ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 7, Issue 10, October 2021

Automation of migrating data from USB storages to Cloud

Gaurav Roy, Shreya Bharath Jain

Graduate, Information Science, Dayananda Sagar College of Engineering, Bangalore, India
Graduate, Information Science, Dayananda Sagar College of Engineering, Bangalore, India

***Abstract*— The emergence to store data in the cloud has been receiving a lot of attention recently but yet we see that people who deal with a large amount of data. This paper proposes the development of an automated application/pipeline running on Raspberry pi which when connected to any USB storage would gather the data from the USB storage and store it in the cloud. The idea primarily focuses on the concepts of cloud migrations, providing solutions in migrating the current or previously available data residing locally to the cloud. Thereby providing the user with better accessibility and security of the data.**

***Index Terms*—Internet of Things, AWS Cloud, MQTT, Raspberry Pi.**

I. INTRODUCTION

The invention of computers marked the beginning of a new era in the human lifestyle. Introduced by IBM in 1956, the IBM 305 RAMAC was the first hard disk developed for data storage in general-purpose computers for storing and retrieving digital information. Hard disks have now evolved into a2TB model that combines 7,200rpm platters and high-density data, thereby making the computers outfitted with this drivable to read data extraordinarily fast.
In 2011, the tech giant Sony bespoke what could be the end of the venerable disks. The electronics giant announced that it would stop selling the 30-year-old storage media in Japan. The following year, Sony stopped selling the disks in most international markets due to dwindling demand and competition from other storage formats. The slow death, however "diskette" began in 1998 when Apple decided to not include a drive-in its G3 iMac computer.

With the advent of cloud storage, one wouldn't need to carry around a physical storage device or use the same computer to save and retrieve his information. A cloud storage sys- tem needs just one data server connected to the Internet. A client (e.g., a computer user subscribing to a cloud storage service) sends copies of files over the Internet to the data server, which then records the information. When the client wishes to retrieve the information, he or she accesses the data server through a Web-based interface. The server then either sends the files back to the client or allows the client to access and manipulate the files on the server itself. By using cloud storage service, customers need not invest in storage devices, even technical support is not required for maintenance, storage, backup, disaster recovery.

The Internet of Things (IoT) is a large distributed network consisting of highly dynamic devices. Small low powered "smart" devices can connect and communicate with one-other with sensors that record real-world data. This data can then be transmitted to other devices allowing them to trigger actions. Now a day, every industry is trying their best to de- ploy IoT-based services to connect everything starting from traditional machine to an industrial machine, this will provide data analysis real-time production control, better management for process, production, machine fault detection, and personal affectation. By using the AWS IOT platform, it's very easy to interface different gateways such as Raspberry Pi, Beagle boon black, Arduino boards and other single board computers.

The Raspberry Pi is a credit card-sized computer developed by the Raspberry Pi Foundation. The RPi's ability to act as a GNU/Linux server and the interfacing services provided by its general-purpose I/O pins make it a popular choice of hard- ware for IoT applications.

II. MOTIVATION

A typical hard drive consists of an axis on which one or more flat discs generally known as plates are mounted. The data is recorded on concentric circles, which are known as tracks. External hard disk drives usually connect through USB. Furthermore, in recent times external hard drives with 6TB of storage space are fairly smooth to discover in the market. Al- though hard disks have an excessive storage capacity, it has its own unique disadvantages. Hard drives have slower data transfer rates and are more susceptible to mechanical disasters due to the fact that they include moving parts. Further- more, an average Hard disk

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 7, Issue 10, October 2021

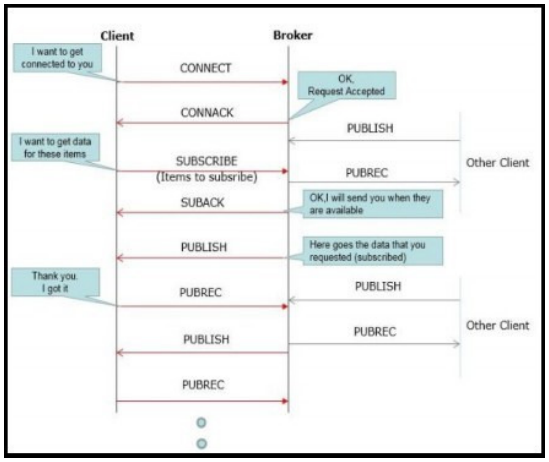
storing 1TB of records expenses everywhere among INR 3000-10,000.

