

Informatik

Geodatenprozessierung mit Budget Instanzen

Semesterarbeit



Abbildung 1: Eine Tour mit Budget Instanzen [Web10]

Departement:	Informatik
Kurs:	CAS CLD FS20 – Cloud Computing
Autor:	Tobias Reber
Experte:	Jörg Thomann
Ansprechpartner:	Christoph Böcklin
Datum:	14. 08. 2020

Perplexity
is the beginning of knowledge.

- Kahlil Gibran

[AD19, S. 33]

1 Management Summary

In der Kürze liegt die Würze

Inhaltsverzeichnis

1 Management Summary	ii
2 Einführung	1
2.1 GIS - Geographical Information Systems	1
3 Vorgehen	2
3.1 Arbeitsmethodik	2
3.2 Projektplan	2
4 Ausgangslage	3
4.1 swisstopo bei AWS	3
4.2 Publikation von Geodaten	3
4.3 Web-Services	3
4.4 Exkurs 3D Daten	4
5 Der Use Case	5
5.1 Problemstellung	5
5.1.1 Spot Instanzen	5
5.2 Ist-Zustand der 3D Datenpublikation	5
5.2.1 Prozess der Publikation	5
5.2.2 Aufwand der Prozessierung	6
5.2.3 Technische Komponenten	7
6 Architektur	8
6.1 Analyse des Ist-Zustandes	8
6.1.1 Bereitstellung der Rohdaten und Sicherstellung dessen Qualität	8
6.1.2 Bereitstellung der Infrastruktur für die Prozessierung	8
6.2 Bewertungskriterien	8
6.3 Architekturentscheid	8
6.3.1 Angebote für die Prozessierung	9
7 Prototyp	10
7.1 Realisierung	10
7.1.1 Da	10
8 Evaluation	10
8.1 Erfahrungen	10
8.2 Wirtschaftlichkeit	10
8.3 Kritische Punkte	10
9 Ausblick	11
10 Schlusswort	11
Literaturverzeichnis	12
A Fachbegriffe und Abkürzungen	I
B Konfigurationen	II
C Kennenlernen von AWS	II
D Kanban	II

E	Projektplan	III
F	Konfigurationen und Kommandos	III
F.1	EFS auf EC2-Instanz mounten	III
F.2	Von der SPOT Instanz aus abfragen, was ihr Status ist	VI
G	Für die Semesterarbeit verwendete Software	VII

Abbildungsverzeichnis

1	Eine Tour mit Budget Instanzen [Web10]	1
2	Internetkarte des Bundes <i>map.geo.admin.ch</i> . Hier ein Ausschnitt der Saane bei Kleinbödingen, das Luftaufnahmen von 1946 mit heute vergleicht.	1
3	Im Viewer werden zurzeit Gebäude, Bäume, Seilbahnen, Namen und das Terrain dargestellt. Um aktuell zu bleiben, müssen diese 3D Daten regelmässig nachgeführt werden.	4
4	Arbeitsschritte, die es braucht, um die 3D Daten zu Publizieren.	6
5	Klassisches Kanban auf <i>github.com</i>	II
6	Projektplan. Die orangen Meilensteine wurden von der BFH vorgegeben.	III

2 Einführung

Ich arbeite als GIS-Spezialist bei der swisstopo¹, dem Bundesamt für Landestopografie, in Wabern. Wir machen Karten. Unser Team macht Internetkarten - wie Google Maps², jedoch von der Schweiz für die Schweiz; und für alle anderen auch. Unsere Internetkarte, der Viewer, erfreut sich relativ grosser Beliebtheit und beinhaltet ca. 800 Themen wie Wanderwege, Solarkataster und Luftfahrthindernisse. Lieber Leser³, falls dir map.geo.admin.ch noch kein Begriff sein sollte, kann ich dir wärmstens empfehlen, darin zu schmökern. Es gibt viel zu entdecken und es ist gratis - ein Service Public.



Abbildung 2: Internetkarte des Bundes *map.geo.admin.ch*. Hier ein Ausschnitt der Saane bei Kleinbödingen, das Luftaufnahmen von 1946 mit heute vergleicht.

2.1 GIS - Geographical Information Systems

Wie erwähnt, arbeite ich als *GIS-Spezialist*. Wobei mir der Titel *Geo-Informatiker* besser gefällt: weil er die Begriffe *Geografie* und *Informatik* vereint. *Geografie* kommt aus dem Griechischen und bedeutet Erdbeschreibung [SE04, S. 14]. *Informatik* ist die Wissenschaft von der systematischen Darstellung, Speicherung, Verarbeitung und Übertragung von Informationen [10b].

GIS ist ein Akronym für Geographical Information Systems. Es bedeutet im engsten Sinn eine Ansammlung von Computerprogrammen, die zur Bearbeitung von Karten und Geodaten verwendet werden. Geodaten sind nichts weiter als Daten mit einem räumlichen Bezug⁴. In einem weiteren Sinn deckt der Begriff GIS ein ganzes Fachgebiet ab, das sich mit Karten und Geodaten auskennt. Es ist also nicht nur ein Werkzeug, sondern ein Fachgebiet, das Kenntnisse über Datensammlung, Speicherung, Analyse und Darstellung innerhalb von vielen verschiedenen Themen mit einem räumlichen Bezug abdeckt. Typische Geodaten sind digitale Karten, Inventare und Register von Parzellen, Umweltfaktoren, Grenzen, Entwicklung, Planung etc., die einen räumlichen Bezug haben und dadurch in einem geografischen Zusammenhang analysiert und dargestellt werden können [BJB06, S. 15].

