Scalable and Intelligent Oilfield Streaming Data Analytics Platform

Dr. Lei Huang, Darshita Mistry, Yi Zhu
Associate Professor
lhuang@pvamu.edu
Cloud Computing Research Lab
Prairie View A&M University, Texas, USA
http://www.pvamu.edu/cs

Outline

- Introduction
- Platform Overview
- Implementation
- Results
- Conclusion
- Ongoing work

Big data and Deep Learning

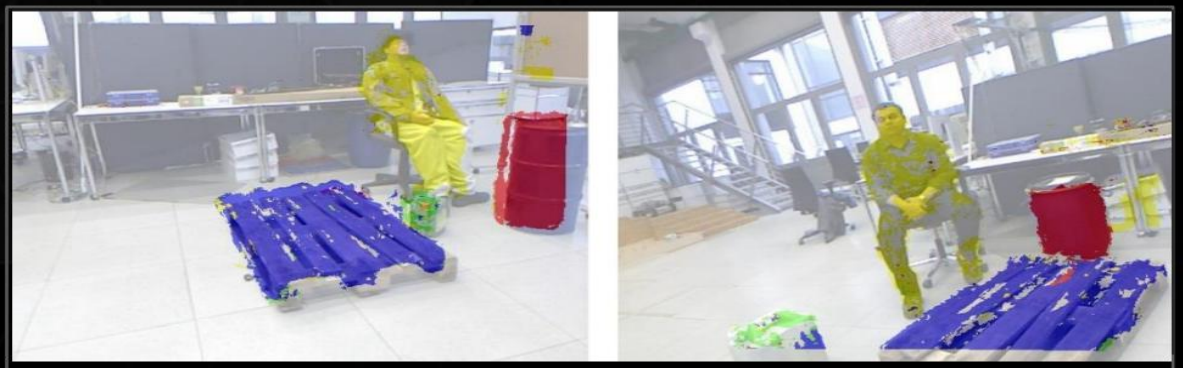
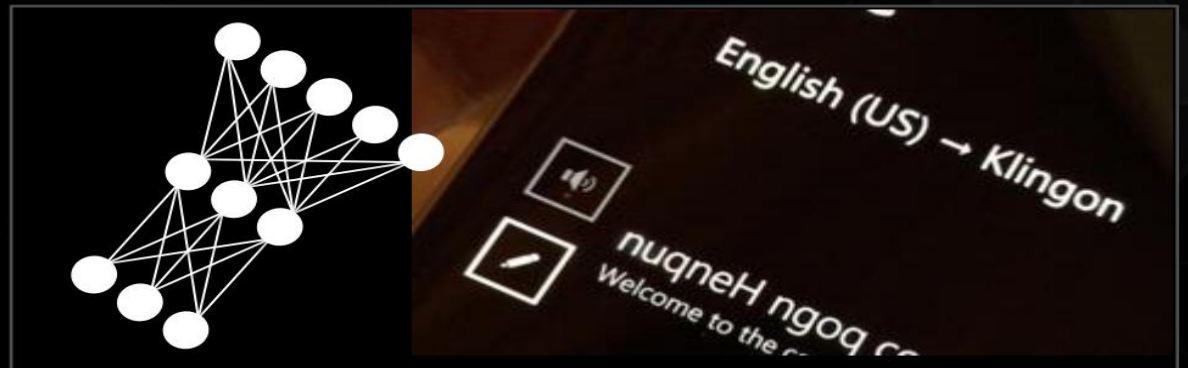
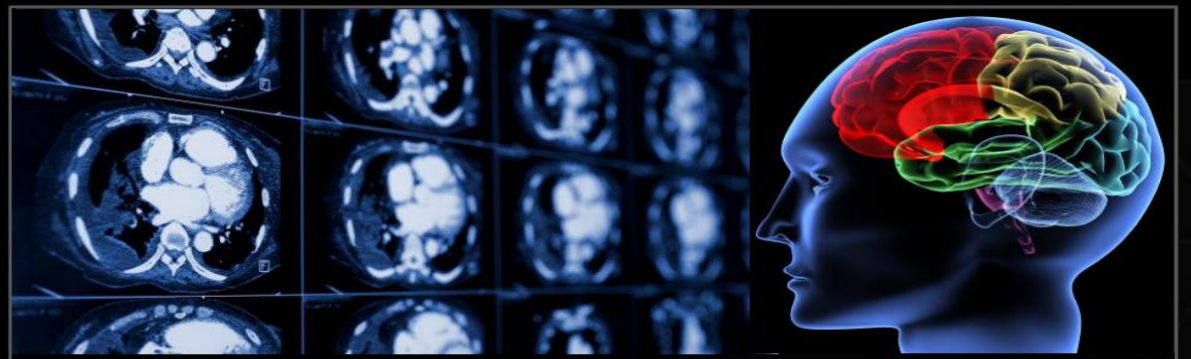


Image Classification, Object Detection, Localization,
Action Recognition, Scene Understanding



Speech Recognition, Speech Translation,
Natural Language Processing

Pedestrian Detection, Traffic Sign Recognition



Breast Cancer Cell Mitosis Detection,
Volumetric Brain Image Segmentation

Oil & GAS Industry

- Geological structure identification
  - Fault/Channel/Salt dome detection
- Reservoir Property Analysis
  - Stratigraphic estimate
- Oilfield Intelligence
  - Production Forecasting/Optimizations

The Cloud Computing Research Lab

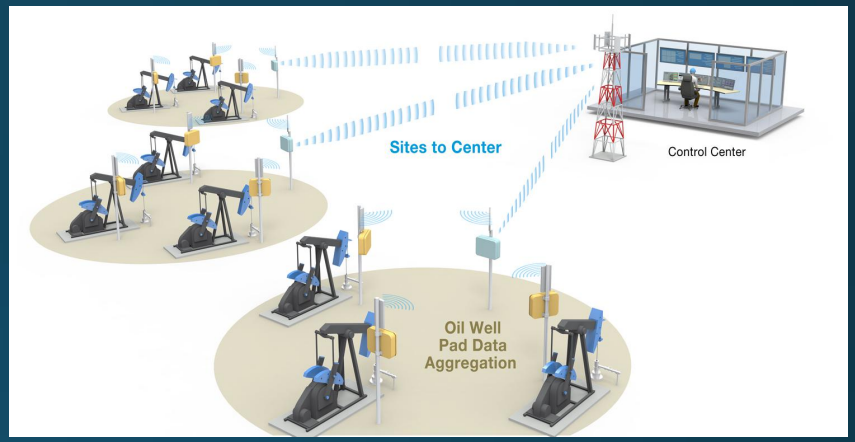
- **Goal:** building a scalable domain-specific big data analytics cloud platform
  - Built on top of Apache Hadoop and Spark
  - Big data storage, analytics and visualization
  - Machine Learning/Deep Learning models
  - Scalable performance & productivity
- Sponsored by National Science Foundation

