

Proseminar "Softwareentwicklung in der Wissenschaft"

Betreuer: Dr. Hermann Lenhart

Cloud Computing in der Wissenschaft

- schriftliche Ausarbeitung -

Vorgelegt von: Maximilian Marquardt

am 31.08.2020

Inhaltsverzeichnis

1. Einleitung	3
2. Wissenschaft ohne Cloud Computing	4
3. Was ist Cloud Computing?	
a. Cloud Arten	4
b. Cloud Services	5
4. Nutzen für die Wissenschaft	6
5. CERN	
a. Cloud Lösung	7
b. ATLAS Experiment	8
6. Pro und Contra für die Wissenschaft	10
7. Fazit	11
8. Quellen	13

1. Einleitung

Cloud Computing ist in der Wissenschaft ein wichtiges Thema. Seit 2008 steigt das Interesse in dieses Thema, mehr und mehr wird das eigene Rechenzentrum durch Cloud Dienste ersetzt oder ergänzt.

Unter dem Begriff Cloud Computing, versteht man die Bereitstellung von IT-Leistung auf externen Servern, diese IT-Leistung können überall genutzt werden, wo eine relativ gute Internetverbindung vorhanden ist. Diese IT-Leistung sind frei skalierbar und in kurzer Zeit verfügbar, also man muss keine großen Investitionen tätigen. Gegen die Verwendung von Cloud Computing kann die Anforderung an den Datenschutz sprechen, von vielen Anbietern wird der Datenschutz nur mäßig erfüllt. In der Wissenschaft wird es vor allem genutzt, um große Datenmengen zu verarbeiten und zu speichern.

Im zweiten Kapitel wird auf die Wissenschaft ohne Cloud Computing eingegangen. Im darauf folgenden Kapitel wird das Thema Cloud Computing behandelt. Es wird gezeigt welche Arten von Cloud und welche Cloud Services es gibt. Danach wird auf den Nutzen für die Wissenschaft eingegangen. Dort werden Aspekte aufgezählt, die durch Cloud Computing in der Wissenschaft verbessert werden können. Im fünften Kapitel wird das CERN behandelt. Zuerst gibt es ein paar Allgemeine Infos, danach wird auf die Cloud Lösung am CERN eingegangen und zum Schluss wird am ATLAS Experiment gezeigt, wie Cloud Computing dort umgesetzt wird. Daraufhin werden im nächsten Kapitel die Pro und Contra Argumente für Cloud Computing aufgezählt und ein wenig erklärt. Im letzten Kapitel kommt dann ein kurzes Fazit zum Thema Cloud Computing in der Wissenschaft.

2. Wissenschaft ohne Cloud Computing

Die Wissenschaft ist sehr vielseitig, es wird mit verschiedensten Systemen gearbeitet, um Rechenleistung zu nutzen. Große Institute haben eigene Rechenzentren und kleinere Wissenschaftler Gruppen müssen sich ihre Rechenleistung von woanders beziehen.

In der Wissenschaft wird mit großen Datenmengen gearbeitet, diese müssen gespeichert und ausgewertet werden. Um die Datenmengen zu speichern ist viel Speicher nötig, das ist mit großen Kosten verbunden. Die Kosten entstehen nicht nur bei der Anschaffung, sondern solche Systeme sind auch mit Servicekosten verbunden. Man braucht nicht nur viel Speicher, sondern auch genügend Rechenleistung, um z.B. Klimamodelle zu berechnen.

Des Weiteren arbeiten Wissenschaftler selten allein, meist wird in Gruppen gearbeitet. Diese sind teilweise über die ganze Welt verteilt, der Austausch von Daten spielt dabei eine wichtige Rolle.

Außerdem gibt es keinen Ort, wo alle wissenschaftlichen Erkenntnisse gesammelt werden. Man muss also immer verschiedenen Quellen durchsuchen um die Erkenntnisse zu sammeln die man benötigt.

3. Was ist Cloud Computing?

Cloud Computing ist die Bereitstellung von IT-Ressourcen über das Internet. Dazu gehören Speicher, Rechenleistung und Software.

3.a Cloud Arten

Es gibt drei Arten von Clouds:

Public Cloud

Die Public Cloud, diese stellt seine Dienste offen in das Internet und ist somit für jedermann zugänglich. Ein Beispiel dafür wären die Google Dienste.

