Impact Factor (SJIF): 3.632

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

(Special Issue for ITECE 2016)

Splunk : for the Internet of Things

Darshan Thoria[1], Drashti Hirani[2], Vishal Kansagara[3]

[1]*CSE, SLTIET*
[2]*M.E.C.S.E.*
[3]*CSE, SLTIET*

Abstract — The Internet of Things (IoT) has been defined in a number of different ways. Generally speaking, it refers to a global, distributed network (or networks) of physical objects that are capable of sensing or acting on their environment, and able to communicate with each other, other machines or computers. Such 'smart' objects come in a wide range of sizes and capacities, including simple objects with embedded sensors, household appliances, industrial robots, cars, trains, and wearable objects such as watches, bracelets or shirts. Their value lies in the vast quantities of data they can capture and their capacity for communication, supporting real-time control or data analysis that reveals new insights and prompts new actions. In this paper we have described Splunk software tool, which is ideal for managing and analyzing the high velocity, volume and variety of data generated by IoT.

Keywords-internet of things, industrial data, splunk enterprise, real time system, IoT with splunk analysis

I. INTRODUCTION

The Internet of Things is an emerging topic of technical, social, and economic significance. Consumer products, durable goods, cars and trucks, industrial and utility components, sensors, and other everyday objects are being combined with Internet connectivity and powerful data analytic capabilities that promise to transform the way we work, live, and play. Projections for the impact of IoT on the Internet and economy are impressive, with some anticipating as many as 100 billion connected IoT devices and a global economic impact of more than $11 trillion by 2025. At the same time, however, the Internet of Things raises significant challenges that could stand in the way of realizing its potential benefits. Attention-grabbing headlines about the hacking of Internet-connected devices, surveillance concerns, and privacy fears already have captured public attention. Technical challenges remain and new policy, legal and development challenges are emerging. [2]

The Internet of Things (IoT) is an important topic in technology industry, policy, and engineering circles and has become headline news in both the specialty press and the popular media. This technology is embodied in a wide spectrum of networked products, systems, and sensors, which take advantage of advancements in computing power, electronics miniaturization, and network interconnections to offer new capabilities not previously possible. An abundance of conferences, reports, and news articles discuss and debate the prospective impact of the “IoT revolution”—from new market opportunities and business models to concerns about security, privacy, and technical interoperability.

II. IoT AT A GLANCE

A. IoT Model

The reference model of IoT is as given below and consists of seven different levels:

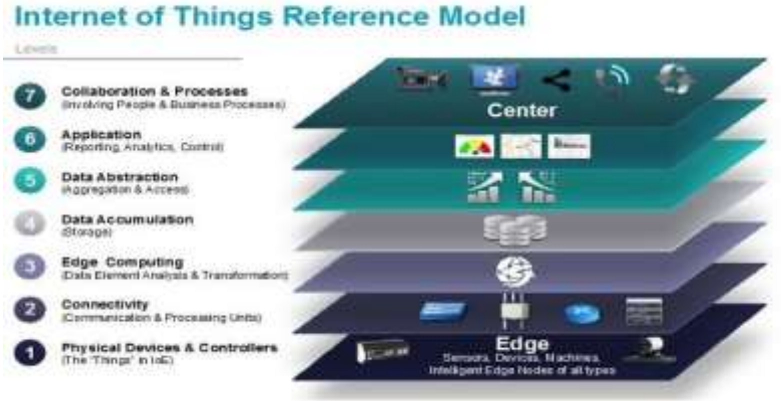

Internet of Things Reference Model
Levels
7 Collaboration & Processes
(Involving People & Business Processes)
6 Application
(Reporting, Analytics, Control)
5 Data Abstraction
(Aggregation & Access)
4 Data Accumulation
(Storage)
3 Edge Computing
(Data Element Analysis & Transformation)
2 Connectivity
(Communication & Processing Units)
1 Physical Devices & Controllers
(The 'Things' in IoE)
Center
Edge
Sensors, Devices, Machines,
Intelligent Edge Nodes of all types

Fig: 1 IoT Reference Model[10]

- The very first level is includes devices, sensors, machines etc.

