

## INDUSTRY EXAMPLES

### IoT

The Internet of Things (IoT) encompasses the billions of physical devices being made ‘smart’—through the addition of sensors and communications technology—and connected to the internet. Inexpensive processing and wireless networks make it feasible to turn practically any objects, large or small, into part of the IoT, where they can communicate independently of human interaction.

### Predictive Maintenance

Machines and equipment are outfitted with sensors that continuously monitor key attributes or performance indicators, such as temperature, pressure, vibrations/second, noise levels, etc. By capturing and utilizing data streaming from sensors and connected devices, businesses can now gain visibility into the condition of their valuable assets and specific components in real time. Predictive maintenance can also be used in refineries, at sites, and for other midstream uses.

### Vehicle and Fleet Management

Using data analytics to optimize fleet operations can help keep vehicles operating longer and more reliably, while minimizing service disruptions or delays in drilling.

### Oil Exploration and Discovery

Oil and gas exploration is the search by petroleum geologists and geophysicists for hydrocarbon deposits such as oil and natural gas beneath the earth’s surface. Oil and gas exploration, grouped under the science of petroleum geology, depends heavily on the ability to ingest and analyze large volumes of data, including streaming data.

# StreamSets for Oil & Gas

## Overview

The oil and gas industry has a long history with data, and constant change throughout the industry is making data increasingly important. Streaming data provides up-to-the-moment understanding of what’s happening, giving oil and gas company executives visibility across the business and leading to real-time, impactful decisions. Streaming data plays a vital role in enabling oil and gas businesses to achieve their goals in exploration, maintenance, quality, and safety. It’s also an important factor in trading and supply activities, both upstream and downstream.

## Challenges

Oil and gas companies have access to massive and diverse data sets, especially with the wealth of new sensor-enabled and connected tools for energy discovery, exploration, and distribution. Still, making full use of this data—and turning it into insights—remains problematic, especially for unstructured and streaming data. In addition, data drift and other complications can erode the quality of the data that’s available for analysis.

Only a few oil and gas companies use cloud technologies, and many are just starting that phase of digital transformation. While automation—with or without the cloud—would help get data to the right places at the right time, combining streaming data with existing legacy data poses particular challenges. For instance, the industry faces a shortage of specific coding skills and a tight talent market for these skills. Increased competition, dwindling areas of ‘easy’ oil and gas extraction, less favorable economic conditions, volatile geopolitical forces, and other challenges put even more pressure on companies to harness and use data to best advantage. Intelligent pipelines, 4D seismic technology, and other advanced technologies are a double-edged sword: They provide data crucial to oil and gas companies’ success, but they also exacerbate the data management challenges.

When it comes to data, oil and gas companies need to be able to:

- Enhance oil and gas exploration and discovery.
- Implement proactive equipment maintenance.
- Modernize cybersecurity defenses.

Managing data effectively within oil and gas companies is difficult because:

- The sheer volume of data generated by modern exploration and production (E&P) processes can overwhelm legacy data management systems.
- Processing data at the edge can be more efficient and reliable than sending data to a cloud or centralized data center, but edge technologies are still new and often require advanced and costly development.
- Data movement platforms must be able to easily ingest, manage, and process the full array of streaming data at an affordable cost.
- While machine learning and predictive capabilities can help deliver competitive advantage, deploying machine learning at scale is tricky and requires specialized expertise with tools like Apache Spark.
- Data ingestion systems also must handle diverse data structures and evolving schemas, including everything from intermittent readings on temperature, pressure, or vibrations per second to fully unstructured data (e.g., images, video, text, spectral data) and thermographic or sound/noise data from drilling equipment.
- For regulatory compliance, companies must not only protect their intellectual property (IP) and secure their data enterprise-wide, but also document and evaluate data security across all their systems. Many open-source frameworks for ingesting data in real time leave gaps in security.

## INDUSTRY EXAMPLES

### R&D

Because constant innovation is needed in the oil and gas industries, companies are using years of historical data to enhance the development of their products and services.

#### **Seismic Trace Identification**

Seismic trace identification, which is the storing and processing of seismic data that can be used to discover seismic trace signatures that were previously overlooked, is highly dependent on the careful analysis of streaming data.

While domain expertise has long guided exploration and discovery activities, new strategies for oil and gas exploration are now reliant on data and analytics, with increasing emphasis on leveraging streaming data.

## Solution

StreamSets enables oil and gas companies to continuously ingest sensor data from the edge, transmit it to clouds and data centers, then turn that data into insights that help enhance energy exploration and discovery.

**Enhance oil and gas exploration and discovery.** Using StreamSets, energy companies can ingest fleet, weather, supply, seismic, subsea, production, trading, and other data for up-to-the-moment visibility into potential new oilfield opportunities. StreamSets' Control Hub lets companies troubleshoot issues with their Internet of Things (IoT) pipelines, while managing topologies to optimize IoT spending and deployment.

**Implement proactive equipment maintenance.** StreamSets Data Collector Edge is easy to deploy within an application architecture, where it provides oil and gas companies with real-time, actionable data on their equipment operation and performance—whether in oilfield services, refineries, drilling and exploration, or production processes. By enabling a proactive approach to equipment maintenance, StreamSets helps lower companies' expenses while avoiding costly downtime.

**Modernize cybersecurity defenses.** StreamSets allows oil and gas companies to ingest data faster, for improved cyber threat detection and forensics. StreamSets enables companies to uncover risks for drilling, operation, and delivery. The solution can also expand to manage additional threat analysis outside of traditional Security Information and Event Management (SIEM) systems..

## ABOUT STREAMSETS

StreamSets transforms how enterprises flow big and fast data from myriad sources into data centers and cloud analytics platforms. Its DataOps platform helps companies build and operate continuous dataflow topologies, combining award-winning open source data movement software with a cloud-native Control Hub. Enterprises use StreamSets to enable cloud analytics, data lakes, Apache Kafka, IoT, and cybersecurity.

Founded by Girish Pancha, former chief product officer of Informatica, and Arvind Prabhakar, a former engineering leader at Cloudera, StreamSets is backed by top-tier Silicon Valley venture capital firms, including Battery Ventures, New Enterprise Associates (NEA), and Accel Partners.

For more information, visit [streamsets.com](http://streamsets.com)

## StreamSets Benefits

StreamSets enables organizations to:

- Ingest and analyze all data regardless of form or semantics.
- Respond more quickly to analytics outputs.
- Design topologies with common destinations and processing stages.
- Ingest data from the edge for real-time insight from endpoints, as events occur.
- Take advantage of numerous open-source, industry-specific StreamSets connectors for protocols such as OPC, WITSML, LAS, and others that enable standards-based industrial interoperability and application development.
- Scale to meet the demands of expanding cyber threats.
- Apply machine learning directly to streaming data to identify key indicators of equipment breakdown and develop predictive models to alert users.

## Closing

Oil and gas companies are more dependent than ever on data, but their legacy pipelines often can't handle the variety and real-time speed of that data. The StreamSets platform helps energy companies turn their real-time and big data into insights that can improve the entire energy production process.

Find out more about how StreamSets can help oil and gas companies optimize sensor data to enhance energy exploration and discovery. [Contact a StreamSets representative today.](#)