INTERNATIONAL JOURNAL OF ADVANCED STUDIES IN COMPUTER SCIENCE AND ENGINEERING
IJASCSE VOLUME 6 ISSUE 9, 2017
09/30/2017

Farm Data Analytics Using Splunk

Hiral Raveshiya
Information Technology Department,
Universal College of Engineering,
Mumbai, Maharashtra, India

***Abstract*— Everything is getting connected to the Internet in current era. Internet of Things (IoT) is the hottest trend giving rise to development of lots of applications leading to smart livings and smart cities. This paper presents use of modern technologies and trends in the field of farming leading to smart farming. This paper represents use of splunk for data analysis in farming. Manual methods of data collection are replaced by sensors and gathered data is fed to the splunk software for some analysis and prediction. Prediction related to farming and application development for farmers is currently under development. The aim of this paper is to give overview of use of splunk and machine lerning to the readers and students in particular who are interested in research and contribution towards smart farming.**

Keywords—data analytics ; Splunk; MQTT ; CoAP

I. INTRODUCTION

Agriculture and farming forms the vital role in economy of any country. According to BI Intelligence estimates 2015, the demand of food will increase around 70% by 2050 than current demand and consumption of food. To meet this growing need the traditional methods of farming need to be updated. With the advancement in network technologies, cloud computing and data analytics techniques, the focus are to adapt new methods in farming leading to smart farming. Smart farming is the traditional farming updated with new methods using Information and Communication Technologies. This paper introduces the concept of applying big data analytics techniques to farming. Smart farming is currently the topic of focus by researchers and also by government but little literature is available in this field. This paper illustrates use of splunk for farming data analysis. The work is done using references available on internet and some white papers as less standard papers are available in this domain. The project is developed halfway and currently under progress but the intension of this paper is to throw some light on the possibility of development in smart farming field. The paper covers layered view and block diagram of the system, farming data collection, some sample analytics using splunk , machine learning toolkit in splunk followed by conclusion and future work.

II. LAYERED OVERVIEW OF THE SYSTEM ARCHITECTURE:

This section describes farm analytics system in layers terminology. Fig. 1 shows the main layers of the system. We will discuss all the layers one by one to have shallow Understanding of the system. Technical specifications and other details are briefed in later sections.

Layer 6: UI -Alerts and suggestions to farmers
Layer 5: Data Analytics - Splunk
Layer 4: Sensors to database –Opensensordata.net
Layer 3: Communication protoco –MQTT , CoAP
Layer 2: Wireless transport: wifi , Bluetooth
Layer 1: Data collection: Sensors and other IoT devices

Fig. 1. Layered view of Farm Analytics System.

Layer 1: Data collection in smart farming: the manual methods of the data collection in farming are outdated. Sensors offer solution for controlling and monitoring crop performance. Sensor networks use different types of sensors such as LIDAR sensors, image based sensors, specific soil sensors and such other sensors for different variables sensing like light, temperature, humidity and climatological and anthropological events. Some example systems using combinations of multiple sensors are there such as Unmanned Aerial Vehicles (UAV) uses ground based sensors, in addition to regular sensors, for crop monitoring. Some other trends in the same field which uses collection of sensors for measuring farming data are Agricultural robotics systems and autonomous or remotely guided agricultural tractors [1]. In addition to the sensors modern farms also uses data from cameras, past existing databases, human inputs collected through mobile apps, GPS and such other sources. Modern farms generate data in volumes of TB that is combination of time-series data, spatial data, images and many more semi-structured and unstructured data.

The continuous data generated from IoT devices (sensors, events, mobile devices) forms machine data. This machine data is the most valuable resource to gain insights of activities around the farms. The processing of machine data can generate some alerts and alarms around the farms as well can predict many important things regarding crop health, crop maturity and many more. Machine data can be considered as one of the fastest emergent area of big data.

Layer 2 – Wireles Transport: WiFi and Bluetooth Low Energy (BTLE) are used as transport layer protocols. This paper does not discuss these protocols in detail as sufficient material is available on the Internet.

Layer3 – Communication Protocols - MQTT and CoAP :

"Add everything to the Internet "is the hottest trend in current scenario. IoT applications are getting more and more popular , giving rise to the concepts of smart cities and comfortable living. Millions of new low-cost and low-power devices are joining the club of IoT world every day. These IoT devices need to communicate with each other, service providers and cloud. These devices have some characteristics that make them different from our traditional communication devices like computers, routers and such which are rich in memory, power and connectivity. IoT devices are generally small in size, remotely deployed, limited in computing resources and battery powered. In all IoT applications these things or IoT devices need to communicate using Internet to recognize as a node. Internet protocol suite for our traditional applications is not suitable for IoT devices because of their characteristics. To deal with this light weight devices, we need light weight protocols. There are two popular IoT data connectivity protocols:1) Message Queuing Telemetry Transport (MQTT) and Constrained Application Protocol (CoAP). Table 1 compares MQTT and CoAP.

