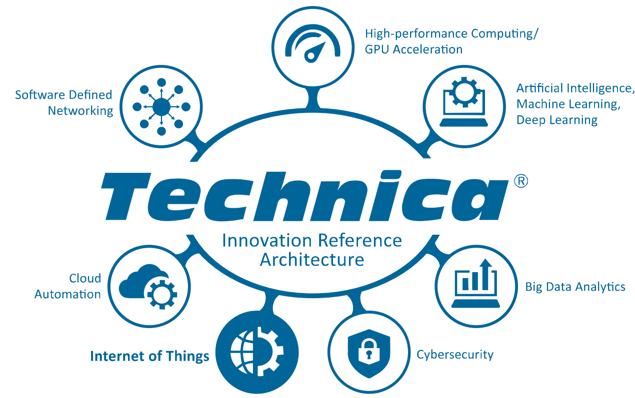
The plethora of IoT devices is already adding to the exponentially increasing volumes, variety, and velocity of Big Data. This paper examines IoT analytics and provides an analytics reference architecture that can incorporate IoT. A future paper will examine Technica's GPU database, SQuadron, within this IoT analytics ecosystem.

The Technica Innovation Platform White Paper Series presents advanced topics that will drive competitive advantage for next-generation IT over the next three-to-five years.



IoT ANALYTICS IN THE ENTERPRISE WITH FUNL

The "Internet of Things" (IoT) became one of the most hyped buzzwords in 2014. Reminiscent of the Big Data buzzword, analysts breathlessly tried to outdo each other's predictions regarding the size and scope of IoT.

It is difficult to verify Cisco's prediction of 50 billion connected devices by 2020 or Gartner's guess of $1.9 trillion in economic value add. What is clear is that a significant movement is afoot, and major players like Google, with its acquisition of Nest, are positioning themselves to ride the IoT wave.

IoT is simply interconnected devices that generate and exchange data from observations, facts, and other data, making it available to anyone. This includes devices that generate data such as a sensor; devices that combine data to deduce conclusions; and devices or services that tabulate, store, or present the data. IoT solutions are designed to make our knowledge of the world around us more timely and relevant, making it possible to get data about anything from anywhere at any time. IoT devices have one thing in common: they can be accessed via the Internet to either display data or interact with the devices directly.

ENTERPRISE IoT

IoT affords enterprises an opportunity to collect data about every physical operation of their business. This extends to public and private enterprises. Military organizations can peer through the fog of war to see soldiers and tanks in real-time.

Smart cites can route traffic and electricity effortlessly. Intelligence agencies can gain previously unimagined real-time situation awareness of targets.

Some Big Numbers:

14 bn
Connect Devices (Bosch SI)

50 bn
Connected Devices (Cisco)

309 bn
IoT Supplier Revenue (Gartner)

1.9 tn
IoT Economic Value (Gartner)

7.1 tn
IoT Solutions Revenue (IDC)

Some Small Numbers:

By 2020, component costs will have come down to the point that connectivity will become a standard feature, even for processors costing less than

$1

Data from http://postscapes.com/internet-of-things-market-size

Figure 1 – IoT Predictions

IoT USE CASES

It is nearly impossible to enumerate all of the types of connected devices, and there will be many we have not even imagined yet. However, experts predict the biggest impact in these areas:

- Sensor Networks (e.g., traffic, weather, military, etc.)
- Fleet Management
- RFID Tags (e.g., supply-chain management)
- Home Automation
- Connected Health
- Connected Vehicles
- eHealthcare Systems
- Smart Homes
- Environmental Monitoring
- Industrial Automation
- Smart Grids

This chart portrays the anticipated growth of IoT devices over the coming years:

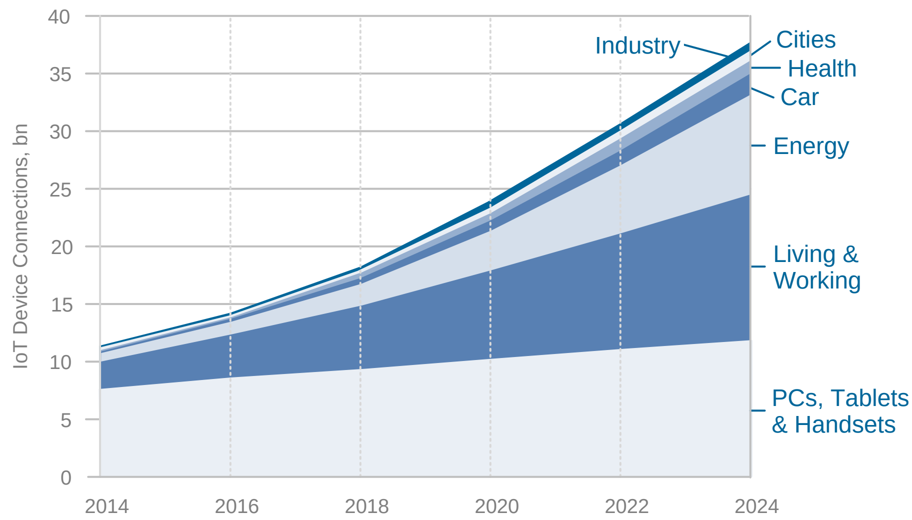


Figure 2 – IoT Forecasts by Key Domains

While PC, tablets, and smartphones are expected to rise gradually, dramatic growth is anticipated in the areas of connected living and working.

Looking at a picture of what the IoT world may look like, there will eventually be many more IoT devices than people on the planet
!

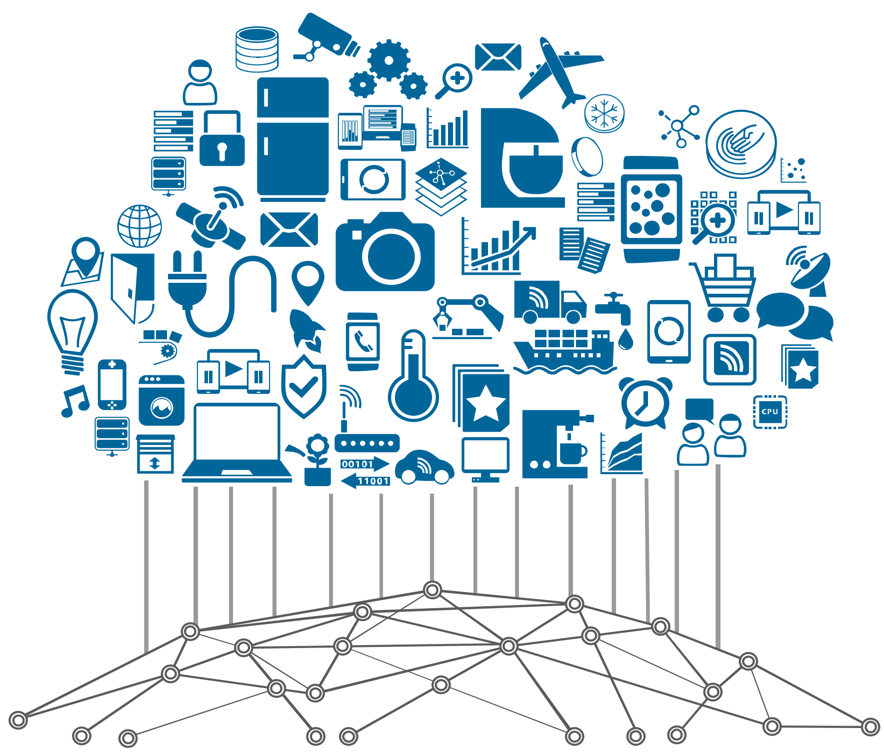


Figure 3 – Futuristic View

IoT DATA FLOW

The data generated by all of these IoT devices will be truly staggering. The data will move through three phases: Data Acquisition, Data Transport, and Data Analysis.

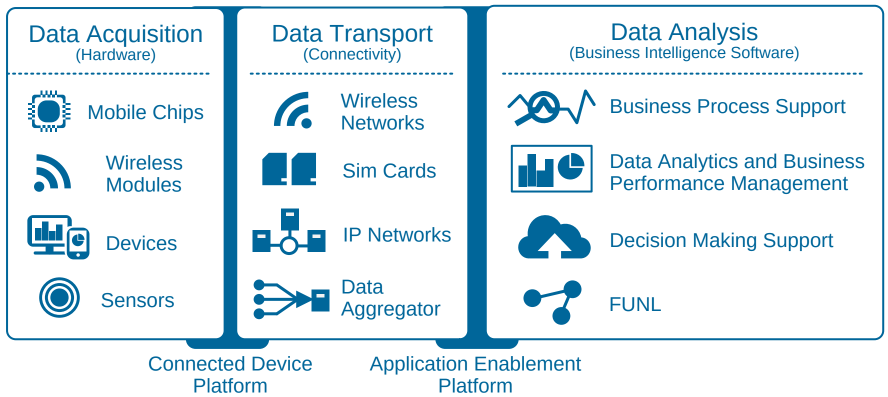


Figure 4 – IoT Data Flow

1. **Data Acquisition:** Data is collected by various sensors and other hardware.
2. **Data Transport:** IoT data moves through the cloud via Hyper-Text Transport Protocol (HTTP) or a messaging protocol leveraged by IoT devices such as Message Queue Telemetry Transport, more commonly known as MQTT. Depending upon the use case, data aggregators roll up data from many devices.