Es ist zwar weit von der klassischen Geografie zu Zeiten Alexanders von Humboldt⁵ entfernt, aber es liegt auf der Hand, dass auch die Geografie Cloud Computing nutzt.

¹www.swisstopo.ch

²maps.google.com

³Im vorliegenden Dokument wird durchwegs der männliche Singular (Leser, Benutzer) als Ansprache verwendet. Diese Ansprache bezieht sich auf beide Geschlechter sowie gegebenenfalls mehrere Personen. Sie dient lediglich der leichteren Lesbarkeit der Semesterarbeit

⁴mit Koordinaten (Nord/Ost, x/y)

⁵Ein Forschungsreisender des 19. Jh. mit einem weit über Europa hinausreichenden Wirkungsfeld. Mehrjährige Forschungsreisen führten ihn nach Lateinamerika, USA und nach Zentralasien [Keh05]

3 Vorgehen

In diesem Kapitel wird das Vorgehen beschrieben, wie die Arbeit geplant und realisiert wurde.

3.1 Arbeitsmethodik

Da die Arbeit nur von einem Autor umgesetzt wurde, musste bezüglich Arbeitsmethodik nur wenig definiert und koordiniert werden.

Um zu starten, hat dem Autor geholfen, dass der Experte Jörg Thomann bereits anfangs Sommer (16. Juli) einen ersten Besprechungstermin angesetzt hat. Um eine Diskussionsbasis für diesen Termin zu haben, musste sich der Autor ernsthaft mit der Semesterarbeit auseinandersetzen. Dazu hat er aus einer BFH Vorlage ⁶ ein Gerüst der Arbeit erstellt und stichwortartig abgefüllt. Nach gründlicher Diskussion mit dem Experten wurde dieses Gerüst etwas angepasst und diente von da an als Basis für die Arbeit; wo sich die Kapitel zunehmend von losen Notizen zum finalen Zustand festigten.

Um die Übersicht nicht zu verlieren, wurde beschlossen, nach Kanban zu arbeiten, respektive ein Kanban Board zu verwenden. Weil für die Arbeit sowieso Repository auf github.com⁷ angelegt wurde, konnte dort auch gleich ein Kanban erstellt werden. Das Kanban Board wurde vor allem auch dafür genutzt, fortlaufend neue kleine Aufgaben aufzulisten und darauf zu achten, die Menge der angefangenen Arbeit zu begrenzen.

Dadurch konnte im Alltag (wenn gerade etwas Zeit zur Verfügung stand) eine solche Aufgabe genommen und erledigt werden. Es wurde nicht strikte nach Kanban gearbeitet, aber einige Prinzipien daraus wurden verwendet. Zum Beispiel wurde das erste Grundprinzip angewendet *"Beginne mit dem, was du gerade tust: In diesem Grundprinzip stecken zwei Dinge. Indem man mit dem beginnt, was man gerade tut, beendet man die aktuelle Arbeit, bevor etwas Neues begonnen wird. Genauso ist hier aber auch die Aussage enthalten, dass Kanban einfach eingeführt werden kann."*[10c]. Eine Momentaufnahme des Kanbans befindet sich im Anhang D.

3.2 Projektplan

Um den übergeordneten Rahmen der Arbeit nicht aus dem Fokus zu verlieren und um sich bewusst zu machen, wieviel Zeit generell bis zur Abgabe zur Verfügung steht, war es hilfreich, einen Projektplan mit Meilensteinen konsultieren zu können. Durch die Berner Fachhochschule waren die grundlegenden Meilensteine bereits vorgegeben. Der Projektplan befindet sich im Anhang E.

⁶Herzlichen Dank an Andreas Habegger und Lukas Studer fürs Bereitstellen der Vorlage

⁷Projekt auf github.com

4 Ausgangslage

Hier werden das Arbeitsumfeld und die Aufgaben beschrieben, aus denen sich die Problemstellung dieser Semesterarbeit ergeben hat.

4.1 swisstopo bei AWS

Es liegt auf der Hand, dass die swisstopo in ihrer Rolle als *Geoinformationszentrum* auf Cloud Computing setzt. Die swisstopo nutzt Cloud Computing mit AWS⁸ seit mehr als 10 Jahren für den Betrieb des Geoportal des Bundes.

"Mit der BGD⁹ unter AWS können wir derzeit ca. eine Million Internetbenutzer pro Monat bedienen. Dank AWS können wir die zur Zuordnung neuer Server benötigte Zeit erheblich verkürzen und unseren Fokus auf echte Kundenanforderungen verstärken." [Chr20].

4.2 Publikation von Geodaten

Wie bereits erwähnt, können auf dem Viewer ca. 800 Themen wie Wanderwege, Solarkataster und Luftfahrthindernisse angesehen werden. Unser Team publiziert diese Daten. Der Nachführungszyklus wie auch der Aufwand zur Aufbereitung der Daten fürs Web sind unterschiedlich. Einige Daten werden manuell aufwändig aufbereitet, andere stündlich automatisch nachgeführt.

