# Document Scanner and PDF Generator Using Cloud Platform

**Dirgh Jani**

Smt. Kuntaben Dinsha Patel Department of Information Technology
Chandubhai S. Patel Institute of  Technology, CHARUSAT

*Abstract-* The project gives you a production ready document scanning application. The project let the users  manage projects, upload images, and generate a PDF from detected text. . The developed project could be used as sample for development of expense tracking applications, handling legal documents and forms, or for digitizing notes and books. It can be accessed by using an API endpoint. You can use the service in Mobile as well as Computer. In mobile, It uses back camera to scan the document whereas it used File system in Personal Computer to upload the file.

*Index Terms*- aws, aws amplify, aws s3, aws cloudformation, cloud computing, document scanner, pdf generator

## I.  INTRODUCTION

This project demonstrates document scanning application and pdf generator. It lets the end point user manage projects, also they can upload multiple images and the project also has functionality of generating a PDF from text that is detected. The developed project could be used as sample for development of expense tracking applications, handling legal documents and forms, or for digitizing notes and books. It is implemented in cloud platform.

## II.  PLATFORM DETAILS

### CLOUD COMPUTING:

Cloud computing platforms delivers the services related to computing which includes software, servers, databases, networking, servers, intelligence, analytics, and storage on the cloud, not on the local machine, and it provides faster innovation, flexible resources, and also use economies of scale. You only pay for only the amount and section of cloud services you use, helps you lower operating costs used by you, efficiently use your infrastructure and increase or decrease as your business need change.

Advantages of using Cloud Computing Platform:
- Give capital expense instead of variable expense – You don't have to invest large amount in data centers and setting up servers as you have no idea of how you're going to use the hardware By using cloud computing, you need to pay only if computing resources are used by you.
- Advantage by being part of massive economies of scale - You can get lower variable cost than you can get on your own setup of environment by using cloud computing. When you set up a datacenter you need to pay the fix amount even if you use or not, while in cloud, hundreds of thousands of users are present and all usage are aggregated, so you can get benefit of the huge economy of scale in cloud like AWS .
- No need to worry about capacity – You do not need to worry about the capacity of resources your infrastructure would require. When you yourself deign a architecture, even after so much calculation and thinking of the requirements, you end up either by reserving a lot more resource or sometimes less resources. It can be expensive to you in both the scenario. By using cloud computing, these is not a problem. You can use as much resource you need. You can get even a small amount of resource whenever needed, and u can get huge chunk without overpaying.
- Much more agility and speed - In an environment of cloud-computing, new IT resources are too fast. That is because you reduce the time of resources, which is to be allocated to the development team. Traditional manner, you need to wait for resources but in cloud, it is just a click away. The output is a drastic increase in speed and agility of the team as you can get the resources earlier, it saves time and the cost even.
- No need to spend money on setting up and maintaining the datacenter – Instead of shifting the corrupted drives and shifting the racks, cloud computing helps you with it. You just need to focus on project and the customer, everything else related to hardware maintenance is managed by the cloud provider.

- Reach worldwide quickly – No need to worry if you have intercontinental business. You can deploy your developed software in multiple regions in just few clicks and by doing that, you can give your customer faster low latency experience at a very low cost.

## AMAZON WEB SERVICES(AWS):

Globally considered as the most comprehensive and broadly adopted cloud platform, Amazon Web Services (AWS) offers nearby 170 plus completely featured services from so many data centers there are present globally. Customer of Amazon Web Services (AWS) includes the fast-developing startups, big government agencies and bigger enterprises and they are using this platform to experience lower costs, be more agile and work faster.