PRAIRIE VIEW A&M UNIVERSITY

MEMBER OF THE TEXAS A&M UNIVERSITY SYSTEM

Oilfield Data Analytics

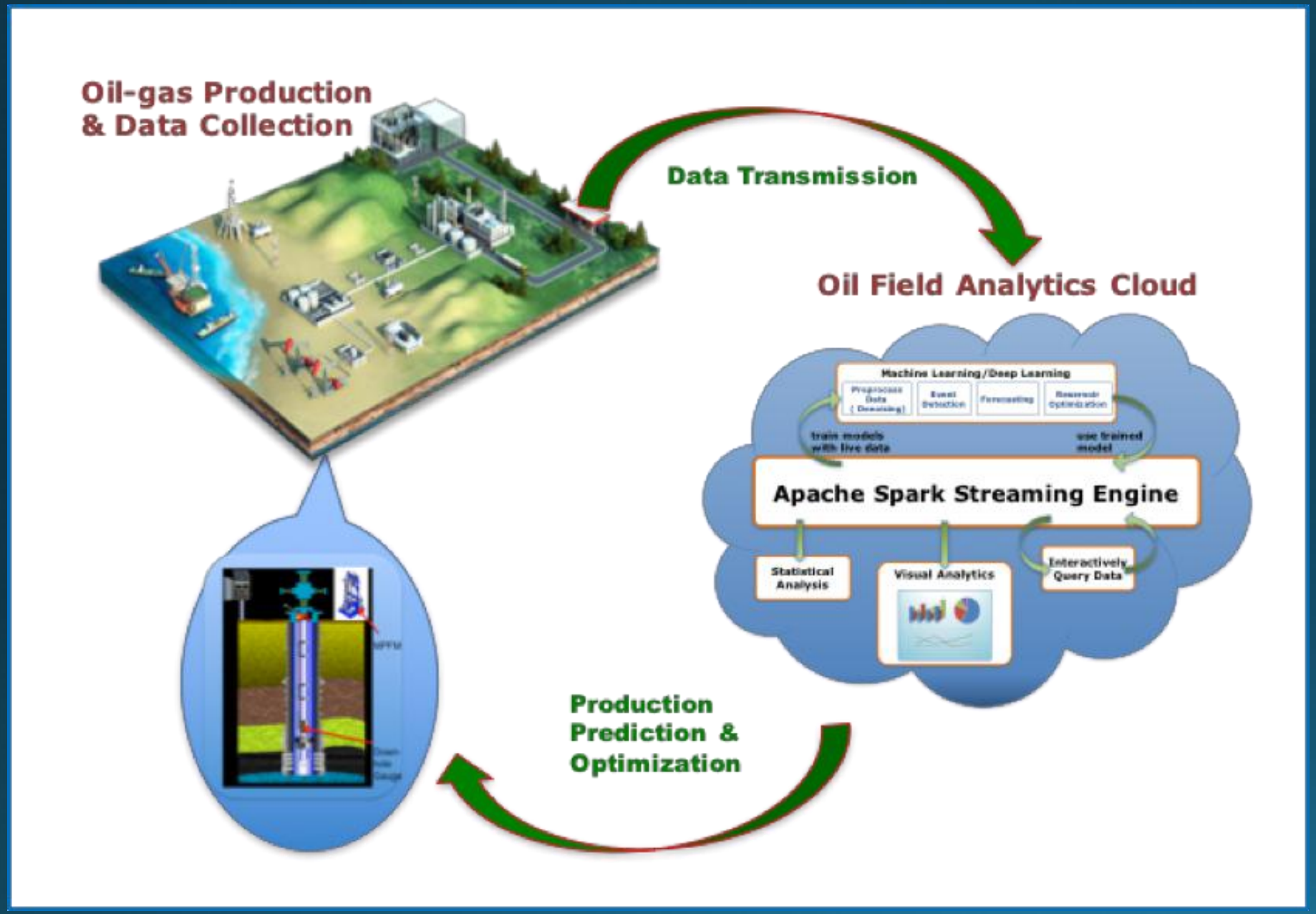
- Huge amounts of data collected using electronics sensors like Permanent Downhole Gauges(PDG) from thousands of wells to determine the wells' performance and reservoir characteristics.



- Collected data are streamed to a data center for storage and processing.
- Value extracted from data in real-time is crucial for production optimization and operations of wells.

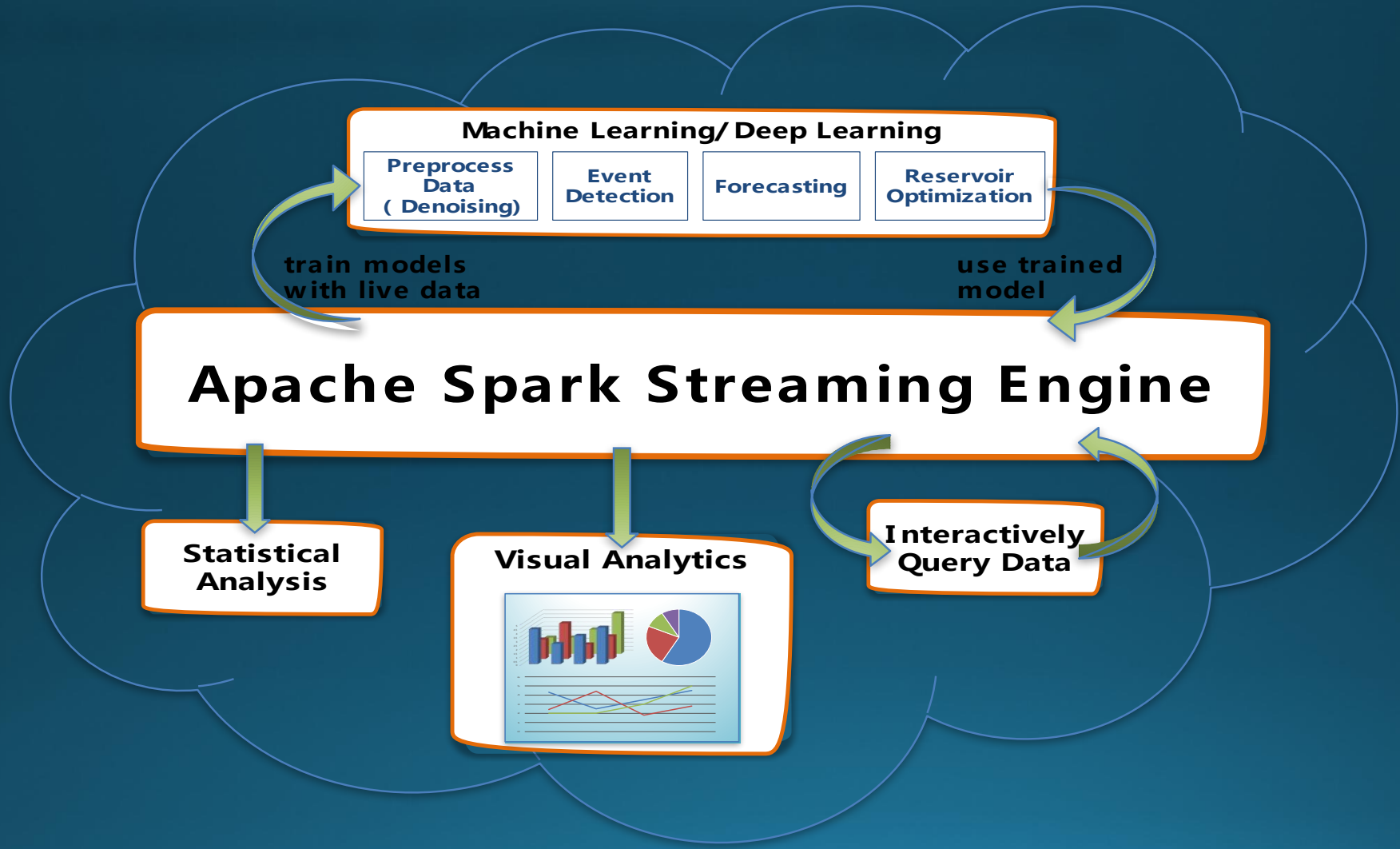
Challenges in Existing Oilfield Data Analytics System

- High rate of sensor data expansion
- Unique characteristics of each well
- Complex analytics algorithms (Innovative deep learning models)
- Real-time analytics