- The second level includes communication and processing units.
- The third level is responsible for data element analysis and transformation.
- The fourth level does data storage.
- The fifth level is responsible for data abstraction for aggregation and access.
- The sixth level includes applications like reporting analytics and control.
- The last seventh level is responsible for collaboration of people and business processes.

B. IoT Applications

Internet of things is used for smart-X applications like smart cities, smart energy and smart grid, smart mobility and transport, smart home, smart building and infrastructure, smart factory and smart manufacturing, smart health, food and water tracking and security, smart logistic and retail etc., from which applications are given below with examples. [4]

User Element	Description	Examples
Human	Devices attached or inside the human body	Devices (wearable's and ingestible) to monitor and maintain human health and wellness; disease management, increased fitness, higher productivity
Home	Buildings where people live	Home controllers and security systems
Retail Environments	Spaces where consumers engage in commerce	Stores, banks, restaurants, arenas – anywhere consumers consider and buy; self-checkout, in-store offers, inventory optimization
Offices	Spaces where knowledge workers work	Energy management and security in office buildings; improved productivity, including for mobile employees
Factories	Standardized production environments	Places with repetitive work routines, including hospitals and farms; operating efficiencies, optimizing equipment use and inventory
Worksites	Custom production environments	Mining, oil and gas, construction; operating efficiencies, predictive maintenance, health and safety
Vehicles	Systems inside moving vehicles	Vehicles including cars, trucks, ships, aircraft, and trains; condition-based maintenance, usage-based design, pre-sales analytics
Cities	Urban environments	Public spaces and infrastructure in urban settings; adaptive traffic control, smart meters, environmental monitoring, resource management
Outside	Between urban environments (and outside other settings)	Outside uses include railroad tracks, autonomous vehicles (outside urban locations), and flight navigation; real-time routing, connected navigation, shipment tracking

Table 1 : IoT Applications

C. IoT Challenge

The IoT is a natural evolution of the world's networks. Just as people became more connected by devices and applications during the explosion of the social media revolution, devices, sensors and industrial equipment are also becoming more connected—and are consuming and generating data at an unprecedented pace.

Disparate and deployed connected devices can provide a unique touch point to real-world operations and conditions. But collection, storage and insight of the machine data generated by the Internet of Things can be a challenge. Few architectures and applications are designed to handle the constant streams of real-time events, sensor readings, user interactions and application data produced by massive numbers of connected devices.

There are two distinct segments of the IoT—the largely siloed, enterprise-managed industrial systems, consisting of legacy applications and heavy equipment (e.g., oil and gas, power distribution and manufacturing systems); and consumer-focused, cloud-connected devices like connected vending machines, wearable's and smart home appliances. Whether in industrial systems or the consumer-focused IoT, the constant processes, applications, interactions, sensor readings, geolocation information and streams of data can be easily managed and analyzed using Splunk software.

III. SPLUNK AT A GLANCE

A. Intoduction to Splunk

Splunk software is a scalable and versatile platform for machine data generated by all of the devices, networks, applications and end users connected by today's networks. Use Splunk to collect, index and harness the power of the machine data generated by connected devices and machines deployed on your local network or around the world.[8]

Splunk helps to make sense of machine data. Splunk handles both the form and the semantics of machine data. It accomplishes this through a unique approach of universally indexing any machine data from across any element of the infrastructure. Splunk does this without requiring costly connectors or agents or filtering or parsing the data to load it into a database. By providing users an index of all the machine data generated by all systems and infrastructure, Splunk enables users to ask any question and find answers quickly to the most simple or strategic propositions and develop role based views and dashboards for real-time visibility and analytics for operational intelligence.

