# What to know about the Google Cloud: A non-technical guide for academic staff

Joep Steegmans

Utrecht *University* School of Economics

October 30, 2020

# Contents

# 1 Introduction

In 2018, I started an ambitious project into online housing search behaviour with microdata of Funda (see Steegmans and de Bruin, 2019). The 10 terabytes of Google Analytics data forced us to work with the Google Cloud Platform (GCP). This created a challenge as there was no centrally available information on working with the Google Cloud Platform at Utrecht University.

This report aims to provide insights into GCP research projects. Most of all, the report will focus on what there is to know *before* you start your research; that is, we hope to prevent UU researchers from reinventing the GCP wheel over and over again. To the extent possible, we try to provide a Utrecht perspective. We include, for instance, publications that are available through our own university library as much as possible.

The report covers the main lessons that we have learnt until now. These lessons relate to obtaining Google Cloud funding, Google Cloud conferences, the GCP Faculty Community, etc. The report is mostly non-technical in nature as technical information is generally available elsewhere. We hope that this document will provide a useful steppingstone for those UU researchers who want to innovate by using the Google Cloud Platform. Details on the available services and products on the Google Cloud Platform can be found on cloud.google.com. Just as information on getting started immediately.

Finally, it is good to realize that the Google Cloud research support, the main focus of this document, is not static. On the contrary, the Google Cloud initiatives and programmes are as dynamic as cloud technology itself. Support programmes are regularly upscaled, optimized, downscaled, or even discontinued. At the same time new pilot projects see the light of day frequently. Nevertheless, this document should be useful for the years to come, especially when you keep in mind that support programmes and eligibility criteria are likely to show some changes.

# 2 Cloud technology

## 2.1 Terminology

Cloud technology is being adopted by more and more companies (Riti, 2018, p. 19). But what exactly is it? According to Krishnan and Gonzalez (2015, p. 4): "Cloud technology standardizes and pools IT resources and automates many of the maintenance tasks done manually today. [...] *Cloud* in this context refers to cloud computing architecture, encompassing both public

and private clouds". Public clouds are the conventional model in which "cloud service providers make available their computing infrastructure and products for general use by other enterprises and individuals" (Bisong, 2019, p. 3). Private clouds, on the contrary, are "intended for internal use only" Riti (2018, p. 22). Cloud technology provides options for storage, computing, applications, services, etc. Our focus is on public clouds and, in particular, the Google Cloud Platform.

Krishnan and Gonzalez (2015, p. 4) name three technical benefits of using a public cloud. (1) Uptime: public cloud providers offer uptime of over 99.9 percent. GCP even claims a 99.99 percent uptime. (2) Resource utilization: the resources (CPU, network traffic, etc.) that are required often demonstrate large variability. Cloud providers rely on large numbers of users and economies of scale. Cloud users can deploy or release resources instantly through the pay-per-use business models of the providers, leading to efficient resource utilization. (3) Expertise: cloud service providers provide knowledge and experience in the form of their system and network administrators, hardware maintenance personnel, available documentation, etc. By using the cloud you do not need experts on-site, or at least less of them.

Most often, the term 'cloud computing' is used to describe the wider range of cloud technologies that are referred to above. According to Bisong (2019, p. 3): "Cloud computing is the practice where computing services such as storage options, processing units, and networking capabilities are exposed for consumption by users over the Internet (the cloud)." According to Krishnan and Gonzalez (2015, p. 3): "Cloud computing is about abstracting the computing infrastructure and other associated resources and offering them as service, usually on a pay-per-use basis, over the Internet." It is thus good to realise that the term cloud computing often includes (more restrictive concepts such as) cloud storage and cloud infrastructure as well.

