Solution Showcase

HPE Solutions for Internet of Things Analytics

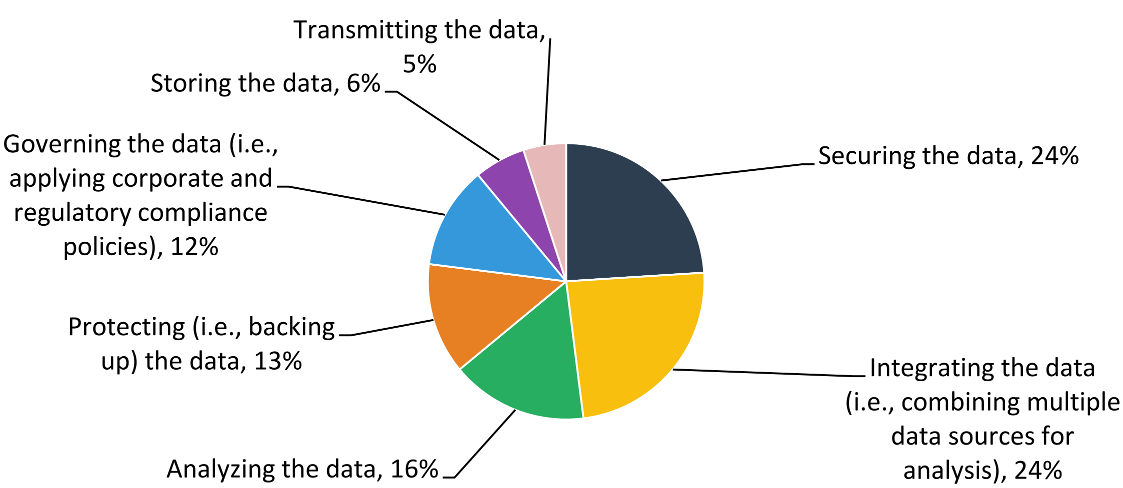
Date: June 2016 **Authors:** Eugene Signorini, Senior Analyst & Nik Rouda, Senior Analyst

Harnessing the Power of Data from the Internet of Things (IoT)

Any organization deploying IoT will need analytics. Sensor, CCTV, machine, and social media data is only useful if it can be understood and acted upon. Yet one of the unique challenges in utilizing IoT data with respect to analytics is the way it is distributed. Most data will be generated at the edge, by vehicles, in manufacturing plants, from mobile devices, and by large machinery. Really just about *anything,* any*where* could be producing raw information in an IoT world, but this data must be collected, processed, and analyzed to unlock its value. If a signal falls in the forest and no one is around to hear it, does it make a difference? Instrumentation of devices is only the starting point. IoT data can be an organizational asset if the business can build an intelligent feedback loop from things to ideas to actions, spanning remote locations and central data centers. There is an inevitable intersection of analytics and IoT, and it's not always an easy one to navigate, as shown in Figure 1.[1]

FIGURE 1. Data Presenting Organizations with the Biggest Challenges from a Technology Perspective

When you consider all of the data your organization collects or will collect as part of IoT initiatives, which of the following would you consider to be the biggest challenge from a technology perspective? (Percent of respondents, N=501)



Source: Enterprise Strategy Group, 2016

[1] Source: ESG Research Report, *2016 IT Spending Intentions Survey*, February 2016.

According to ESG research, 24% of organizations believe data integration will be the biggest challenge for their IoT initiative, with the same amount of respondents citing security concerns. Analyzing the data is close behind, with 16% concerned about how to most effectively understand what their IoT system is telling them.

Data Considerations for IoT

There are a number of key challenges for analytics of IoT data, which may require different approaches than other use cases. Some of these differences stem from the nature of the data being generated in IoT environments, including industrial applications, surveillance infrastructure, and social media. Here are some common considerations:

- Data may be generated from a different set of sources, often machine-generated log files or instantaneous signals from sensors.
- The outputs can occur at set frequencies; on specific events, thresholds, or exceptions; or may even take the form of continuous state information.
- The monitored signals might even be analog and need to be translated to digital records for storage and analysis.
- Further, the data volumes can be daunting, depending on the number of monitored inputs, the frequency, the average size of readings, and many other variables.
- These sources can export information in many different formats, requiring standardization and integration of each.
- Data quality can also be an issue, with environmental factors causing inaccurate measurements, interruptions, and other problems.
- Text and rich media data tend to be contextual and require deep understanding for effective interpretation. For example, the word “accident” has a very different meaning when it refers to spilled coffee compared with when it describes a train derailment. Similarly, video of a person entering an office at 8am will likely be interpreted differently than footage of the same person entering an office at midnight.