4.3 Web-Services

Nebst der Publikation der Daten ist unser Team für den Betrieb und der Weiterentwicklung der Web-Services und des Viewers verantwortlich. Der ganze Technologie Stack wurde schon länger nicht mehr grundlegend erneuert. Zurzeit wird die gesamte Architektur analysiert und überarbeitet, um eine gestaffelte Migration auf eine neue Lösung zu ermöglichen. Einige Rahmenbedingungen dieser zukünftigen Architektur sind bereits klar: Das Geoportal des Bundes wird weiterhin in der AWS Cloud betrieben werden, die Migration wird vor allem über Microservices gestaffelt erfolgen, diese Services werden als Docker Container laufen, Amazon Elastic Kubernetes Service wird die Orchestrierung der Container übernehmen; und für Continuous Integration wird AWS Codebuild/Pipeline zum Einsatz kommen.

⁸Amazon Web Services

⁹Bundesgeodateninfrastruktur: Viewer und andere Services

4.4 Exkurs 3D Daten

Die swisstopo erfasst und aktualisiert Daten mit einem räumlichen Bezug. Diese Geodaten sind die Basis für die Ableitung in eine Vielzahl von Produkten, wie die Landeskarten 1:25'000. Nebst Karten gibt es die Produktpalette der Landschaftsmodelle. Diese geben die Objekte der Landschaft im flexiblen Vektorformat wieder. Sie bestehen aus thematischen Ebenen (Bsp. Gebäude). Jede Ebene umfasst georeferenzierte Punkt-, Linien-, Flächen- oder 3D-Objekte. Jedes Objekt enthält Attribute und Beziehungen [10d].

Zu den Landschaftsmodellen gehören Produkte wie swissTLM3D und swissBuildings3D. Im Viewer wird eine Auswahl von Themen aus eben diesen Landschaftsmodellen dargestellt: Zurzeit Gebäude, Bäume, Seilbahnen, Namen und das Terrain. Vor wenigen Jahren wurden diese 3D Daten mit einem grossen Effort medienwirksam publiziert.

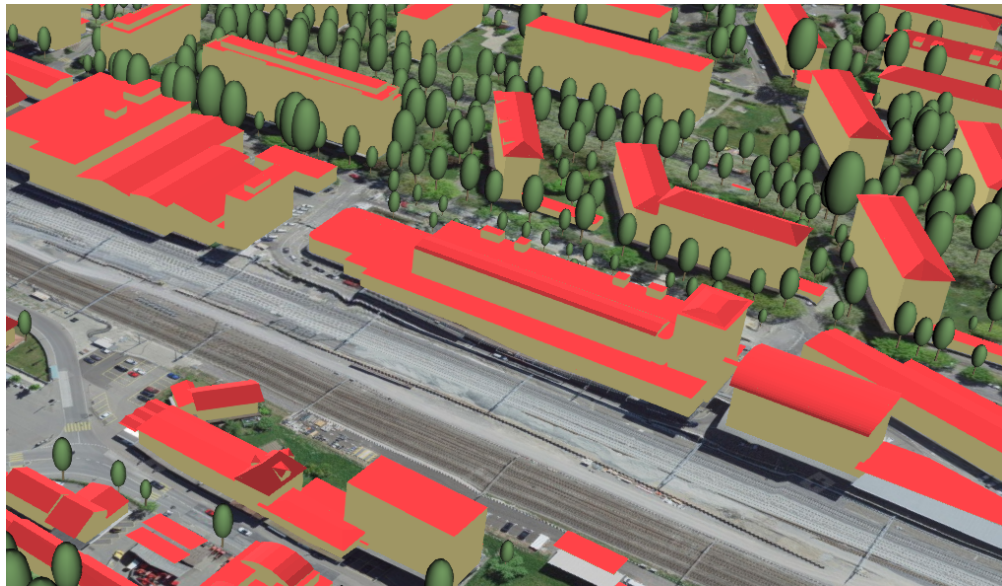


Abbildung 3: Im Viewer werden zurzeit Gebäude, Bäume, Seilbahnen, Namen und das Terrain dargestellt. Um aktuell zu bleiben, müssen diese 3D Daten regelmässig nachgeführt werden.

Seit dem Erfolg der Erstpublikation ist inzwischen etwas Zeit vergangen. Als die ersten Aktualisierungen der Daten anstanden, wurde den Beteiligten bewusst, dass sich diese nicht einfach so *auf Knopfdruck* realisieren lässt: Seit der Erstpublikation hat es personelle Wechsel gegeben und punkto Dokumentation und Automatisierungsgrad wurden Lücken identifiziert.

Es gibt immer gute Gründe für *technische Schulden*, wie in diesem Fall positive medienwirksame Reaktionen¹⁰. Aber früher oder später müssen diese abgebaut werden, weil es einen direkten Einfluss auf die Wartbarkeit des Produktes hat [10e].

¹⁰Wie Bsp. auf watson.ch oder Twitter [18]

5 Der Use Case

5.1 Problemstellung

Es ist der swisstopo schon lange ein Anliegen, dass die Publikationsprozesse von Geodaten optimiert werden sollen. Wann immer möglich, soll der Automatisierungsgrad erhöht werden.

Besonders aufwändig erweist sich zurzeit die Publikation von 3D Daten. Die manuelle Publikation der 3D Daten benötigt eigene Tools, die auf einem performanten und somit teuren Rechner laufen müssen. Ausserdem erfordert die Bereitstellung einen hohen Koordinationsaufwand zwischen der Infrastruktur und der Entwicklung. Dabei passiert es, dass Mängel in den 3D Daten erst nach beendeter Webpublikation bemerkt werden; und der ganze Publikationsprozess muss wieder von vorne gestartet werden. Auch dem Hersteller der 3D Daten (dem Bereich Topografie) wäre es ein Anliegen, wenn er diese Daten selbst automatisch publizieren und prüfen könnte.