Both MQTT and CoAP are gaining popularity in IoT market. Both have their own strengths and issues. It is upto the developer to choose one which best suits to the need of application. We have used MQTT as it is easy to get MQTT network running up because of maturity of the standard.

TABLE 1. MQTT AND CoAP

	MQTT	CoAP
Strengths	Better scalability and power efficient because of *Publish/subscribe model*	Supports small transmit cycles , small packets and less overhead because of native support of UDP
	Space ,time and synchronization decoupling	built on top of IPv6, which enables multicast addressing for devices
	Uses SSL/TLS over TCP for security	DTLS over UDP
	Provides guaranteed message delivery with three levels of quality of service	Uses RESTful model to support integrated resource discovery
Issues	Needs Always on connections For TCP Communications	Lacks reliability
	Central broker may lead to failure of entire system	Less mature standard than MQTT

Layer 4- data from sensors to database: OpenSensorData.net is used which is free hosting service which uploads data in streaming form and allows downloading data under valid licence.

Layer 5: Splunk in smart farming: "Splunk" is the product by an American multinational corporation named Splunk. splunk makes machine data meaningful by identifying patterns from it in real time scenario. Splunk can capture, index and correlate data in a searchable repository. Data from searchable repository can be represented in multiple forms such as graphs, reports, alerts, dashboards, and visualizations[4]. Splunk provides operational intelligence as it provides real time analytics of data across IT infrastructures.

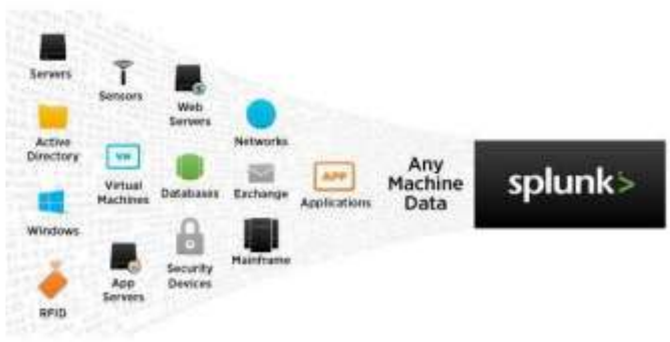

Fig. 2. Splunk and machine data [19]

Splunk provides following features:

- Collect and index anything in real-time: splunk can collect and index data from different IoT devices in smart farming.
- Flexible data input: splunk supports almost all types of data because of its variety of standards and custom methods. Data from IoT devices can be directly ingested using event collector API.
- Flexible schema support: splunk has no predefined schema. Any interpretation of data is done at search time.
- Time-based event chronology: data that has no time stamp can be inferred from the context
- Search and investigate across all logs: splunk provides splunk search processing language (SPL) , real-time search , time-range search , transaction search and interactive results.
- Correlate and analyse across all systems : easily find relationships between events and activities. Supports machine learning toolkit and algorithms for analyses.
- Visualize and report : supports wide range of charts , visualization tools and reporting tools in real-time.

- Monitor and alert: continous monitoring of events and KPIs is possible through dashboard. Custom alerts can be generated for some unusual events.
- Security and administration: splunk support secure
- data access and transport, granular access and audit controls, user authentication and data integrity.
- Scalability: support horizontal and vertical scalability.
- Developer resources: developers can create apps to solve specific problems that can be integrated with splunk.

III. DESIGN AND SAMPLE RESULTS

Implementation of this project is done on small scale currently. Once all the functionalities are implemented the project will be expanded for large scale. The main steps of the project are as follows:

- **Acquire** farming data using sensors and human input through mobile applications
- **Push** the collected data into server
- **Analyse** the collected data using splunk
- **Develop** mobile app for farmers for crop related predictions and alerts

Currently the project is in third phase that is analysis of collected data. Fig. 3. Show block diagram of the system.