For smaller research projects, like mine, it is an enormous advantage that the cloud takes care of all hardware and maintenance issues. Similarly, I do not have to worry about software used behind my user interface. It is good to note that this corresponds to the standard Cloud Computing (Service) Models: Software as a service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). For additional information on these service models, see for instance: Riti (2018, pp. 20-21) and (Bisong, 2019, pp. 5-6). Last but not least, I want to stress again that (instant) 'scalability' – the buzzword of cloud computing – is one of the most important advantages of cloud computing. Uncertain futures for research projects, or its spinoffs, can easily be incorporated when the need arises.

## 2.2 Cloud service providers

The three main competitors amongst the cloud service providers are Amazon Web Services, Microsoft Azure, and Google Cloud Platform (Riti, 2018, p. 19). In 2006, Amazon Web Services (AWS) was the first to start offering cloud services (Garfinkel, 2007). In 2010, Microsoft Azure became publicly available (Hill et al., 2010). In 2011, Google Cloud Platform was first offered to the public (Riti, 2018, p. 23). Its predecessor, Google App Engine, dates back to 2008 (Qian et al., 2009).

The market shares of the cloud service providers are not entirely clear as Microsoft does not report revenue for Microsoft Azure separately (Vellante, 2020). It is clear, however, that AWS is the largest player by far (inVerita, 2020; Vellante, 2020). The most cited market shares are reported by Canalys. For the second quarter of 2020 they report a market share of 31 percent for AWS, 20 percent for Microsoft Azure, and 6 percent for Google Cloud (Canalys, 2020). Of the remaining providers Alibaba Cloud is the largest.

It is not the purpose of this document to compare strengths and weaknesses of the providers. Still, AWS is generally seen as the "the world's most comprehensive and widely adopted public cloud platform" (inVerita, 2020). It has been around the longest and it is considered the most mature. GCP is still seen as a "newcomer" (inVerita, 2020) and "challenger" (Vellante, 2020). The infrastructure and analytics of the Google Cloud Platform are seen as its strengths (Vellante, 2020). This is hardly surprising as Google uses the same infrastructure for its own services (including, for instance, Google Search, YouTube, and Gmail).

The choice for a particular cloud service provider, however, is not necessarily a technological one. Early adopters might simply prefer to continue working with the cloud they started with. Similarly, the accumulated expertise of Microsoft Azure at our university is a valid argument for the use of Azure. Our Funda project illustrates the situation where there is no choice; the data in our project originate from Google Analytics and are stored in the Google Cloud by Funda. Using an alternative cloud service provider would, therefore, make little or no sense at all.

# 3 Privacy and data protection

Privacy and data protection are more important than ever. This is illustrated, for instance, by the General Data Protection Regulation (GDPR), which was introduced in April 2016. The GDPR (in Dutch: *AVG, Algemene verordening gegevensbescherming*) protects the data and privacy of individ-

uals in Europe. The GDPR is enforced since May 2018 (Steegmans and de Bruin, 2019).

According to Google Cloud: "Compliance with the GDPR is a top priority". In practice, Google Cloud facilitates compliance with the GDPR by providing guidelines, implementation options and taking care of the technological side to the extent possible. Still, in the end, compliance with the GDPR remains the responsibility of the customer (i.e., data owner, data controller or researcher): "Data controllers are responsible for implementing appropriate technical and organisational measures to ensure and demonstrate that any data processing is performed in compliance with the GDPR" (Google Cloud, 2018, p. 3).

Google Cloud's data protection commitments include, among other things, data processing according to clients' instructions only, employee confidentiality commitments (mandatory confidentiality agreements and training), encryption for data in transit and at rest,[1] layered protection of hardware and software, and vulnerability management (continuously scanning for software vulnerabilities and addressing those) (Google Cloud, 2018, pp. 5-7).

Apart from the aforementioned default protection services, other services need user implementation/involvement as is the case for reducing the risk of unauthorised access by setting up 2-step verification (Google Cloud, 2018, p. 9). This will be discussed in detail in Section 4. Similarly, Google Cloud Identity and Access Management (Cloud IAM) provides the opportunity to manage detailed permissions (and roles) when more people are involved in a GCP project. Although the GDPR does not require personal data to be stored within the EU, Google Cloud does provide the option to choose the location where your data is stored. All in all, the cloud user is in charge when choosing the level of data protection and privacy.