Einerseits soll in dieser Arbeit mittels POC¹¹ aufgezeigt werden, wie der Automatisierungsgrad erhöht werden könnte. Andererseits soll untersucht werden, ob anstelle von EC2 direkt Spot Instanzen für die Prozessierung verwendet werden können und wie deren Einsatz aussehen könnte.

5.1.1 Spot Instanzen

TODO: Theorie - was sind Spot Instanzen und wie stehen diese im Verhältnis zu Problemstellung? Wie soll damit umgegangen werden, dass die Instanz jederzeit verschwinden könnte?

- ▶ REST Abfrage
- ▶ Cloud Watch
- ▶ Marke Eigenbau für diesen Use Case

5.2 Ist-Zustand der 3D Datenpublikation

5.2.1 Prozess der Publikation

Das Aufzeigen des Ist-Zustandes der 3D Datenpublikation soll helfen sich einen Überblick, eine Ausgangslage, zu verschaffen. Es bildet die Grundlage, um die Frage zu beantworten, welche Arbeiten erledigt werden müssen, um die 3D Daten im Web zu publizieren? Welche Arbeitsschritte könnten automatisiert werden?

Eine Datenpublikation läuft folgendermassen ab: Sobald der Auftrag für eine 3D Datenpublikation erteilt wurde¹², müssen zurzeit folgende Schritte, die in der Abbildung 4 referenziert sind, erledigt werden:

¹¹Prove of Concept

¹²Vom Bereich Topografie

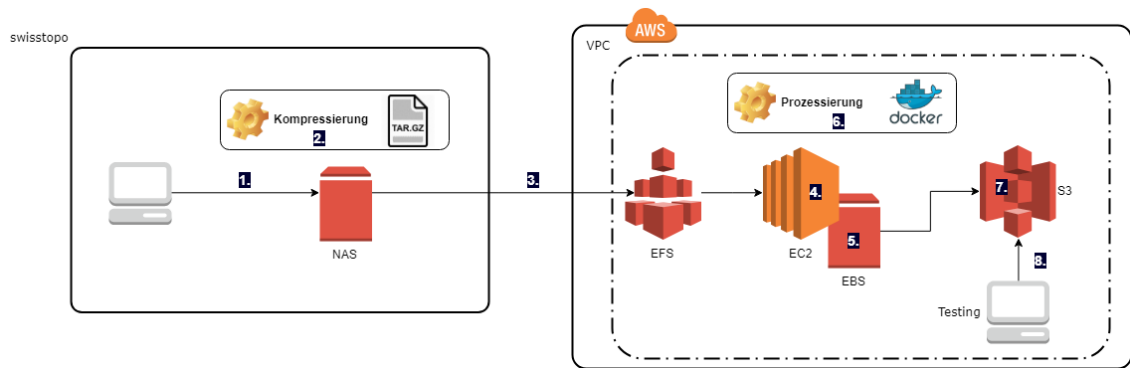


Abbildung 4: Arbeitsschritte, die es braucht, um die 3D Daten zu Publizieren.

1. Die Rohdaten¹³ werden vom Auftraggeber (Bereich Topografie) auf einem NAS bereitgestellt.
2. Da es sich um Millionen Dateien handelt, werden diese je Kartenblatt erst einmal gezippt, um so (weil bedeutend weniger Dateien) schneller kopiert werden zu können.
3. Kopieren der gezippten Dateien vom swisstopo Netzwerk in die Amazon Cloud (AWS VPC¹⁴) kopiert¹⁵
4. Parallel dazu wird die IT via Ticket gebeten einen Server mit dem entsprechenden Image¹⁶ bereitzustellen.
5. Kopieren und Entzippen der Rohdaten auf die gemountete Festplatte¹⁷ des Servers.
6. Die Daten auf dem Server via Docker prozessieren¹⁸.
7. Kopieren der Web-optimierten Daten auf S3.
8. Die Daten visualisieren, um inhaltlich testen zu können. Ein Codepen Projekt¹⁹, dass auf die 3D Tiles zugreift.

5.2.2 Aufwand der Prozessierung

Folgende Schritte sind besonders aufwendig:

- **Abb. 4, Schritt 2. und 5.:** Das Kopieren / Zippen (Packen und Entpacken) der Rohdaten nimmt lange.
- Es kommt immer mal wieder vor, dass Daten korrupt sind, was zu einer Nachlieferung führt, mit der Gefahr, dass es mit Versionen der Lieferung zu einem Durcheinander kommen könnte.
- **Abb. 4, Schritt 4.:** Unsere IT muss für die Prozessierung eine EC2 Instanz mit EFS bereitzustellen. Um laufende Kosten zu verringern, wird diese Instanz nach getaner Arbeit²⁰ wieder gestoppt. Falls mit den Daten etwas nicht in Ordnung ist, muss dieser Schritt von der IT wiederholt werden. Nebst der Bemühung der IT, muss auf der Instanz selber anschliessend das eine und andere manuell installiert und konfiguriert werden.