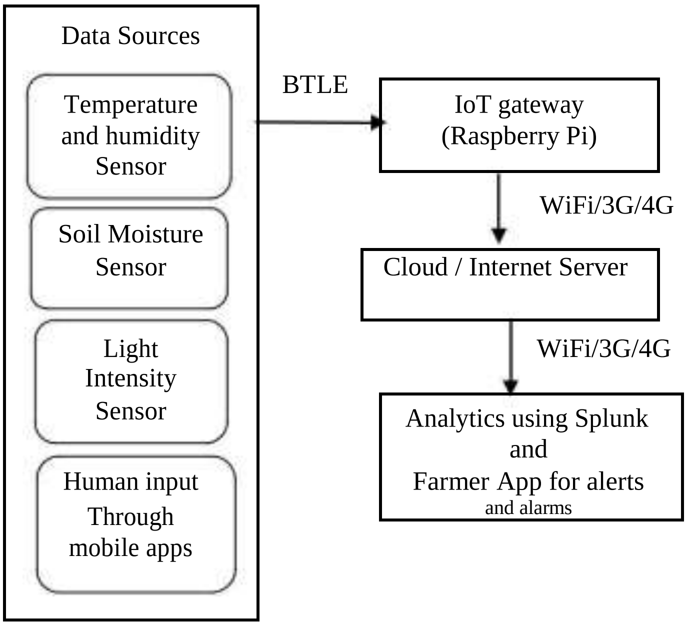

Fig. 3. Block diagram of farm analytics sytem

The main components of the system are explained below:

A. Data Sources

Temperature and humidity, light and soil moisture sensors are used. Human input which consists of their past experinces is also considered through mobile phone applications. Table 2 shows details of the sensors used.

TABLE 2. Sensors Details

Sensor	Name	Manufacturer
Humidity and Temperature	HDC 1000	Texas Instruments
Light	OPT3001	Texas Instruments
Soil moisture	HDC 1080	Texas Instruments

B. Raspberry Pi Technical Specification [18]:

TABLE3. Raspberry Technical Specification

Component	**Specification/Purpose**
Processor	Broadcom BCM2837 Arm7 Quad Core Processor
RAM	*1 GB*
CSI Camera Port	Connecting camera to Raspberry Pi
DSI display *Port*	Connecting touch screen display to Raspberry Pi
Micro SD port	String Data and Loading OS
GPIO	40 pin extended
Power Source	Micro USB
HDMI	Full size HDMI

C. Raspberry Pi and mobile phone interface

This interface is created with the help of rasbian OS and debian packages - ppp , usb-modeswitch and usb-modeswitch-data.

D. Raspberry Pi and sensor interface

Raspberry Pi is interfaced with sensors using BTLE (Bluetooth Low Energy) 4.0 adapter.

IV. DATA ANALYTICS IN SPLUNK

After gathering data from different sensors the aim is to get analysis from combined data. Free hosting service is used to upload stream data from sensors and to download the same data in CSV file format. DataStream are defined and data is uploaded to opensensordata.net using java scripts. Uploaded data is later downloaded in .CSV (crop_data.csv) file with the help of datastream id.

The "crop_data.csv" contains data from all the sensors and mobile phone app used in this experiment. Following output shows data in splunk software.

Fig. 4. Crop_data.csv in splunk software

Output of the the query "Source="crop_data.csv" | max (DailyAverageRelativeHumidity) " which finds maximum from collected dailyaverage humidity is shown in Fig. 5.

Fig. 5. Maximum from daily average humidity from Crop_data.csv

Output of the maximu temperature day wise is shown in chart form in Fig. 6. These are sample output shown. Multiple variables from light , temperature and huidity and human input can be combined to keep watch of many interesting factors.

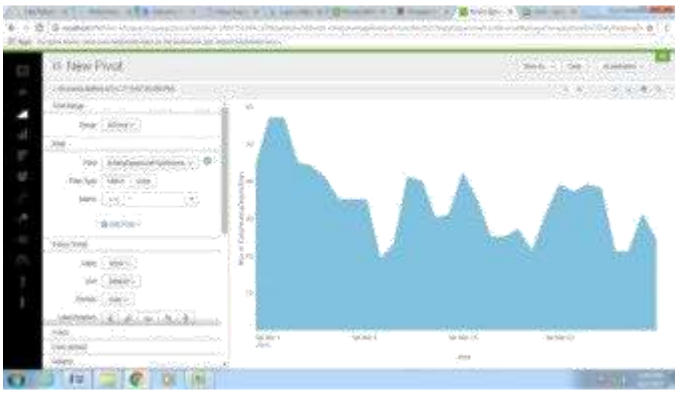

Fig. 6. Maximum daily temperature from Crop_data.csv

V.MACHINE LEARNING IN SMART FARMING

Current work focuses on crop disease detection using Machine Learning (ML). splunk has machine learning toolkit to support ML algorithms. Here brief introduction of use of ML in farming is provided for the interested readers who want to work in this field. ML is one of the major areas under artificial intelligence which is set of techniques and algorithms that makes machine capable of taking decisions. Machine uses data from past and then uses this data to take decisions in future. ML is multidisciplinary field that supports three types of learning: supervised, unsupervised and reinforcement learning.