Cloud storage lets their clients think about computing as limitless, having lower expenses, trustworthy, a cheap solution for backup and replication. There are many advantages to storing records within side the cloud over local storage. Companies only pay for the storage they use; the records are quickly accessible and reliable, there's higher safety in case of a disaster, cloud vendors provide hardware redundancy and automatic storage failover, absolutely endless storage capacities, there are workload balance and a unified view of the storage.

III. RELATED STUDIES

A. *MQTT*

MQTT is a standardized publish/subscribe Push protocol that was launched by IBM in 1999. MQTT was planned to send a piece of information accurately under the long network delay and low bandwidth network condition. For communication, it exchanges a range of control packets in a specific manner. There are fourteen control packets.



Basic Concepts of MQTT:

1. Publish/subscribe:

In the MQTT protocol, publishers pub- lishing messages, and customers subscribing to topics that are usually considered as a Publish/Subscribe model. The sub- scriber subscribes to particular topics which might be related to them and through that, they receive every message that is published to those topics. On the other hand, customers can publish messages to topics, in such a way that permits all subscribers to access messages of those topics.

2. Topics and subscriptions:

In MQTT, publishers post messages to topics that may be taken into consideration as message subjects. Subscribers, thus, join topics to get specific messages. The Subscriptions

of topics may be express, which restricts the data which might be accumulated to the particular topic. Topics contain two wildcard degrees, to get data for a range of related topics.

3. Quality of service levels:

This protocol describes the Quality of Service (QoS) levels that are a deal between two parties of a message with respect to the assurance of distribution of data. It helps in 3 degrees of Quality of Services which might be defined as below.

- QoS0 (At most once): In this Quality level of service, the message is sent at most once and it does not offer guaranteed delivery of a message.

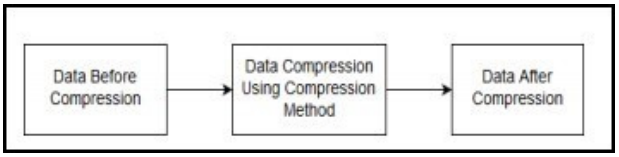
- QoS1 (At least once): In this Quality level of service, the data is sent at least once and it is viable to supply a message more than once by setting the value of the duplicate flag to 1.

- QoS2 (Exactly once): In this Quality level of service, the message is dispatched exactly once using the 4-way handshaking.

The choice of the QoS level relies upon the machine- like if a machine needs constant data delivery, it adapts to QoS2 for transmission of data even if there may be a time delay.

B. *COMPRESSION*

Compression is the process of converting a data set into a code to save the need for storage and transmission of data making it easier to transmit a data. With the compression of a can save in terms of time and storage that exist in memory (storage). Many compression algorithm techniques can be performed and function properly such as the Huffman, Lempel Ziv Welch, Run Length Encoding, Tunstall, And Shannon Fano methods.



Gzip is a file format and software application used on Unix and Unix-like systems to compress HTTP content before it's

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 7, Issue 10, October 2021

served to a client. This process is known to shrink files by up to 80 percent, resulting in improved page load time, de-

creased bandwidth consumption and reduced SSL overhead.

File types associated with gzip include:

- gz – Indicates a file extension compressed by the gzip algo- rithm.
- .tar file, tarball – A format used to store multiple files for archiving, but not for compression. Gzip can be used to com- press .tar files.
- .tgz, .tar.gz, .gz file – Indicates a .tar file that's been com- pressed by gzip.

Enabling gzip greatly improves transmission efficiency.

decompression speed (fast > slow): gzip, zip > 7z > rar > bzip2

compression speed (fast > slow): gzip, zip > bzip2 > 7z > rar

availability (unix): gzip > bzip2 > zip > 7z > rar

C. CLOUD STORAGE

The usage of Cloud Storage means that a user/customer/company will save their data within the cloud instead of on a local system. The access to the data is consummated by network connectivity and a client service. One benefit of Cloud Storage is that clients can access their data from any location, even though they do not have access to their organization's network. Stronger wireless networks and wider use of cell gadgets under BYOD (Bring-your-own-device) have accelerated the reliance on online storage.

Amazon is the most popular provider of cloud storage products for business. There are three types of data cloud storage and solutions that Amazon provides.