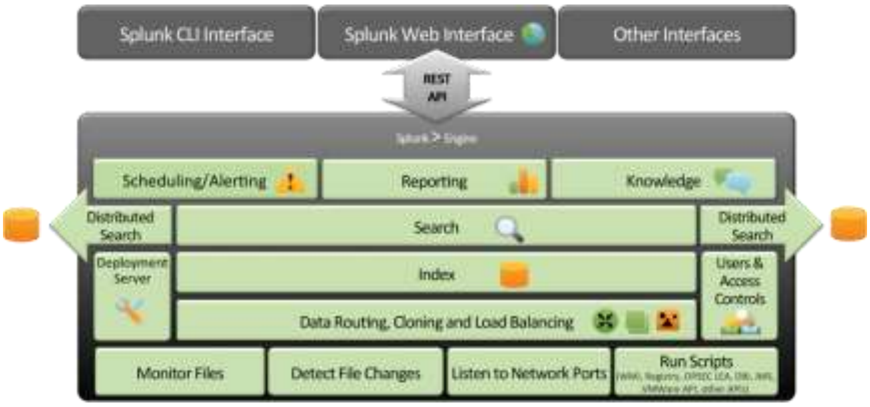


Fig: 2 Splunk Architecture

The Splunk architecture has mainly four components as given below:[12]

1. **Forwarder :** The Forwarder agent is installed wherever data needs to be collected directly from endpoints in real-time. It can monitor local application logfiles, capture the output of status commands on a schedule, grab performance metrics from virtual or non-virtual sources or watch the file system for configuration, permissions and attribute changes. Forwarders come in two flavors.
   - The **Universal Forwarder** is a lightweight agent, which just can forward data to an indexer or another forwarder (intermediate forwarding)
   - The **Heavy Forwarder** has a smaller footprint than a Splunk indexer but retains most of the capability, except that it lacks the ability to perform distributed searches. Unlike the universal forwarder, a heavy forwarder parses data before forwarding it and can route data based on criteria such as source or type of event. It can also index data locally while forwarding the data to another Splunk indexer.

2. **Indexers:** Splunk indexers, or index servers, provide indexing capability for local and remote data and host the primary Splunk data store, as well as the Splunk Web interface. The main duty of these systems is transforming raw data into events and placing the results into a search index. The transformation process creates a constant high I/O data stream, which is written to a disk and causes high CPU utilization. When an indexer processes the data, it creates two main types of files: the rawdata file containing the original data in compressed form and the index files that point to this data. Typically, the compressed rawdata file is approximately 10% the size of the incoming, pre-indexed raw data. The associated index files range in size from approximately 10% to 110% of the rawdata file. This value is affected strongly by the number of unique terms in the data.

3. **Search Head**: A search head is a Splunk instance configured to distribute searches to indexers, or search peers. Search heads can be either dedicated or not, depending on whether they also perform indexing (if collocated with an indexer). Dedicated search heads don't have any indexes of their own (other than the usual internal indexes). Instead, they consolidate results originating from remote search peers. Search heads are both CPU and Memory intensive.

4. **Deployment Server**: A deployment server distributes configuration information to running instances of Splunk via a push mechanism, which is enabled through configuration. A key use case for the deployment server is to manage configuration for groups of forwarders.

B. Data Movement in Splunk[12]

Data in Splunk Enterprise transitions through several phases, as it moves along the data pipeline from its origin in sources such as log files and network feeds to its transformation into searchable events that encapsulate valuable knowledge. The **data pipeline** includes these following segments:

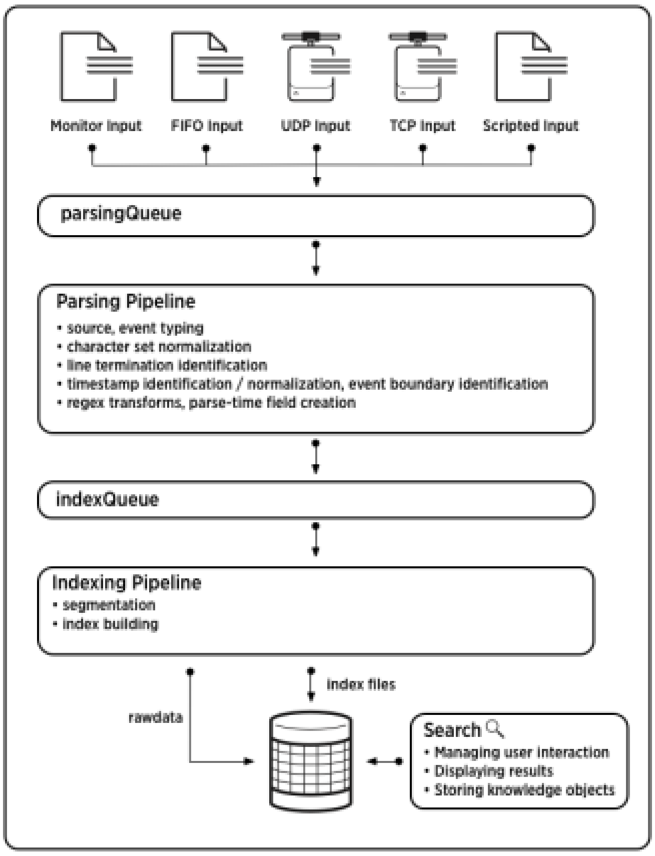

Monitor Input
FIFO Input
UDP Input
TCP Input
Scripted Input
parsingQueue
Parsing Pipeline
• source, event typing
• character set normalization
• line termination identification
• timestamp identification / normalization, event boundary identification
• regex transforms, parse-time field creation
indexQueue
Indexing Pipeline
• segmentation
• index building
rawdata
index files
Search
• Managing user interaction
• Displaying results
• Storing knowledge objects

Fig: 3 how data move through splunk

1. **Input:** In the input segment, Splunk Enterprise consumes data. It acquires the raw data stream from its source, breaks it into 64K blocks, and annotates each block with some metadata keys. The keys apply to the entire input source overall. They include the host, source, and source type of the data. The keys can also include values that are used internally by Splunk Enterprise, such as the character encoding of the data stream, and values that control later processing of the data, such as the index into which the events should be stored.During this phase, Splunk Enterprise does not look at the contents of the data stream, so the keys apply to the entire source, not to individual events. In fact, at this point, Splunk Enterprise has no notion of individual events at all, only of a stream of data with certain global properties.

2. **Parsing:** During the parsing segment, Splunk Enterprise examines, analyzes, and transforms the data. This is also known asevent processing. During this phase, Splunk Enterprise breaks the data stream into individual events. The parsing phase has many sub-phases:
   - Breaking the stream of data into individual lines.
   - Identifying, parsing, and setting timestamps.
   - Annotating individual events with metadata copied from the source-wide keys.
   - Transforming event data and metadata according to Splunk Enterprise regex transform rules.

3. **Indexing**: During indexing, Splunk Enterprise takes the parsed events and writes them to the index on disk. It writes both compressed raw data and the corresponding index files.For brevity, parsing and indexing are often referred together as the indexing process. At a high level, that's fine. But when you need to look more closely at the actual processing of data, it can be important to consider the two segments individually.

4. **Search:** Splunk Enterprise's search function manages all aspects of how the user sees and uses the indexed data, including interactive and scheduled searches, reports and charts, dashboards, and alerts. As part of its search function, Splunk Enterprise stores user-created knowledge objects, such as saved searches, event types, views, and field extractions.

C. Splunk for IoT[7][8]

Disparate and deployed connected devices can provide the enterprise a unique touch point to real-world operations and conditions. But collection, storage and insight of the machine data generated by the IoT can be a challenge. Splunk software ingests, analyzes and visualizes real-time and historical machine data from any source—including industrial control systems and connected devices—enabling you to improve operations, ensure safety and compliance, perform preventative maintenance and better manage the lifecycle of assets. Use Splunk to collect, index and harness the power of the machine data generated by connected devices and machines deployed on your local network or around the world.

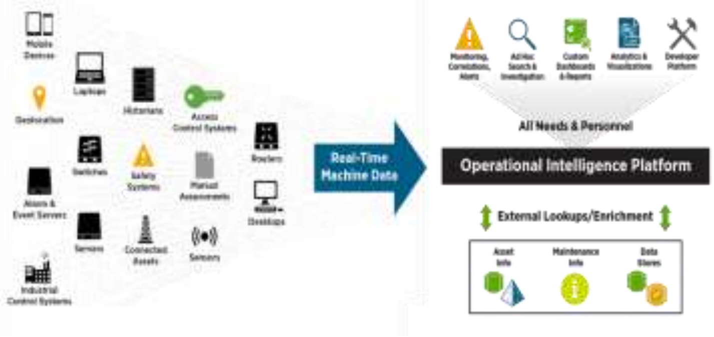


Fig: 4 Internet of Things kepware diagram by Splunk

1. Monitoring and Diagnostics

Ensure that equipment in the field operates as intended. Monitor and track unplanned device or system downtime. Understand the cause of failure on a device to improve efficiency and availability. Identify outliers and issues in device production or deployment.