Private Cloud

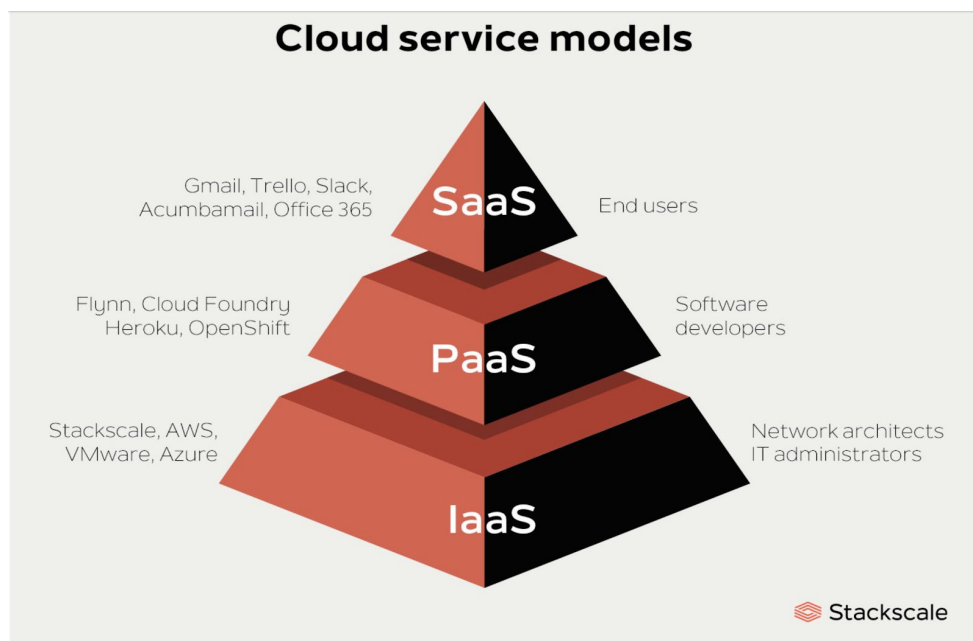
Die Private Cloud, diese wird meist von Unternehmen genutzt, damit sie den Datenschutz besser kontrollieren können. Diese Dienste sind ausschließlich für Mitarbeiter der Firma verfügbar.

Hybrid Cloud

Die Hybrid Cloud, dies ist eine Mischform der Public- und Private-Cloud. Da werden die Dienste getrennt, manche Services laufen über die Public Cloud und bei Services, wo der Datenschutz entscheidend ist wird die Private Cloud genutzt.

3.b Cloud Services

Desweiteren unterscheidet man zwischen drei Arten von Cloud Services:



Quelle: <https://www.stackscale.com/blog/cloud-service-models/>

Infrastructure as a Service (IaaS)

IaaS stellt die unterste Ebene dar. Hier werden IT-Ressourcen zur Verfügung gestellt. Dazu gehören Rechenleistung, Speicher und Netzwerkkapazitäten. Der Nutzer kann selbst die benötigten Ressourcen zusammenstellen.

Platform as a Service (PaaS)

Die mittlere Ebene stellt PaaS dar. PaaS ist ein Typ von Service, wo Programmiermodelle und Entwicklerwerkzeuge bereitgestellt werden.

Software as a Service (SaaS)

SaaS ist die oberste Ebene. Hier werden Applikationen zur Verfügung gestellt und diese können online genutzt werden.

4. Nutzen für die Wissenschaft

Es gibt mehrere Vorteile für die Wissenschaft durch Cloud Computing, aber nicht alle Vorteile sind überall anzuwenden, denn z.B. die Kostenvorteile sind hauptsächlich bei kleineren Wissenschaftler Gruppen zu sehen. Bei großen Instituten wäre dieser Vorteil nicht anzuwenden, denn durch enorme Rechenleistung entstehen natürlich mehr Kosten.

Einer der Vorteile ist das Outsourcing der IT-Ressourcen, man muss sich um wenig kümmern und man gibt es in die Hand eines Profis auf diesem Gebiet, also man kann sich mehr auf die Wissenschaft konzentrieren. Dazu gehört auch die Kostenersparnis, man muss sich nicht um Wartungen kümmern oder überhaupt IT-Ressourcen anschaffen. Bei der Cloud zahlt man nur das, was man wirklich benötigt. Auch wenn man Software benötigt, kann man diese aus der Cloud beziehen, diese ist immer auf dem neuesten Stand. Die eigene Hardware ist bei Cloud Computing nicht entscheidend, man benötigt lediglich einen Computer mit einem Internetzugang.

