:

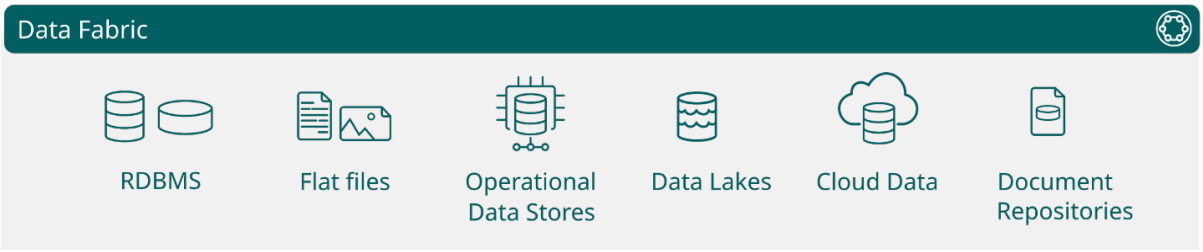
- Bring together disparate data sources from across the enterprise to provide a unified view
- Bridge data silos at any level
- Respect the distributed nature of grid data (i.e. not try to centralize it)
- Enable data-driven automation and coordinated decision making across transmission, distribution, and the edge to manage multi-directional energy flows.

Data fabric 101

At its most basic level, a data fabric is a data management solution that enables integration and data access across disparate data sources – within the enterprise and beyond. It plays a critical role in helping users "see" across data silos – and by extension, discover, govern, and utilize the data within them.

It is important to note that a data fabric does not store data – it should not be confused with a data lake or data warehouse. Rather, a data fabric can best be thought of as a foundation that brings together decentralized data of all types into a centralized view that is virtual. Such a distributed data foundation makes it easy for applications, programs, and users to access and leverage the right data – any type, any format, from any source, at any time.