¹³Das Format der Rohdaten ist KML/COLLADA

¹⁴AWS: Amazon Web Services, VPC: Virtual Private Network. Ein Kopieren

¹⁵Via *rsync*

¹⁶Eine EC2 Instanz eine *m4.10xlarge* aus einem bereits vorhandenes AMI

¹⁷Ein EBS Volume

¹⁸Umwandeln ist das Web-Format *Cesium3DTiles*

¹⁹SaaS: Eine Webseite, um Front-End Code zu schreiben, zu testen, und bereitzustellen (codepen.io)

²⁰Der Prozessierung

5.2.3 Technische Komponenten

Auflistung der technischen Komponenten:

- ▶ **Abb. 4, Schritt 2., 3. und 5.:** Das Komprimieren und Kopieren der Rohdaten erfolgt mit Linux Bordmitteln (*cp*, *rsync*, *tar*)
- ▶ **Abb. 4, Schritt 4. und 7.:** Erfolgen via AWS CLI
- ▶ **Abb. 4, Schritt 6.:** Via Docker. Der Container wurde von der Firma Analytical Graphics Inc. [10a] bereitgestellt. Das Tool, das die Rohdaten in ein Web-Format umwandelt, wird mit Node.js ausgeführt.
- ▶ **Abb. 4, Schritt 7.:** Ein Projekt, um die Daten im Browser betrachten und inhaltlich Testen zu können, erfolgt über die Webseite codepen.io.

6 Architektur

6.1 Analyse des Ist-Zustandes

6.1.1 Bereitstellung der Rohdaten und Sicherstellung dessen Qualität

Basierend auf der Aufwandeinschätzung des Ist-Zustandes des Kapitels 5.2.2 geht hervor, dass vor allem das Bereitstellen der Rohdaten und das Sicherstellen dessen Qualität aufwändig ist.

Das Bereitstellen der Rohdaten nimmt aufgrund der Datenmenge²¹ viel Zeit in Anspruch. Hier gäbe es folgende zwei Lösungsansätze:

1. Der Datenlieferant könnte direkt aufs EFS schreiben (Abb. 4, Nr. 5)
2. Automatischer Prozess via Cronjob oder oder Trigger, der die Kopier-Schritte (Abb. 4, Schritte Abb. 4) regelmässig bereitstellt.

Wobei der Lösungsansatz simpler und somit weniger fehleranfällig zu sein scheint. Dieser Ansatz kann, sobald die swisstopo eine Hybrid Cloud hat, weiterverfolgt werden.

Sicherung der Qualität: Weil es dem Lieferanten an Tools fehlt, um die bereitgestellten Daten inhaltlich zu Prüfen, werden Fehler häufig erst nach der Publikation entdeckt. Dies sicherlich auch, weil die Daten im Web an ein breites Publikum gelangen. Da keine inhaltliche Prüfung der Daten gemacht werden kann, könnte man sich überlegen, wie die Daten sonst noch geprüft werden könnten.

6.1.2 Bereitstellung der Infrastruktur für die Prozessierung

Die Rohdaten werden via Docker Image prozessiert. Jedes Mal wenn das Image fürs Prozessieren ausgeführt werden soll, muss die IT gebeten werden die nötige Infrastruktur bereitzustellen. Idealerweise findet sich eine Lösung, die diesen Schritt überflüssig macht.

6.2 Bewertungskriterien

Kosten einsparen	Änderungen des Prozesses sollen nicht teurer sein, als die bisherige Lösung
Automatisierbar	Die Lösung soll den Grad der Automatisierung möglichst weit vorantreiben und dadurch so wenig Personalaufwand wie möglich beanspruchen.
Einfach	Ein neuer Mitarbeiter soll die Lösung rasch verstehen und warten können.
Bestehende Technologie	Obwohl der Anwendungsfall anders ist, soll der Technologie Stack möglichst demjenigen der Microservices entsprechen

6.3 Architekturentscheid

Für den Architekturentscheid wurde auf eine Entscheidungsmatrix verzichtet. Dies Aufgrund der grossen Auswahl und der vielen Entscheidungen, die getroffen werden mussten. Die oben erwähnten Kriterien wurden in der

²¹Weil es sich um mehrere Millionen Dateien handelt

Entscheidungsfindung mit einbezogen und weitere Argumente sind in qualitativer Form festgehalten:

6.3.1 Angebote für die Prozessierung

Um Geodaten mit AWS Spot Instanzen zu prozessieren, hat der Autor drei Möglichkeiten identifiziert: *Direkt auf der Spot Instanz*, *Kubernetes (EKS)* und *AWS Batch*. Gerne hätte der Autor alle drei Lösungen als POC weiterverfolgt, aber die Zeit dazu reichte leider nicht. Aus diesem Grund beschränkt sich der Autor lediglich auf eine Beschreibung der drei Möglichkeiten. Wobei die erste Möglichkeit als POC umgesetzt werden wird.

Direkt auf Spot Instanz Diese Möglichkeit bedingt nur geringe Anpassungen, weil es im Prinzip eine EC2 Instanz ist; mit dem Nachteil, dass diese jederzeit durch eine andere ersetzt werden könnte.

- ▶ Kein eigenes AMI, aber eine Konfiguration: TODO: Slides von Vorlesung mit Pro - Contra
- ▶ Ubuntu Base AMI, weil bereits das bestehende AMI auf Ubuntu war und wir sowieso mit Debian und dessen Derivaten arbeiten.

Kubernetes (EKS) auf Spot Instanzen erweitern TODO: Weil die Microservices in Zukunft auf EKS laufen werden, wäre es sicher ein Möglichkeit, abzuklären, inwiefern sich ein Dataprocessing in dieser Umgebung einrichten lässt. In der Literatur hat der Autor folgende Elemente gefunden.

- ▶ Job
- ▶ CronJob
- ▶ Eigenes ReplicaSet?