1. Object Storage - Amazon Simple Storage Service (S3): For applications that require scale and flexibility. To import existing data stores for analytics, backup, or archive.
2. File Storage - Elastic File System (EFS): For applications which need to access shared files and require a file system. For use cases such us large content repositories, development environments, media stores, or user home directories.
3. Block Storage - Elastic Block Store (EBS): For applications such us databases or ERP systems which often require dedicated, low latency storage for each host.

Amazon Simple Storage Service (Amazon S3).

Amazon S3 provides a cloud storage with security, reliability and high scalability. Amazon S3 has web service interface to

store and retrieve the data from the internet. The customer pays only for the storage that they use with no minimum fee and no initial cost.

Amazon S3 is recommended for use cases such us content distribution, disaster recovery, dig data, backup and recovery, and cloud applications.

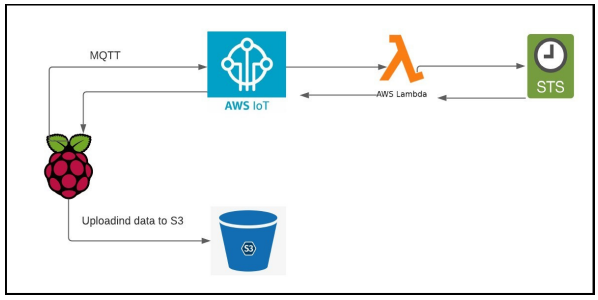
Common operations

- Create a bucket – Create and name a bucket in which one can store objects. A bucket is a container for gad-gets saved in Amazon S3. Every object is present in a bucket.
- Write an object – Objects are the essential entities saved in Amazon S3. An object encompasses object data and metadata. The data element is opaque to Amazon S3. The metadata is usually a set of name-value pairs that describe an object. These encompass a few default metadata like the date last modified, and standard HTTP metadata, and Content-Type. In S3, one can control what they want in the object. In this step, we store data by creating or overwriting an object. When you write an object, you specify a completely unique key in the namespace of your bucket. This is also a good time to specify any access if needed.
- Read an object – Read data back. You can download the data through HTTP. A key is a unique identifier for an object inside a bucket. Every object in a bucket has precisely one key. The aggregate of a bucket, key, and version ID uniquely discover each object.
- Delete an object – Delete some of your data.
- List keys – List the keys contained in one of your buck- ets. You can filter the key list based on a prefix.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 7, Issue 10, October 2021

IV. PROPOSED SYSTEMS



Main components used in this application:

A. *RASPBERRY PI*

The Raspberry Pi is a low cost computer, the size of a credit-card that uses a standard keyboard and mouse and plugs into a computer monitor or TV.
It is capable to enable people of all ages to explore computing, and to learn how to program in languages like Scratch and Python.

Device(Raspberry Pi) infrastructure comprises of Raspberry pi with Linux operating system, openJDK JRE and python 3. The device when connected to an usb storage helps to collect the data the usb storage to the cloud, thereby helping the user to retrieve the data from any particular place at any time.

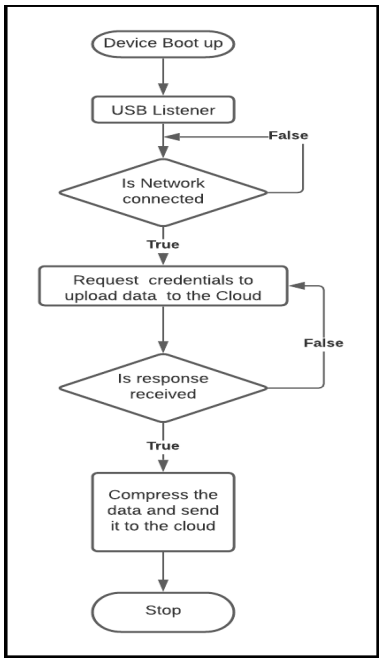
B. *AWS IOT*

In an order to make the device(Raspberry Pi) connect to the cloud, we have used MQTT Protocol of AWS IoT which is an MQTT message gateway that can send and receive MQTT messages to and from devices. Each device is termed as a thing in AWS IOT console.
For a device(Raspberry Pi) to be connected to the AWS IOT core, a "thing" should be created in the AWS console. After the creation of a thing, security certificates for the "thing" is to be created. Once the certificates are created, we'll have to download them and inject them in our Raspberry Pi for making it capable of connecting to the cloud. Cloud infrastructure comprises of AWS IOT core, Aws lambda, S3 and STS.

