CS 6381 Final Report – Google Cloud Pub Sub Team:

1. Overview

\subsection{Google Cloud Pub Sub}

GCPS (Google Cloud Pub Sub) is a message-centric, decoupled messaging service, that utilizes the HTTP delivery protocol and transport. The system is structured to deliver message from point to point as since communication is handled through multiple terminals on the same computer. The underlying system structure utilizes a hub and spoke system, where there is connection from the sending to the receiver and back. You can create a VPC (Virtual Private Cloud) within google cloud and connect this service to other projects or google based resources.

When transporting messages across GCPS, the system handles TCP and UDP when it comes to balancing how the messages are sent and received. Messages received from GCPS can be sent to any application that utilizes a protocol like TCP or UDP such as a python-based sending/listening application. Data that is sent through these protocols must be structured in JSON format and utilizes various dating types such as strings and time stamps. Publishers make and send messages to a topic with data, ordering key, and associated metadata, whereas subscribers have message ordering property options.

GCPS supports sending messages through MQTT bridge to cloud IoT core, which supports multiple QOS levels. GCPS possesses the ability to dynamically tweak publisher and subscribers based on message throughput and system load capacity. This allows the most efficient use of time and memory to allow faster delivery of messages across the system. When sending messages, publisher will utilize a publisher feature that pushes out whatever message is being delivered and the subscriber will pull that message and decode it and utilize it, as necessary.

On the backend for the abstraction of implementation details, there exists a service proxy among clients and data senders that helps optimize the connection details for clients. This control plane assigns the publishers and subscribers to the servers. Then this data plane allows transmissions of data between the publishers and subscribers. When utilizing GCPS, the up-front struggle stems from: type of software to use, message quotas and usage, horizontal scalability, publishing and error handling, redundancy, rollback, traffic limits, incremental changes, capacity management, and handling failed publications.

When handling large scale implementations, GCPS caps out at 10000 topics per project, 10MB of data per message, and 10 KB per schema, etc. Across a load-balanced set of subscribers, only one subscriber can be in partition at a time, so there is a limit on parallel processing. To effectively utilize a large-scale implementation, you will need to create a GC (Google Cloud) account and setup the appropriate features using website UI, set up the appropriate topic(s) for the publisher, and the subscribers. Install the appropriate language and libraries and then generate the multiple terminals that will handle the point-to-point connection between the publishers and subscribers.

~\cite{gcps\_documentation}

~\cite{how\_to\_make\_a\_proper\_connect\_between\_raspberry\_pi\_and\_gcloud\_iot}

~\cite{crate\_google\_pubsub1}

~\cite{log\_streaming\_google\_cloud\_pubsub}

~\cite{implementing\_the\_hub\_and\_spoke\_model}

~\cite{publisher\_subscriber\_model\_for\_apps\_and\_data}

~\cite{internal\_tcp\_udp\_load\_balancing}

~\cite{publishing\_messages\_to\_topics}

~\cite{google\_cloud\_pub\_sub\_ordered\_delivery}

~\cite{cloud\_google\_pubsubmessage}

~\cite{publishing\_over\_mqtt}

~\cite{pub\_sub\_google\_scale}

~\cite{cloud\_google\_subscriber\_overview}

~\cite{cloud\_google\_what\_is\_pub\_sub}

~\cite{cloud\_google\_pub\_sub\_quota\_limits}

~\cite{publish\_with\_error\_handling}

~\cite{cloud\_google\_reliability}

~\cite{cloud\_google\_publishing\_messages\_to\_topics}

~\cite{cloud\_google\_install\_sdk}

~\cite{google\_cloud\_tech\_pub\_sub}

~\cite{cloud\_pub\_sub\_in\_action\_ep\_3}

~\cite{cloud\_google\_pub\_sub\_client\_libraries}

~\cite{build\_kotlin\_spring}

~\cite{building\_one\_to\_many\_pub\_sub\_system}

~\cite{webservice\_google\_client}

~\cite{watch\_google\_bucket}

~\cite{google\_cloud\_pubsub\_241}

~\cite{google\_cloud\_pubsub\_pricing}

~\cite{contact\_us\_google\_cloud}

1. Experiment Criteria and Design

\subsection{Google Cloud Pub Sub Experiment Criteria and Design}

We created a set of applications to test the message sending/receiving efficiency of the Google Cloud Pub Sub service based on our overall understanding of the experiment designers’ guidelines and the time available. We established pub sub topics, project and topic ID’s, and a JSON key with the help of the Google Cloud Pub Sub user interface online. On Ubuntu 18.04, we set up the Google Cloud SDK and configured information based on the project and topic ID’s and the JSON key. With this, we were able to run our sets of applications.