Take, for example, monitoring the performance of shipboard diesel engines. There could be a number of business reasons to do this: Better engine operating efficiency can significantly reduce fuel costs; more detailed records of service usage can determine the optimal parts and fluids maintenance schedule, minimizing downtime of an expensive asset; and failures can be predicted and resolved at a planned dock time versus stranding a ship at sea with cargo aboard. To achieve these goals, sensors can monitor oil pressure, engine temperature, RPMs, current voltage, fuel consumption, and many more direct factors. Some sensors may uncover normal behavior patterns that are interrelated—for example, upon startup, a cold engine may have higher oil pressure, lower temperature, and poorer fuel economy. That doesn’t necessarily mean the engine isn’t running well, just that it’s still warming up.

Context is important, and analytics must interpret both normal patterns and significant deviations. Yet some factors may be external. If the engine temperature doesn’t rise quickly, it could be a problem with a thermostat stuck open and circulating water from the radiator, or it could instead be that the engine is being started in Alaska in January. Running the engine cold for long periods will affect friction and fuel usage, so that temperature data should be combined with ambient weather information. To gather data for correlation, the organization may need to track longer term patterns, like routes, storms, tides, and similar info from NOAA or Department of Commerce data sets, not just the engine itself. Even an infrequent event, such as the scraping of barnacles off the ship’s hull may be relevant in determining how hard an engine has to work to maintain a desired oceangoing speed. Yet the service records might include a technician’s free form notes—unstructured data that still needs to be understood.

Follow the Data

If the nature of IoT data is somewhat different from other traditional analytical use cases, then so is the way that data flows across an enterprise. Much IoT data is created in remote locations, far from the data center, with one notable exception being machine log files used for IT operations analytics. Indeed, IoT data often comes from sensors that are embedded with operational technology (OT) equipment that might be monitoring production quality in factories, sensing environmental conditions for scientific purposes, or tracking performance of logistics and transportation.

This distributed data generation means that there are some unique requirements for analytics activities to be effective. Data must be captured, collected, aggregated, and integrated within and across locations' boundaries. Sometimes, the analytics determine simply which data is important, separating significant signals from a lot of noise. Then the process becomes about applying an appropriate response, either sending this information back to a data center for further analysis or responding locally.

In the former "phone home" case, the meaning of a condition might be better determined in the context of a broader view, utilizing information collected from additional sites, over a longer time span, or from other relevant data sources that are housed and accessed centrally. Yet this approach may have some inherent limitations. It isn't always trivial, timely, or economical to send all data back to the data center. IoT networks may span mobile networks, satellites, or other expensive, high-latency, low-bandwidth connections[2] and those factors, with storage considerations, often serve as drivers for the use of analytics to determine what data to send back.

The second "local response" option (not mutually exclusive) is to perform analytics at the edge to generate an immediate reaction. This can be faster, since data doesn't need to make a roundtrip, less expensive for the same reasons, and more reliable in the case of disconnected operations due to a network fault. Edge-based IoT analytics can be performed in a local server, in a gateway, or even in the instrumented device itself, depending on the appropriate vantage point for the decision and the available resources. For an example, even without getting into the excitement of self-driving cars, having an onboard computer in your car to appropriately manage traction control and apply anti-lock brakes are simple examples of local IoT analytics applications where it's imperative to perform functions without delay and without dependence on a multistep connection to a data center.

Analytics Form Follows Function

Across an IoT environment, analytics can be deployed to useful effect at many points. Enterprises must deploy the right tool set to fit each step in the data journey from device to data center.

Yet the combination of both of these approaches can yield some of the most interesting insights, by employing holistic, long-term, and advanced analytics in the data center core, after local and immediate analytics have already taken place on the edge. Here, for example, one might get a view into each time anti-lock brakes were engaged, correlated and enhanced with information on speed, tire pressures, weather conditions, driving routes, driver habits, traffic levels, and more. From this viewpoint, more effective (safer) products can be designed, tested, and serviced.

Design Considerations for Analytics Infrastructure

Each of the above forms of IoT analytics will have its own nuances when it comes to choosing the right tools for the task. It's impossible to characterize all the possibilities in a short overview, but thinking about the considerations for analytics by position in the data flow is a great place to start. Here are some defining characteristics that should be considered at each stage of the whole environment, with illustrative examples of Hewlett Packard Enterprise (HPE) offerings.

[2] See ESG Solution Showcase, *HP Enterprise and IoT Connectivity*, March 2016.

