An Application of Sensor & Streaming Analytics to Oil Production

Chetan Gupta

Krishnamurthy Viswanathan, Choudur Lakshminarayan, Ming Hao, Umeshwar Dayal[1]

Ravigopal Vennelakanti, Paul Helm, Sumitha Rangaiah, Harikrishnam-Raju Sagiraju, Sunil Doddmani[2]

[1]HP Labs Palo Alto, [2]HP Enterprise Business

COMAD, Bangalore, December 21st 2011

Introduction

•Diverse complex systems such as are increasingly fitted with large number of sensors and other measuring devices

- Energy production and exploration, Transportation, Smart grid, computer networks, etc.,

•This data is primarily being stored in “historians”

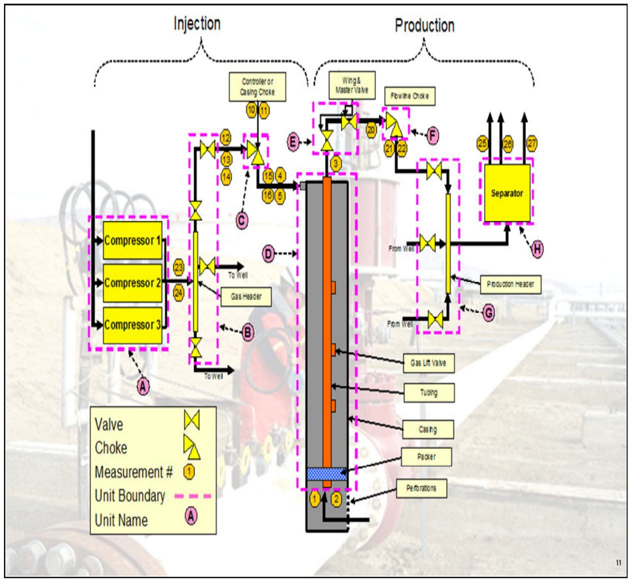
- Sub-optimal use of resources
- Delayed detection of emerging problems
- Reactive rather than proactive management

• Under the rubric of the “Live Operational Intelligence (Live OI)” program, at HP Labs we are interested in building solutions in various verticals to make sense of the data.

•We are bringing together techniques from various disciplines both in terms of processing techniques and the algorithms.

Oil Production – An Example Problem

- The oil and gas industry collects massive amounts of data from operations via sensors and operational logs.
- In oil production, there is a direct correlation between flow rate and revenue generated.
- Highly variable flow results in unpredictable conditions   and investigation shows that this can lead to a reduction in overall productivity of up to 5%.
- Onset of such unpredictable flow rates cannot be done with simple threshold based methods.
- Furthermore, once the onset has been detected, it is useful to help the operator respond in the correct fashion based



Basic Challenges

•Data of many disparate types – structured and unstructured, streaming and historical – has to be integrated, managed, and analyzed.

