

CS CAPSTONE TECH REVIEW DOCUMENT
NOVEMBER 8, 2019

**AUTOMATE THE SETTINGS THAT CONTROL A
MILLION-DOLLAR PRINTING PRESS**

PREPARED FOR
HEWLETT-PACKARD, INC
PIETER VAN ZEE

PREPARED BY
GROUP62
PROPRIETORS OF THE PRESS
KUAN-YU LAI

Abstract

The document includes three pieces from the term project "Analysis of input data for generation of outputs", "Hosting the feature on the web", and "Storage of outputs and profiles". In each piece, it contains three different technology that is related to the pieces. The technology is considered to be a temporary option in the project.

1 INTRODUCTION

Our team project is to design an assisting feature for the user of the HP PageWide printing presses. This assisting feature will help the user to find the best set of each individual printing tasks they submit into the system. The decision of the setting will be based on factors such as page type, printing quality, or ink usage.

The core of this feature is an engine that can consider every input and generate the most optimal set of settings. The feature can be broken down into several pieces which are "Analysis of input data for generation of outputs", "Hosting the feature on the web", and "Storage of outputs and profiles".

2 PIECE1: ANALYSIS OF INPUT DATA FOR GENERATION OF OUTPUTS

The analysis of input data is the main piece of the project, it determines the reliability and the meaning of the assisting feature. It is an engine that will take all of the input into consideration based on the ruleset we decide. Eventually, generate the optimal set of setting for the users.

2.1 Drools

Drools, also known as knowledge is everything(KIE), is a business rules management system(BSMS) that allows the user to generate the result based on the ruleset of their own. It is also a Java rules engine with forwarding and backward chaining. The engine uses an algorithm called the Phreak rule algorithm to evaluate the rules. The algorithm is an enhanced version of the Rete algorithm, a pattern matching algorithm invented in 1979. It's better at speed and supports more evaluation methods. A benefit of using this technology is that it the same concept with our assisting engine. Therefore the implementation will be easier and possibly more time-efficient.[1]

2.2 Google Cloud Platform

Google cloud platform is a platform announced by Google in 2008. It's a cloud-computing platform similar to Amazon Web Services(AWS). It allows you to use the well-trained machine learning engine from Google. The user will be able to input data into the engine and gets the analytical data from the engine. It supports Rest API so it's very friendly for the user who doesn't want to train their own engine. However, Google does support customize engine. Users can either use the built-in algorithm or create their own training application and run it on the cloud. A benefit of using this technology will be the advantage of the pre-trained engine. The engine already has high accuracy on the picture analysis which might be very helpful for our input analysis.[2]

2.3 TensorFlow

TensorFlow is an open-source platform specifically for machine learning from Google. It allows the user to build the machine learning application with a variety kind of computer languages. The high-level API makes the building and training steps easier and the user can deploy the application on to their product eventually. A benefit of using this technology will be a friendly development environment for the beginner. At the same time, it is also very powerful enough to generate a decent machine learning engine.[3]

3 PIECE2: HOSTING THE FEATURE ON THE WEB

To provide a convenient service to the client of the presses. Our project will have a cloud application that allows users to submit their printing tasks with the setting they want on the cloud. The cloud application will then sent the input data from the user to the engine and generate the result for them. The cloud application and the engine will be communicated through API. As a result, we have to find a reliable and customizable web server.

3.1 Oregon State University engineering server

Oregon State University allows students to publish their own applications using their servers. Each student will be assigned a cloud space with limited storage. Students then can publish the application they made by simply move the data into the publish folder. After that, anyone can access the application with the link http://web.engr.oregonstate.edu/~your_engineering_username. However, the service isn't very customizable but they do provide the other paid server which costs 88 dollars monthly. A benefit for using this server will be the server is very stable and is maintained by the university. One disadvantage of using this server is the free version might support the engine we built. Also, we will have to pay monthly if we want to have an advanced server.[4][5]

3.2 Heroku

Heroku is a cloud application platform that allows you to put your application on the cloud and publish it to the world. Its popularity is growing for the past few years due to its simplicity and highly customizable property. The reason why it's simple to use is that the developer will not have to worry about the infrastructure of the server. Heroku handles both hardware and server for you so you can focus on perfect your own application. This is very friendly to people who don't have lots of experience in setting up and maintaining the server. As a result, it's being used by large non-tech related companies such as Toyota. The benefits of using Heroku will be, eliminate all the issues from the server and hardware since this is handled by Heroku. Some disadvantages will be the collaboration issue and the active time of the application. The free version of Heroku only allows one person to use and will put the application into sleep mode if it doesn't receive any traffic within one hour. Although the manager will wake the server up when it receives a request. It will still cause a significant delay for the application to restart.[6]