Data flows between our applications, pub.py, broker.py, and sub.py. Pub.py sends messages to broker.py via ZMQ continuously. This involves establishing a ZMQ context, getting the IP address via a socket command, and binding the socket to a TCP path that includes the IP address and a given port number. Then with a “while True” loop, we loop through the files in the messages folder, get the text from there, append it to the message size, which is how many characters there are in that text, current timestamp, and iteration count, and send off the combined message through the socket. The message files in the messages folder contain text that vary in lengths and are created with the help of messageGenerator.py, which generates a string of random characters to a specified length and saves it to the messages folder as a message file.

Broker.py receives messages from pub.py via ZMQ and sends messages to sub.py via Google Cloud Pub Sub continuously. This involves passing the project and topic ID’s to a class. In this class, a given IP address and port number are used to connect to the socket and a filter is established based on the topic keyword. In a “while True” loop, a call on the poll() function occurs and when a topic match is found, incoming messages are received, a timestamp is taken, and this information is passed to the measureTime() function that parses the message, finds the time difference between the receipt time and the time the message was sent from pub.py, and assembles a new one based on the message size, current timestamp, this calculated publisher to broker time, success rate, and message content. This new message is converted into bytes and then sent out via Google Cloud Pub Sub.

Sub.py receives messages from broker.py via Google Cloud Pub Sub and saves the messages into CSV files when specified numbers of messages have been received. This involves listening for incoming messages via Google Cloud Pub Sub. When the message is received, a timestamp is taken and both are sent to the extractMessage() function. This function parses the message, calculates the time difference between the receipt time and the time the message was sent from broker.py, appends the message ID, publisher to broker time, and this calculated broker to subscriber time to a message list, and returns the message size and publisher to broker success rate in the form of a dictionary. When the size of the message list equals the given message limit, the messages from the message list, containing the time measurements, are saved into a time measurements CSV file that includes in its name the message limit, message size, publisher to broker success rate, and UUID for clear distinction. Messages contain message ID’s and times it took to get from pub.py to broker.py and from broker.py to sub.py.

We ran tests using shell files to examine the effects of varying producer/consumer conditions based on the following: We treated broker.py as the producer since it sent messages via Google Cloud Pub Sub and sub.py as the consumer since it received messages via Google Cloud Pub Sub. Within a test shell file, we fetched the JSON information and configuration values for Google Cloud Pub Sub in each command that involved running a producer or consumer or pub.py. The order of running from 1st to last is as follows: sub.py, broker.py, and pub.py. Each test included variation in message length based on the messages in the messages folder. The message files vary in terms of total number of characters as follows: 22, 149, 6815, 155355, and 761135. There was some sleep time applied to wait for processes to finish before attempting to delete the processes for broker.py, then pub.py, and finally sub.py and running clear.py to clear future messages in a way since we have observed that it seems like Google Cloud Pub Sub sends old messages that the recipient has not received the next time it connects for receiving messages. The other way we found to essentially empty out the old messages in a way was to start up sub.py and check to see that all the old messages have been received as logs and the output stops expanding.

Our expectations for running the tests are as follows: The 1 producer vs. 1 consumer test involves 1 broker.py and 1 sub.py, and we run this once. The 1 producer vs. 5 consumers test involves 1 broker.py and 5 sub.py, and we run this once. The 5 producers vs. 1 consumer test involves 5 broker.py and 1 sub.py, and we run this once. The 8 producers vs. 8 consumers test involves 8 broker.py and 8 sub.py, and we run this 3 times. The 25 producers vs. 25 consumers test involves 25 broker.py and 25 sub.py, and we run this 4 times.

We used calculateStatistics.py to read the data from the time measurements CSV files and save a CSV filled with statistics and calculateStatisticsAll.py for similar purposes with the additional feature of generating graphs based on the calculations. Here, we set a for-loop to check for CSV files in a folder. When there’s a match, we call readCSV(), which takes in the file and opens it, iterates through it, and appends the write and read times to lists as well as write and read success rates, which are based on how many time rows there are divided by the total iterations for a file. For latency and throughput, using this collected information, we get the minimum and maximum values for times by using the min() and max() functions respectively, average by summing the tuple-converted data list of float-converted times and dividing that by the length of the data list, and standard deviation by getting the variance for the result of dividing the sum of the data list by the length of the data list and setting it to the power of 0.5. These findings are added to a calculations CSV file with the current timestamp and UUID in the title to distinguish the files apart. calculateStatisticsAll.py featured generating graphs by passing in the file name that is parsed from the path name. Using matplotlib and pandas, graph settings are configured with the X-axis showing the various test conditions and the Y-axis showing the values for the parameter of interest as graphs are generated.

We used hardwareMeasure.py to continuously see the latest hardware statistics, including CPU, RAM, and storage. In a “while True” loop, we printed the functions that return a dictionary including the max average values for each of these parameters. In each function, we appended the parameter fetching using psutil to its own list and from that list, we find the maximum value using the max() function and average value by summing up the tuple-converted data list and dividing that by the length of the list.