EBook auf meinem Reader: Job und CronJob aber auch Kubernetes on AWS S. 88

AWS Batch auf Spot Instanzen Diese Möglichkeit wäre wohl die Naheliegendste, weil AWS Batch für genau diese Art von Aufgaben gebaut wurde. Im Internet hat der Autor einen schönes Beispiel gefunden, wie sich AWS Batch einsetzen lässt. <https://aws.amazon.com/de/blogs/compute/creating-a-simple-fetch-and-run-aws-batch-job/>

7 Prototyp

7.1 Realisierung

7.1.1 Da

8 Evaluation

8.1 Erfahrungen

Die Anregungen aus dem Studium²² haben den Autor motiviert, die komplette Infrastruktur für den POC selber in einem eigenen AWS Testaccount aufzubauen. Da die Infrastruktur und das Processing deklarativ festgehalten wurden, kann ein Grossteil dieses Codes auch im Account des Betriebes ohne grosse Anpassungen wiederverwendet werden.

Im Betrieb hätte das Team und die IT Abteilung das nötige Know-How gehabt, um mit dieser Arbeit weiter zu kommen, als es jetzt der Fall ist. Aber der Autor hat sich bewusst dazu entschieden, die Arbeit möglichst selbständig zu realisieren und diese Ressource zu schonen, indem er sie so wenig wie möglich genutzt hat. Es war für den Autor schon nur eine Hilfe, zu Wissen, dass die Möglichkeit einer Unterstützung da war. Dies führte dazu, dass der Autor Momente der Ratlosigkeit erleben musste. Was nicht weiter schlimm war, da *"Perplexity is the beginning of knowledge"*[AD19, S. 33].

8.2 Wirtschaftlichkeit

Personalstunden Dank den Ansible Skripten konnte der Grad der Automatisierung erhöht werden. Dadurch fallen Personenstunden weg. Vor allem diese Kosten rechnen sich.

Einsparungen Spot im Vergleich zu On-Demand Um die Kosten von On-Demand mit Spot Instanzen zu vergleichen, werden hier die Ergebnisse des POCs aufgelistet. Prozessierung der Gebäude mit der Mindestanforderung: 16 CPUs und 60 GByte Memory. Die Gesamtrechnenzeit war ca. 18 Stunden und es konnten 76% der Kosten eingespart werden.

Pro Stunde	On-Demand	Spot
1h	0.29 \$	1.19 \$
Gesamtrechnenzeit	On-Demand	Spot
18 h	21.42 \$	5.22 \$

Schaut man die Kosten relativ an, dann ist das Sparpotential enorm: 76%! Jedoch für einen einzelnen Prozessierungsauftrag²³ ist das Sparpotential nicht riesig: ca. 15 \$.

8.3 Kritische Punkte

Bezüglich Authentifizierung hat der POC Technische Schulden.

²²CAS in Cloud Computing an der BFH BERN: Bsp. die Vorlesungen PaaS, Docker und Kubernetes

²³ohne Wiederholungen, weil Fehler in den Daten oder sonst gefunden wurden.

9 Ausblick

10 Schlusswort

Literaturverzeichnis

- [10a] (2010). Analytical Graphics Inc., Adresse: <https://www.agi.com> (besucht am 09.08.2020).
- [10b] (2010). Informatik, Adresse: <https://de.wikipedia.org/wiki/Informatik> (besucht am 16.07.2010).
- [10c] (2010). Kanban (Softwareentwicklung), Adresse: [https://de.wikipedia.org/wiki/Kanban_\(Softwareentwicklung\)](https://de.wikipedia.org/wiki/Kanban_(Softwareentwicklung)) (besucht am 19.07.2010).
- [10d] (2010). swisstopo Onlineshop, Adresse: <https://shop.swisstopo.admin.ch> (besucht am 17.07.2020).
- [10e] (2010). Technische Schulden, Adresse: https://de.wikipedia.org/wiki/Technische_Schulden (besucht am 18.07.2010).
- [18] (2018). Diese 3D-Landkarte gibt dir einen völlig neuen Blick auf die Schweiz, Adresse: <https://www.watson.ch/digital/schweiz/674619561-diese-3d-landkarte-gibt-dir-einen-voellig-neuen-blick-auf-die-schweiz> (besucht am 18.07.2010).
- [AD19] J. Arundel und J. Domingus, *Cloud Native DevOps with Kubernetes*. O'RELLY, 2019.
- [BJB06] T. Balstrøm, O. Jacobi und L. Bodum, *GIS og geodata*. Forlaget GIS og Geodata, 2006.
- [Chr20] H. Christ. (9. Juli 2020). swisstopo Fallstudie, Adresse: <https://aws.amazon.com/de/solutions/case-studies/swisstopo-hpc/>.
- [Keh05] D. Kehlmann, *Die Vermessung der Welt*. Rowohlt Verlag GmbH, 23. Sep. 2005, 304 S., ISBN: 3498035282. Adresse: https://www.ebook.de/de/product/3563778/daniel_kehlmann_die_vermessung_der_welt.html.
- [SE04] M.-H. Schertenleib und H. Egli-Broz, *Grundlagen Geografie: Aufgaben des Fachs, Erde als Himmelskörper und Kartografie : Lerntext, Aufgaben mit Lösungen und Kurztheorie*. Zürich: Compendio Bildungsmedien, 2004, ISBN: 3715591714.
- [Web10] Webpage. (2010). Keep calm and make your choice..., Adresse: <https://www.grenke-40-one.de/excursions-detail/> (besucht am 11.07.2010).