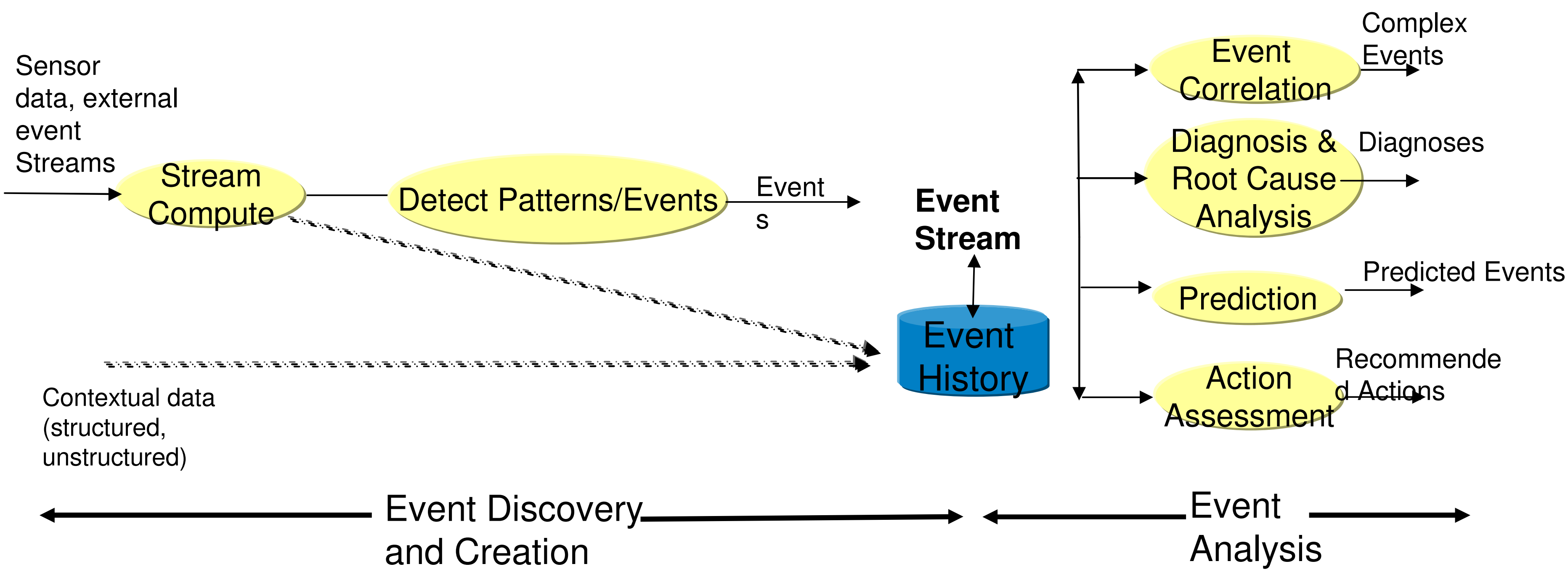
•Data from different sources has to be combined and aligned.

•Variability induced due to differences in calibration, data collection procedures, sampling rates, and terminology needs to be properly comprehended and adjusted.

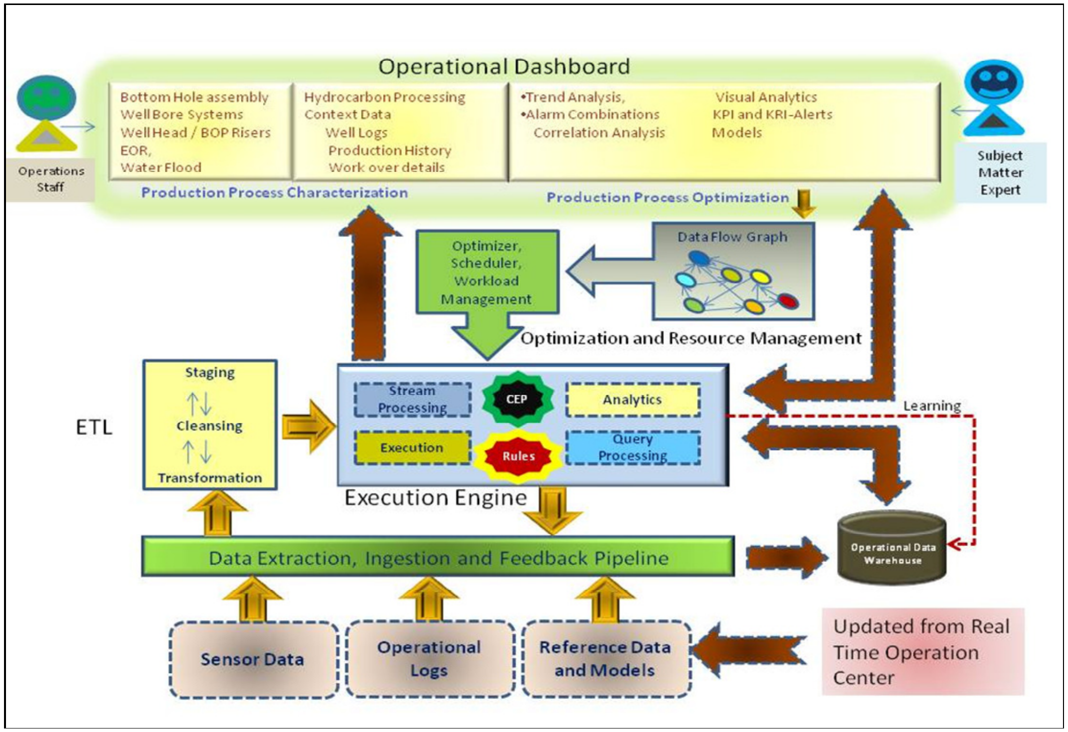
•Events such as sensor malfunctions, equipment failures, missing data, bogus values, and a myriad others, impose challenges.