1. Experiment Results

\subsection{Google Cloud Pub Sub Experiment Results}

The following includes our observations of trends among the graphs, which are stored in the “graphs\_GoogleCloudPubSub1” folder. In terms of standard deviation, average, max, and min for publisher to broker success rates and broker to subscriber success rates, the 1producer1consumer test yields low results, the 1producer5consumer test yields moderate results, and the other conditions as well as when all data is considered yield the highest results. For broker to subscriber success times, this trend appears similar for standard deviation, average, and max, but for min, 1producer5consumer, 25producer25consumer, and 5producer1consumer tests yield moderate results as so do publisher to broker times in terms of standard deviation, average, and max. For publisher to broker times min, there is less variation between the different conditions with slightly less results for the 8producer8consumer tests and when all data is considered. One thing though is that the 25producer25consumer test did not seem to work to full efficiency on the computer we tested on as the computer froze and we only have 20 results for 1 run of that test.

From these observations, it seems that Google Cloud Pub Sub throughput, which corresponds to the broker to subscriber success rate, appears mostly high when there are more producers and especially consumers. When we increase the number of producers and especially consumers, the latency increases. Perhaps this is due to increased competition for access to Google Cloud Pub Sub, which may create a bottleneck effect. A similar pattern occurs with ZMQ, but the results are more moderate for some conditions. When there are too many producers/consumers or more producers than consumers, this yields moderate results at least in our case. For the most part though, it seems Google Cloud Pub Sub can send messages efficiently in certain aspects like ZMQ.

For the base file, it appears the hard disk memory stays roughly stable while the readIteration.py test is run. RAM and CPU increase from the 1st to the 2nd load. The computer we used had other activities going on outside of the VM, so perhaps this contributed to the RAM and CPU fluctuations. The computer has a lot of hard disk memory, so changes to the memory should appear minimal as long as large-sized items are not added or removed.

graphs from “graphs\\_GoogleCloudPubSub1” folder:

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/brokerToSubscriberSuccessRatesAverage\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/brokerToSubscriberSuccessRatesMax\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/brokerToSubscriberSuccessRatesMin\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/brokerToSubscriberSuccessRatesStandardDeviation\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/brokerToSubscriberTimesAverage\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/brokerToSubscriberTimesMax\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/brokerToSubscriberTimesMin\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/brokerToSubscriberTimesStandardDeviation\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/publisherToBrokerSuccessRatesAverage\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/publisherToBrokerSuccessRatesMax\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/publisherToBrokerSuccessRatesMin\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/publisherToBrokerSuccessRatesStandardDeviation\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/publisherToBrokerTimesAverage\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/publisherToBrokerTimesMax\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/publisherToBrokerTimesMin\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/publisherToBrokerTimesStandardDeviation\_04-23-2021\_195539\_9f7b0c37-4d13-4bde-876b-6366cc505eeb.png}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/HardwareMeasurements\_CPU.PNG}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/HardwareMeasurements\_HardDiskMemory.PNG}

\includegraphics[width=50mm,scale=0.5]{./fig/graphs\_GoogleCloudPubSub1/HardwareMeasurements\_RAM.PNG}

1. Repository
   1. <https://github.com/jamesvu1/cs6381finalprojectgooglecloudpubsub>
2. References

@inproceedings{gcps\_documentation,

title={Google Cloud Pub/Sub Documentation},

url={https://cloud.google.com/pubsub/docs}

}

@inproceedings{how\_to\_make\_a\_proper\_connect\_between\_raspberry\_pi\_and\_gcloud\_iot,

title={How to Make a Proper Connect Between Raspberry Pi and GCloud-IOT},

author={Reshani P},

url={https://www.dckap.com/blog/how-to-make-a-proper-connect-between-raspberry-pi-and-gcloud-iot/},

year={2020}

}

@inproceedings{crate\_google\_pubsub1,

title={Crate google\\_pubsub1},

url={https://docs.rs/google-pubsub1/2.0.0+20210322/google\_pubsub1/}

}

@inproceedings{log\_streaming\_google\_cloud\_pubsub,

title={Log streaming: Google Cloud Pub/Sub},

url={https://docs.fastly.com/en/guides/log-streaming-google-cloud-pubsub},

year={2021}

}

@inproceedings{implementing\_the\_hub\_and\_spoke\_model,

title={Implementing the hub and spoke model},

url={https://cloud.google.com/managed-microsoft-ad/docs/hub-spoke}

}

@inproceedings{publisher\_subscriber\_model\_for\_apps\_and\_data,

title={Publisher/Subscriber Model for Apps and Data Ingestion Flows},

author={Daniel martinez Bielostotzky},

url={https://towardsdatascience.com/publisher-subscriber-model-for-apps-and-data-ingestion-flows-b8ba7e85e992},