C. *AWS LAMBDA AND STS*

AWS Lambda is an event-driven, serverless computing platform provided by Amazon as a part of Amazon Web Services. Raspberry Pi that will publish data to MQTT Topics in AWS IOT Core, in turn will trigger an AWS Lambda function to check if such a Raspberry Pi is registered in the cloud, if yes: It fetches the temporary token credentials of the S3 bucket associated to the Raspberry Pi and sends the credentials back to the Raspberry Pi.
The credentials received by the Raspberry Pi is then used to upload the files of the usb storage to the AWS S3 bucket.
Once stored in the S3 bucket, the files and data can be easily accessible from any location and time. A validation in order to make sure the bucket and it's contents are secure will be required.

V. ALGORITHMS



VI. DETAILED FLOW

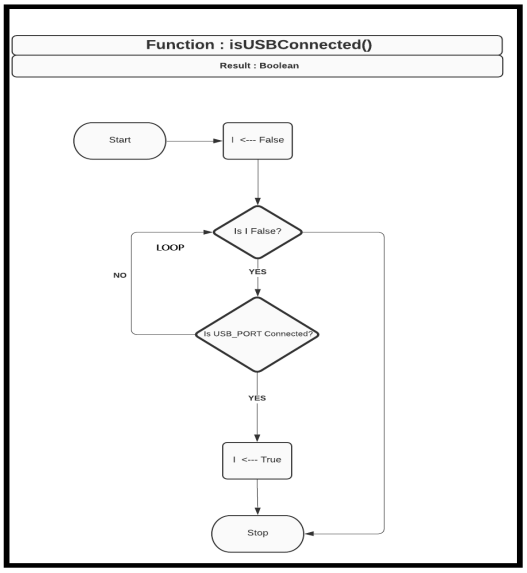
On boot up, the devices runs 2 scripts.

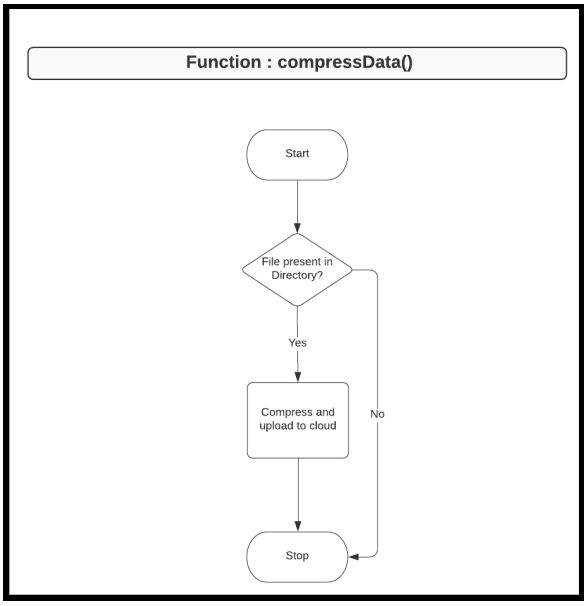
1) *First Script*

Python script to check if any usb storage is being inserted to the usb port. This script runs continuously after boot up and once an usb device is connected, it notifies the device to the 2nd Script.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

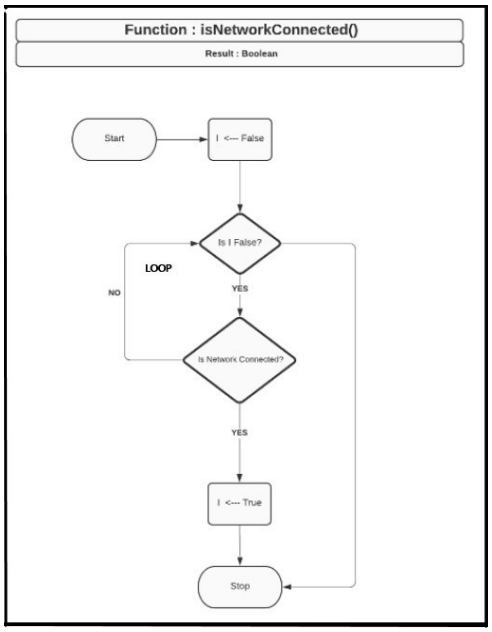
International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 7, Issue 10, October 2021





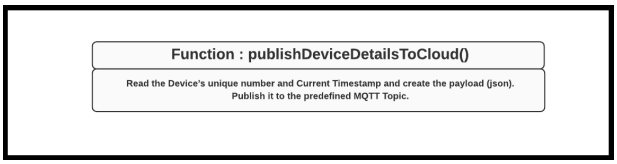
2) Second Script

The Second script checks if the Network is connected and if yes: it notifies the device to run the java application which has the logic for compressing and sending the data to the cloud.



