WHITE PAPER

FAST ANALYTICS OVER SLOW DATA

VITRIA®

EXECUTIVE SUMMARY

Fast analytics over slow data might sound like an oxymoron. When we talk about fast analytics we somehow relate to its applicability only for data that is very fast and where we need faster response time for actions. That might make sense in some use cases but the Internet of Things (IoT) brings a different challenge. In IoT we will end up dealing with enormous amount of data that has a high degree of variance over speeds, feeds and data cycles. As we see millions and billions of devices in IoT being connected, each passing moment we see a significant amount of new data that can be used to detect anomalies, predict problems early, mitigate any disruption of service or provide new customer experiences. As a matter of fact it can be the linchpin in the Internet of Things (IoT) to deliver timely insights and outcomes.

This paper highlights how fast analytics over any data speed – fast, slow or in between – can help you improve operational performance and monetize your IoT network.

Contents

Today's asset-intensive companies are replete with equipment and systems with embedded sensors generating data. Thousands of sensors and controllers continually measure, report and record temperatures, pressures, flows and vibrations — sometimes at sub second intervals — and time stamp the data (often referred to as time series data). Huge volumes are collected and stored, but most of it is never touched again, since it falls outside the traditional confines of IT systems, and is essentially "dark data" to the IT department.

Source: Gartner, 2014.

THE INTERNET OF THINGS

The Internet of Things (IoT) is poised to change everything – economy, industries and lifestyle. There are different categories and sub-categories that are morphing such as machine to machine (M2M), the Internet of Intelligent Things, the Industrial Internet and the Internet of Everything. No matter what we want to call them, it basically consists of sensors, machines or network of things communicating with other things or people resulting in huge volumes of data. It is this untapped volume of data, in different forms and shapes, which can have a transformational impact if analyzed in a timely fashion to provide insights and actions to deliver impactful outcomes.

Value Potential

The industry estimates over 25 billion IoT devices by 2020[1] and $15 trillion of global GDP by 2030[2]. Directionally these numbers are huge and even a fraction of that is too big to ignore. The proliferation of such IoT devices will be across industrial, enterprise and consumer segments. Whether it's a smart grid or smart home, the data from sensor or device is continuously flowing to the network and to the end-user, and in some situations back to the device. This connectivity of devices, equipment, factories, products and supply chain to the network or to the cloud is leading to massive volumes of data every second. And, it is these volumes of data combined with the advancements in technological capabilities in the area of advanced analytics to provide insights and patterns that can bring timely outcomes across 3 major segments of IoT (Figure 1) – industrial, enterprise and consumer:

1. **Safety and Security** – This is a mission-critical aspect for all segments of IoT addressing cyber security, physical security, fraud detection and personal safety.
2. **Operational Efficiency –** In the Industrial IoT alone, 1% of improvement in operational efficiency, such as predictive maintenance and asset optimization, translates into $300 billion savings over 15 years[2].
3. **Revenue Growth –** With new business models and services, service providers can enable better customer engagement models and predictive 1-1 marketing to provide end-user centric services based on the lifestyle of target consumers.