# 4 Securing your Google account

If you are going to use the Google Cloud Platform, make sure to increase the security of your Google account. We highly recommend the use of a hardware authenticator, a YubiKey to be precise. Nevertheless, we go over

[1] The latest development, which I only mention as an illustration, is confidential computing (see Rashid (2020) for a short overview). Confidential computing allows customers to encrypt data that is in use. This real-time encryption guarantees, as stressed by Google Cloud, that users' "data is not being exposed to cloud providers or their own insiders" (Potti and Manor, 2020). Confidential computing is available in beta on the Google Cloud Platform since July 2020. Similarly, Microsoft has started offering Azure confidential computing recently too.

all available options and go into some of the pros and the cons. Before we do so we touch upon some security terminology.

## 4.1 Two-factor authentication (2FA)

We want to stress that it is crucial to set up two-factor authentication in order to secure your account. Multi-factor authentication is "an authentication method in which a computer user is granted access only after successfully presenting two or more pieces of evidence (or factors) to an authentication mechanism: knowledge (something the user and only the user knows), possession (something the user and only the user has), and inherence (something the user and only the user is)." Two-factor authentication, thus, adds a *second* authentication method to the use of your password for an account log-in.

## 4.2 SMS, phone calls, Google Authenticator

There are various option for 'Google Authentication in two steps'. For instance, Google can provide you with a verification code through SMS or a phone call. However, these are considered the weakest options for 2FA (Klosowski, 2020).

A higher level of security is provided by the (free) Google Authenticator app, which is available in Google Play and the Apple App Store. Like other authenticator apps, Google Authenticator is installed on your phone. The authenticator app generates a one-time password (OTP) that is the second method of authentication. The Google Authenticator app also works when your phone is *not* connected to a network.

## 4.3 USB security key: YubiKey

The last option is the use of a hardware authenticator. As noted in PCWorld by Ian Paul (2019): "The absolute safest way to lock down your accounts with two-factor authentication is to use a physical security key." The security keys that I talk about here connect to the USB port of your computer. For use in combination with your phone you need a security key that also supports NFC (near-field communication); these should work with both Android phones and iPhones although checking the model particularities seems worthwhile. It is possible to register your computer (or phone) so that you do not have to identify yourself again when you access your account through a familiar device. I, however, prefer not to use that option.

Even though Google has its own security key, the Titan Security Key, a Yubico YubiKey is to be preferred. The Titan Security Key drawbacks include that it supports only FIDO and FIDO2F (Paul, 2019). Apart from that, it is not available in the Netherlands as it cannot be ordered in the Google Store. More information on the Yubico YubiKey 5 Series can be found on the Yubico website. Again, consider the NFC option when you want to use the key for your phone too.

Finally, it is highly recommended that you buy (and register) two security keys as losing your key may result is losing access to your account altogether. Back-up telephone numbers and back-up codes may provide an alternative to regain access but a second security key is by far the safest and easiest way to recover access in case of loss of a key. Note that the two-factor authentication – particularly the security of the hardware key – provides an additional advantage of working in the cloud!

# 5 Google Cloud Research Credits

Researchers can individually apply for Google Cloud research credits, which can be used for most services on the Google Cloud Platform. The main exception are Maps Platform services. The credits expire 365 days after the coupon has been redeemed – research credits have to be redeemed within 60 days after receiving them – although it is possible to request an extension. Extensions beyond 365 days in total are not possible.

Currently, there are two research grant programmes: the COVID-19 Research Grant and the (standard) Academic Research Grant. Eligibility for the academic research grants is limited to faculty researchers and PhD candidates from accredited academic institutions from certain countries (among which the Netherlands). Most importantly, Utrecht University is one of the eligible institutions.

