Adopting technology to revolutionise and accelerate the flow of seismic data from sensor to customer

Erik Ewig[1]*, John Brittan[1], Cerys James[1], John Oluf Brodersen[1] and Sverre Olsen[1] present the PGS journey in enbracing the cloud and digitalsation to ensure the company remains highly competitive.

Introduction

Transforming a traditional industrial player into a cloud-native, data-driven energy data company is a monumental journey, one that PGS embarked on in 2020. Faced with ageing on-premises equipment that required substantial capital expenditure, the company found itself at a crossroads, especially coming out of a challenging economic downturn. It was in this context that a pivotal decision was made, to leverage the need for equipment renewal as an opportunity to look at our entire seismic data flow (from sensor to customer) and overhaul our IT and high-performance computing (HPC) landscape. The primary goal was to create a sustainable, cost-effective, and scalable solution that would not only future-proof the company but also deliver added value to our customers.

We quickly realised that this was more than just a technology project; it requires the whole organisation to adapt to significant changes for both people and processes. At the onset of 2021, PGS consolidated its technological and digitalisation resources and projects into a unified organisation, empowered with the mandate and capabilities essential for executing this transformative initiative.

The ultimate objective was articulated as the establishment of a sustainable and future-proof data ecosystem, ensuring the optimal flow of high-quality subsurface data with maximum efficiency, economy, and minimal environmental impact.

The initiative was guided by the key principle ‘cloud first’, i.e., always looking towards a cloud-based solution first to evaluate if it is technically and commercially viable. This gave us the opportunity to use cloud-native technologies and build on the inherent superior security, flexibility and scalability of the cloud. At the same time, we wanted to minimise data duplication, eliminate data handover points, and facilitate seamless and swift customer access to the subsurface information.

In the initial stages, we outlined key focus areas related to seismic data in the cloud. The most intricate was the migration of our on-premises High-Performance Computing (HPC) workload from Cray computers to the Google Cloud Platform (GCP). Simultaneously, we embarked on constructing a cloud-based data delivery environment for our multi-client seismic data library. A third initiative in our digital transformation journey was to use operational data from our vessels to improve efficiency and asset maintenance. Concurrently we also lifted all our enterprise applications into the cloud thus fully taking advantage of this technology.

Upon successful validation of the low earth orbit satellite’s maturity in the beginning of 2023, we introduced a fourth focus area concentrating on reducing operational costs through satellite utilisation.

The projects have had a transformational impact on the organisation and impacted all parts of PGS. We started to work on new

Figure 1 Traditional data flow (top) vs. optimised data flow (bottom). The traditional flow contains many data duplications and physical data handover points. The PGS optimised data flow transfers the subsurface information as soon as possible via satellite into the cloud for QC, processing and final delivery.

[1] PGS
* Corresponding author, E-mail: Erik.Ewig@pgs.com
DOI: 10.3997/1365-2397.fb2024016

business models, opened new business lines, changed workflows and best practice, and this resulted in reduced turnaround time, a reduced cost base, more satisfied customers and a happier workforce.

The following article will provide more insight into the key projects as well as exploring the changes we had to apply to the underlying platforms and infrastructure, which enabled this journey to be successful.

Revolutionising maritime seismic data transfer from tapes to the cloud

The initial phase in the seismic data journey in PGS generally starts after the seismic data has been recorded by the streamer spread and the onboard recording system. After initial quality control (QC) of both the data and the associated navigation information it is stored on the vessel until it is physically shipped to ours or our client's onshore facility for further data conditioning and imaging (see Figure 1). Historically, this workflow has remained unchanged due to insufficient satellite bandwidth to facilitate immediate data transfer of the seismic data, despite having internet connectivity via satellite on all vessel for decades.

However, this paradigm has now been disrupted with the arrival of Low Earth Orbit (LEO) satellite technology. LEO satellites provide a compelling alternative to traditional geostationary (GEO) satellites, offering enhanced data transmission capabilities and reduced latency.

While we are still in the early stages of exploiting this technology, it is relevant to explore the benefits LEO satellites offer, their potential applications, and the lessons learnt so far.

- **Reduced latency**: LEO satellites, operating at a lower altitude than GEO satellites (Figure 2), significantly reduce signal latency, facilitating near-real-time data transfer between seismic vessels and onshore data centres. This real-time data availability enables faster decision-making and operational adjustments.
- **Increased bandwidth**: LEO satellites can handle a large volume of data, meeting the high-bandwidth requirements of seismic data transmission. This enhanced bandwidth facilitates the transfer of high-quality seismic data onshore.
- **Reduced operational costs**: Seismic data transmission via LEO satellites reduces or even removes the need for on-board data processing, streamlines operations and potentially minimises costs in the long term.