Note:	A forthcoming paper will describe how Technica's SQuadron can serve as an IoT Data Aggregator.

3. **Data Analysis:** This phase is where the information collected from sensors and devices and transported over the network is analyzed, interpreted, and acted upon or presented to humans to act upon. The data analysis phase is garnering strong attention because Big Data and real-time analytics are rapidly becoming core areas of differentiation and sources of innovation for many enterprises.

FUNL, Technica's Graphics Processing Unit (GPU)-accelerated analytic solution, operates in the data analysis phase. FUNL gives a voice to IoT's silent intelligence by allowing improved tracking and tracing capabilities, optimizing workflows, and automating systems.

PROPERTIES OF IoT DATA

Real-Time/Streaming

- Ex: Machine log data, messages, news alerts
- High velocity data requires architectural changes for capture and processing
- Analysis must be in real-time or near real-time

High-Volume

- Necessitates archiving, data management, and analytics at scale
- Key is to "productionalize" large data environments
- High-performance analytics must allow reporting near real-time results to make data useful

Semi-Structured

- Not modeled to fit well into relational databases
- Requires additional parsing to ensure data quality and context to fit into a structured environment
- Language and keyword frequency are needed to fully analyze

Real-Time/Streaming

- Ex: Images, videos, voice, binaries
- Requires image recognition or audio analysis to support
- Requires transformation and parsing

IoT ANALYTICS

The goal of analytics is to derive "meaning" from data, specifically by creating value from raw data. In the context of IoT, "noisy" data is generated by individual devices and sensors. As noisy IoT data flows over the IoT network to data aggregators, the data can be filtered. This filtered data moves through the value chain to be processed as information, knowledge, and wisdom. As the analytics improve over time, greater understanding can be derived within the enterprise.

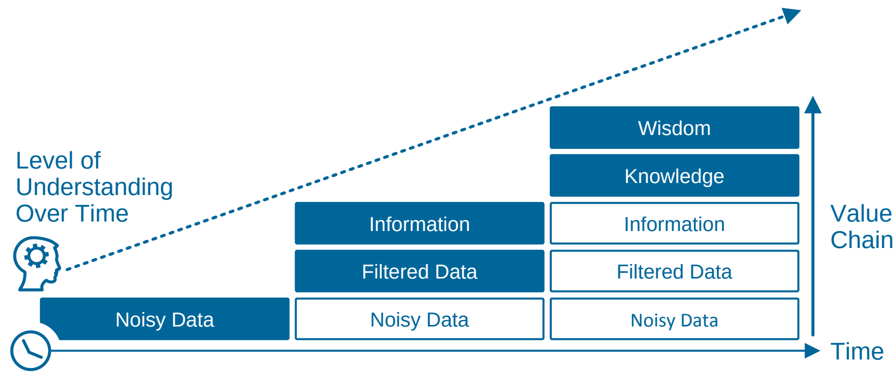


Figure 5 – Understanding Increased Over Time

IoT data has fundamentally different characteristics than traditional batch-oriented enterprise data. This sidebar list highlights some of the properties of IoT data[1]. In comparison to IoT data, traditional data is more structured, with less of a need for real-time processing. Note that characteristics vary by IoT application, i.e., voice or image recognition may not be relevant IoT data in your enterprise.

One of the key factors that differentiate IoT analytics from traditional analytics is the real-time nature of IoT. While not overlooking past investments in Big Data solutions, Enterprises must extend their traditional analytic data solutions to incorporate IoT data with a streaming analytic capability. Transactional and historical data, however, is not disappearing. Big Data that has typically been housed within data warehouses or, increasingly over the last 10 years, Hadoop clusters, will remain a major component of the analytics architecture. The objective then, is to augment traditional data warehouse and Hadoop solutions with streaming process capabilities.

IOT ANALYTICS REFERENCE ARCHITECTURE

To accommodate both IoT data and traditional data, a new analytics architecture is necessary to create value from raw data.