A Fachbegriffe und Abkürzungen

AWS	Amazon Web Services
AWS CLI	Das Command Line Interface, um AWS Ressourcen zu verwalten
BGDI	Bundesgeodateninfrastruktur
EC2	Virtuelle Server (TODO)
EBS	TODO
EFS	TODO
Hybrid Cloud	Eine Computerinfrastruktur, die Public Cloud und Private Cloud kombiniert nutzt
PaaS	Plattform as a Service
POC	Proof of Concept: Die Machbarkeit eines Produktes oder einer Idee aufzeigen
S3	TODO
SaaS	Software as a Service
21	22
31	32
41	42
51	52
61	62
71	72
81	82
91	92
101	102

B Konfigurationen

C Kennenlernen von AWS

D Kanban

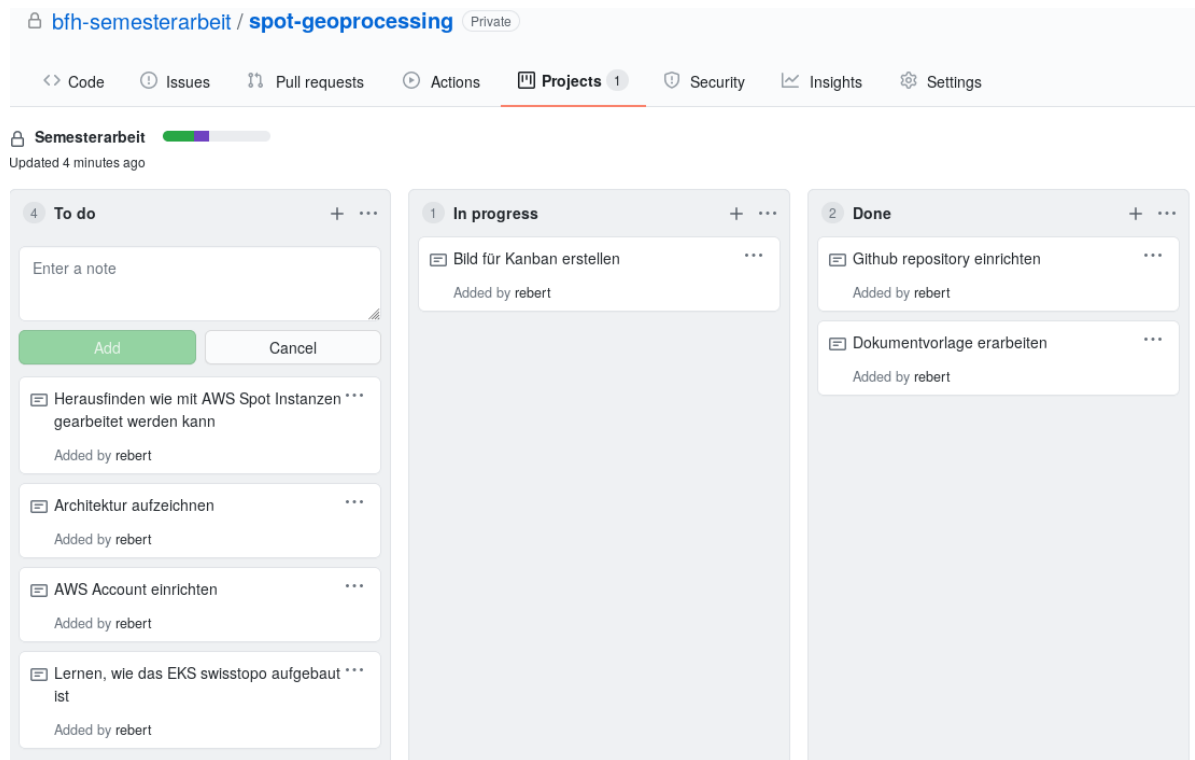


Abbildung 5: Klassisches Kanban auf *github.com*

E Projektplan

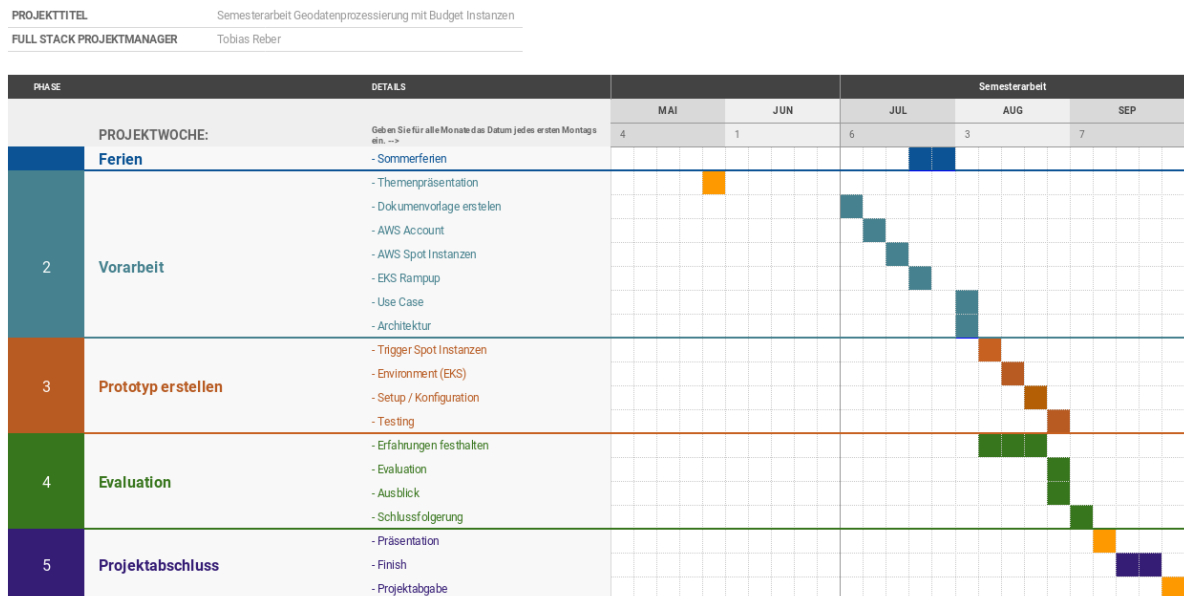


Abbildung 6: Projektplan. Die orangen Meilensteine wurden von der BFH vorgegeben.