[1] Gartner
[2] General Electric, Wikibon

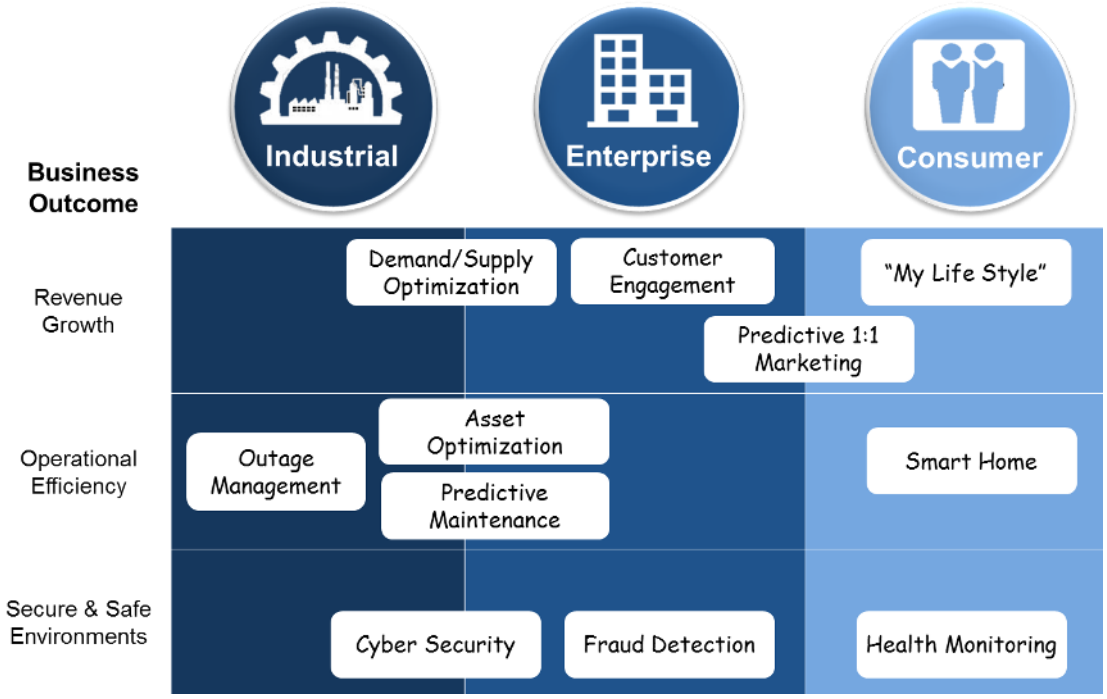

Figure 1: IoT use cases across market segments

Business Imperative

Over the past 20 years, response time of 2-3 days to resolve customer issues in supply chain was considered the norm. However, with the advancements of IP-based technologies for ubiquitous connectivity, mobility and cloud services, customers are expecting 24x7 always-on service availability with minimal service disruption. More importantly, time to act in real-time is also becoming a key SLA (service level agreement) in secured and safe environments. With IoT this window for time to act is even going to shrink from days to minutes to seconds to milliseconds, and the value will diminish rapidly beyond that time window, if not acted upon.

For example, as show in Figure 2:

- In providing electric service, the time window to detect an electricity shortfall and respond will be less than 30 minutes.
- In customer contact center, the time window to provide a better customer service to avoid churn will be less than 30 seconds.
- In smart grid, the time window to detect and respond to a cyber-security breach will be in milliseconds.

