

DataCloud develops cloud based prescriptive analytical solutions for the energy sector

Case Study



Overview

The company has over a decade of experience in applying analytics and real-time data models to oil well drilling operations. Using this experience, and industry knowledge, the company has developed the DataCloud platform, which enables logic operations to be performed on time-based stream data and for this to be expressed in a standardized language. This engine forms that foundation of the company's Drillytics application.

Drillytics is a real-time analytics application that integrates with the industry standard Wellsite Information Transfer Standard Markup Language (WITSML) API to extract wellsite data, either from rig sensor or third party sources, and provide analytical information with prescribed actions back to the rig and enterprise data store. Drillmetrics complements Drillytics by providing a web-based application to track alerts and metrics produced by analytics and measure their effect across multiple wells. This provides a user interface to complement by closely monitoring & evaluating data collected by the Drillytics application.

Highlights

- Technology: Apache Cassandra® and Apache Spark™
- Sector: Oil and Gas
- Use case: IoT and real-time analytics.

Instaclustr has provided us with a managed environment for getting underway quickly with Apache Cassandra® and they are delivering excellent support and the expertise necessary to build and scale our application.

Frode Sormo,
CTO & Co-Founder, DataCloud

Challenge

The oil & gas industry stores sensor data in an industry-specific document database, where data access is only available through a proprietary API based on SOAP and XML. This method of data storage and access does not provide the performance necessary to perform real-time analytics solutions for this dataset.

To highlight the challenge faced by the DataCloud engineering team, one of the company's customers has over 200 wells. Accessing the data for those wells from the industry specific WITSML store can take up to several hours per well. To perform deep analytics on the full set of wells using this data access method was simply not feasible.

The DataCloud engineering team identified the need to be able to stream real-time data into a database, load historical data as necessary and access that data quickly on a time-based index. The team also identified the need to re-analyze historical wells periodically to test new real-time analytics, which drives write speed from tens of rows per second to many thousands of writes per second.