Oilfield Data Analytics Cycle

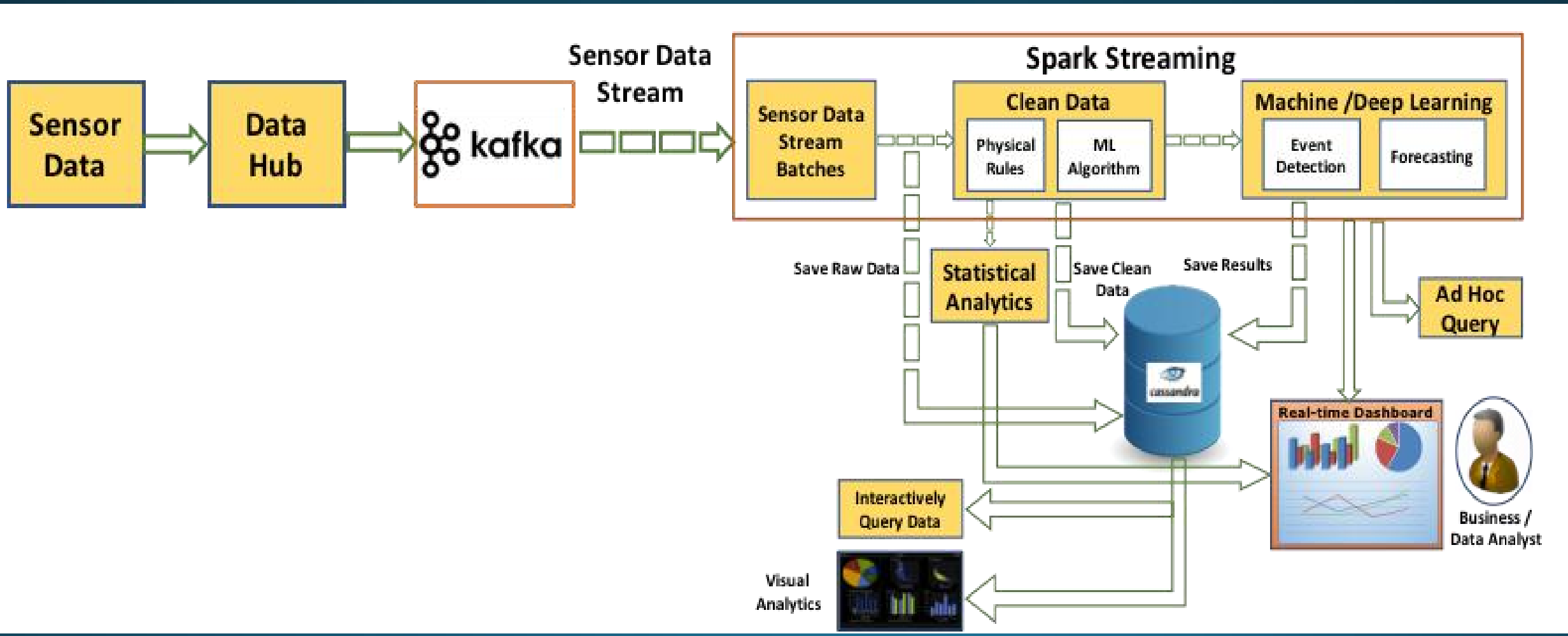
Our Oilfield Analytics Cloud



Platform Modules

Data Ingestion	Data Cleansing	Real-time Dashboard	Interactive Querying
Production Forecasting	Event Detection	Visual Analytics	Statistical Analysis

Platform Architecture



Software Stack

Oilfield Streaming Data Analytics Applications		
Visualization	Workflow	Data Services

OpenCV/Breeze	Spark MLib	Scikit-learn	TensorFlow
Spark Batch	Spark Streaming	Spark Interactive	

Hadoop HDFS	YARN	Mesos	Cassandra	Kafka

Data Ingestion

- Multivariate time series data from the PDGs of multiple wells collected and streamed to the Kafka cluster in real-time
  - Guaranteed to store in order of receiving
  - Guaranteed to be processed one time
  - Guaranteed to not miss any message
  - High throughput

PDG Data Cleansing

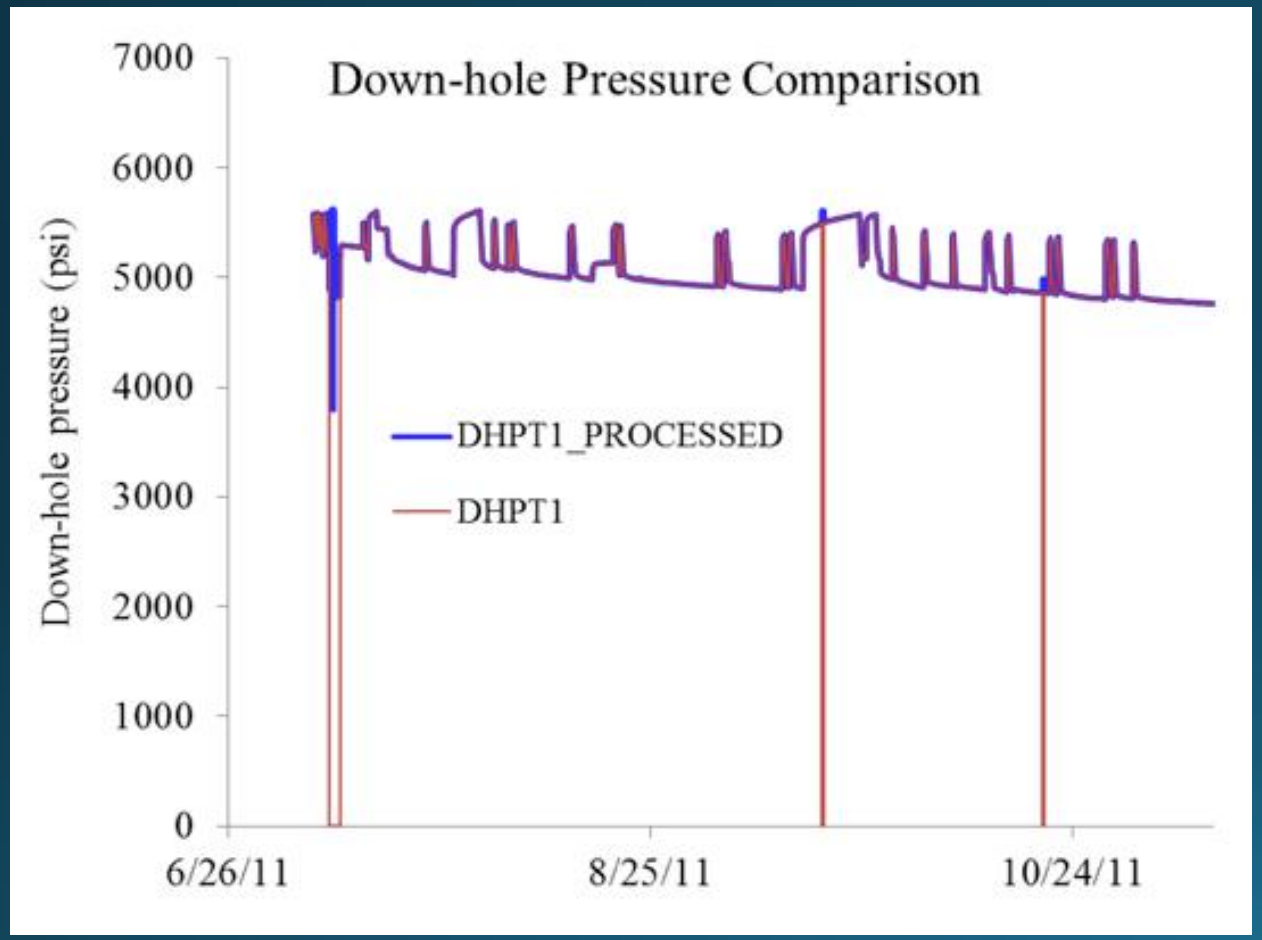
- Problems within raw data
  - Missing values
  - errors
  - conflictions among data attributes, e.g., wellhead pressure higher than down-hole pressure
- Preprocessing techniques
  - ✓ Supervised machine learning to calculate the error, missing or conflicting data
  - ✓ Physical rules of Petroleum engineering to check the results of machine learning, improve the results if necessary

Physical Rules Used For Preprocessing

- Obey the oil production equation, which is controlled by pressure difference between reservoir and downhole
- Oil rate and downhole pressure can not climb simultaneously
- When the well is shut down, the downhole pressure should be growing to maximum
- When production increases, the downhole pressure is supposed to be dropping
- Downhole pressure should be higher than wellhead values and can never be zero

$$q_o = \frac{4\pi k h k_{ro}}{\mu_o \left[\ln\left(\frac{4A}{\gamma C_A r_w^2} \right) + 2s \right]} (\bar{P} - P_{wf})$$