Schematic illustration of a data fabric and diverse data sources



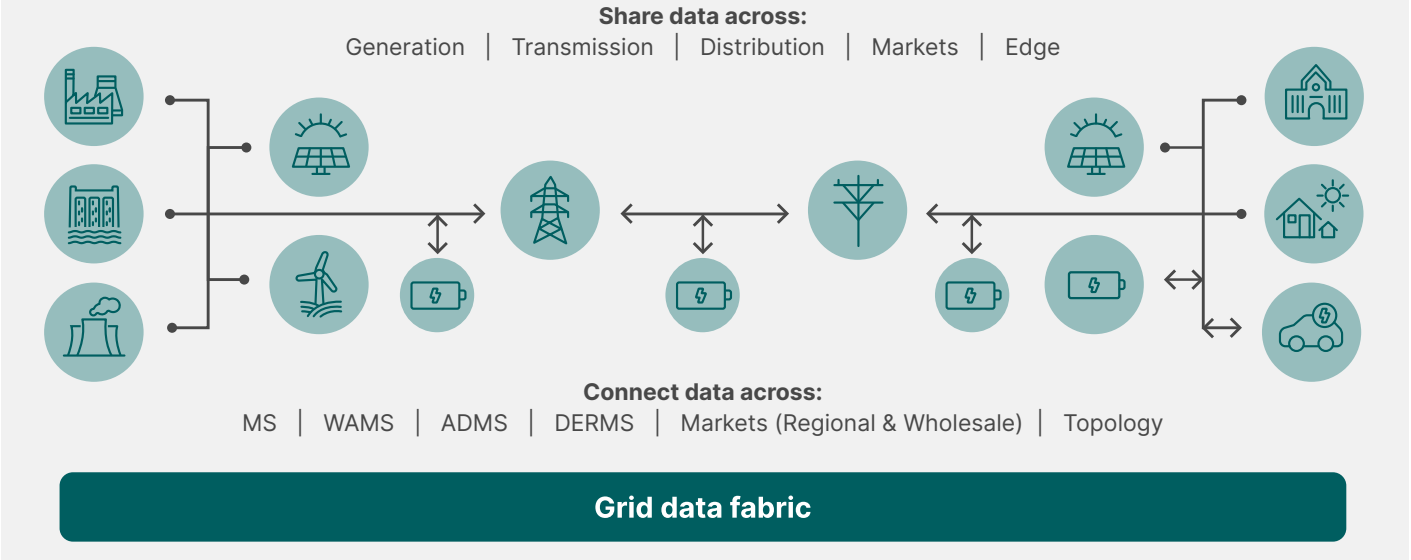
Data fabric for grids

Every grid can benefit from a grid data fabric. A grid data fabric, as the name suggests, is engineered to accommodate the unique challenges, characteristics, and uses of grid data. A grid data fabric helps utilities by:

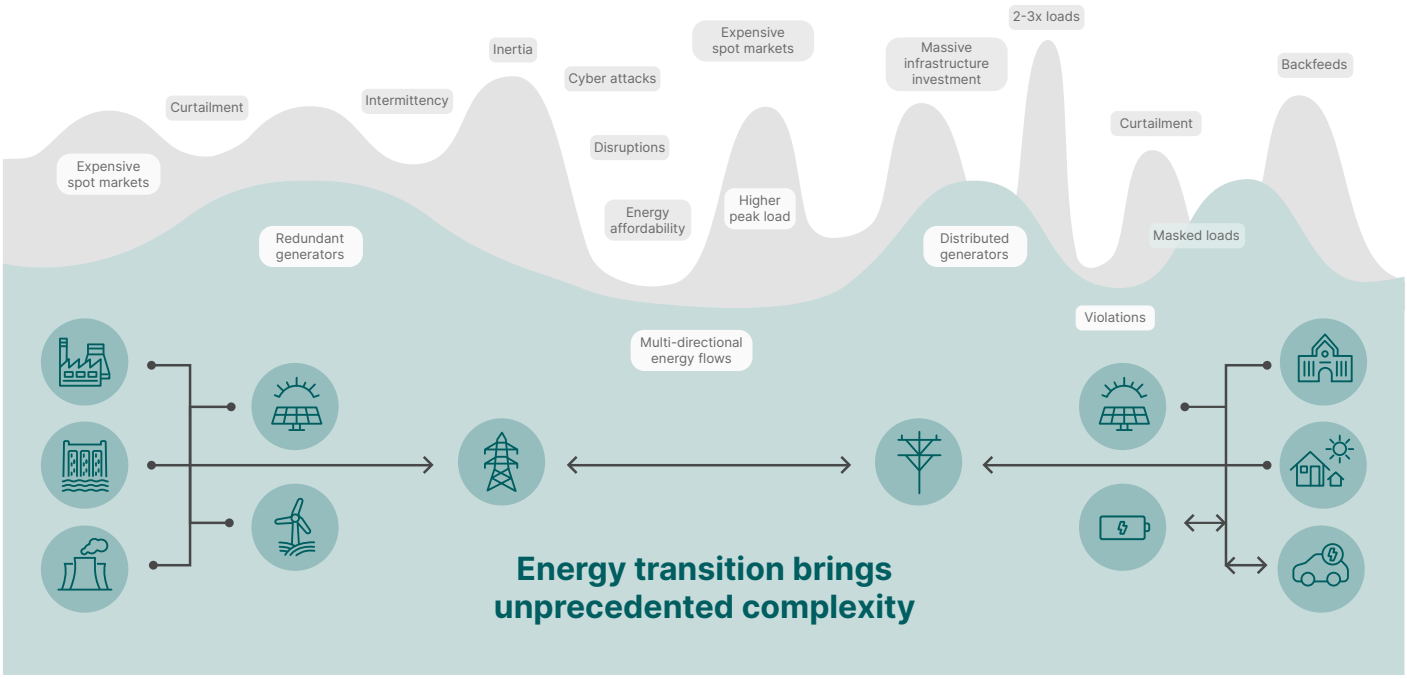
- Securely and seamlessly connecting data between grid applications across OT and IT, such as EMS, ADMS, and DERMS, energy market management, and asset management.
- Facilitating the controlled sharing of data between grid participants across generation, transmission, distribution, the edge, market operators, aggregators, and DER operators.
- Creating a data foundation for AI and ML to drive grid automation use cases, as robust training data is critical for leveraging them effectively (see "Grid data fabric use cases" section, example #1)

Overall, a grid data fabric ensures that data can be more easily found, accessed, and leveraged by applications, programs, and users. This improved access to data helps unlock data-driven automation and coordinated decision-making across transmission, distribution, and the edge to manage multi-directional energy flows.