•In addition to automated data analysis, the application must incorporate knowledge from human experts

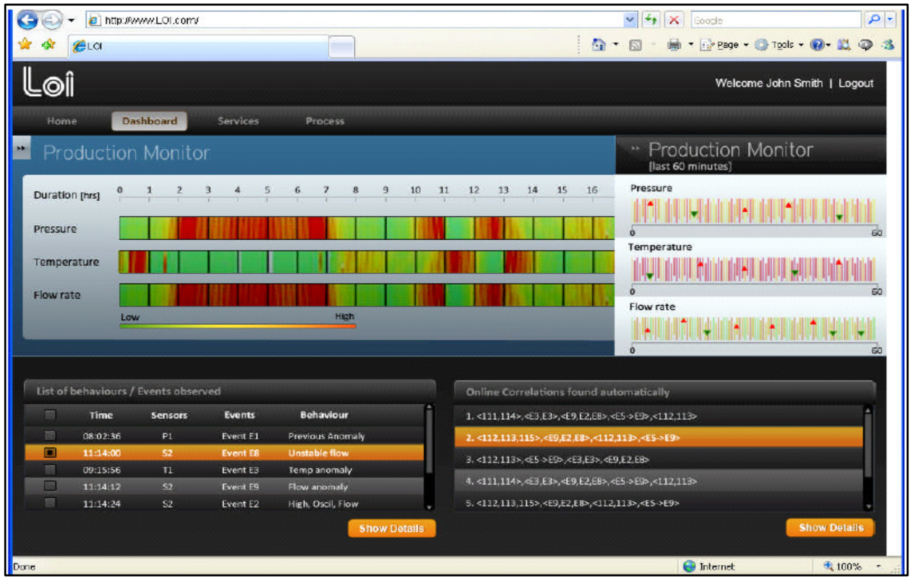
Stream & Event Analytics – General Model