F Konfigurationen und Kommandos

F.1 EFS auf EC2-Instanz mounten

Anhand einer Anleitung, einem sogenannten Walkthrough, wurde via AWS CLI²⁴ ein EFS an eine EC2-Instanz gemountet und die 3D Daten wurden dorthin kopiert.

Listing 1: EFS auf EC2-Instanz mounten

```
#!/bin/bash
# from https://docs.aws.amazon.com/efs/latest/ug/wt1-create-ec2-resources.html

# security group 4 ec2
aws ec2 create-security-group \
--region eu-west-1 \
--group-name efs-dataprocessing1-ec2-sg \
--description "Amazon EFS dataprocessing 1, SG for EC2 instance" \
--vpc-id vpc-87ad55fe

# security group 4 efs
aws ec2 create-security-group \
--region eu-west-1 \
--group-name efs-dataprocessing1-mt-sg \
--description "Amazon EFS dataprocessing 1, SG for mount target" \
--vpc-id vpc-87ad55fe

# access to ec2 instance group from everywhere
aws ec2 authorize-security-group-ingress \
--group-id sg-098669727548dcedd \
--protocol tcp \
--port 22 \
```

²⁴AWS Command Line Interface: ein kommandozeilenorientiertes Werkzeug

```
--cidr 0.0.0.0/0 \
--region eu-west-1

# describe security group
aws ec2 describe-security-groups \
--region eu-west-1 \
--group-id sg-098669727548dcedd

# access to efs storage from ec2 group
aws ec2 authorize-security-group-ingress \
--group-id sg-02778494bc39601d4 \
--protocol tcp \
--port 2049 \
--source-group sg-098669727548dcedd \
--region eu-west-1

# get subnet id
aws ec2 describe-subnets \
--region eu-west-1 \
--filters "Name=vpc-id,Values=vpc-87ad55fe"

# run ec2 instance
aws ec2 run-instances \
--image-id ami-047bb4163c506cd98 \
--count 1 \
--instance-type t2.micro \
--associate-public-ip-address \
--key-name bfh_root.pem \
--security-group-ids sg-098669727548dcedd \
--subnet-id subnet-f66512ac \
--region eu-west-1

aws ec2 describe-instances \
--instance-ids i-09cb26ed64cdde683 \
--region eu-west-1

# ec2-54-75-53-151.eu-west-1.compute.amazonaws.com
# EFS =====
aws efs create-file-system \
--creation-token FileSystemForDataprocessing1 \
--tags Key=Name,Value=Dataprocessing1 \
--region eu-west-1

# create mount target
aws efs create-mount-target \
--file-system-id fs-5aceld90 \
--subnet-id subnet-f66512ac \
--security-group sg-02778494bc39601d4 \
--region eu-west-1

# On instance amazon linux (ec2-user)
sudo yum -y update
sudo reboot # dont know why... like windows
sudo yum -y install nfs-utils

# On instance ubuntu (ubuntu)
sudo apt-get update
sudo apt-get install nfs-common

mkdir ~/efs-mount-point
cd ~/efs-mount-point
sudo chmod go+rw .

sudo mount -t nfs \
-o nfsvers=4.1,rsz=1048576,wsz=1048576,hard,timeo=600,retrans=2,noresvport
fs-5aceld90.efs.eu-west-1.amazonaws.com:/ \
~/efs-mount-point
```

```
# CP from home to ec2 instance
scp -i bfh_root.pem \
/media/saibot/vortrag/bfh/buildings/*.tar \
ec2-user@54.75.53.151:/home/ec2-user/efs-mount-point/buildings/
```

F.2 Von der SPOT Instanz aus abfragen, was ihr Status ist

Von der Instanz aus kann via REST API Metadaten der Instanz abgefragt werden. Bezüglich Determinierung einer Spot-Instanz kann der Zustand *none*, *hibernate*, *stop* oder *terminate* sein. *none*, wenn nichts ansteht. Von da an, wo klar ist, dass die Instanz abgestellt werden wird, kann der Zeitpunkt ausgelesen werden.

Listing 2: Status der Instanz abfragen

```
TOKEN=$(curl -X PUT "http://169.254.169.254/latest/api/token" -H  
"X-aws-ec2-metadata-token-ttl-seconds: 21600") && curl -H "X-aws-ec2-metadata-token: $TOKEN"  
http://169.254.169.254/latest/meta-data/instance-action
```


G Für die Semesterarbeit verwendete Software

- ▶ JabRef: Verwaltung des Literaturverzeichnisses (BibTeX)
- ▶ Gummi: LaTeX Editor
- ▶ AWS CLI: Für das Bereitstellen der AWS Infrastruktur
- ▶ jq: Für das Filtern von JSON (vor allem von AWS CLI Antworten)
- ▶ git: Versionsmanagement der Textdateien
- ▶ Google Spreadsheet: Für den Projektplan