The qualifying criteria and the grants have changed significantly over the past two years. Most importantly, the research grants have become smaller and the credits expire quicker than before. Currently (September 2020), the maximum in GCP credits that may be requested by academic staff is 5,000 USD while the maximum for PhD candidates is 1,000 USD in GCP credits. It seems that the standard validity period is going to decrease – again – to 180 days (extensions to in total 365 days remain possible). According to Google Cloud employees the lower amounts and shorter standard periods increase the efficiency of the total research credits that they are allowed to distribute. On the Google Cloud Faculty forum (see below), this was referred to as "Google's low usage initiative".

Finally, it is good to realise that projects are eligible only once. Researchers can apply with a different project at a later time; however, a proposal can only be submitted after the previous project has finished. The review and decision process should take a maximum of 4-6 weeks. Nevertheless, the decision for my last proposal took longer. All in all, back to back funding is not possible with GCP research credits. For the latest information, and the application form, one can check the GCP research credits website.

# 6 Online courses

The Google Cloud (Google Education) offers a wide range of training and certification opportunities. In some cases it might be worthwhile to invest in obtaining certification, in other cases a training will be more than sufficient.

## 6.1 Training and certification

The Google Cloud offers a Career Readiness Programme. The starting point is the GCP Associate Cloud Engineer certification. For faculty the fees are covered by the Google Cloud, students still have to pay 50 percent of the exam fee (62.50 USD). Additional certification is available in the Professional certification track and the User certification track.

Apart from the most formal certification the Google Cloud offers the possibility to obtain Google Cloud Skill Badges by following training courses. The courses are divided into three separate tracks: (1) Cloud infrastructure, (2) Data, machine learning, and AI, and (3) Application Development. The courses can be accessed through Qwiklabs; credits to freely do so are available for academic staff. Details will be provided below.

Independent of the above, Grow with Google offers Google IT Support certifications, such as the Google IT Automation with Python Professional certificate.

## 6.2 Access to Coursera and Qwiklabs

The courses and labs offered by Google Education (for details see the programme overview) are available through Coursera and Qwiklabs. One can apply for (free) credits to access Google's Coursera and Qwiklabs courses. Faculty, students, and IT administrators are eligible to apply.

Applications are generally processed within two weeks although I have previously received notices that the pandemic has increased the processing time. Nevertheless, in my case it took less than five hours before I received

both 200 Qwiklabs Credits (representing a value of 200 USD) and access to Google's Coursera courses. The credits expire after one year. Questions can be sent to CloudEduTraining@google.com although, once again, an answer may take a while.

To get familiar with the possibilities I did the following Qwiklab 'courses', which all consist of a series of labs:

- BigQuery Basics for Data Analysts (5 modules, 4 hours, 2 Credits).
- Insights from Data with BigQuery (6 modules, 5 hours, 11 Credits).
- BigQuery For Data Analysis (12 modules, 1 day, 35 Credits).

When starting a Qwiklab lab you are provided with "temporary credentials that you use to sign in and access Google Cloud for the duration of the lab." That gives you (temporary) access to all required cloud platform services that you need to finish the tasks. It is important to note that you must complete all activities within the allocated time: between 40 and 90 minutes depending on the lab. When you *really* want to understand all code, you might want to read the Lab's material before you actually start the lab (and time starts running).

Some of the courses and quests can only be completed by doing a final 'Challenge Lab'. In a Challenge Lab, contrary to the other modules, you are not provided with step-by-step instructions. Instead you are provided with a scenario and tasks related to it. An automated scoring system checks whether you have completed a task. To pass you need to score 100 percent within the time provided. I did the Insights from Data with BigQuery: Challenge Lab. The problem here was that a daily updated Covid-19 data set was used, and that over time the data set no longer corresponded to the queries of the automated scoring system. In the end, I got enough answers wrong to pass the Challenge Lab successfully... That took, however, more time and credits than hoped (and expected). Google Cloud Training Support has promised me to correct the Challenge Lab.