Once the java application is running, if checks the number of files present in the device, and then starts to compress the files present in the device.

Compressing is done by the Apache Commons Compress library. After the compression is done, files are being sent to the S3 Bucket by using the Multi File upload functions provided by the java sdk of AWS IOT.



VII. RESULTS, DISCUSSION AND FURTHER SCOPES

A human on an average creates around 1.7 MB of information every second. This enormous growth in data generation requires a secure facility to store the data. Although hard disks can store large volumes of data and are easily avail- able, they pose a major threat to data corruption and are not very feasible due to high costs. The proposed system aims to mitigate such limitations by opting for cloud storage amalgamated with IOT devices to make the process of backup and recovery easily accessible. On boot up, network availability along with the USB connection is checked. Once connected to both, a send-receive process is carried wherein JSON is sent and the credentials to access the S3 bucket are received and thus begins the process of cloud backup. By using Gzip for compressing the data before uploading it to the cloud we are reducing the space of the data by approximately 68 per- cent. We compressed data of approximately 1400 MB by various compression

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 7, Issue 10, October 2021

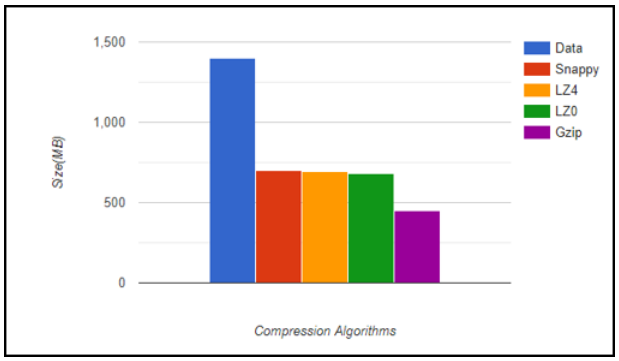
algorithms such as Snappy, LZ4, LZO, Gzip and observed that the memory of data gets reduced to

68 percent using Gzip. The system therefore eliminates the use of hard disks as it is:

- •Cheaper than other storages.
- •More safety and security.
- •More accessibility.

Further scopes are as follows:

1. Once data is received in the cloud, various ML models could be used in classifying and organizing the data with respect to their dates, sizes, etc.
2. Notification to the user when uploading fails or when the cloud storage limit is reached.



VIII. CONCLUSIONS

Proposed device (i.e. Raspberry pi along with AWS cloud) which when connects to an USB storage, gathers all the data from the USB storage and sends it to the cloud which could then be retrieved by the users from websites/applications from anywhere, at any time. It provides the users better security and accessibility when compared with storing data in USB storages. Our proposed set up is also cost efficient with contrast to the generic USB storages which provides inbuilt cloud facilities. Our proposed device when connected to any usb storages would provide the functionality of storing that data to the cloud. Once bought can be used with any storage devices.

IX. REFERENCES

[1] Ali Al Dahoud, Mohamed Fezari, " Use a Raspberry Pi and Amazon AWS For IoT Application Developement."
[2] Luluk Anjar Fitriya, Tito Waluyo Purboyo, Anggun-meka Luhur Prasasti, "A Review of Data Compression Tech- niques." in International Journal of Applied Engineering Re- search ISSN 0973-4562 Volume 12, Number 19 (2017) pp. 8956-8963.
[3] AWS IoT: Developer Guide, Copyright © 2019 Amazon Web Services.
[4] Dipa Soni, Ashwin Makwana, “A SURVEY ON MQTT: A PROTOCOL OF INTERNET OF THINGS(IOT).”
[5] Santiago Lucas Obrutsky, “Cloud Storage: Advantages, Disadvantages and Enterprise Solutions for Business."
[6] Bindu Trikha, “A journey from floppy disk to cloud storage” in International journal of Computer Science and Engineering Vol 02, No. 04, 2010, 1449-1452.
[7] Faisal Khalil, “Raspberry Pi Personal Cloud Storage”.
[8] Thomas Lee Scott, Amna Eleyan, “CoAP based IoT data transfer from a Raspberry Pi to Cloud” in September 2019