3.3 Amazon Web Services

Amazon Web Services(AWS) is a well-known cloud platform from Amazon since 2006. It's being used by many large companies such as General Electric. The platform provides thousands of features which also include hosting the application. The server provided by AWS is extremely customizable and stable. It also supports some other popular platform like docker, a set of platforms that allows you to build your application from scratch rapidly. It also supports machine learning computation similar to the Google Cloud Platform. A benefit of using AWS will be highly customizable and supported properties. We won't have to worry about almost any cloud-related technology that doesn't support our server. A disadvantage for this is the services isn't free but they offer a free trial for 12 months after the first sign-up.[7]

4 PIECE3: STORAGE OF OUTPUTS AND PROFILES

To build an application, it's essential to have a storage space that can keep all the user data and the ruleset we made. This is the fundamental piece of the project since all of the features is based on the data stored in the database. Thus, it's important to choose a reliable database that is suitable for our data.

4.1 MYSQL

MYSQL is a traditional open-source database developed by a tech company called Oracle. It's a database that uses structured query language(SQL) and stores the data as tables. MYSQL is widely used in the industry, it's being used by companies like Walmart or Spotify. Some benefits of using MYSQL will be it's widely support property and powerful query feature. A disadvantage of this is that our rules don't have a significant relationship. This might make the query a little bit complicated when pulling the rules.[8]

4.2 MongoDB

MongoDB is a modern database that is considered to be the most popular database nowadays. Unlike the traditional row and column-based structure, it uses the document-based structure and is built for the cloud era. The data will be formed as a JSON file, so developers can modify the JSON document to edit the data in the database. The powerful query feature is another reason why it's extremely popular. Furthermore, it provides a statistical feature that allows the developer to visualize the data stored in the database. Due to its outstanding ability in this era, it is being used by giant tech companies like Google and Facebook. A benefit of using MongoDB will be the usage of powerful query feature. Since our ruleset doesn't have a significant relationship with each other. The powerful query feature will simplify the query compared to the traditional SQL database.[9]

4.3 Redis

Redis is also an open-source database. Unlike the traditional row and column-based structure and document-based structure. Redis uses key-value based on its structure. That is, developers can pull out the specific data by query the key of that data. Also, Redis stores the data in the memory instead of disk, so the speed is extremely fast compared to the database like MongoDB or MYSQL. A benefit of using Redis is that we will be the property of easily pull out specific data. The data has a corresponding key so the relationship between the data won't affect the query complexity significantly. A disadvantage of using Redis will be the in-memory storing structure because it's not stored permanently. Thus, it will have to rebuild the database whenever the server restart.[10]

5 CONCLUSION

In this modern world, there are thousands of technology that exist and more and more coming out in the future. In addition, finding and choosing the appropriate technology will increase the performance and the development time of the application. Therefore, the technology present in each piece will not be the only option to consider.

REFERENCES

- [1] Drools Documentation. [online] Docs.jboss.org. Available at: <https://docs.jboss.org/drools/release/7.29.0.Final/drools-docs/htmlsingle/index.html> [Accessed 9 Nov. 2019].
- [2] Google Cloud. (2019) Training Overview AI Platform Google Cloud. [online] Available at: <https://cloud.google.com/ml-engine/docs/training-overview> [Accessed 9 Nov. 2019].
- [3] TensorFlow. (2019). Introduction to TensorFlow — TensorFlow. [online] Available at: <https://www.tensorflow.org/learn> [Accessed 9 Nov. 2019].
- [4] Information Services. (2019). Server Management — Applications Access and Deployment, Shared Infrastructure Group — Information Services — Oregon State University. [online] Available at: <https://is.oregonstate.edu/service/server-management> [Accessed 9 Nov. 2019].

- [5] It.engineering.oregonstate.edu. (2019). Where do I put my personal webpages? — Information Technology and Computing Support — Oregon State University. [online] Available at: <https://it.engineering.oregonstate.edu/where-do-i-put-my-personal-webpages> [Accessed 9 Nov. 2019].
- [6] Devcenter.heroku.com. (2019). How Heroku Works — Heroku Dev Center. [online] Available at: <https://devcenter.heroku.com/articles/how-heroku-works> [Accessed 9 Nov. 2019].
- [7] Docs.aws.amazon.com. (2019). What Is Amazon EC2? - Amazon Elastic Compute Cloud. [online] Available at: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html> [Accessed 9 Nov. 2019].
- [8] Dev.mysql.com. (2019). MySQL :: MySQL 8.0 Reference Manual :: 1 General Information. [online] Available at: <https://dev.mysql.com/doc/refman/8.0/en/introduction.html> [Accessed 9 Nov. 2019].
- [9] MongoDB. (2019). What Is MongoDB?. [online] Available at: <https://www.mongodb.com/what-is-mongodb> [Accessed 9 Nov. 2019].
- [10] Redis.io. (2019). Redis. [online] Available at: <https://redis.io/documentation> [Accessed 9 Nov. 2019].