Abstract— Big data analytics is the journey to turn data into insights for more informed business and operational decisions. As the Chemical Engineering community is collecting more data (volume) from different sources (variety), this journey becomes more challenging in terms of using the right data and the right tools (analytics) to make the right decisions in real-time (velocity). This paper highlights recent advancements in the big data analytics journey at The Dow Chemical Company in the areas of Enterprise Manufacturing Intelligence, multivariate analysis, on-line fault detection, inferential sensors, and batch data analytics.

Keywords: Big data analytics, Internet of things, Data-driven modeling, inferential sensors, Machine learning

I. INTRODUCTION

The big data era is driven by the explosion of data in all fields in terms of new data generation (e.g., social media), new measurement capability (e.g., internet of things, smart digital devices), improved data storage power (e.g., cloud computing), and improved computing technology for analytics (e.g., machine learning, artificial intelligence, cognitive computing).

Early discussion of big data date back to 2001, when an analyst of META (currently Gartner) used 3 V’s to describe the characteristics of the growth in data [1]

- Volume: the ever increasing amount of data generation and collection
- Velocity: the need for faster collection, processing speed to deal with large volume of data
- Variety: the need to contextualize all types of data including structured data and unstructured data (such as texts, audio, video, webpage, report, etc.)

Since the term “Big Data” was coined, there have been more debates on its precise definition. As of October 2016, Wikipedia’s definition is “Big data includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data size is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data.” [2]. ARC advisory group points out that this is a common misconception that big data is a thing [3]. To clear the misconception, ARC defines big data as “a journey toward more informed business and operational decisions” [3]. Simply put, a big data journey has started if volume, velocity, and variety characterize your data challenges and steps are taken to seize the opportunities. To emphasize more on the potentials of big data, a more preferable industry term big data analytics is used in this article.

In a chemical process industry context, Qin [4] commented that for well understood chemical mechanisms, first principles approaches can be used effectively to develop mechanistic models for process operations. For complex processes where first principles are not well understood, process data analytics are valuable assets to provide insights on process improvements. Big data has room to grow into a new paradigm for process industries to enhance data-driven operations and control. In Chemical Engineering Progress’s March 2016 special issue on big data analytics, a four-article series outlines the Why (why you should care about big data) [3], the What (success stories in the process industries) [5], the How (Getting started on the journey) [6], and the Future (Challenges and future research directions) [7]. Big data analytics is a fast growing research area in the Chemical Engineering community, challenges and directions are discussed in details in a review paper [8]. The goal of this paper is to share big data analytics implementation examples at The Dow Chemical Company.

II. ADVANCES IN BIG DATA ANALYTICS

A. Enterprise Manufacturing Intelligence

In process monitoring, univariate control charts have been the de facto standard in ensuring processes are operating within safe limits. As plants become more heavily instrumented with thousands of sensors and actuators, the large volume of data can easily overwhelm plant personnel. As a result, many of the data streams are often not examined. Enterprise Manufacturing Intelligence (EMI) is a platform to contextualize critical plant key performance indicators (KPI) into dashboard for real-time visualization. Based on process knowledge, a troubleshooting guideline can then be built into the EMI platform, facilitating data and knowledge-driven decisions. The EMI platform has shown measurable successes at Dow [6, 9].

B. Multivariate analysis

Multivariate analysis is another way to contextualize a large volume of data. Continuous processes result in dense (as opposed to sparse) and structured data streams. Multivariate analysis is an approach to exploit the naturally occurring correlation structures in these dense data sets due to flow, mass transfer, energy transfer and basic
A comprehensive overview of the state-of-art fault detection methodologies are available in [10].

Among the data-driven modeling methods, projection based methods such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) are used in multi-effect fundamental problem solving at Dow. Large scale chemical plants do not operate at true steady state because of changes in raw materials, product grade, production rate, and operating strategy. While advanced process control strategy is used in Dow plants to compensate disturbances effectively for normal operations, multi-effect process faults often negatively impact product quality and production volume. To resolve complex multi-effect process issues quickly, multivariate analysis combined with chemical engineering knowledge is the key to turn data into new leading hypotheses, which eventually pinpoint the root cause.

C. On-line fault detection

To be proactive in preventing multi-effect process faults, multivariate analysis can be implemented on-line to detect fault. The first on-line Dow application to improve product quality was published in [11]. After more than 10 years since the model was debuted, the model continues to function as designed. Today plant personnel rely on the model for product release as part of the standard operating procedure. The next milestone of on-line fault detection at Dow is in the application of process monitoring of large-scale continuous processes with hundreds to thousands of process variables. On-line fault detection is shown to be effective for plant personnel to detect process issues early.

D. Inferential sensor

Inferential sensors (or soft sensors) predict important variables on-line using available process data. The important variables being predicted are often difficult or uneconomical to measure online [10].

Dow inferential sensor applications in Predictive Emissions Monitoring System (PEMS) started in early 2000s. These models were approved annually by Environmental Protection Agency (EPA) as a reliable mean to monitor emissions. [12].

Inferential sensors are also used for advanced control or quality monitoring and allow plants to react faster to process excursions to prevent off-grade products. As a result of improved process control, plant can operate closer to operating limits to maximize production rate. Over 300 inferential sensors are deployed globally across Dow, an active research focus at Dow is to increase life span of these soft sensors through robust model building [13] and model maintenance strategy [14].

E. Batch data analytics

Unlike data analytics applications in continuous processes (Sections A to D), consistent and structured data sets are rare in industrial batch processes, in which, vector, matrix and sometimes tensor data can exist. The data quality is often poorer due different sampling rates, missing context and limited instrumentation. Unfolding, warping and feature extraction are common techniques to preprocess the data prior to modeling [15]. Batch processes are often multi-step or multi-stage batch processes, inputs can also be further grouped by phase or blocks to increase interpretability of the model results. Batch data analytics implementation experiences at Dow are published in [16].

III. CONCLUSION

Dow is an early adopter of big data analytics. Online instrumentation, centralized control systems, and storage of process data enable big data analytics implementation. EMI, multivariate analysis, on-line fault detection, inferential sensors, and batch data analytics have found success at Dow. With recent advancements in both computer hardware and numerical algorithms, big data analytics have room to grow in Dow to enterprise-wide implementations using data with more volume, variety and velocity.

REFERENCES