I have also done a (more formal) Coursera course, Exploring and Preparing your Data with BigQuery, which is part of the specialization 'From Data to Insights with Google Cloud Platform'. It is more formal because, instead of a digital badge, you are provided with a (verifiable) certificate. The Coursera courses have a more traditional classroom setting; that is, an instructor generally demonstrates how things are done in short tutorial video's. Proficiency has to be shown through short quizzes and a (limited) number of practice labs. The "hands-on labs" are offered by "a third-party tool", which – as you might have guessed – refers to Qwiklab. The Coursera sandbox account,

which also gives you access to the Qwiklab components, is independent of the direct Qwiklab access. This meant in my case that I had to re-do a lab in order to also get credits in Coursera. I would recommend doing the Qwiklab training *or* the Coursera course(s) as there is overlab within the topics. It seems that the most hands-on approach is found in Qwiklab whereas the broader perspective is provided more extensively in Coursera.

# 7 Google Cloud Faculty (forum)

The Google Cloud Faculty forum has turned out to be a valuable source of information as questions are posted, and answered, on a daily basis. The Google Cloud Team and expert users are highly active on the forum. Eligibility is limited to verified researchers that have received Research Credits and teachers that use Teaching Credits in their classes and/or labs. I strongly recommend joining this community. One should receive an email invitation to join after being granted the credits. Alternatively, you can request access via Connect with fellow faculty and researchers by using your UU mail address. Both ways will make you a member of the Google Cloud Faculty Community on Google Groups, which will make it possible to access the GCP Faculty forum.

The forum's focus is on using the Google Cloud in teaching so many question come from the field of Computer Science. However, I recommend other researchers to join the community too. The topics covered by the forum range from error messages when setting up a VM to best practices when redeeming GCP research credits coupons. Apart from that information on the Virtual Office Hours are posted, as well as occasional links of recordings of live sessions.

# 8 Google Cloud conferences

Google Cloud organize a variety of conferences where it is possible to "explore new ideas, learn from industry experts, and engage with peers" (Carty, 2017). The conferences include: Google Cloud Next, Google Cloud Summit and Faculty Institute. Most of the conferences are held annually. Google Cloud Next and Google Cloud Summit are basically event series that are organised in major cities worldwide throughout the year. Google Cloud Summit, for instance, visited Amsterdam in 2018 and 2019 (although the Google Cloud Summit Amsterdam 2018 was actually held in Zaandam).

Google Cloud Next and Google Cloud Summit bring together customers,

developers, engineers and Google partners (Carty, 2017). It is good to note that these events have a marketing focus. Although keynotes, breakout sessions and learning opportunities can be very interesting getting customers to use the Google Cloud Platform, and GCP partner products, remains the main goal. Due to the pandemic Google Cloud Next '20 has gone digital: Next OnAir EMEA. The main advantage is that past sessions can be seen on demand as well.

For academics, Google Cloud Faculty Institute is – undoubtedly – most relevant. The Faculty Institute London '18 had a particularly appealing line-up. Sadly, I did not get closer to London than the waiting list. The Faculty Institute 2020 has also gone digital. It is my personal impression that this has negatively affected the content of the sessions (as well as their number). Nevertheless, when you are not entirely familiar with the GCP, it provides sessions that can help getting your project started.

# 9 Project funding

Cloud service providers have a pay-per-use business model, implying that you will need funds for your GCP research project. You may simply pay for these services from your research budget but alternatives are available. Each new Google Cloud customer can, for instance, use 300 USD credit and free products. This should get you started. There is also the option to apply for the (low-barrier) Google Cloud Research Credits, which have been discussed in Section 5.

Apart from the above, Google offers research grants that cover more than just GCP costs. Best known was the Faculty Research Awards Programme (2005-2019). However, this programme has recently been discontinued. It has been replaced by the Research Scholar Programme, which shall be discussed in a moment. In the last part of this section I will describe another Google research funding initiative, the Google PhD Fellowship Programme.