2. Security, Safety and Compliance

Help protect mission-critical assets and industrial systems against cyber security threats. Gain visibility into system performance or set points that could put machines or people at risk, and satisfy compliance reporting requirements.

3. Preventative Maintenance and Asset Lifecycle Management

Gain real-time insight into asset deployment, utilization and resource consumption. Recognize patterns and trends, and use operational data to proactively approach long-term asset management, maintenance and performance.

As businesses build and deploy connected devices, they are also deploying a new generation of commercial IoT platforms and services. These platforms and services enable device connectivity, visibility and simple provisioning and remote device management; they act as both a gateway to device operations and provide a platform for interaction with remote device operations and performance.

Splunk software enables powerful machine data analytics for the Internet of Things, and eliminates the need to build them from the ground up. Leading IoT platforms including Xively by LogMeIn, Citrix Octoblu, and AWS IoT are already integrated with Splunk software, enabling fast time to value for developers and end users.

Splunk Integrates with Leading IoT Platforms and Services As businesses build and deploy connected devices, they are also deploying a new generation of commercial IoT platforms and services. These platforms and services enable device connectivity, visibility and simple provisioning and remote device management; they act as both a gateway to device operations and provide a platform for interaction with remote device operations and performance.

Splunk software enables powerful machine data analytics for the Internet of Things, and eliminates the need to build them from the ground up. IoT platform providers are already integrating Splunk technology with their platforms and making Spunk’s platform for operational intelligence available to their own customers.

Use Cases for the Consumer & Emerging Industrial IoT
Device data can be used to gain unique business insight. Customers are already using Splunk software to analyze data from devices already in users’ hands. Real-time access to data from connected products provides unprecedented intelligence into consumer behavior and allows businesses to provide proactive services to customers.

As the number of connected vending machines and product dispensers grows, businesses can adopt digital intelligence strategies to better understand user preferences and activities. Splunk customers like Coca-Cola are on the leading edge of understanding how real-time insight into these processes adds value to a business analytics strategy. Additionally, as devices are connected, manufacturers and service providers can harvest valuable information on field performance and user interaction. They can proactively service and update the products before the customer is even aware of any issues.

D. Example of IoT with Splunk[9]

Monitor your own Smart Home – three top tips from Splunk

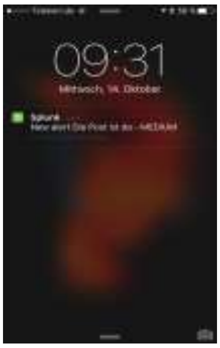
One of the great things at Splunk is that there are so many devices you can collect data from and make something meaningful out of.

1. **Splunk the postbox:** How does it work? built an infrared photo sensor into the postbox that gets activated if someone drops in a letter. This sensor sends a signal to HomeMatic server and from there we can collect the data via RestAPI easily into Splunk for reporting. However it's not just reporting on what days and times we gets post. Thanks to the Splunk Mobile App we can access those dashboards on our iPad and iPhone when we would travel.

Postbox

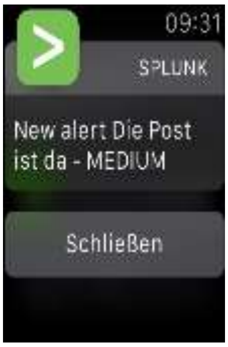
infrared sensor built into the postbox

2. **Apple Push Notifications:**
Set up an alert if a letter drops into the postbox, and the Splunk Alert is pushed directly via the Splunk App to his screen. That push notification is also displayed on our iWatch.



