Big Data and Apache Hadoop for the Oil and Gas Industry

The oil and gas industry has relied on data to run their businesses for many years. As one of the first sensor-based industries, oil companies have long collected sensor data from their oil wells to monitor operations.

Today oil and gas companies are collecting a much larger volume of vastly different types of data, and at a faster pace. This includes drilling and production data, GPS and spatial data, seismic data, business data, weather, and operations logs. Much of this data is unstructured so it’s difficult to store, integrate and access using traditional database technology.

Apache Hadoop is enabling oil and gas companies to develop a “digital oilfield” where they can integrate real-time production, maintenance, and engineering data to improve decision-making, efficiency and safety.

The top-ranked MapR Distribution including Hadoop offers the best enterprise-grade, low-risk big data solution to build a digital oilfield. MapR can help you cost-effectively integrate and analyze a wide variety of unstructured data to increase drilling and production performance while preventing environmental and safety problems. The MapR high availability (HA) features are particularly valuable for helping you reliably run remote operations. And multi-tenancy features provide the flexibility to run multiple applications on the same cluster.

Below is a representative architecture that shows the various data sources of relevance in the oil and gas industry.

**MapR Distribution including Hadoop Highlights**

**High availability and disaster recovery.** Business continuity and higher business-level service level agreements.

**Multi-tenancy.** Supports multiple business groups and applications in one cluster without conflicts.

**Integrated security.** Built-in data access controls.

**Direct Access NFS.** Direct data ingestion, familiar access methods, existing tools/libraries continue to work.

**High performance.** Fast, responsive access to data, and higher throughput.

**Volume support.** Disparate user groups and data by logical volumes.

**Job placement control and resource management.** Jobs run simultaneously in the same cluster.

**Data protection.** Consistent snapshots with point-in-time audits and recovery.

**Support for structured, semi-structured, and unstructured data.** All data in the enterprise data architecture.

**Architecture for Oil and Gas**

![Architecture for Oil and Gas](image-url)
Real-time Analytics for Oil & Gas Use Cases

**Oil Production**

Analyzing seismic, drilling and production data can help optimize oil recovery from existing wells. Big data techniques can also be used to forecast oil production. If the forecast does not meet a determined production level, that well can be remediated. MapR can help engineers integrate and analyze data to increase throughput from existing wells.

**Equipment Maintenance**

Equipment failures occur due to multiple variables, such as usage patterns, usage conditions, oil pressure or purity. Sensor data from equipment and geological data can be analyzed to predict equipment failure and understand what equipment works best where. MapR enables real-time capture, storage and analysis of high-resolution sensor data to better predict potential equipment failure.

**Supply Chain**

Oil and gas companies can analyze data streams from suppliers to evaluate performance and optimize supply chain operations. Sensors and RFID tags can be used to capture supply chain data to quickly understand inventory levels for purchase planning. They can use analytics to improve “just in time” delivery of materials, reducing inventory levels and supply chain costs.

**Safety and Environment**

By using Hadoop to analyze data from a variety of sources, anomalies in drilling can be identified in real time. MapR can be used to increase the environmental health and safety of oil rigs and drills through identification of patterns and outliers before any catastrophic incidents take place.

**Security**

Oil and gas companies need to identify events or patterns that could indicate an imminent security threat or cyber-terrorist act. Predictive analytics identifies patterns that can help detect these threats in advance. MapR can help you uncover threats in real-time through machine learning and anomaly detection techniques can reduce the likelihood of such incidents.

**About MapR**

MapR delivers on the promise of Hadoop with a proven, enterprise-grade platform that supports a broad set of mission-critical and real-time production uses. MapR brings unprecedented dependability, ease-of-use and world-record speed to Hadoop, NoSQL, database and streaming applications in one unified distribution for Hadoop. MapR is used by more than 500 customers across financial services, government, healthcare, manufacturing, media, retail and telecommunications as well as by leading Global 2000 and Web 2.0 companies. Investors include Google Capital, Lightspeed Venture Partners, Mayfield Fund, NEA, Qualcomm Ventures and Redpoint Ventures.