Leveraging the cutting-edge capabilities of LEO satellites, PGS conducted a series of tests throughout 2023. These tests demonstrated the technology's exceptional effectiveness, with PGS successfully transmitting full-integrity 4D seismic data from two surveys directly to the cloud. This eliminated the need for physical data transfer, revolutionising the speed and efficiency of seismic data transfer. The data delivery time was reduced from an average of nine days to just one day (Figure 3).

Following these successful tests and recognising the transformative potential of LEO satellite technology, a strategic decision was taken to implement a base Starlink (LEO) service level alongside our existing VSAT service (GEO). For the time being, we treat both services as complimentary to each other based on available bandwidth, latency and commercial terms depending on the use case.

Figure 2 Left, LEO and GEO orbits relative to earth. Right, Starlink (front) and VSAT dome (back) satellite antenna on board the PGS vessel Vanguard.

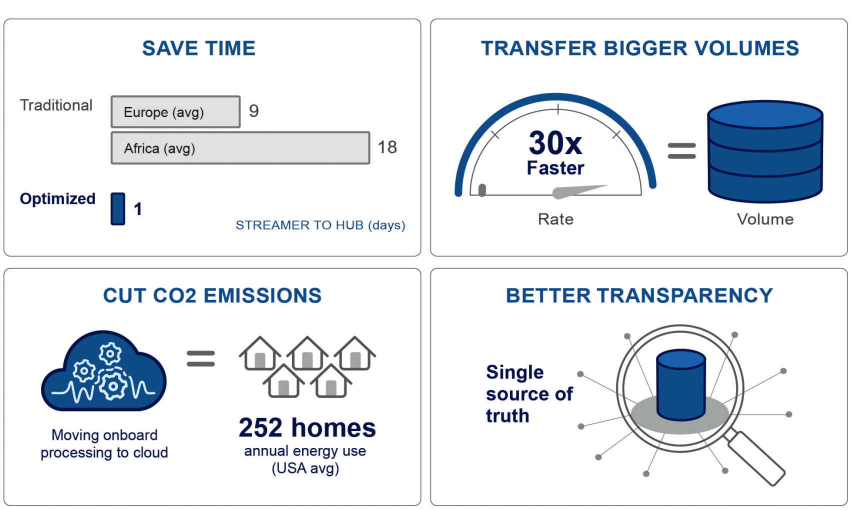

Figure 3 Key benefits of sending seismic data via satellite link rather than physical media transport.

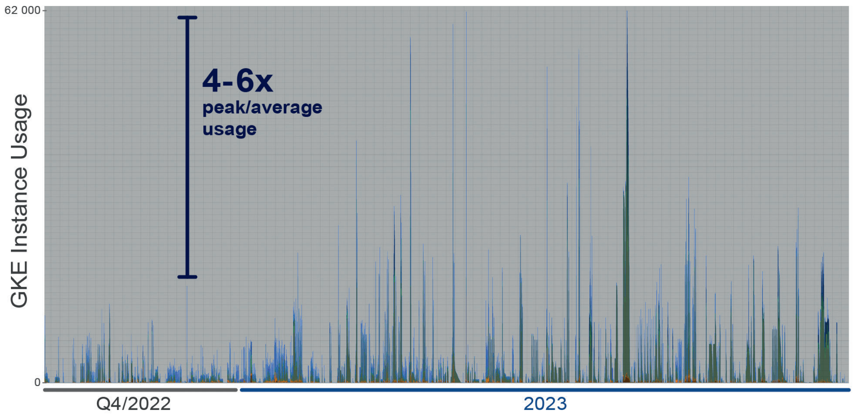

Figure 4 The use of GKE (Google Kubernetes Engine) instances in late 2022 and 2023 for our cloud-based imaging. The graph shows the spikiness of usage, at peak points we are using far more capacity than is ever available to us on-premises. Howeverk, outside these times we do not pay for any unused capacity.

The successful validation of the new satellite technology provides a huge opportunity for PGS and the seismic industry, both in regard to how we provide clients with the acquired data but also on how we enable them to follow the project during the acquisition phase. It has the potential to perform some of the tasks onshore and gives the onboard crew a better connection to the onshore world.