### VUE.JS

Framework Vue is progressive used for constructing user interfaces. It is not like other monolithic frameworks. This framework is designed from the scratch to be incrementally adoptable. The view layer is only the focused core library and is easy to pick up and merge with other libraries or existing projects. With modern tooling and supporting libraries, Vue can be used for handlepowering sophisticated Single-Page Applications also.
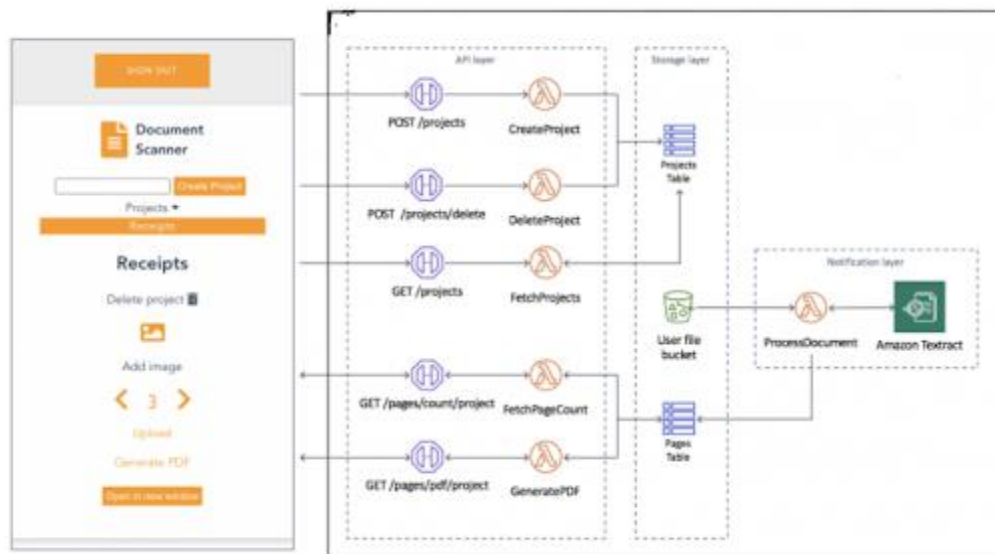
### SERVERLESS

The native architecture of the cloud is Serverless and it enables you to shift more of your operational responsibilities to cloud Service Provider i.e. AWS, it increases innovation and your agility. It gives you power to build and run applications and services without managing servers. Using Serverless also doesn't need management of infrastructure tasks such as server or cluster provisioning, patching, capacity provisioning and operating system maintenance. Developers can use the serverless platform to build anything for so many types of applications or it can also be used for backend service, and things or resources required to run and scale the developed application is handled for you by the service provider.

## III. ARCHITECTURE

The architecture consist of mainly two parts:
- Front-end Application
- Serverless Backend



Architecture Diagram of project

The application frontend application is developed in Vue.js framework and uses the AWS Amplify Framework. AWS serverless technologies are used to built the backend and contains Amazon API Gateway REST API that calls AWS Lambda functions. Analyze text from uploaded images is extracted using Amazon Textract. It is later stored to a Amazon S3 bucket. Whereas the Detected text from the AWS Textract is stored in Amazon DynamoDB.
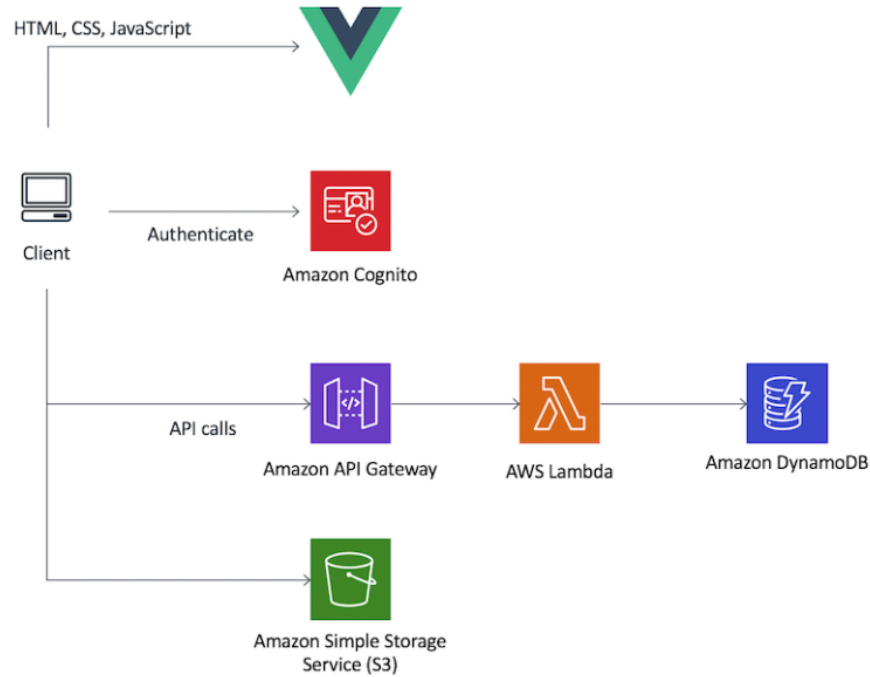
- NPM and Node.JS installed on desktop
- An AWS Account