Related to funds and costs, there is a practical piece of advice: when you start working with the Google Cloud Platform, make sure that you set budgets and budget alerts. Running unintended queries, for instance, may turn costly quickly. Thus, regardless of whether you use research credits or money, make sure that you do not blow your budget (or more!) in one instance. At the minimum, keep track of the costs that you incur in your project through the billing pages, which provide real time information.

## 9.1 Research Scholar Programme

The Research Scholar Programme focuses on early-career faculty, i.e., researchers that received their doctorate within the last seven years. To be eligible researchers must be "doing impactful research in fields relevant to Google" (Johnson, 2020). This implies a (broad) Computer Science perspective; the full list of "cutting-edge research categories" can be found online. Apart from the categories in the list, social science research may qualify as well.

The programme is meant to encourage "new collaborations" and "long term relationships" (Johnson, 2020). The Research Scholar programme provides unrestricted gifts up to 60,000 USD. The grant is intended to "support the professor's research efforts" and should be used in the academic year in which it is provided. It is "not intended for overhead or indirect costs", which implies that the full grant amount should be spent on direct project costs; the university is thus not allowed to withhold overhead or indirect costs from the provided grant. Still, what the grant is intended for remains somewhat vague, especially because the FAQ mentions that the size of the grant is "depending on the cost of student tuition on a regional basis" thereby – if correct – implying student involvement.

The Fall 2020 application round is open from November 5, 2020 to December 3, 2020. At this time there is no information on when the next round will be. However, I expect it to be Fall 2021. As it has already been made public that: "Faculty may apply up to a maximum of 3 times within the 7 years they received their PhD". It seems that the (new) Research Scholar Programme is going to be around for many more years.

In short, the application to the Research Scholar Programme is open to (permanent) assistant professors, associate professors, etc. Utrecht University qualifies as an eligible institution. Applicants should have received their doctorate within the last seven years (i.e., 2013 or later for the Fall 2020 round). And, applicants can submit no more than one application per round (and maximally three over the years). The full list of eligibility requirements can be found on the Google Research Scholar Programme website. Furthermore, make sure that you read the FAQ if you intend to apply. The FAQ includes a suggested 'Proposal Format' as well as an 'Open advice to Google Research Awards proposal writers'. It seems wise to take advantage of both.

## 9.2 Google PhD Fellowship

Google Research also provides PhD Fellowships to graduate students in Computer Science and some related field. This initiative is known as the Google

PhD Fellowship Programme. Students, however, cannot apply directly to the programme. They have to be nominated by the university instead. Utrecht University and its students are eligible. Up to three students can be nominated: "If more than two are nominated, then [...] the additional nominee must self-identify as a woman." This is part of an initiative to provide opportunities for underrepresented groups.

It is only possible to nominate graduate students that are enrolled full-time at the university. Furthermore, students should have finished all courses before the PhD Fellowship begins (Fall of the award year, i.e., beginning of the new academic year). Applications must be submitted by a graduate school official, department chair or similar.

The funds are paid to the university as unrestricted gifts. They are meant to "cover the student's expenses and stipend as appropriate" and cannot be used for overhead. Although the programme is global, differences exist between regions. For Utrecht University the European criteria apply, implying that the Fellowship is up to three years. The yearly bursary can be used "towards stipend/salary, health care, social benefits, tuition and fees, conference travel and personal computing equipment".

This year's (European) application window has already closed. The application window for the PhD Fellowship 2022 opens September 1, 2021 and closes September 30, 2021. Awardees will be notified before the end of January 2022.