ML algorithms can be used in various field of farming. Crop selection and crop yield prediction can be done using classification and artificial neural network algorithms. Weather forecasting can be done using support vector machines. Smart irrigation systems can be implemented using general machine learning algorithms. Crop disease prediction can be perfomed using ANN, support vector machines and pattern recognition [6][7][8]

TABLE4 ML algorithms in splunk

Category	Usefulness	Example algorithms
Feature extraction	Feature extraction algorithms transform fields For better prediction accuracy.	FieldSelector , PCA , kernelPCA , TFIDF
Preprocessing	Preprocessing algorithms are used For preparing data and P help With prediction accuracy	StandardScaler
Cluster	Partition	KMeans ,

Numeric	events with multiple numeric fields into clusters.	DBSCAN , BIRCH, SpectralClustering
Anomaly Detection	Find events that contain unusual combinations of values.	OneClassSVM
Forecasting	Forecast future values given past values of a metric (numeric time series)	ARIMA
Predict Numeric	Predict the value of a numeric field using the values of other fields in that event	LinearRegression, Lasso , Ridge ,elasticnet , KernelRidge , SGDRegressor , DecisionTreeRegressor , RandomForestRegressor
Predict Categorical	Predict the value of a categorical field using the values of other fields in that event	LogisticRegression, SVM, BernoulliNB, GaussianNB, SGDClassifier, DecisionTreeClassifier, RandomForestClassifier

VI. FUTURE SCOPE

The work represented in this paper can be extended to detect crop yield, crop diseases and crop maturity that can help farmers to take better decisions. An easy to use mobile application can be developed for farmers which can generate alerts related to extreme conditions of rain and weather.

VII. CONCLUSION

Farm analytics system using Splunk is implemented on small scale. The work of data acquition using different data sources and data analytics using splunk has given succesfull outputs and results. Once developed fully the next step is to implement the same on large scale.

REFERENCES

[1] Advances in Sensors Applied to Agriculture and Forestry , Gonzalo Pajares , Department of Software Engineering and Artificial Intelligence, Facultad de Informática, Universidad Complutense, 28040 Madrid, Spain;

[2] http://www.oodlestechnologies.com/blogs/Is-MongoDB-a-good-choice-for-your-Application accessed on 10/4/2017

[3] http://uk.businessinsider.com/ accessed on 12/4/2017

[4] https://en.wikipedia.org/wiki/Splunk accessed on 12/4/2017

[5] https://www.greymatterindia.com/importance-of-nosql-over-relational-databases accessed on 20/4/2017

[6] http://www.ijarcce.com/upload/2016/april-16/IJARCCE%2087.pdf

[7] http://www.clearag.com/machine-learning-in-agriculture accessed on 22/4/2017

[8] https://www.linkedin.com/pulse/machine-learning acceseed on 22/4/2017

[9] indian-agriculture-shubhra-saxena accessed o 25/4/2017

[10] https://docs.splunk.com/images/e/ee/MLTKCheatSheet.pdf

[11] https://data.gov.ie/dataset/area-yield-and-production-of-crops/resource/f0d8968d-6fff-4b91-af41-8b75aca5b240 accessed on 25/4/2017

[12] http://opensensordata.net/ accessed on 25/4/2017

[13] http://tekslate.com/tutorials/splunk accessed on 25/4/2017

[14] http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html accessed on 26/4/2017

[15] http://www.hivemq.com/mqtt-essentials-wrap-up/ accessed on 26/4/2017

[16] https://tools.ietf.org/html/rfc7252 accessed on 26/4/2017

[17] http://coap.technology/ accessed on 26/4/2017

[17]http://www.prismtech.com/sites/default/files/documents/MessagingComparsionNov2013USROW_vfinal.pdf

[18] https://www.element14.com/community/docs/DOC-73827/l/raspberry-pi-2-model-b-1gb-technical-specifications accessed on 27/4/2017

[19] https://www.splunk.com/ accessed on 27/4/2017