V. DETAILED ARCHITECTURE

## A. *Front-end Application*

This part of application is developed in Vue.js and it also uses the AWS Amplify Framework.
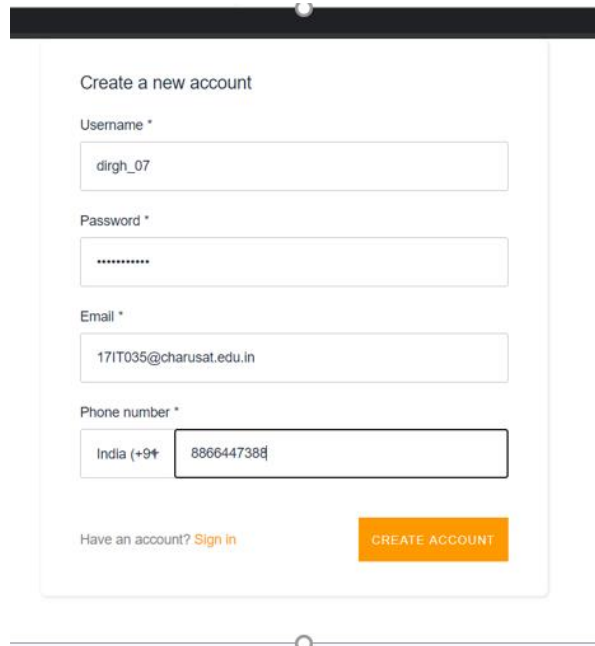


## B. Serverless Backend

The backend would be built using serverless technologies and would consists of an Amazon API Gateway REST API that would invokes AWS Lambda functions. Amazon Textract is used to analyze text from uploaded images to an Amazon S3 bucket. Detected text would be stored in Amazon DynamoDB.
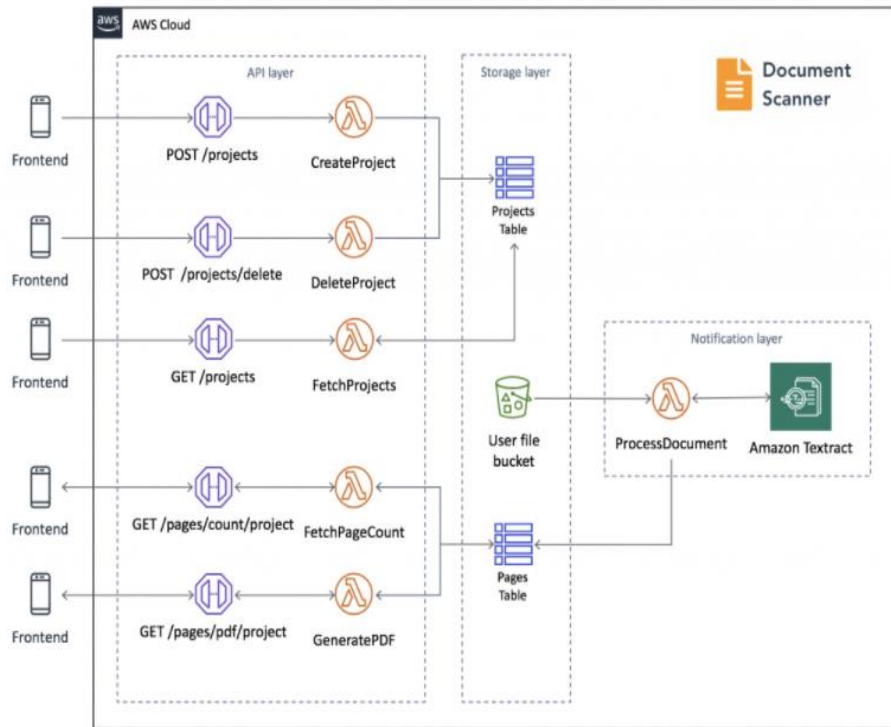
## VI. SERVICES USED

- AWS Amplify – It is a framework service that enables mobile and front-end web based application developers to develop and deploy scalable full stack applications, which is secure, powered by AWS. By using AWS Amplify, developers can setup up configuration of app backends in minutes, make its connection with developed app by just writting few lines of code, and deploy static web apps easily in three simple steps.

- Amazon Cognito – This service allows you to add user sign-up, sign-in, and give authority to login to your web and mobile apps easily and quickly. Amazon Cognito can have thousands of users and provides facility to sign-in with social media application like Facebook. Also by some of the social identity accounts like Google, and Amazon. It also supports the enterprise identity providers via SAML 2.0.

- Amazon API Gateway - It is a completely managed service which makes the task easy for us to create, publish, monitor, maintain and give security to APIs at any scale. APIs functions as the entrance for applications for accessing data, or functionality the backend services of your application.