# 10 GCP research in Utrecht

Most of the information that is bundled in this document originates, directly and indirectly, from my housing market research done in the Google Cloud. More on the technical side of the project can be found, for instance, in Steegmans and de Bruin (2019). I have, however, learnt that there have been a few more Google Cloud research-related projects at Utrecht University. An interesting initiative that I have come across, for instance, was the "Centre for Digital Humanities online seminars: Three Jupyter Notebooks" of June 19, 2020 (followed by a hands-on session a week later). In this seminar, given by Thomas Smits and Ruben Ros, Jupyter notebooks using the Vision API of the Google Cloud were introduced.

Another interesting GCP initiative has been done by the Dutch cardiovascular research consortium CVON-AI, which aim is to build "a generic AI cloud platform to catalyse the application of AI by Dutch [cardiovascular] scientists". Among the 12 partners involved in the consortium are researchers of the University Medical Center Utrecht. The Utrecht researchers co-authored

an article in which the Google Cloud Platform was used "to train and test a deep learning model" (Benjamins et al., 2019, p. 417). This was done in a pilot experiment "using cardiac magnetic resonance data" (Benjamins et al., 2019, p. 414), thereby demonstrating feasibility of artificial intelligence applications in cardiovascular research. Admittedly, however, the research project seems driven by the University Medical Center Groningen.

Apart from the above-mentioned initiatives, the Department of Information and Computing Sciences is a recurring user of Google Cloud services. This is evidenced, for instance, by the use of the Google Cloud Platform in Master's theses. To name just a few: Lucas de Haas (2017; Master of Artificial Intelligence) used the Google Cloud Platform to train his Recurrent Neural Network (RNN) models, used for 'extractive summarization' (de Haas, 2017, p. 50). Alejandro Barredo Arrieta used the Google Cloud's Natural Language API for his thesis (Barredo Arrieta, 2018, p. 19).

Another thesis that grabbed my attention was "To Catch a Thief: Fraud detection with reliable machine learning" written by Rob de Wit (2019; Master of Business Informatics). He studies applicability of Reliable Machine Learning techniques for fraud detection using data of Bol.com. Although he does not work in the Google Cloud Platform himself, "the Google Cloud platform [...] is part of the existing IT infrastructure at bol.com" (de Wit, 2019, p. 21). It is almost inevitable that the eCommerce data that he uses was exported from the GCP by Bol.com. Operationalization of the thesis' findings would thus have to be done in the Google Cloud too.

# 11 Final words

I would like to end with some advice that did not fit elsewhere in this document. The first piece of advice is the possibility to reach out to Google researchers. Before the Covid-19 pandemic I would at least have suggested to consider Google's (academic) conferences. However, attending those physically is currently not possible.

A practical alternative is the Google Research People Directory, where information on research area's and publications of the researchers can be found. Research area's include, for instance, 'Algorithms and Theory' and 'Data Mining and Modeling'. Nevertheless, it also includes categories such as 'Economics and Electronic Commerce', 'Health & Bioscience' and 'Education Innovation'. Until recently 'Google sponsors' – Google employees supporting your research – were a requirement for, for instance, the Faculty Research Awards Programme. However, this requirement seems to be dropped for the Research Scholar Programme.

Finally, in this 'non-technical guide' I have tried to give an overview of the most fundamental aspects of starting a Google Cloud Platform research project. As it turns out, there is a lot to know about. And although this document only scrapes the surface, I truly hope that it helps to get your GCP research project started. Speaking for myself, I know that it would have been beneficial for me if I would have known all of this at the beginning of my first GCP project.

# Acknowledgements

This report has been made possible through financial support of Utrecht University's Research IT innovation programme. A special thanks goes to Martin Schuckman, Wolter Hassink and Jonathan de Bruin for their valuable comments.

# References


Barredo Arrieta, A., 2018. Association rules for grammar inference. Master's thesis. Faculty of Science, Utrecht University. http://dspace.library.uu.nl/handle/1874/368147.

Benjamins, J., van Leeuwen, K., Hofstra, L., Rienstra, M., Appelman, Y., Nijhof, W., Verlaat, B., Everts, I., den Ruijter, H., Isgum, I., et al., 2019. Enhancing cardiovascular artificial intelligence (AI) research in the netherlands: CVON-AI consortium. Netherlands Heart Journal 27, 414–425.