Building a distributed grid data system



Grid data fabric and the energy transition



Today's grid is dramatically different than the grid of the past (and will become more so in the future). For instance, the load and consumption profile is much more difficult to predict. This can be attributed to increased DER generation on the distribution side and much higher peak loads due to EV charging. Those two challenges often result in violations, masked loads, DER curtailment, backfeeds, and other disruptions.

Additional challenges stem from the higher proportion of renewable power generation on the transmission side. Both wind and solar energy sources are intermittent, and solar further complicates inertia management due to its non-spinning nature. These characteristics may result in curtailment and having to activate costly backup generators. Moreover, severe weather has become more common than ever before, putting new pressures on grids to increase resilience and keep the lights on.

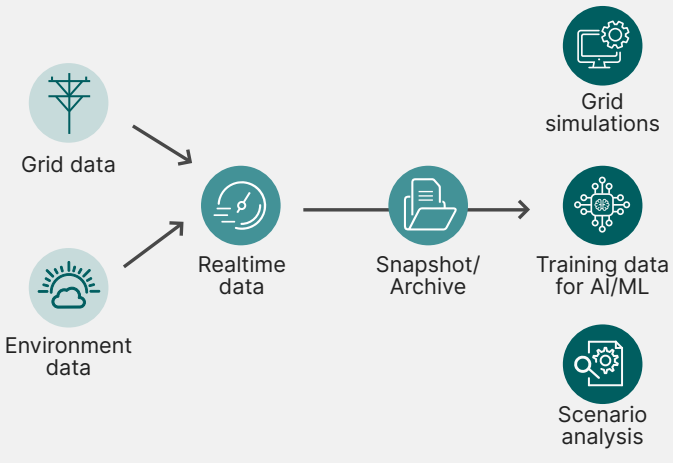
Traditional grid management is no longer adequate for dealing with such immense challenges. **Grid orchestration** is the new way forward.

To orchestrate the grid – including managing renewables at scale, keeping track of millions of DERs, and proactively managing disruptions from climate change – utilities need to unlock the power of data with a grid data fabric. Only then can they access and leverage all the data they need to make grid orchestration a reality.

Grid orchestration- It's the process of coordinating actions across multiple systems, leveraging integration, automation, and modern technologies to streamline and optimize the flow of electricity from generation through consumption.

It offers a proactive approach to drive stability in a complex environment resulting from increased penetration of renewables and severe weather, cyber threats, and other disruptions.

Planning use case data foundation



Grid data fabric use cases for utilities

Let's look at use cases that can benefit from a grid data fabric:

1 Planning

For proper scenario planning, utilities must satisfy two needs. First, they need to bring together real-time grid OT data with sensor, environment, and IT data – each of which typically resides in different places.

Second, utilities must also be able to snapshot or archive their real-time grid and environment data. To understand the importance of this, consider that AI- and ML-powered grid planning applications add value by making predictions about the future.

They make those predictions by studying past events, and the conditions and factors that made them happen. Thus, any utility that wants to leverage AI and ML must have an archive of historical "training" data from which the technologies can learn. A grid data fabric can archive internal and external data to provide this training data source."

A data fabric can also use this data to help with other planning use cases like scenario analyses and grid simulations.