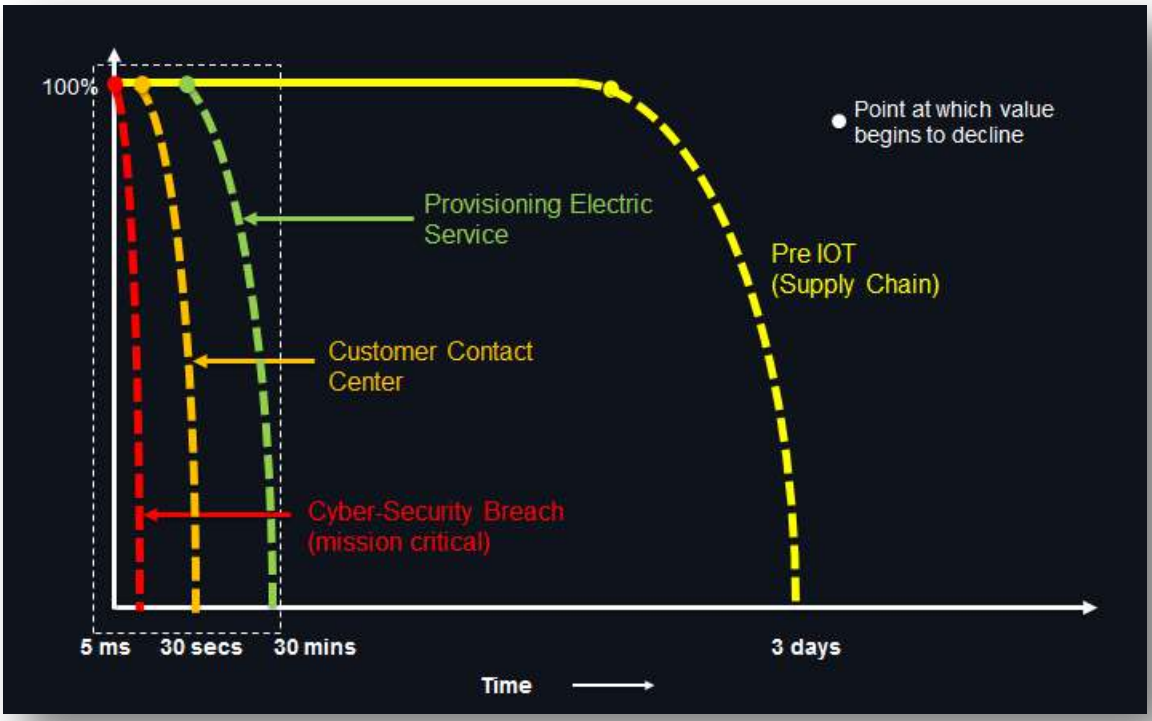

Figure 2: Time to Action

With the growing volume of real-time data in IoT and with the reduced time for decision making, companies need to leverage advanced real-time analytics with predictive and historical models to rapidly assess situations of opportunity or threat before they occur.

Fast Analytics

IoT contributes significantly to the rising volume of sensor data. It is estimated that by 2020, over 40 percent[3] of all data will come from sensors alone. Considering the sheer volume of data from sensors and devices, traditional batch analytics alone will not be able to provide real-time business outcomes. To achieve timely-actions, as outlined above, we need to process data as it streams in real-time using emerging fast analytics, also referred as streaming analytics. Key areas where fast analytics differs fundamentally from traditional batch analytics are:

> As IoT use cases proliferate, tools and techniques such as streaming analytics to handle continuous incoming data, machine learning to enable performance improvement of IoT applications over time (by "learning" from IoT data), and data visualization capabilities may rise in importance.
>
> Source: Cisco, 2014

[3] Trillions of Sensors Feed Big Data, Signal Online, 2014

- Ability to ingest and analyze immediately, in real-time.
- Ability to update analysis continuously as data flows by incremental algorithms.
- Ability to deliver insights for the next actions to capitalize on an opportunity or to mitigate a threat.

FAST ANALYTICS OVER FAST DATA

Intuitively it makes sense to apply fast or streaming analytics to high velocity sensor data, inflowing at millions of events per second. If you take a smart grid as an example, power outages alone cost the US economy over $150 billion[4]. In most situations, severe weather combined with the equipment behavior is the main reason for such outages. Such service disruptions can be significantly minimized by leveraging advanced analytics such as streaming, predictive and prescriptive analytics.

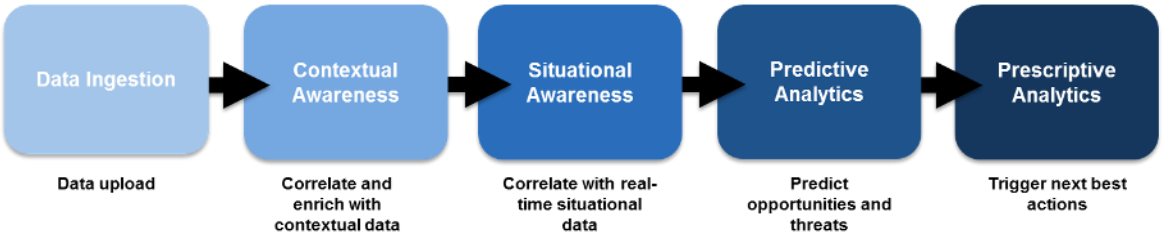

Figure 3: Time to Action

As shown in Figure 3, predicting equipment failure can be achieved by addressing:

1. Ingestion of the data at speed and volume.
2. Correlation of all the contextual data needed for advanced analytics, such as equipment profiles, recent maintenance history and topology.
3. Correlation with real-time situational data such as weather.
4. Prediction of equipment failures using advanced machine learning and predictive models.
5. Prescription of actions to minimize failures or to address any outages using prescriptive analytics.

[4] www.smartgrid.gov

Streaming analytics for fast data speeds can be processed, as outline in the above steps, within a second of data inflow to enable predictive and prescriptive actions in time to impact an outcome. This is an excellent example of applying fast analytics on fast data.

FAST ANALYTICS OVER SLOW DATA

Fast analytics over slow data might sound like an oxymoron. When we talk about fast analytics we somehow relate to its applicability only for data that is very fast and where we need faster response time for actions as outlined in the previous section. However, in IoT we will end up dealing with enormous amount of data that has a high degree of variance over speeds, feeds and data cycles.