The future success of this technology in our industry depends on how we, as individuals, companies, and as an industry, adapt our traditional processes. Nevertheless, PGS embraces this technological change and is committed to driving sustainable seismic data transfer, ultimately enhancing the efficiency and effectiveness of seismic operations worldwide.

A new dawn for handling huge imaging workloads

After the seismic data have been acquired offshore and moved onshore the next step is to produce a 3D image of the subsurface. This typically requires multiple steps using different algorithms, often with significant use of high-performance compute (HPC). The HPC workload typically comprises large datasets (>1 TB and <100 TB) but a small number of files within each project (< 1000). Before 2019, our imaging software stack, comprising 300+ algorithms, utilised around 200,000 cores of Cray hardware and a 70 PB online parallel file system. Many of the most HPC-intensive codes were written specifically for this on-premises hardware and orchestrated by a heavily homegrown and highly customised job management system as well as proprietary and self-supported 3D data visualisation.

Initially, our cloud approach aimed at a hybrid environment, prioritising on-premises compute and utilising cloud compute for excess capacity via a 'Lift and Shift' strategy. This meant trying to move on-premises applications to the cloud without redesigning them and this proved unsuccessful both in terms of user experience and cost.

A fundamental reassessment was essential, extending beyond technical considerations to encompass our overall approach to the project. We started to work closer with our partner Google, reviewing our project set-up as well as our strategic short- and long-term goals.

We developed a minimum-viable-product for one algorithm (3D SRME) on the cloud, integrating it with on-premises infrastructure for an enhanced user experience. Despite initial scalability and reliability challenges, this experience addressed key usability issues.

The onset of the Covid-19 pandemic and subsequent business retrenchment prompted a shift to prioritise Opex over Capex, leaving us with no other option than to pursue a cloud-centric model. The goal was to decommission the Cray platform by mid-2022, focusing on cloud scalability, storage optimisation, and platform independence.

This was achieved by focusing our reengineering effort on only seven to eight key algorithms which make up more than 80% of our compute capacity and initially keeping bulk storage on-premises.

Critical reengineering steps included checkpointing within key algorithms, replacing Lustre with Object Store, and adopting Kubernetes and GKE for cloud-native scheduling. The move to Kubernetes proved transformative, giving the scalability and the stability we needed (Figure 4).

Here are some key benefits of moving processing to the cloud:

- **Scalability and flexibility**: In autumn 2022 we managed to sustain a peak of 1.2 million vCPUs across 12 GKE clusters, three times greater capacity than we previously had access to, allowing us to compete for larger jobs and run tasks in parallel. We can now routinely push through the 1 million vCPU barrier.
- **Improved turnaround time:** Tailoring compute for each large run, enjoying virtually unlimited capacity to run jobs at scale and in parallel, and leveraging the latest software and hardware stack significantly reduces compute time, from weeks to hours in extreme cases.
- **Reduced exposure**: Outsourcing tasks such as procuring equipment, maintaining the computer centre and negotiating power supply minimised our risk exposure associated with running a HPC compute centre.
- **Greener compute:** Our selected data centre with the preferred cloud provider runs on 100% renewable energy, contributing to our sustainability ambitions.
- **Potential to allow us to access new levels of geophysics:** Cloud scalability facilitates the development of new algorithms, such as those using elastic wave propagation, by providing short-term access to compute resources otherwise unaffordable for long-term use.

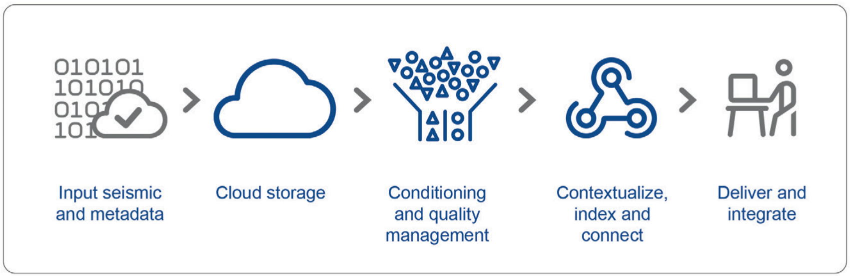

Figure 5 High level process flow for making seismic data available in PGS' cloud solution, providing users fast and secure cloud-based access to contextualised and quality-controlled subsurface data.