- Amazon S3 - Amazon Simple Storage Service (Amazon S3) provides the service of object storage. It provides security, , data availability, industry-leading scalability, and performance. The meaning of this is the users of various capacities and large or small industries can use the service for storage and protect large or small amount of data for a various use cases.

- Amazon DynamoDB - Amazon DynamoDB is database service for key-value and document which delivers very minute millisecond performance at smll and large scale. DynamoDB, being a completely managed, multimaster, multiregion, highly durable database with inbuilt security features, restore and backup.

- Amazon Textract - It is a managed ML service provided by AWS which has capability to extracts data and text from documents which are scanned. The service goes beyond simple Typical optical character recognition (OCR) to understand, identify, and also extract data from images, forms and tables.

## VII. SNAPSHOTS OF WORK(OUTCOMES)



FIGURE 1 SIGN UP PAGE
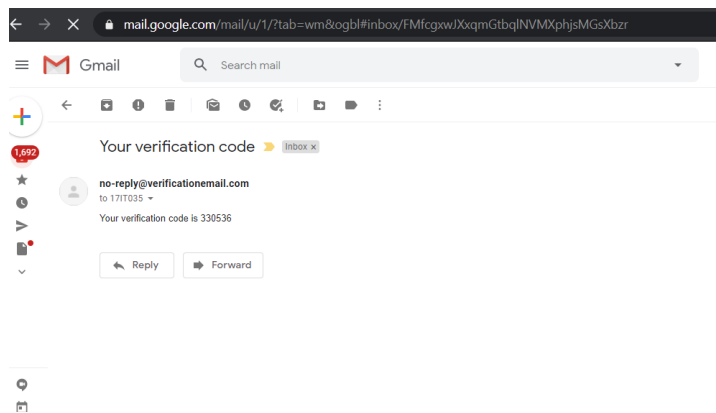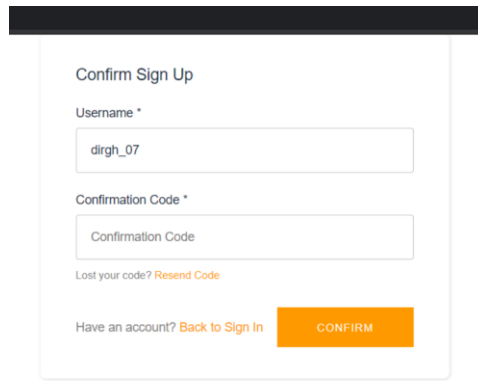


FIGURE 2 CODE SENT TO ENTERED EMAIL TO CONFIRM SIGNUP
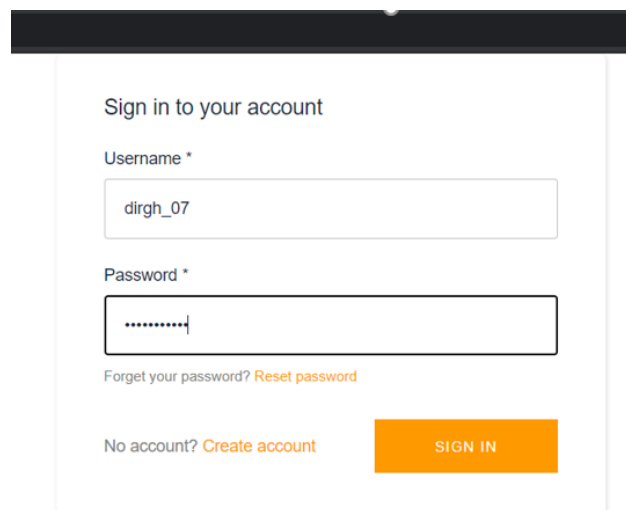
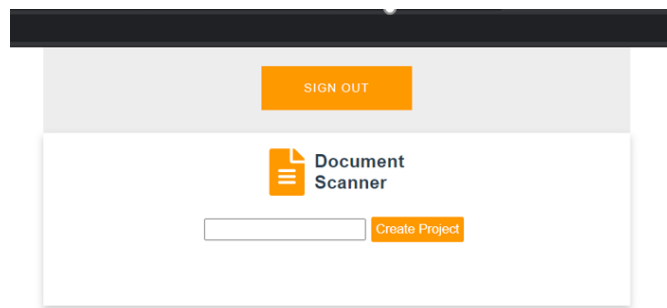FIGURE 3 ENTER CONFIRMATION CODE SENT TO EMAIL