Bisong, E., 2019. Building Machine Learning and Deep Learning Models on Google Cloud Platform. Springer.

Canalys, 2020. Global cloud services market Q2 2020. July 30, 2020. https://canalys.com/static/press_release/2020/Canalys-cloudq220.pdf (last visited on September 29, 2020).

Carty, T., 2017. Google's cloud summit: How google is changing the way we utilize the cloud. ETech 7 Blog. https://blog.etech7.com/author/taylor-carty (last visited on October 8, 2020).

Garfinkel, S.L., 2007. An Evaluation of Amazon's Grid Computing Services: EC2, S3 and SQS. Technical Report TR-08-07.

Google Cloud, 2018. General Data Protection Regulation (GDPR). Google Cloud Whitepaper May 2018. https://cloud.google.com/security/gdpr/resource-center/pdf/googlecloud_gdpr_whitepaper_618.pdf (last visited on October 7, 2020).

de Haas, L., 2017. Extractive summarization using sentence embeddings: Automatic summarization of news articles at Blendle. Master's thesis. Department of Computing and Information Sciences, Utrecht University. https://dspace.library.uu.nl/handle/1874/357917.

Hill, Z., Li, J., Mao, M., Ruiz-Alvarez, A., Humphrey, M., 2010. Early observations on the performance of windows azure, in: Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, pp. 367–376.

inVerita, 2020. AWS vs. Azure vs. Google: Detailed Cloud Comparison. Last updated: May 21, 2020. https://levelup.gitconnected.com/aws-vs-azure-vs-google-detailed-cloud-comparison-b075a35fc8b8 (last visited on September 29, 2020).

Johnson, M., 2020. Exploring new ways to support faculty research. Google AI Blog March 26, 2020. https://ai.googleblog.com/2020/03/exploring-new-ways-to-support-faculty.html (last visited on October 9, 2020).

Klosowski, T., 2020. The best two-factor authentication app. Wirecutter. Updated September 15, 2020. https://www.nytimes.com/wirecutter/reviews/best-two-factor-authentication-app/ (last visited on October 7, 2020).

Krishnan, S., Gonzalez, J.L.U., 2015. Building your next big thing with google cloud platform: A guide for developers and enterprise architects. Springer.

Paul, I., 2019. What is two-factor authentication, and which 2fa solutions are best? PCWorld. Updated June 5, 2019. https://www.pcworld.com/article/3225913/ (last visited on October 7, 2020).

Potti, S., Manor, E., 2020. Expanding google cloud's confidential computing portfolio. Google Cloud Product News September 8, 2020. https://cloud.google.com/blog/products/identity-security/expanding-google-clouds-confidential-computing-portfolio (last visited on October 8, 2020).

Qian, L., Luo, Z., Du, Y., Guo, L., 2009. Cloud computing: An overview, in: IEEE International Conference on Cloud Computing, pp. 626–631.

Rashid, F.Y., 2020. The rise of confidential computing: Big tech companies are adopting a new security model to protect data while it's in use. IEEE Spectrum 57, 8–9.

Riti, P., 2018. Pro DevOps with Google Cloud Platform: With Docker, Jenkins, and Kubernetes. Springer.

Steegmans, J., de Bruin, J., 2019. Online housing search and gravity models. USE Working Paper Series 19-17. https://www.uu.nl/sites/default/files/rebo_use_wp_2019_1917.pdf.

Vellante, D., 2020. Breaking analysis: How AWS, Azure and GCP compete for cloud leadership in the 2020's. February 9, 2020. https://wikibon.com/ (last visited on September 29, 2020).

de Wit, R.C., 2019. To Catch a Thief: Fraud detection with reliable machine learning. Master's thesis. Department of Computing and Information

Sciences, Utrecht University. http://dspace.library.uu.nl/handle/1874/381895.