[1] Source: Blue Hill Research, September 2015

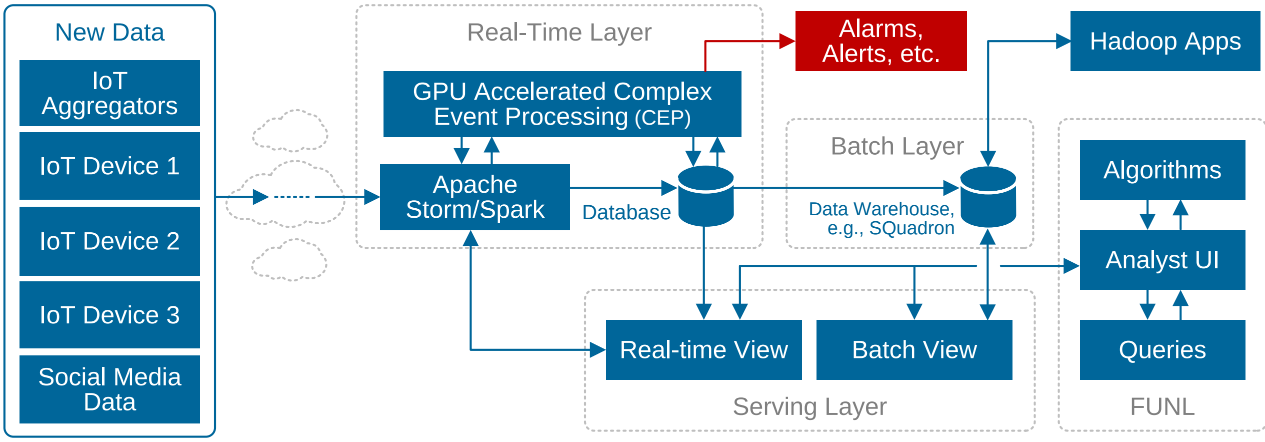


Figure 6 – IoT Analytics Reference Architecture

The components of this architecture, which incorporates streaming data include:

- **New Data:** Data produced from IoT devices, IoT data aggregators, and other sources of streaming data, such as raw social media information. The data passes through the cloud to the enterprise, typically via the aforementioned HTTP or MQTT protocol.
- **Real-Time Layer:** Integrates and processes streaming data from the cloud.
  - Depending upon the use case and skillsets that exist within the enterprise, Apache Storm, Apache Spark, or other proprietary stream processing engines.
  - An open source or proprietary Complex Event Processing (CEP) engine can be employed to trigger alarms, alerts, or other workflows. Technica is developing a GPU-accelerated CEP engine. As with other Technica GPU-accelerated solutions, including FUNL and SQuadron, the GPUs allow higher bandwidth processing at a lower hardware cost and power consumption. This allows the CEP engine to aggregate, enrich, classify, annotate, filter, join, and infer complex patterns over a stream of composite events more quickly than traditional CPU-centric solutions.
  - The streaming data can then be persisted in a database solution that can become a rich datasource.
- **Batch-Layer:** The domain of the enterprise data warehouse, or more recently, a Hadoop cluster. As shown in the architecture, the persisted streaming data can loaded into the data warehouse utilizing the same Extract, Transform, and Load (ETL) process incorporated by other enterprise transactional datasources. The ETL phase of moving sales data, human resource data, logistics data, etc. is not portrayed in the Figure, but assumed.
- **Hadoop Apps or Other Data Warehouse Apps**: Written to analyze the batch data and produce reports or power business intelligence dashboards.

- **FUNL:** Affords a rich set of machine learning, graph analytic, and deep learning algorithms, initiated from a configurable user interface (UI).
- **Serving Layer:** Allows the Analyst UI to interact with the Real-time View or Batch view, depending upon the nature of the query.

The extensible components of the IoT Analytics Reference Architecture can grow as the volume, varieity, and velocity of of data from IoT devices increases.

SUMMARY

IoT data will increase Big Data's volume, variety, and velocity at an exponential rate. Even without IoT, most enterprise Big Data was already increasing exponentially. IoT data characteristics differ from data produced inside the corporate firewall. It therefore requires an architecture that supports the real-time streaming characteristics of IoT information.

The IoT Analytics Reference Architecture allows the enterprise to ingest, process, and react to real-time, streaming IoT data in a manner that is not possible with the traditional, batch-oriented analytics architecture.

Technica's GPU solutions: FUNL, SQuadron, and the under-development CEP solution allow the massively parallel processing capabilities of GPUs to speed-up processing, at a lower costs and power consumption.

Technica provides professional services, products, and innovative technology solutions to the Federal Government. We specialize in network operations and infrastructure; cyber defense and security; government application integration; systems engineering and training; and product research, deployment planning, and support.

Technica®
22970 Indian Creek Drive, Suite 500
Dulles, VA 20166
703.662.2000