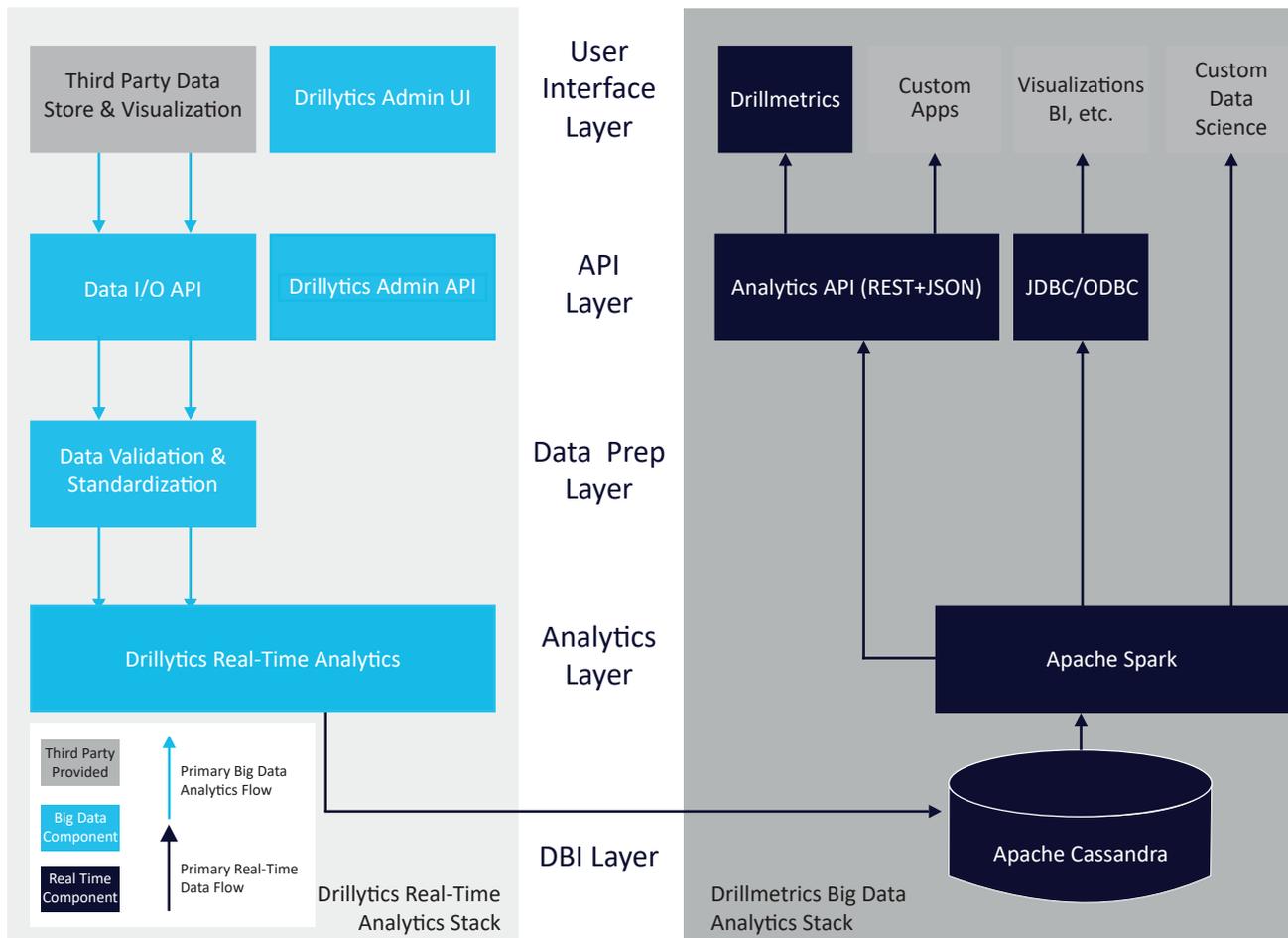
Solution

Cassandra is the answer.

DataCloud's solution was to have the Drillytics application transfer this data into an Apache Cassandra® database cluster. The database provided the performance required to deal with the many thousands of writes per second. The Drillytics application is then capable of performing the detailed analytics required to target drilling solutions modules to address common industry problems. Reducing the frequency of these drilling problems has a material impact on improving safety and lowering production costs.

Drillytics contains a data validation and standardization module to ensure that units, curve names and data quality is standardized for use by drilling solutions modules. These are implemented in a rule-based engine that allow complex logic on time-based stream to be expressed in a standardized language.

The Drillmetrics application then draws from the Apache Cassandra® database that stores output from the Drillytics application to provide metrics, direct visualization and time/depth view of wells updated in real time. The application also uses Apache Spark™ to do more complex data science jobs, used on specific analytic tasks. For example, a customer might want to know how much time it takes to drill a single well, or in more complex scenarios, how specific drilling parameters correlate with drilling speed. This has been very difficult to do in the past, but with Apache Spark's statistics and machine learning capabilities, this is relatively straight forward.



About Instacluster

Instacluster delivers reliability at scale through our integrated data platform of open source technologies such as [Apache Cassandra®](#), [Apache Kafka®](#), [Apache Spark™](#) and [Elasticsearch](#).

Our expertise stems from delivering more than 25+ million node hours under management, allowing us to run the world's most powerful data technologies effortlessly.

We provide a range of managed, consulting and support services to help our customers develop and deploy solutions around open source technologies. Our integrated data platform, built on open source technologies, powers mission critical, highly available applications for our customers and help them achieve scalability, reliability and performance for their applications.

Apache Cassandra®, Apache Spark™, Apache Kafka®, Apache Lucene Core®, Apache Zeppelin™ are trademarks of the Apache Software Foundation in the United States and/or other countries. Elasticsearch and Kibana are trademarks for Elasticsearch BV, registered in U.S. and in other countries.