Durch Cloud Computing hat man einen flexiblen Arbeitsplatz, man kann von überall arbeiten. Falls Veränderungen nötig sind und man dadurch mehr Rechenleistung

benötigt, ist das schnell realisierbar. Außerdem kann man weit nach oben expandieren, denn große Cloud Computing Anbieter haben viele Ressourcen zur Verfügung. Desweiteren sind Kooperationen zwischen Wissenschaftlern durch Cloud Computing vereinfacht, denn die Daten müssen nur in die Cloud hochgeladen werden und jeder hat sie dann zur Verfügung. Es müssen nicht nur Kooperationen sein, denn vorhandene Methoden, Theorien und Kompetenzen können über die Cloud geteilt werden, ein Beispiel dafür wäre die "EOSC", wo Wissenschaftler Zugang zu wissenschaftlichen Daten haben. Generell ist Cloud Computing sehr dynamisch, man kann alles so anpassen wie man es benötigt.

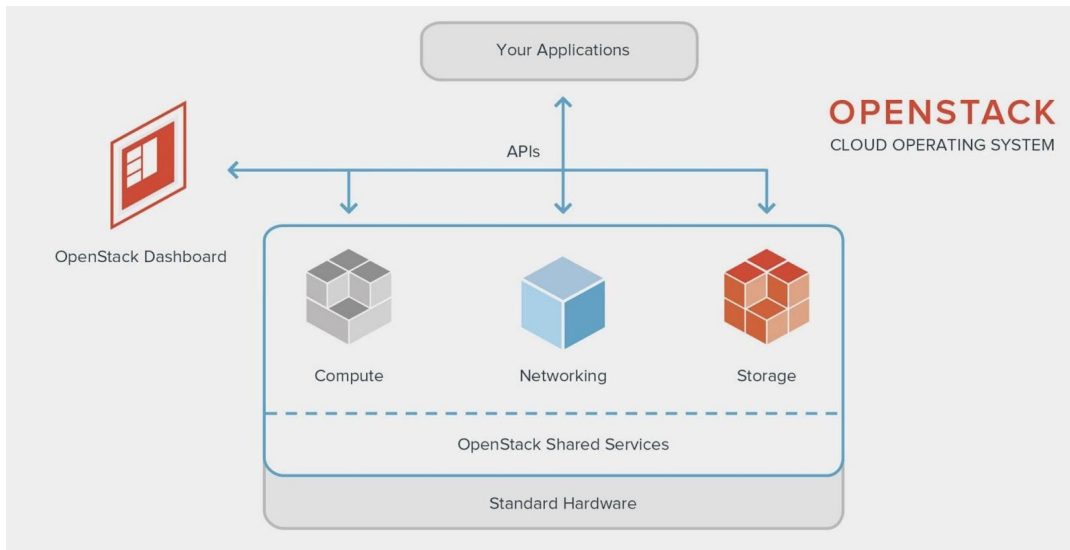
5. CERN

Das CERN ist die Europäische Organisation für Kernforschung, dort wird nach fundamentalen Gesetzen des Universums gesucht. Es ist in der Nähe von Genf lokalisiert. Es arbeiten dort etwa 12.000 Wissenschaftler aus den verschiedensten Ländern. Außerdem steht am CERN der leistungsstärkste Teilchenbeschleuniger, der Large Hadron Collider (LHC). Des Weiteren ist Deutschland einer der Großen Mitfinanzierer, sie fördern das CERN mit ca. 21% der Gesamtsumme.

5.a Cloud Lösung

Das CERN arbeitet hauptsächlich mit Grid Computing, dem "Worldwide LHC Computing Grid". Dieses Computing Grid verarbeitet die Daten des LHC, ungefähr 25 Petabyte pro Jahr.

Das CERN besitzt auch eine Private Cloud, diese deckt den IaaS Service ab. Diese Cloud basiert auf Openstack, dies ist ein "Cloud-Betriebssystem". Openstack besteht aus 9 Basiskomponenten, diese managen die Computing-Services. Die Komponenten machen das "Cloud-Betriebssystem" auch sehr komplex.



Quelle: <https://www.purestorage.com/de/solutions/infrastructure/openstack.html>

Des Weiteren besitzt das CERN zusammen mit anderen Institutionen eine Hybrid Cloud, diese deckt auch den IaaS Service ab. Das Ziel war es externe IT-Ressourcen in das eigene System mit einzubinden. Die Hybrid Cloud heißt Helix Nebula Science Cloud. Die Cloud agiert europaweit und stellt IT-Ressourcen für die verschiedensten Bereiche der Wissenschaft bereit. Die Cloud hat außerdem ein großes Netz an Versorgern, welche die IT-Ressourcen zur Verfügung stellen.