Device-based Analytics

In-place, device-based activity requires lightweight, real-time/streaming analytics that operate on event-based, time-series, raw inputs, usually with no explicit need for human intervention. What this means will vary depending on the device itself. Obviously a modern smartphone has much more processing power than a smart light bulb. Yet each device is capable of measuring certain circumstances, acting in response to thresholds, and communicating with other devices near and far. Most of the device's data will be piped elsewhere for the heavier analytics. Many of these devices will come directly from equipment manufacturers and are designed for specific applications in industrial, health care, or consumer technology settings. Both open source and proprietary toolkits can run directly in these devices. HPE's Aruba networking offerings can encapsulate devices and also serve as a networking platform. Aruba MeridianService and Analytics and Location Engine software can facilitate device-centric applications, as well.

Edge Gateway-based Analytics

Gateways act as the conduit for data collection and processing before communicating the data back to a central data center. These are often servers in their own right, though they have the ability to tie into various networks. These components need to handle I/O even as they analyze what data to prioritize, what to send off-peak, and what to discard entirely. Gateway analytics should be flexible enough to accommodate a wide variety of data types, while still being cost-efficient enough to be widely deployed. HPE's Edgeline IOT Systems 10 and 20 gateways are designed for these needs.

Edge Server-based Analytics

End server analytics will often involve some persistence of data, often in a structured relational SQL database. Additionally, some may be less defined or unstructured and thus better served by NoSQL offerings or Hadoop. More memory and processor resources can be called on to perform data transformation and preparation alongside more robust analytics, including streaming, batch, and even interactive querying. A manufacturing plant may have quite advanced analytics functionality and capabilities, and even act as a regional data center, though often the local staff expertise is more oriented toward operating the environment than analyzing it. HPE Moonshot and ProLiant Servers are good examples of server analytics platforms.

Data Center-based Analytics

Data centers are where the largest and most powerful computing will come into play. A much wider variety of techniques and infrastructure can be leveraged. Large clusters of Kafka (for high-speed ingest), Spark (for streaming and querying), and Hadoop (for economical storage) may run on dedicated servers. Data warehouses could be deployed on purpose-built appliances. Analytics approaches might include machine learning, text for log files, graphing for relationships, and many more. Outcomes may range from merely descriptive through predictive to ultimately prescriptive insights.

The HPE Vertica Analytics Platform Portfolio addresses many high-performance requirements as a massively scalable analytical database to power predictive analytics models, using R or Python. Vertica query engine also runs natively on any Hadoop distribution as a fast and comprehensive SQL on Hadoop to directly query and explore larger data lakes.

Additionally, organizations want to analyze data from text and rich media to empower context-aware applications. These data types can range from simple text messages through social media to video footage. HPE IDOL maintains scalability while uncovering trends, patterns and relationships in 1,000 types of text and rich media data, without requiring explicit queries. For example, IDOL can recognize the relationship of seemingly unrelated events, connect the "dots," and issue alerts.

The following are examples of analytics that could be run using IDOL for airport security:

- Scene analysis and behavior detection, like the same car continually passing through an airport terminal.
- License plate and vehicle matching for identification and location.
- Audio recognition to detect a glass breaking noise at a parking lot and deter theft.

HPE also provides several families of applicable hardware platforms like the scale-up SuperDome, scale-out on ProLiant (especially the DL380), and purpose-built appliances such as the ConvergedSystems 300. These platforms can deliver as the underlying infrastructure, depending on the specific needs of the applications and preferences of the IT department.

Cloud-based Analytics

Public cloud services can be very convenient for a geographically distributed IoT environment, landing data from any location and scaling smoothly. ESG research shows that businesses cite cloud analytics as having advantages for security and availability, as well as time to value and elasticity.[3] HPE's Haven OnDemand provides a fast onramp to insights with cloud APIs supporting machine learning and other next-generation analytics applications. Many organizations also run Vertica in the AWS cloud using an easy-to-use AMI to quickly "stitch" together a cluster and run their most demanding analytics at linear scale and with enterprise-class performance. The scale-out MPP nature of Vertica makes it a natural fit for cloud environments, and offers real-time loading and querying, plus a broad suite of advanced analytics capabilities.

The Bigger Truth: Choosing the Right IoT and Analytics Strategy

It's no exaggeration to say that most IoT value comes from analytics of the new data. Industrial use cases can reduce costs, improve service quality, and guide better business decisions. Operational technology can be better understood through real-time and historical analysis taking place at the optimal point in the data flow across an enterprise IT architecture. Capabilities around discovery, decision management, information delivery, governance, and workflow will all derive from the analytics platforms deployed for IoT.

HPE has a broad and deep portfolio of analytics offerings for IoT, enhanced by many technology and service delivery partners, that will deliver meaningful advantages that are optimized to perform well in the most challenging environments. Further, HPE is well positioned with both IT and industry expertise on staff to help you navigate designing and implementing the right IoT analytics solutions for your business.

[3] Source: ESG Research Report, *Big Data Trends: A Midmarket Perspective*, March 2016.

www.esg-global.com contact@esg-global.com P. 508.482.0188