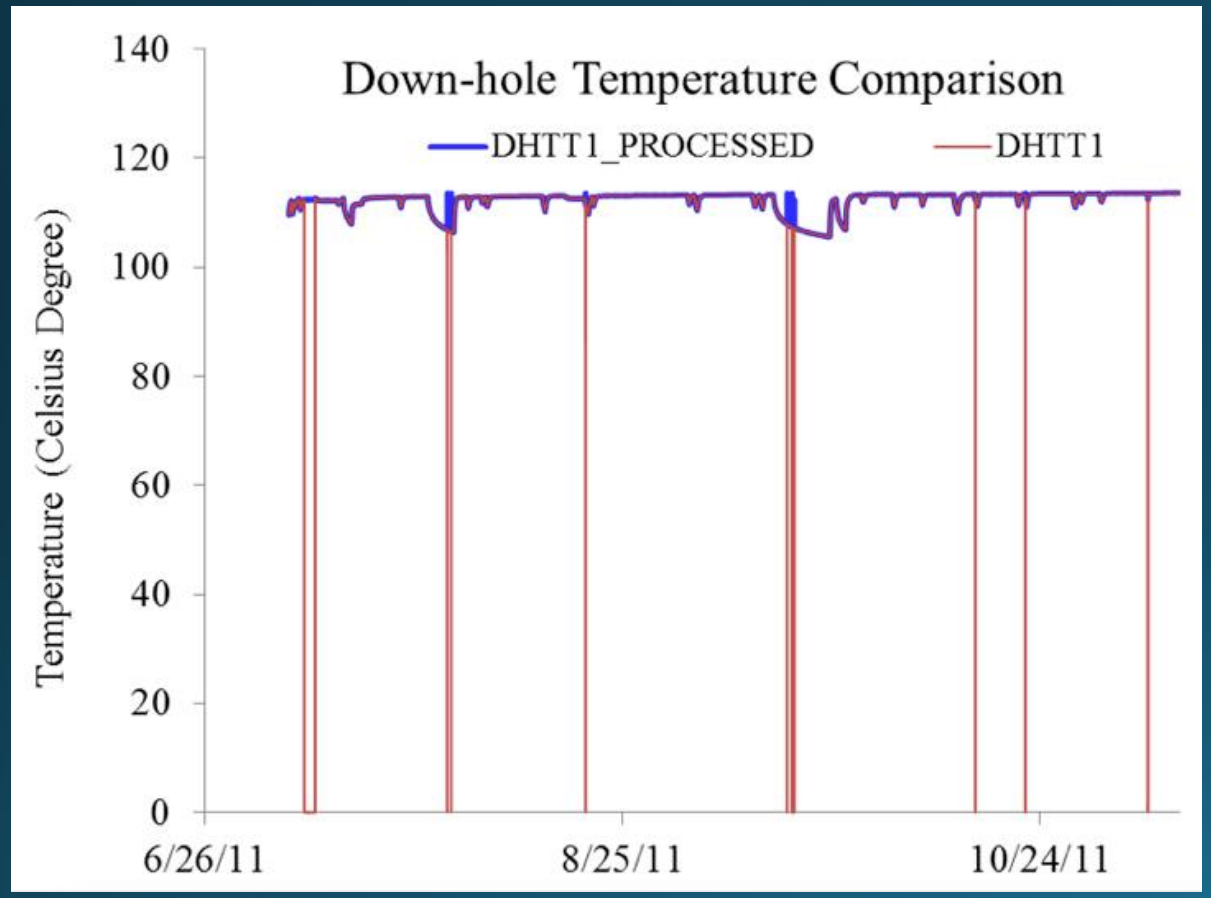
Preprocessed Results



Downhole Pressure

- Downhole pressure can not be zero
- It should be a constant while the well is shut down
- It should be higher than wellhead pressure
- The error or missing data is fixed with other pressure values by machine learning

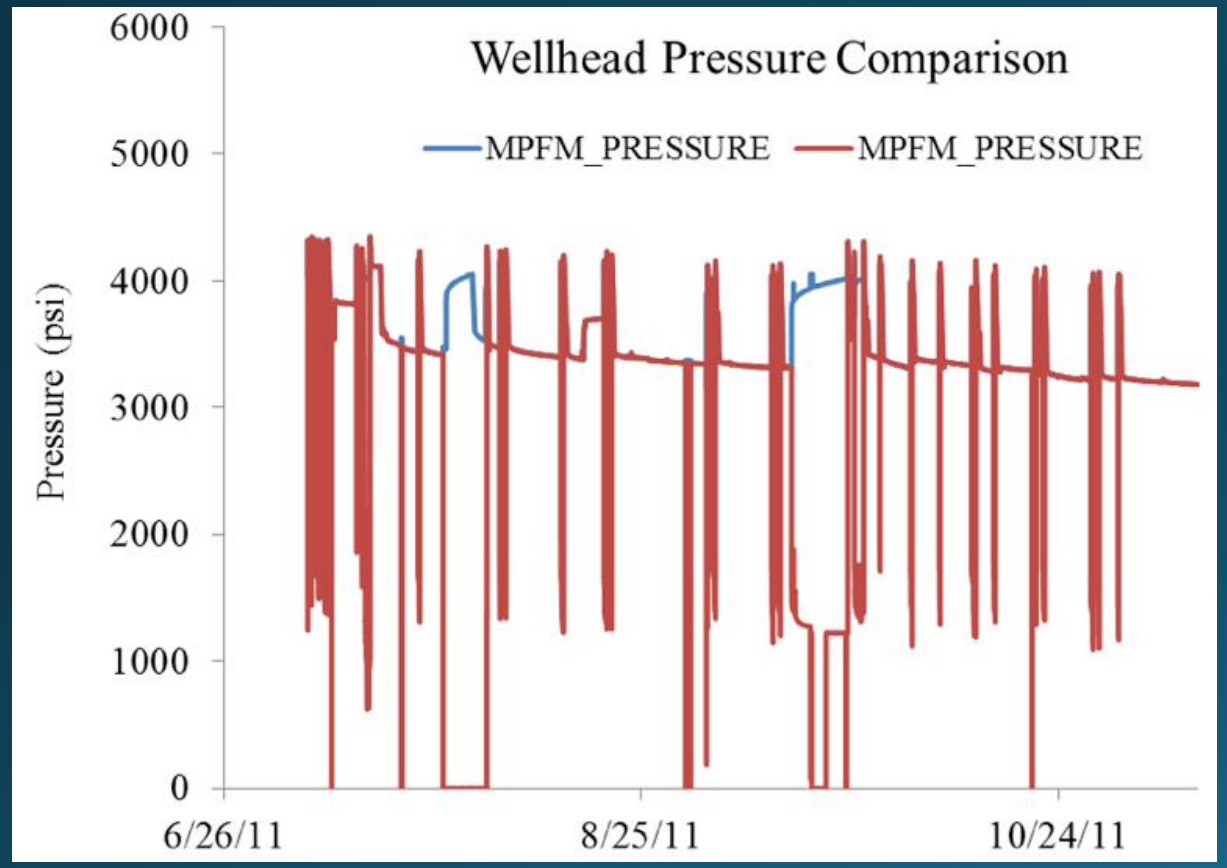
Preprocessed Results



Downhole Temperature

- Downhole temperature can not be zero.
- It is normally higher than wellhead value.
- Using machine learning to correct the downhole pressure and other attributes.