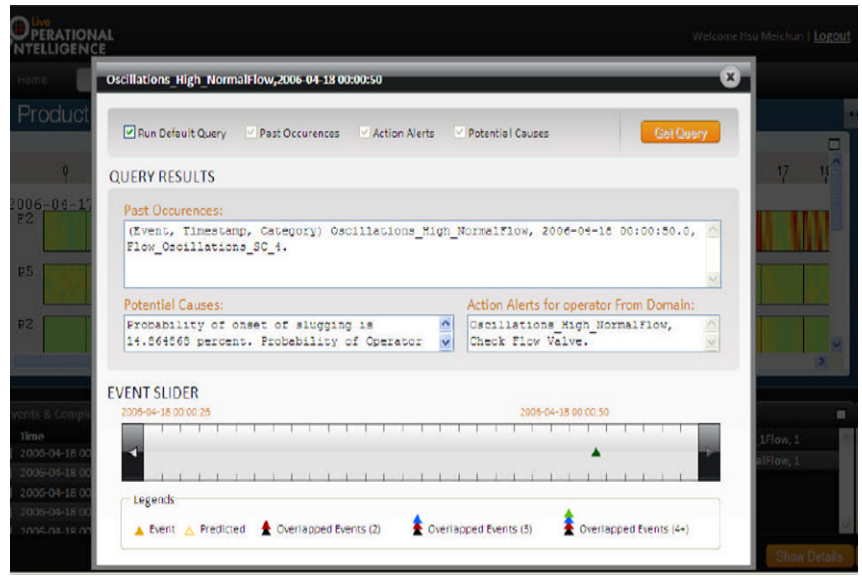


Reference Architecture



Illustrated Use Cases





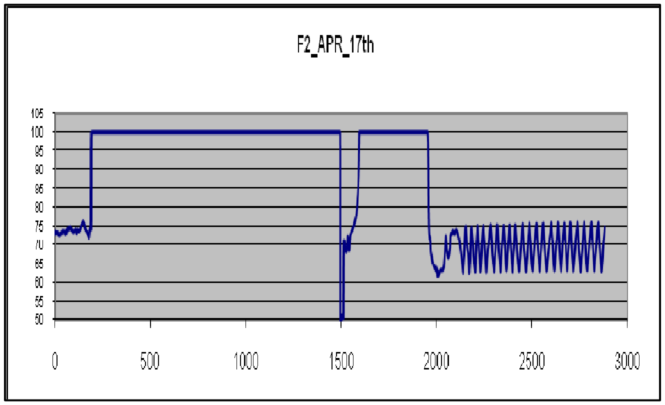
Multivariate Time Series Methods

•The analytics engine for oil and gas production utilizes a variety of algorithms for anomaly detection, pattern detection, and time-series prediction.

•Anomaly detection

- Relies on traditional threshold methods based on the Gaussian distribution.
- More sophisticated method – wavelet based, data mining

- Oscillation Detection
- Typical flow-rates are composed of multiple regimes characterized by normal oscillation, high amplitude oscillation, and low amplitude oscillation
- The data is both non-stationary and non-linear. These inherent conditions render fast detection of regime change of non-linear time series difficult in an on-line fashion



Online Event Correlation

•The purpose of online event correlation is to find correlation between events flowing into the system in real-time.

•Two events are *correlated* if they are in the same “time-space neighbourhood”.

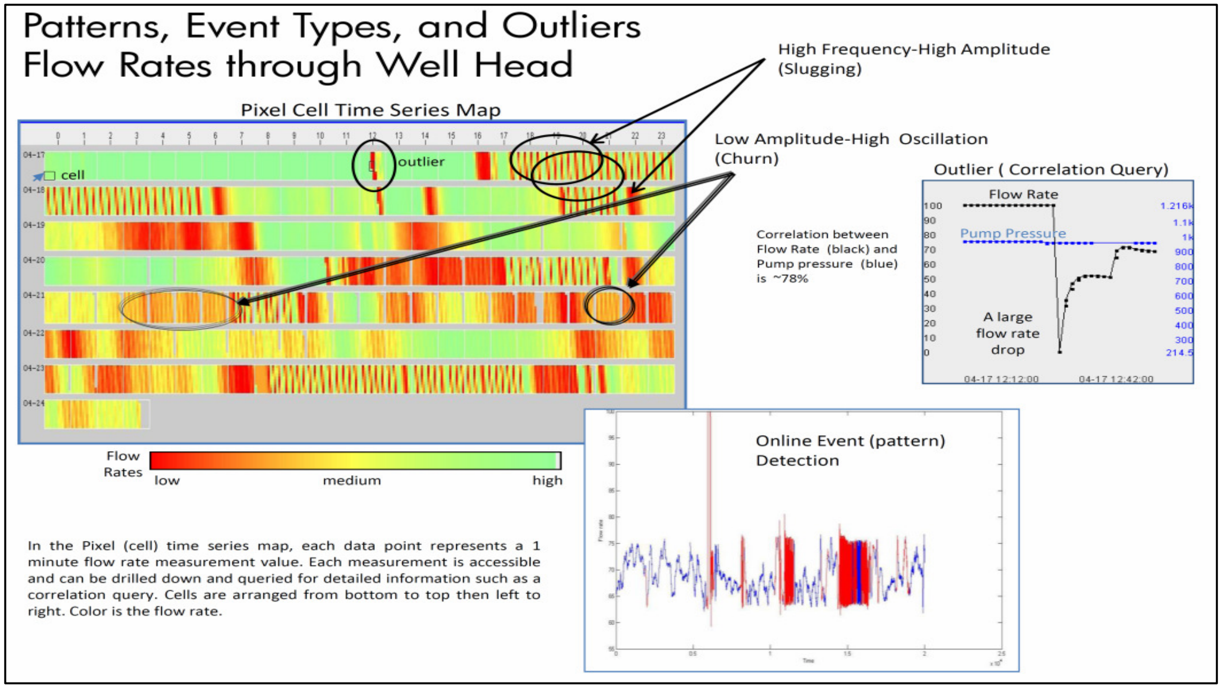
•For finding neighbours in time sliding windows are maintained and for neighbours in space a data structure called *Hierarchical Neighbourhood Tree (HNT) .*