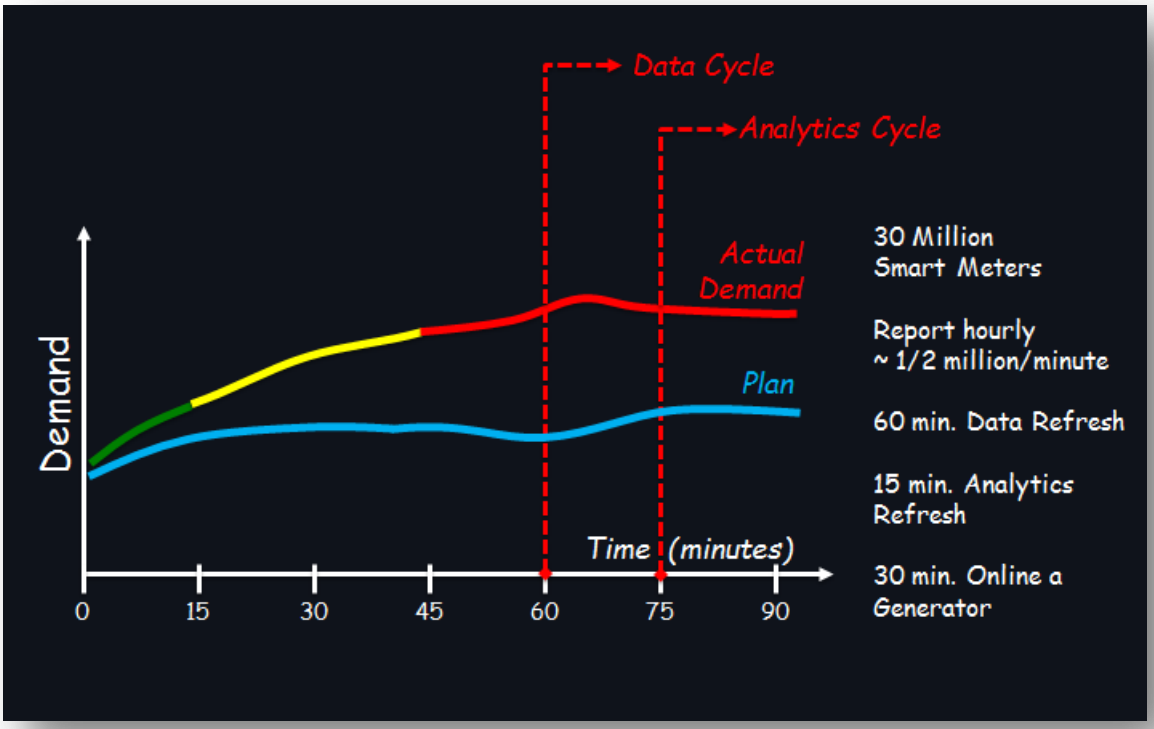

Figure 4: Supply-Demand Forecast with Traditional Analytics over Slow Data

To illustrate this aspect of fast analytics over slow data, let's use the smart grid use case but more on the meter readings. Consider a smart grid with 30 million smart meters with each meter reporting its energy consumption once an hour. It takes 60 minutes to complete a data cycle and additional 15 minutes to compute the analytics with traditional methods resulting in

75 minutes for a full data-to-analytics cycle. Such long cycle times are slow to react to sudden or unexpected changes in weather. Let's take a midsummer day where the temperature is unexpectedly hot and humid. The supply plan would follow the blue line as show in Figure 4, however the actual demand due to sudden surge in temperature will follow green to yellow to red line (Figure 4) and eventually exceed the high-end of the capacity threshold. With the traditional analytics you will not be able to foresee the problem until 75 minutes into the cycle where the actual demand is way above the supply plan. In addition, it would take another 30 minutes to spin up a new generator online to meet the demand and by then we are looking at more than one and a half hour during which the SLAs will not be met.

As a matter of fact, we will be receiving 500,000 updates every minute in this 30 million smart grid network. Statistically this is a significant population that we can leverage to predict trends.