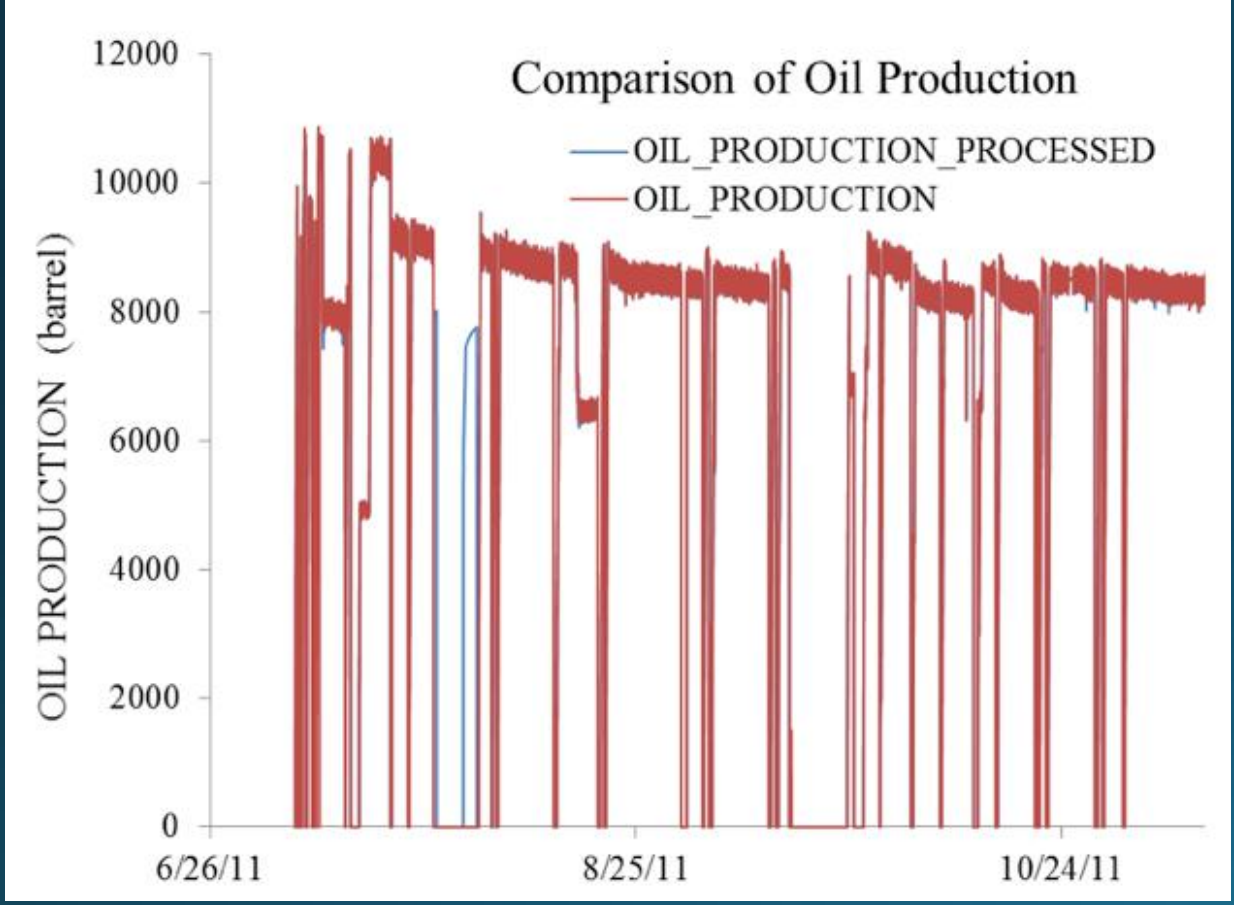
Preprocessed Results



Wellhead pressure

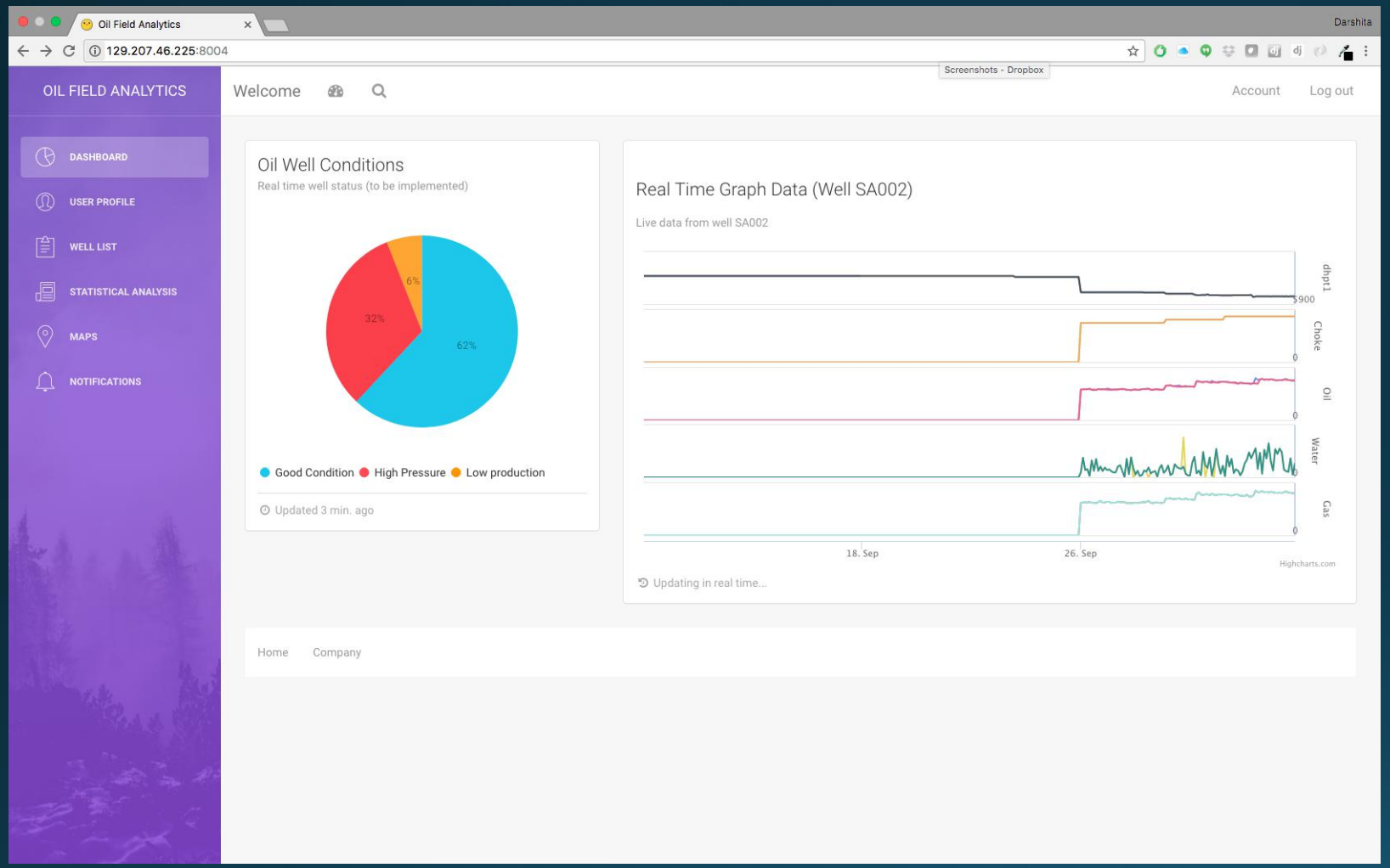
- Wellhead pressure has the most errors and missing ones
- It can not be higher than downhole value
- Use the downhole values to correct it based on the machine learning algorithms.