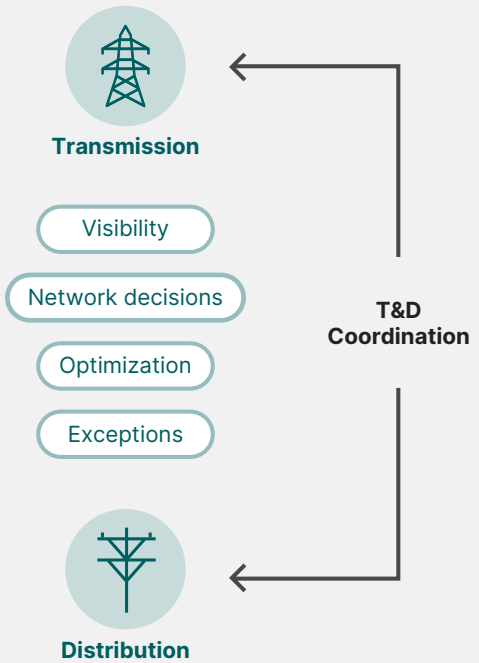
2 Real-time transmission and distribution (T&D) integration

Orchestrating a more sustainable energy grid requires gaining a systemwide view and coordinated operation across T&D and all the way to the edge to manage the multidirectional flow of energy. Only then can utilities unlock advanced, real-time T&D use cases like:

- Outage response and restoration
- Emergency response (e.g. load shed and distributed black start)
- Coordinated Volt/VAR optimization
- Planned outage scheduling
- Operational situational awareness
- Integrated switching

To do so, utilities need to bridge data silos and connect disparate applications and systems like EMS, ADMS, and DERMS. And that can be achieved much more efficiently with a grid data fabric.

Real-time T&D integration use case



3 DER look-ahead analysis and optimization

Utilities with situational awareness of DERs can use forecasting data and grid OT data to enable look-ahead analysis. This is incredibly valuable to modern utilities with high DER penetration, as it allows them to anticipate future grid conditions and behaviors.

For example, on very sunny days, grids with high DER penetration are at risk of backfeed violations, caused by abnormally high voltage levels due to excess generation. Look-ahead analysis enables grid operators to predict potential violations and drive future optimization decisions to avert them. With access to the right data, look-ahead analysis engines can determine the most appropriate optimization measures. The engines can even take into account operational and economic data plus DER program considerations to ensure feasibility of actions and minimize the total cost of operation. With this techno-economic optimization, utilities can schedule DER dispatch decisions in advance, preventing potential future violations and taking full advantage of DER flexibility.

Look-ahead analysis use cases like the above can only be realized by integrating DER asset and program data, forecasting data, grid OT data, economic and market data, and settlement data. This data integration can be best achieved with a grid data fabric, which can connect and provide access to all these data sources better and faster than current manual and/or custom methods.

GridOS®
Data Fabric

> The GridOS Data Fabric is the grid-specific data management layer of the GridOS platform, designed specifically for grid orchestration.

Grid decisions require utilizing data from multiple parts of the energy ecosystem. The GridOS Data Fabric enables the discovery, governance, and utilization of highly distributed data that is increasingly critical for grid orchestration.

The GridOS Data Fabric is designed to:

- Integrate disparate data sources
- Cleanse, pre-process, contextualize and harmonize data
- "Annotate" data with metadata – essentially qualifying data based on unique attributes for easier discovery and governance