year={2020}

}

@inproceedings{internal\_tcp\_udp\_load\_balancing,

title={Internal TCP/UDP Load Balancing overview},

url={https://cloud.google.com/load-balancing/docs/internal}

}

@inproceedings{publishing\_messages\_to\_topics,

title={Publishing messages to topics},

url={https://cloud.google.com/pubsub/docs/publisher}

}

@inproceedings{google\_cloud\_pub\_sub\_ordered\_delivery,

title={Google Cloud Pub/Sub Ordered Delivery},

author={Kamal Aboul-Hosn},

url={https://medium.com/google-cloud/google-cloud-pub-sub-ordered-delivery-1e4181f60bc8},

year={2020}

}

@inproceedings{cloud\_google\_pubsubmessage,

title={PubsubMessage},

url={https://cloud.google.com/pubsub/docs/reference/rest/v1/PubsubMessage}

}

@inproceedings{publishing\_over\_mqtt,

title={Publishing over the MQTT bridge},

url={https://cloud.google.com/iot/docs/how-tos/mqtt-bridge}

}

@inproceedings{pub\_sub\_google\_scale,

title={Pub/Sub: A Google-Scale Messaging Service},

url={https://cloud.google.com/pubsub/architecture}

}

@inproceedings{cloud\_google\_subscriber\_overview,

title={Subscriber overview},

url={https://cloud.google.com/pubsub/docs/subscriber}

}

@inproceedings{cloud\_google\_what\_is\_pub\_sub,

title={What Is Pub/Sub?},

url={https://cloud.google.com/pubsub/docs/overview}

}

@inproceedings{cloud\_google\_pub\_sub\_quota\_limits,

title={Quotas and limits},

url={https://cloud.google.com/pubsub/quotas}

}

@inproceedings{publish\_with\_error\_handling,

title={Publish with error handling},

url={https://cloud.google.com/pubsub/docs/samples/pubsub-publish-with-error-handler}

}

@inproceedings{cloud\_google\_reliability,

title={Reliability},

url={https://cloud.google.com/architecture/framework/reliability}

}

@inproceedings{cloud\_google\_publishing\_messages\_to\_topics,

title={Publishing messages to topics},

url={https://cloud.google.com/pubsub/docs/publisher}

}

@inproceedings{cloud\_google\_install\_sdk,

title={Installing Google Cloud SDK},

url={https://cloud.google.com/sdk/docs/install}

}

@inproceedings{google\_cloud\_tech\_pub\_sub,

title={Cloud Pub/Sub in Action - ep. 3},

author={Google Cloud Tech},

url={https://www.youtube.com/watch?v=f5DOsB7Nlw0&t=254s},

year={2019}

}

@inproceedings{cloud\_pub\_sub\_in\_action\_ep\_3,

title={Cloud Pub/Sub in Action - ep. 3},

author={Erik Fredericks},

url={https://www.youtube.com/watch?v=UKAmZBrR300},

year={2021}

}

@inproceedings{cloud\_google\_pub\_sub\_client\_libraries,

title={Pub/Sub Client Libraries},

url={https://cloud.google.com/pubsub/docs/reference/libraries}

}

@inproceedings{build\_kotlin\_spring,

title={Build a Kotlin Spring Application with Google Cloud Platform},

url={https://codelabs.developers.google.com/codelabs/cloud-spring-cloud-gcp-kotlin#1}

}

@inproceedings{building\_one\_to\_many\_pub\_sub\_system,

title={Building a one-to-many Pub/Sub system},

url={https://cloud.google.com/pubsub/docs/building-pubsub-messaging-system}

}

@inproceedings{webservice\_google\_client,

title={WebService-Google-Client},

author={Steve Dondley},

url={https://github.com/sdondley/WebService-Google-Client},

year={2018}

}

@inproceedings{watch\_google\_bucket,

title={Watch Google Bucket with Elixir: Google Cloud Storage + PubSub + Elixir Broadway},

author={Pavel Tsiukhtsiayeu},

url={https://paveltyk.medium.com/watch-google-bucket-with-elixir-google-cloud-storage-pubsub-elixir-broadway-500279375739},

year={2021}

}

@inproceedings{google\_cloud\_pubsub\_241,

title={google-cloud-pubsub 2.4.1},

author={Pavel Tsiukhtsiayeu},

url={https://pypi.org/project/google-cloud-pubsub/},

year={2021}

}

@inproceedings{google\_cloud\_pubsub\_pricing,

title={Pricing},

url={https://cloud.google.com/pubsub/pricing}

}

@inproceedings{contact\_us\_google\_cloud,

title={Contact Us | Google Cloud},

url={https://cloud.google.com/contact}

}