Preprocessed Results

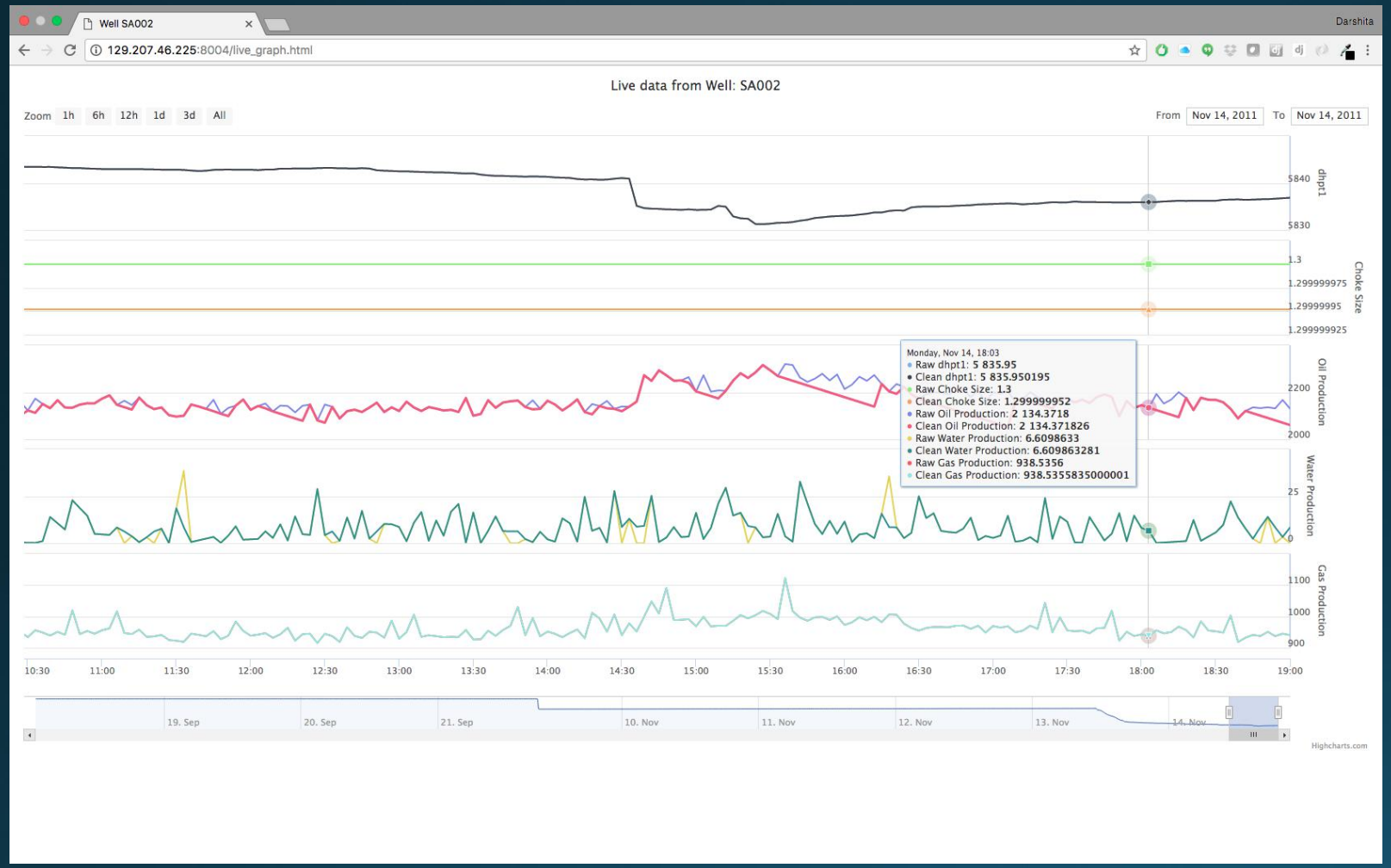


Production Correction

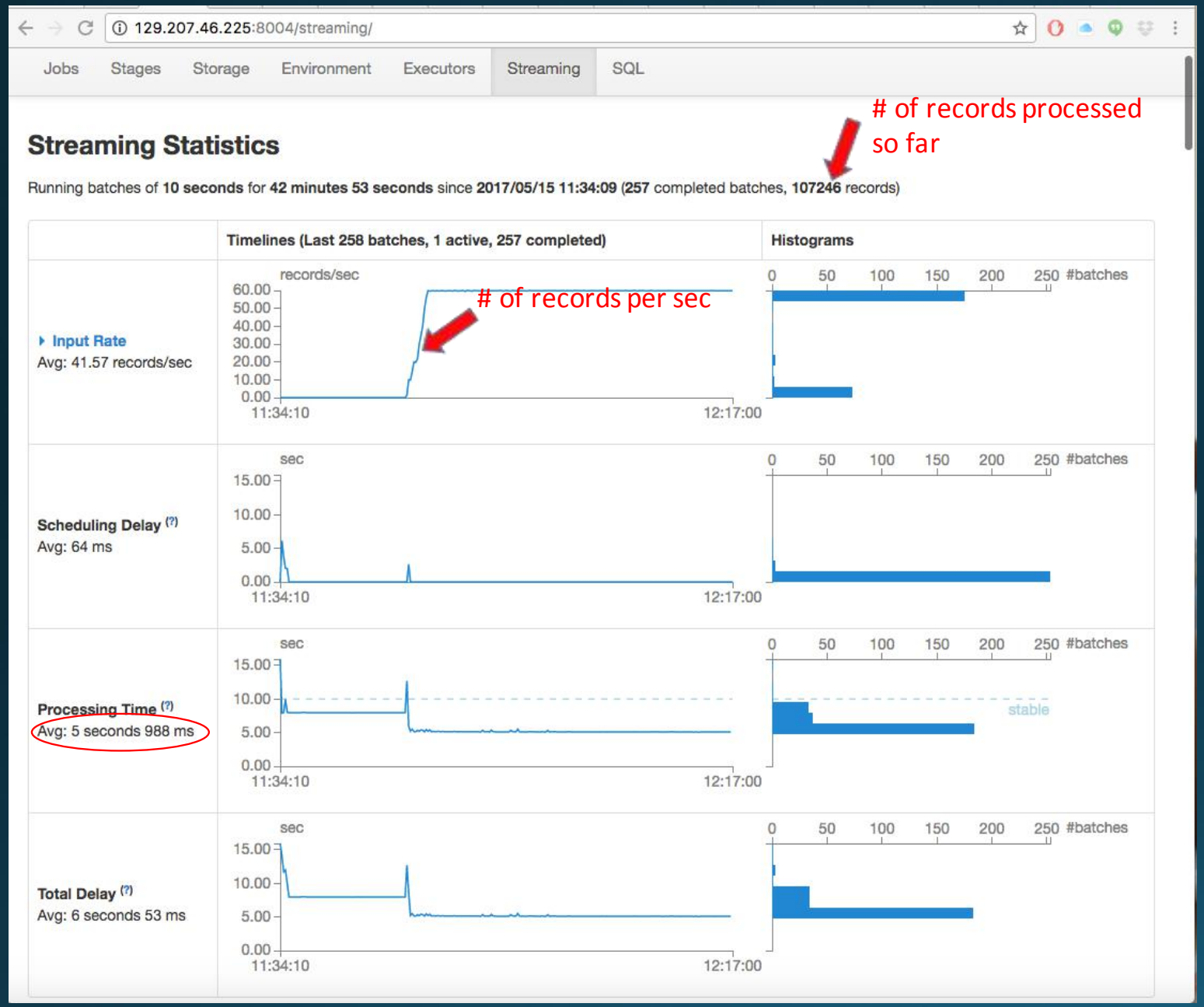
- Production should be zero when downhole pressure is maximum and stable
- And it should not be zero when pressure is dropping
- Values are checked and fixed by machine learning algorithms