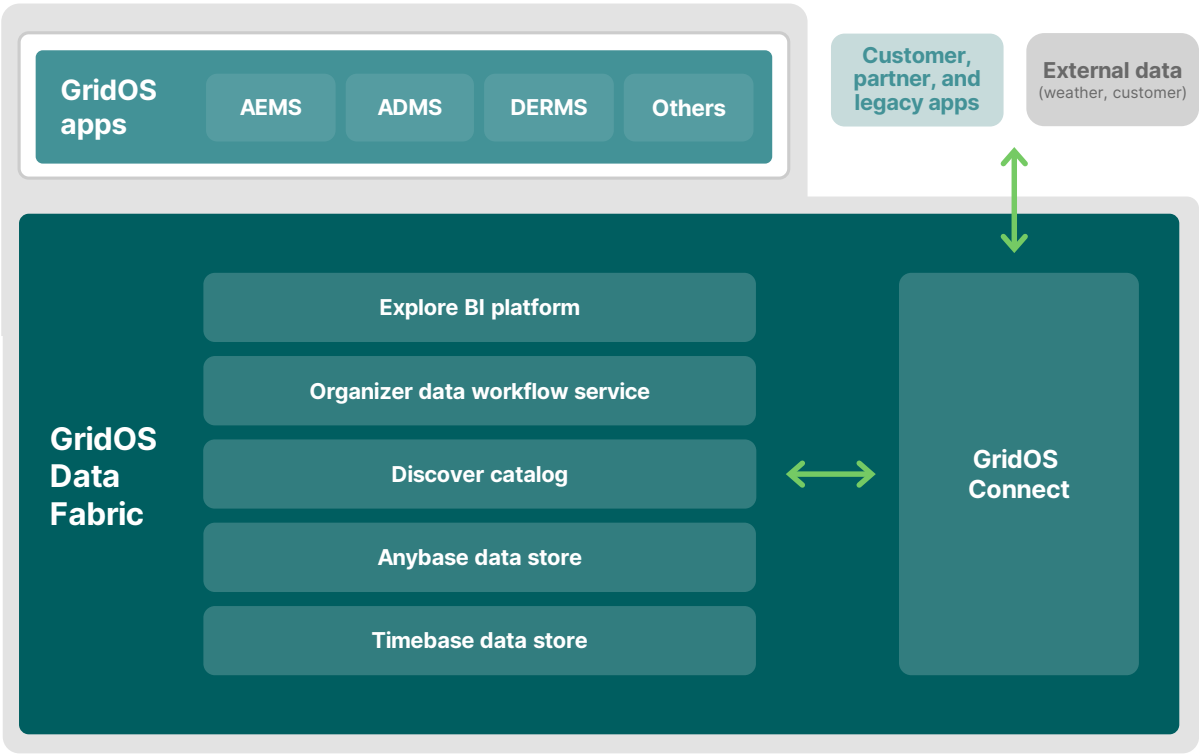
The GridOS Data Fabric serves as a distributed data foundation that can fuel AI- and ML-powered applications, grid automation initiatives, and other data-driven use cases. It is designed to help utilities access, integrate, and contextualize energy data, combined from disparate sources into a unified view capable of spanning T&D and the edge.

The GridOS Data Fabric uses data federation to access disparate data sources. Data federation is a software process that allows multiple databases to function as one. The process takes data from a range of sources and converts it to a common model. This provides a single source of data that front-end applications can leverage.

A federated database is virtual and doesn't physically store any data. Since the data is not centralized, its stewardship still remains with the different applications and functions within the utility, improving data quality and accuracy.

GridOS Data Fabric

The federated grid data management layer of the GridOS platform.



How is the GridOS Data Fabric structured?

The GridOS Data Fabric has the following components:

Component	What is it designed to do?	What benefits is it designed to offer?
GridOS Connect	The data integration engine within the GridOS Data Fabric. This engine connects data from a distributed ecosystem of applications and services across transmission, distribution, and the grid edge.	GridOS Connect enables faster integration of data –structured and unstructured, from inside and outside the enterprise, and across OT and IT – at scale. This makes managing data integration more self-service (via low-code environments) to connect internal and external energy data sources.
Explore BI platform	A self-service business intelligence (BI) platform with user-friendly web tools for performing data analytics.	The self-service environment allows operators and analysts to explore, visualize, and enrich datasets, making data extraction easier so that users can focus on analysis. This can help users get the insights they need for more efficient and effective grid orchestration.