FIGURE 4 SIGN IN PAGE

FIGURE 5 APPLICATION HOME PAGE
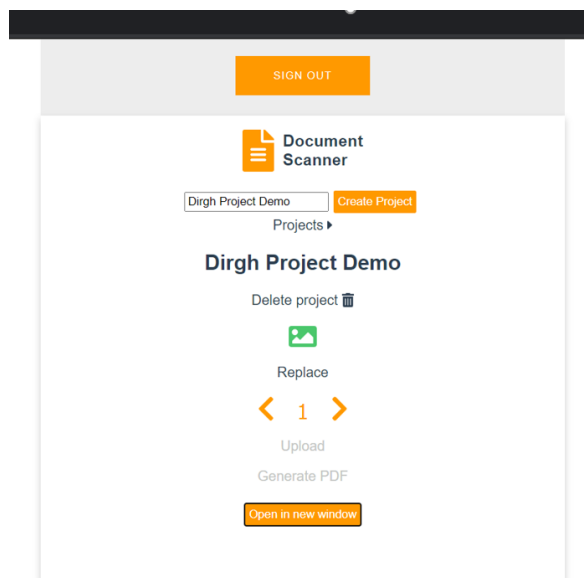


FIGURE 5 CREATING PROJECT
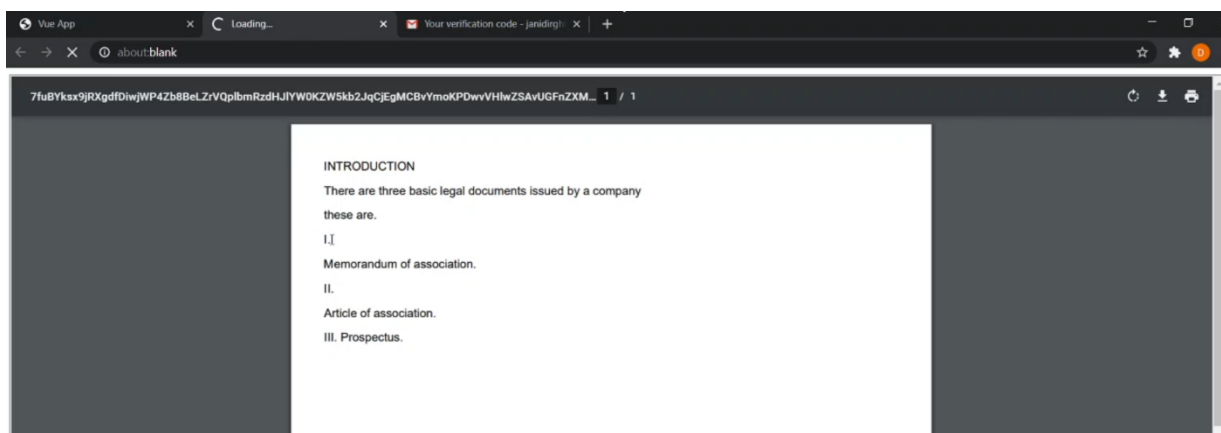
FIGURE 6 GENERATED PFD OF TEXT EXTRACTED FROM THE IMAGE

REFERENCES

[1] https://azure.microsoft.com/en-in/overview/what-is-cloud-computing/

[2] https://docs.aws.amazon.com/whitepapers/latest/aws-overview/six-advantages-of-cloud-computing.html

[3] https://aws.amazon.com/what-is-aws/

[4] https://www.serverless.com/

[5] https://aws.amazon.com/s3

[6] https://aws.amazon.com/dynamodb

[7] https://aws.amazon.com/serverless/

[8] https://aws.amazon.com/amplify/

[9] https://aws.amazon.com/cognito/

[10] https://www.researchgate.net/publication/326073288_A_Review_Paper_on_Cloud_Computing

[11] https://aws.amazon.com/textract

[12] https://www.explainthatstuff.com/how-ocr-works.html

[13] Slavina Ivanova, Georgi Georgiev. "Using modern web frameworks when developing an education application: a practical approach" , 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2019

AUTHORS

**First Author** – Dirgh Jani, B.Tech Student, CSPIT, CHARUSAT

# Paper 3

| **7**% | **5**% | **4**% | % |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

**1**    Slavina Ivanova, Georgi Georgiev. "Using modern web frameworks when developing an education application: a practical approach", 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2019
Publication     **2**%

**2**    docs.aws.amazon.com
Internet Source     **2**%

**3**    tutorialsdojo.com
Internet Source     **1**%

**4**    ijircce.com
Internet Source     **1**%

**5**    neoniche.com
Internet Source     **1**%