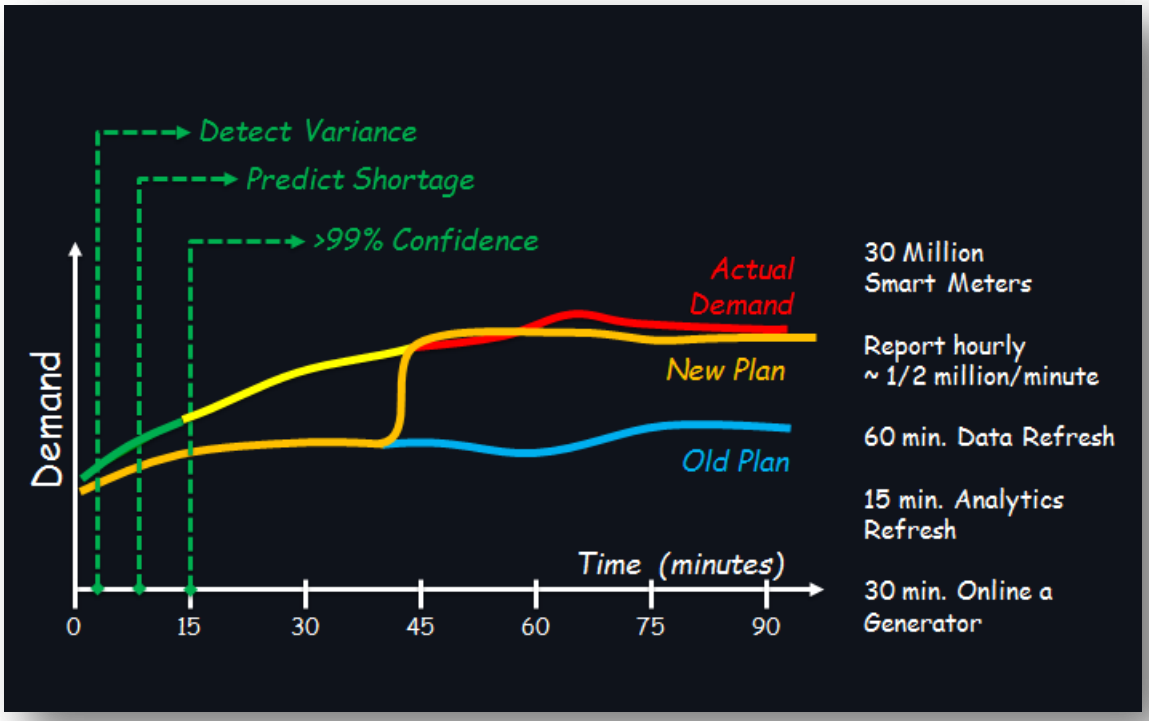

Figure 5: Supply-Demand Forecast with Fast Analytics over Slow Data

By applying fast analytics over a slow data, as show in Figure 5, we can achieve the following:

- Within 5 minutes into the data cycle we can detect a variance from actual demand based on 2.5 million new readings.
- Within 10 minutes into the data cycle we can predict a shortage of capacity based on 5 million new readings.

- Within 15 minutes into the data cycle, we can predict an energy shortfall with over 99% confidence level based on 7.5 million new readings. At this point we can start to spin up a generator.
- Within 45 minutes into the data cycle, we can bring a new gas turbine online, just-in-time to avoid the red state (New Plan in Figure 5).

This scenario is highly relevant and extensible across all IoT segments, where we are dealing with millions of sensors, meters or devices providing continuous feeds at a given time. Applying fast analytics over sub-populations of slow reporting devices, with new data inflowing every second, improves business outcomes across industrial, enterprise and consumer segments.

CONCLUSION

We have seen a use case, as outlined above, where **Fast Analytics over Slow Data** can make a huge impact. We see this kind of use case repeated over and over in IoT, where devices report in on an hourly, daily or weekly basis. However, since there are millions of them, with each passing moment we see a significant amount of new data, from which we can:

- Detect anomalies and deviations from plan early.
- Predict problems early.
- Act in time to avoid or mitigate problems.
- Future-proof the operations against every faster data cycles.

So, as your data cycles increase and your time-to-action decreases, fast analytics can keep up. As a matter of fact, fast analytics can be the linchpin in the Internet of Things (IoT) to deliver timely insights and impactful business outcomes.

www.vitria.com