Component	What is it designed to do?	What benefits is it designed to offer?
Organizer data workflow service	A modular, flexible middleware engine enabling data flow coordination with low code interfaces for backend data exchanges. The service supports and organizes data flows and the exchange of data and actions across the GridOS portfolio and other applications.	The Organizer service pushes data into applications, extracts data out of applications, and integrates external data. This reduces data services complexity, improving application accuracy and supporting dependencies between applications.
Discover catalog	A search engine that uses metadata to enable the discovery of data assets across the GridOS portfolio and other applications.	This centralizes key information associated with decentralized data, making it faster and easier to discover and model data, while enabling policy-based data governance to improve data integrity, consistency, and accuracy. The Discover catalog is the backbone of GridOS Data Fabric.
Anybase grid data store	A highly scalable, available, columnar data store service that enables federated data access to operational data from multiple sources across the grid ecosystem.	Anybase provides operational storage and data federation capabilities that enable quick access to disparate data and efficient delivery of that data to the applications and displays needed for grid operations and analysis. Operational storage scales to meet the sizes and demands of today's grid, while data federation offers decision makers secure, user-friendly tools like SQL query functions and dashboards , thus helping them capitalize on time-sensitive opportunities without the need to perform any extract-transform-load (ETL) effort.
Timebase data store	A specialized database for storage, retrieval, and archiving of temporal and/or time series data.	Timebase is designed to provide APIs and integrations optimized for grid use cases, such as forecasting, schedule management, simulations, scenario planning, and study modes. The database will also provide training data for AI and ML models where historical data is critical to represent the state of the grid and to understand responses to specific environments.

The GridOS Data Fabric provides access to a low-code development platform and software development kits (SDKs). Together, these tools help grid modernization teams accelerate solution development and deployment, bridge data silos, and improve data accuracy and availability. This helps reduce the time and expense needed to put energy data to work.

GridOS Data Fabric helps utilities drive outcomes

The GridOS Data Fabric is designed to make it easier for grid applications and users to **discover, govern,** and **utilize** the data they need to facilitate better, faster decisions that keep the grid reliable and resilient.

Discover

With today's wide assortment of applications and services, there are many places for data to hide. The GridOS Data Fabric is designed to use centralization of metadata to discover this data. A significant barrier to digital transformation is the massive growth of grid data in a monolithic application environment. Data silos, resulting from separate databases tied to individual monolithic applications, make it increasingly difficult at times for data practitioners to use standard data management tools and techniques to access data. The GridOS Data Fabric is designed to address this problem by enabling data federation that allows multiple databases to virtually function as one.

Govern

Inaccurate, fragmented, or "out of context" data can do more harm than good. The GridOS Data Fabric uses a metadata catalog to accelerate data governance workflows, increasing accuracy and confidence in the lineage and quality of data. Utilities often struggle to generate meaningful insights from data that originates from different departments. Very often, an organization's attempt to unify data in centralized locations causes unmanageable workloads for IT teams, leads to inaccurate and siloed data, and obscures valuable insights in a sea of information.

The GridOS Data Fabric helps utilities to implement data governance strategies using the metadata catalog to organize, index, and enable easy access to the data (and its associated metadata) flowing through various applications and databases. With metadata and automation capabilities, the GridOS Data Fabric accelerates governance workflows and can provide API-based diagnostic tools for development and troubleshooting.

Utilize

To explore and analyze data, utilities need a way to work with it. The GridOS Data Fabric enables users and applications to access data with secure, process-driven, self-service (user friendly) tools like SQL query functions and dashboards. It uses federation to collect data from multiple sources and convert it to a unified model or a "single source," consumable or interactable by front-end applications.

Federating and virtualizing data simplifies data queries across sources. This allows utilities to have better organizational buy-in, as federated access brings data together while control of the original databases continue to remain with the owning division or branch (and hence ensures continued accuracy through decentralized ownership).

Data federation architecture allows utilities to confidently embrace disparate ownership of data and hence, reliability of data, while unblocking teams to build insights by querying across siloed sources. With governance discovery and federated access to data, the GridOS Data Fabric allows organizations to be more agile and innovate more efficiently by accessing data across different energy participants (not just within a utility), to enable outcomes like greater markets participation and/or aggregator visibility.