•Store all the events in a time window in the HNT.

•Whenever, two or more events occur in the same hierarchical neighbourhood a correlation event is issued.

•Each correlation is weighted based on the size of neighbourhood and the frequency of events involved.

Visual Analytics

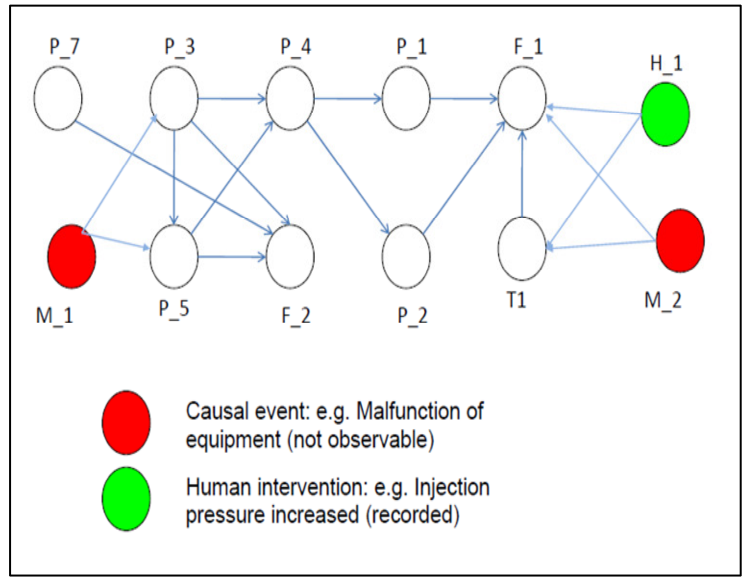


Bayesian Causal Models

•Determining root causes of observed events is a crucial subtask for improving the operations in oil production.

•Given the value of some of the random variables, it is possible to infer the posterior probability of the causes conditioned on the observed variables.

•The structure can be obtained with the help of domain experts and the probabilities can be learned from the data.



Summary

•We have built a prototype of the Live OI system and implemented use cases from oil production.

•We have validated the results with domain experts.

•We have also extended the Live OI prototype to monitor drilling operations and are building an instance of Live OI engine for traffic management.

•Besides building solutions for the various industry verticals, we are working on adding new analytic functionality to our engine with new algorithms for time series matching, and event correlation.

Challenges & Future Work

- There are many architectural challenges before us such as:
  - lack of standards and naming conventions.
  - lack of common integration framework.
- Besides these, there is a need for better algorithms :
  - For outlier detection and event correlation
  - Need to handle uncertainty in data
  - Solving for hybrid queries is also an emerging area in the database community.
- Finally, in large scale deployment of such a system, as we design our execution flows, we will need to optimize for various competing objectives.

Thank You. Questions?