Real-time Dashboard

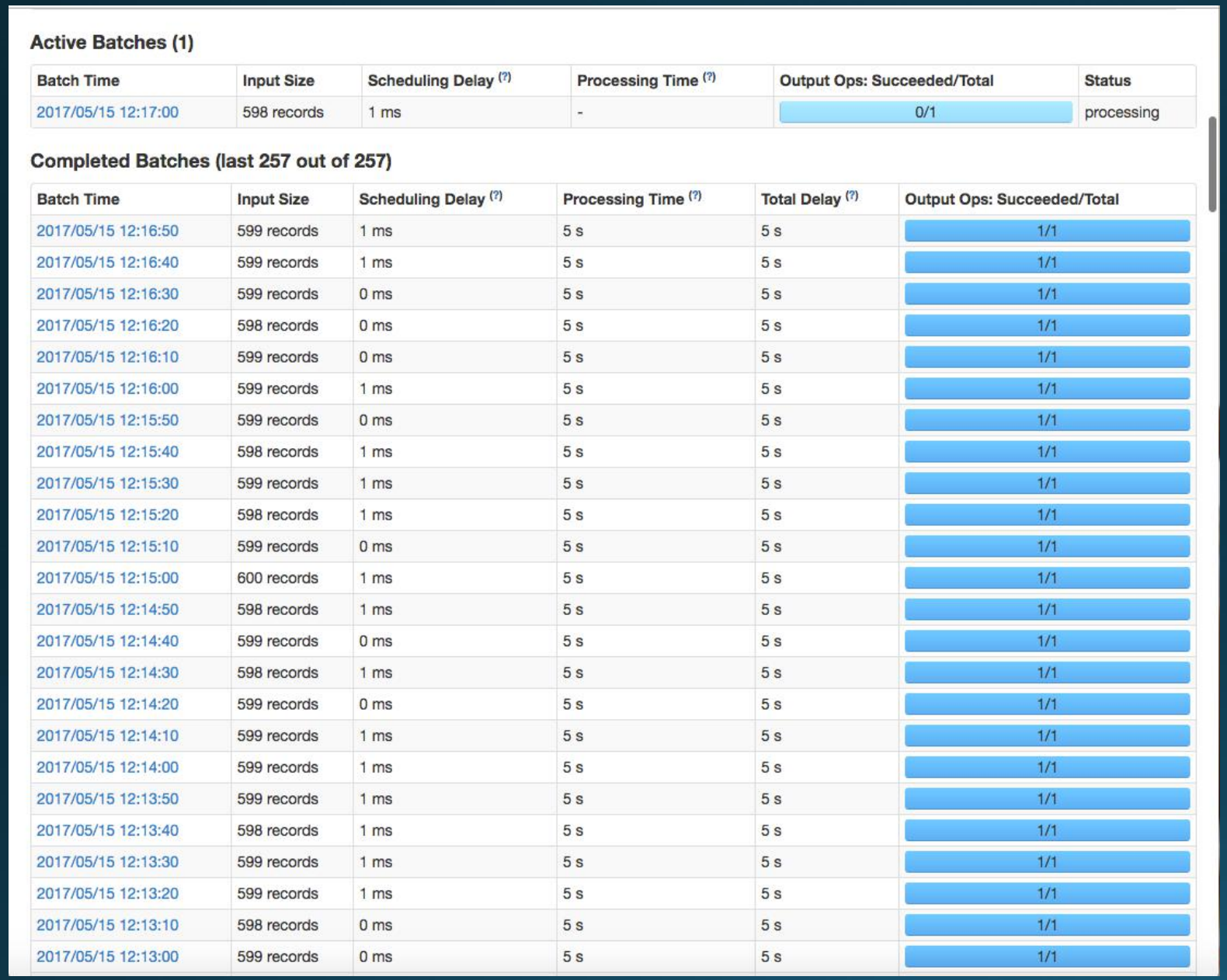


Live Streaming PDG Data



Apache Spark Web UI Screen Shot:

Streaming Statistics



Active Batches (1)

Batch Time	Input Size	Scheduling Delay (?)	Processing Time (?)	Output Ops: Succeeded/Total	Status
2017/05/15 12:17:00	598 records	1 ms	-	0/1	processing

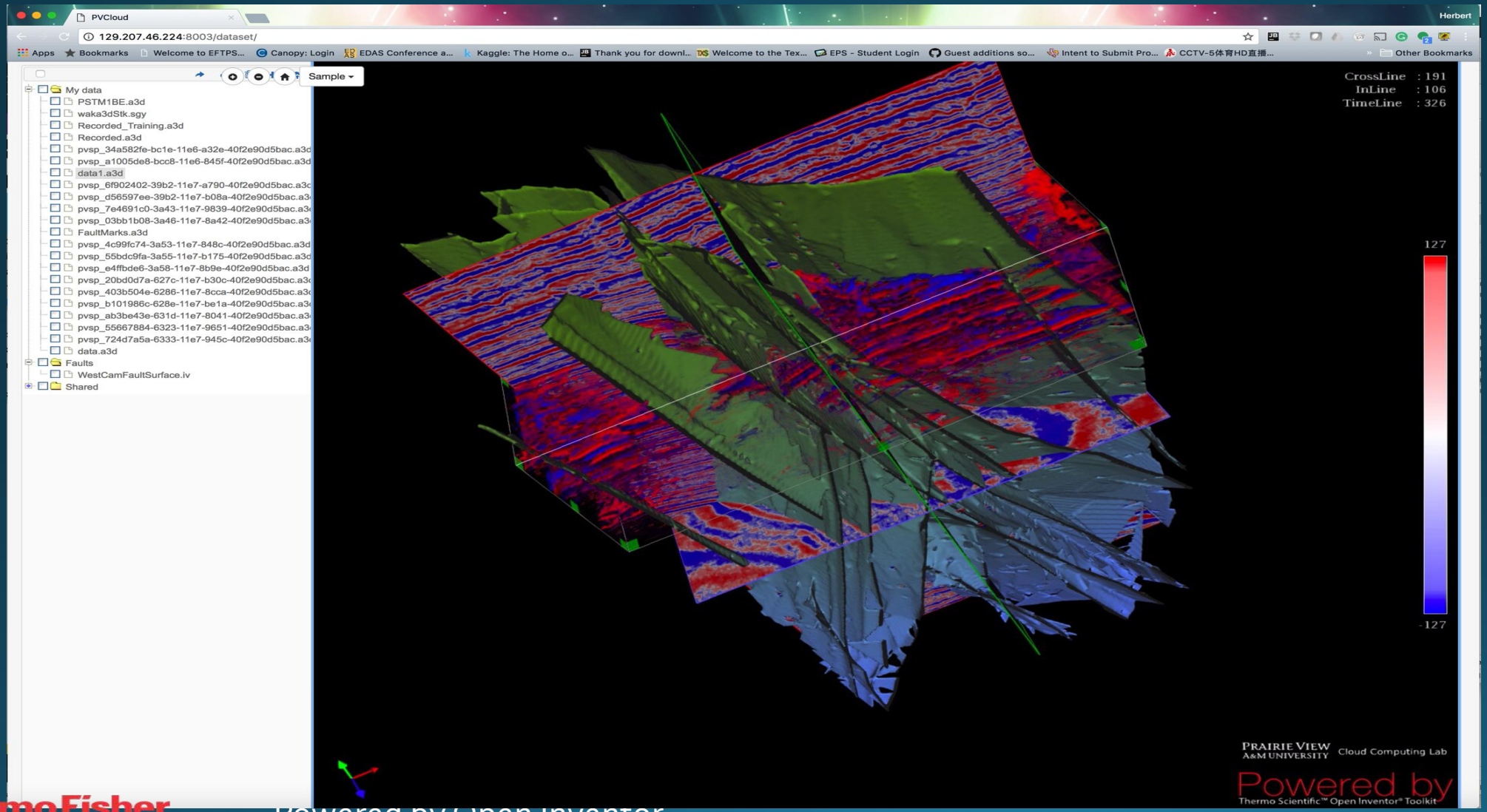
Completed Batches (last 257 out of 257)