The success of this approach can be illustrated with some examples:

- A large reverse time migration (RTM) on an OBN dataset which would have taken 6-8 weeks using our entire legacy on-premises compute capacity took 14 days in the cloud.
- A multi-azimuth streamer RTM which would have taken 40 days using our entire compute capacity was completed in 26 days in the cloud.

A cloud-based solution for seismic storage and accelerated data access

The final leg of our journey involves delivering the final imaged data to our clients, either as part of contracted work or through our multi-client data library for widespread access. Traditionally, this data transfer has occurred via physical media (e.g. tapes), involving multiple phases of global logistics (Figure 1). This created a high risk of data loss, required many points of data duplication and was extremely time-consuming.

Managing our seismic data library, a high value asset for both PGS and data owners, is a complicated task as the petabyte library spans data acquired decades ago up to those acquired today. This complexity introduces challenges, contract terms evolve, companies merge or acquire others, and regulatory bodies alter rules, requiring vigilance in ownership tracking and licence-honouring by vendors.

Creating specific data deliveries, involving the extraction of subsets of data volumes, is a time-consuming process, taking days or weeks and necessitating meticulous planning. This can result in the vendor and end-user storing multiple copies of the data from disparate sources, further complicating matters. Our clients have said they spend a considerable amount of time on data sourcing and validation due to historical data management intricacies.

Recognising an opportunity to revolutionise subsurface data management, PGS, as the owner of one of the world's largest multiclient libraries, in 2019 started to move our multi-client library to the cloud. The goal was to eliminate data duplication, streamline data management through automation, and expedite data delivery. Even though the cloud offers nearly infinite storage capacity with ubiquitous, secure, and granularly controlled access, like our processing in the cloud story above, a mere lift-and-shift approach wouldn't have addressed all the problems, like consistent data conditioning, client trust, or reduction of data duplication.

To tackle these challenges, PGS entered a partnership with Cognite, a company who specialises in industry-scale data-ops challenges, to co-develop a tailored platform for subsurface data, especially multi-client data. The shared objective was to contextualise data in relation to licences and other metadata, providing PGS with a fully trace-indexed library and giving clients easy, secure and fast access to load or even stream the data into their environment (Figure 5).

By late 2022, PGS implemented fully functioning and cloud-native data ingestion pipelines, automating the reading, conditioning, ingesting, and indexing of post-stack and pre-stack data from cloud storage. A well-documented connection endpoint (API) and Software Development Kit (SDK) were made available to clients for 24/7 trace and metadata access. This empowered our sales department to explore diverse commercial models with clients for accessing seismic data.

Key benefits of enabling cloud-based data storage and access:

- **Improved turnaround time:** reducing access time from weeks to less than an hour
- **Direct access to data in the cloud at scale:** data is available via Open Subsurface Data Universe (OSDU)-aligned APIs for further sharing, collaboration, and interoperability within a client's own environment or within the industry. OSDU is becoming an open-source industry standard for storage, retrieval, and sharing.
- **New commercial offers:** this solution allows us to condition, ingest and connect our client's proprietary data, proving a state-of-the-art managed cloud-based solution for storage, retrieval, and seismic data sharing. This enables more rapid and informed exploration and development decisions.

Conclusion

Confronted by ageing infrastructure and economic hurdles, PGS seized the opportunity to revolutionise our seismic data flow, information technology, and high-performance computing landscape to transform to one of the industry's leading cloud-native, data-driven energy companies.

A key learning was that the challenge extended beyond mere technical platform upgrades. It necessitated a profound shift in our workforce mindset. By embracing new project methodologies and technologies, integrating greater business participation into projects, scrutinising and revising numerous workflows, and persuading stakeholders that the cloud provides substantial business potential.

An important role, only briefly mentioned in the article, relates to the IT offering within our organisation. Traditionally positioned at the periphery as a service function, it became an integral component of our transformation journey. IT now stands as a driving force behind deploying, operating, and maintaining a comprehensive IT platform that serves the entire organisation.

Embarking on this initiative, we envisioned an exceptionally ambitious future, uncertain of its technical feasibility at the outset. While we haven't reached all our goals yet, and there's still work to be done in aligning and engaging with our customer base in this transformative journey, noteworthy milestones have already been attained. The initiative revolutionises the methods by which we obtain, process, and utilise seismic data.

Acknowledgements
The journey we have described in this article would not have been possible without the dedication and commitment of a large number of people, both from the Technology and Digitalization group within PGS and across the other business units. We acknowledge that the hard work and innovation from these teams has allowed us to get this far.