Splunk
New alert Die Post ist da - MEDIUM

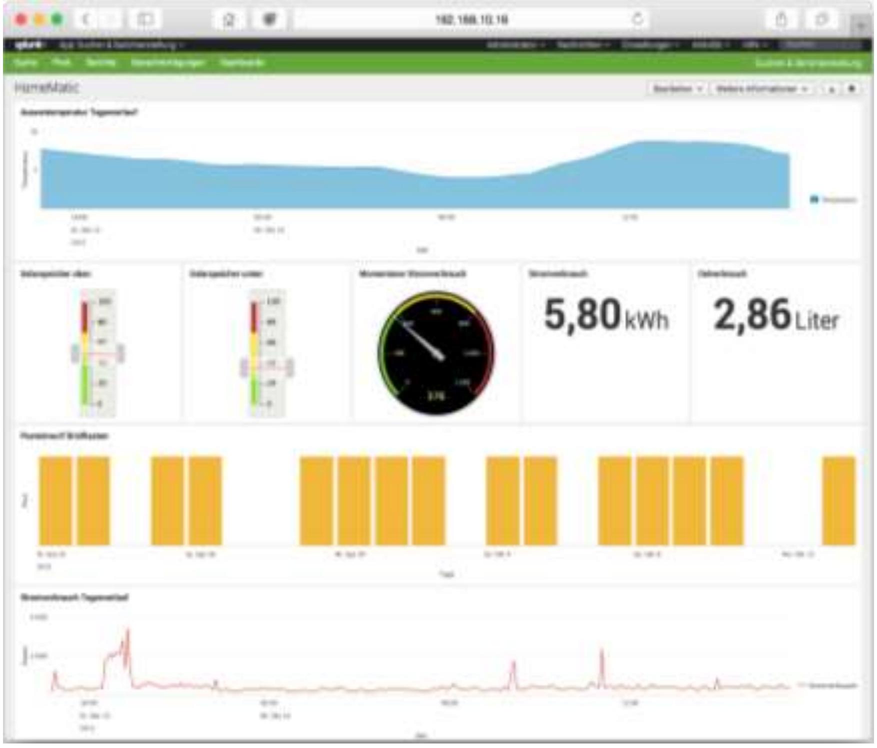
Splunk Push Alert on iOS



New alert Die Post ist da - MEDIUM

Splunk Alert on a iWatch

3. **Your Smart Home on one dashboard:**
We can also monitor our home heating system, warm water solar system and measuring the energy consumption of our house.



SmartHome Dashboard

IV. CONCLUSION

The IoT has created unique opportunities for insight into real-world operations and customer behavior—all through the analysis of machine data. Connected industrial and consumer-focused devices are being connected to the world's networks, creating an IoT. The data from these connected devices can provide valuable insight into processes, security and business operations. Splunk Gains real-time insight into operations with data from the IoT, Manage and analyze the high velocity, volume and variety of data generated by the IoT and Improve monitoring and diagnostics, safety and compliance, preventative maintenance, and asset lifecycle management.

REFERENCES

[1] Ron Davies, "The Internet of Things Opportunity and Challenges", a publication in European Parliamentary Research Service, May 2015.

[2] Karen Rose, Scott Eldridge, Lyman Chapin, "The Internet of Things: An Overview", Understanding the Issues and Challenges of a more Connected World, ©2015 The Internet Society (ISOC).

[3] Cooper, J., James, A., "Challenges for database management in the internet of things." IETE Tech Rev, Vol. 26, No. 5, pp. 320-329, 2009.

[4] Ovidiu Vermation, Peter Friess , "Internet of Things : From Research and Innovation to Market Deployment", ©2014, River Publisher.

[5] "EMC Infrastructure for Splunk", handout by EMC^2, 2015.

[6] Moreno, M., et al. "A holistic IoT-based management platform for smart environments." Communications (ICC), 2014 IEEE International Conference on. IEEE, 2014.

[7] "Splunk User guide", by Splunk Inc., 2014.

[8] "Splunk Solution Guide", by www.splunk.com

[9] Accessed "http://blogs.splunk.com/2015/11/17"

[10] "Managing a Desktop Virtualization Architecture with Splunk", white paper, by www.citrix.com

[11] Accessed "https://www.linkedin.com/pulse/internet-things-objectives-scientific-challenge".

[12] Accessed "http://www.learnsplunk.com/, splunk tutorial".