Batch Time	Input Size	Scheduling Delay (?)	Processing Time (?)	Total Delay (?)	Output Ops: Succeeded/Total
2017/05/15 12:16:50	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:16:40	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:16:30	599 records	0 ms	5 s	5 s	1/1
2017/05/15 12:16:20	598 records	0 ms	5 s	5 s	1/1
2017/05/15 12:16:10	599 records	0 ms	5 s	5 s	1/1
2017/05/15 12:16:00	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:15:50	599 records	0 ms	5 s	5 s	1/1
2017/05/15 12:15:40	598 records	1 ms	5 s	5 s	1/1
2017/05/15 12:15:30	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:15:20	598 records	1 ms	5 s	5 s	1/1
2017/05/15 12:15:10	599 records	0 ms	5 s	5 s	1/1
2017/05/15 12:15:00	600 records	1 ms	5 s	5 s	1/1
2017/05/15 12:14:50	598 records	1 ms	5 s	5 s	1/1
2017/05/15 12:14:40	599 records	0 ms	5 s	5 s	1/1
2017/05/15 12:14:30	598 records	1 ms	5 s	5 s	1/1
2017/05/15 12:14:20	599 records	0 ms	5 s	5 s	1/1
2017/05/15 12:14:10	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:14:00	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:13:50	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:13:40	598 records	1 ms	5 s	5 s	1/1
2017/05/15 12:13:30	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:13:20	599 records	1 ms	5 s	5 s	1/1
2017/05/15 12:13:10	598 records	0 ms	5 s	5 s	1/1
2017/05/15 12:13:00	599 records	0 ms	5 s	5 s	1/1

Apache Spark Web UI Screen Shot:

Completed Batches

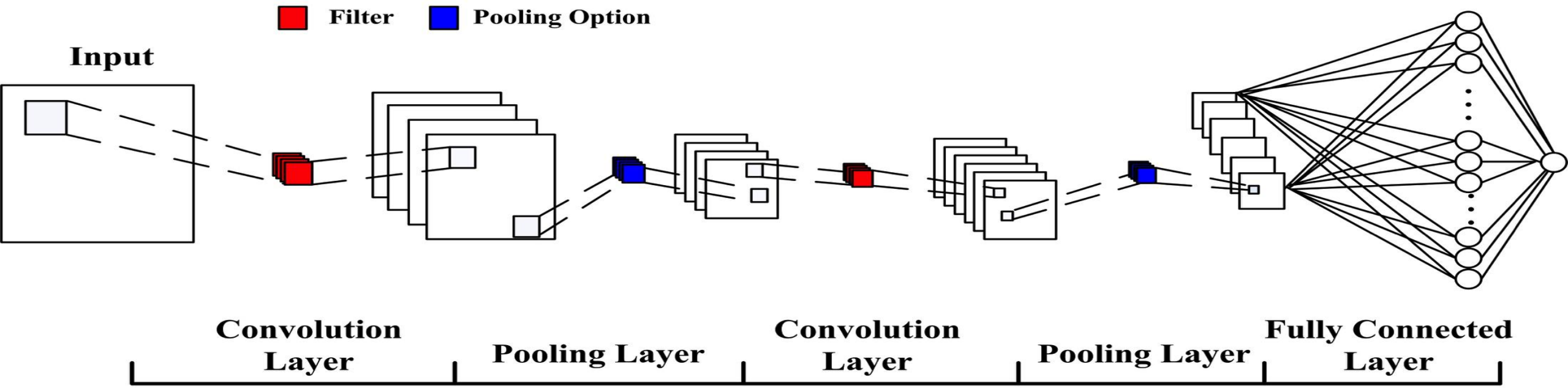
3D Visualization



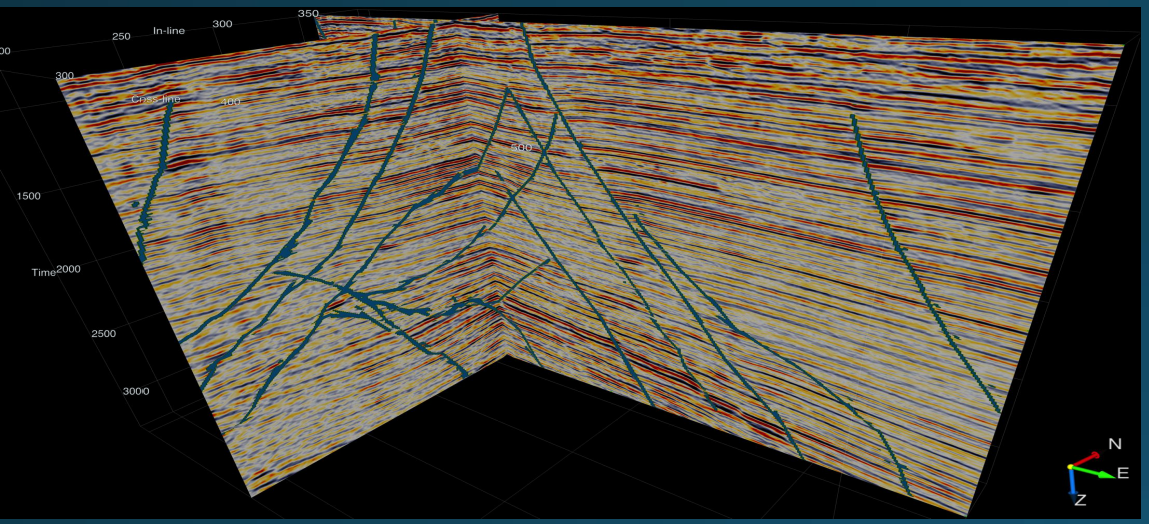
Powered by Open Inventor

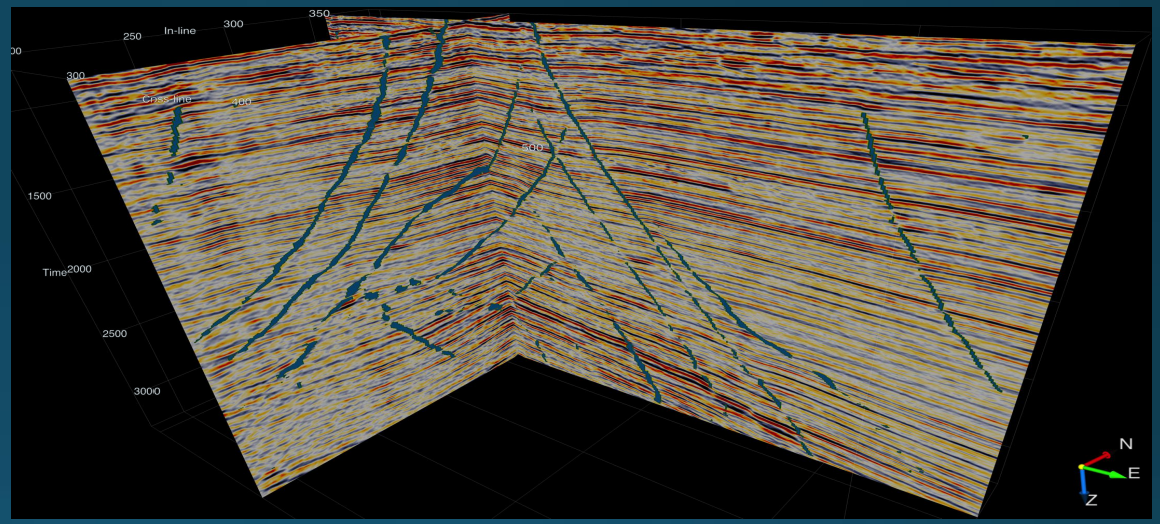
ThermoFisher SCIENTIFIC

Deep LEARNING:
Convolutional Neural Networks (CNN)



West Cameron Data Set





Expert-picked Faults

Machine-picked Faults

Conclusion

Our petroluem data analytics platform is capable of

- ingesting PDG streaming data from multiple wells
- denoising them using physical rules and machine learning algorithms
- processing them to compute basic statistics
- Storing raw streams data and processed streaming data in a NoSQL database in real-time
- Applying deep learning to predict oilfield production, and facilitate seismic interpretation

Ongoing Work

- Designing and integrating machine/deep learning algorithm to detect events in real-time and forecast production for reservoir optimization.
- New deep learning models for seismic interpretation
- Enhance web interface for streaming visual analytics and statistical analysis.

Thank You

lhuang@pvamu.edu
Cloud Computing Research Lab
Prairie View A&M University
http://www.pvamu.edu/cs/