USE CASE: ROOT CAUSE ANALYSIS

Today, humans, spreadsheets, custom applications, manual copy-and-paste operations and point-to-point integrations essentially carry out the functions of data fabric. Such manual methods simply do not work for highly complex grids with exploding amounts of data available to leverage. The GridOS Data Fabric replaces these tedious and manual processes to improve access to data.

The value of the GridOS Data Fabric is especially clear in scenarios like root cause analysis of an incident.

Let's take an example.

> A utility needs to pull together the sequence of events that led to a safety violation and transmission equipment damage at a substation. This requires months of manual data requests from data sources like: EMS alarms, historical voltages and flows, system operator log entries, recorded phone calls, substation security badge reader logs, substation video footage, historical and current equipment drafting diagrams, emails, equipment maintenance records, known or suspected equipment failure modes, etc.

The review also needs to address concerns related to not just technology failures, but human error as well. For example, were control center and substation maintenance personnel busy or multi-tasking at the time of the incident? Answering this aspect of the situation could require even more manual data requests.

Pinpointing an exact root cause requires considering all of these factors and stitching together a time-aligned sequence of events. Without the GridOS Data Fabric, it will be a long, frustrating, tedious job.

The GridOS Data Fabric is designed to vastly simplify data queries like the above, making incident root cause analysis significantly more efficient.

In addition, remember that this is just one single use case for the GridOS Data Fabric. It can do more than just simplify data queries. The GridOS Data Fabric is essential for AI- and ML-powered applications, grid automation use cases, analytics, simulations, and scenario analysis. It can also support a more advanced digital twin of the grid that can help run operations while enabling the grid to be automated and perhaps even autonomous.

CONCLUSION

Utilities need easier access to all their data in order to orchestrate the sustainable energy grid of the future. Such open access can only come from the capabilities of the grid-specific GridOS Data Fabric – a new platform solution within the GridOS portfolio for grid orchestration. The GridOS Data Fabric improves data accuracy and availability – eliminating data silos across the grid and accelerating solution development and deployment for analysis, modeling, AI-driven automation, and other essential use cases.

The GridOS Data Fabric is designed to:

- Help solve the "data discoverability" problem and improve use of data assets across the grid ecosystem. This helps operators and users quickly and easily find, access, integrate, and share data.
- Help ensure that operators and applications have more accurate, consistent, and secure data by deploying governance capabilities that help preserve data integrity and validity.
- Provide "middleware" APIs, configuration files, and data integration adaptors to make it easy for utilities to leverage data from GridOS, customer, and partner applications as well as external and/or data sources. This allows subject matter experts in key fields and departments to participate in the data modeling process.
- Support faster, easier, and more secure integration and access of data, reducing the cycle time for analysis and decision making.
- Help utilities more confidently leverage AI to automate key tasks for more efficient grid operation (because AI is only as good as the data training it).

MEET OUR AUTHORS:

Dr. Andrew Gillies (BSc. MSc. PhD.)
CTO
Grid Software, GE Vernova

BIO: Andrew led the early validation of the modern transformative technologies that now form the core of the GridOS® portfolio. He has expertise in bringing together orchestrated microservice architectures, containers, Kubernetes, Kafka, and other cloud-native technologies. Andrew also demonstrated a full suite of real-time WAMS in 2019.

Jay Shah
Product Marketing Director
Distribution, GridOS, and Data/ Cloud technologies

BIO: Jay is the Product Marketing Director for Distribution, GridOS, and Data/Cloud technologies. Jay has a background in product management and marketing. He has a bachelor's in computer engineering from the University of Mumbai and an MBA from Case Western Reserve University.

GE VERNOVA

GridOS® Orchestration Software

ORCHESTRATING A MORE SUSTAINABLE ENERGY GRID

Connect with a GE Vernova expert today:

Get in touch 1-833-690-5552

ELECTRIFICATION SOFTWARE

GE Vernova's Electrification Software is focused on providing a suite of software products and services to customers aiming to accelerate a new era of energy by electrifying and decarbonizing the energy ecosystem through intelligent and efficient data analytics, monitoring, and management.