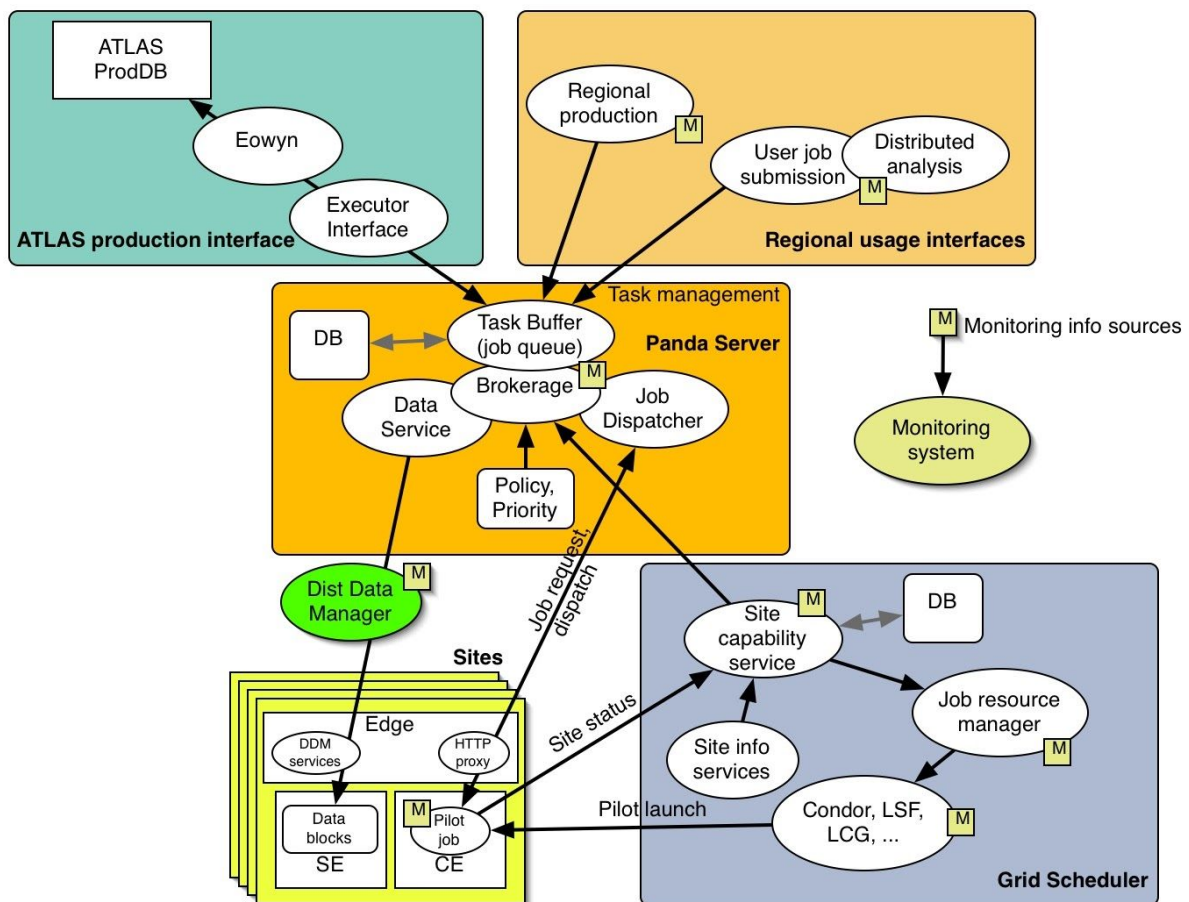
5.b ATLAS Experiment

ATLAS ist ein Teilchendetektor am LHC, dieser produziert enorme Datenmengen. Mit dem Teilchendetektor möchte man die Bestandteile der Materie und die fundamentalen Kräfte der Natur erforschen. Es ist 1 von 4 großen Experimenten am LHC. Für die Forschung sind nur in etwa ein hunderttausendstel, ungefähr 1 GB pro sekunde, der Daten relevant. Die Speicherung und Verarbeitung der Daten wird mithilfe des Grid Computings bewältigt. Mit diesen Daten arbeiten Wissenschaftler aus über 38 Ländern.

Mit Cloud Computing wollte man die Prozesse optimieren. Zum einen wurde die eigene Private Cloud genutzt, aber man wollte auch die Public Cloud mit einbinden. Dazu wurden mehrere Public Clouds getestet und mit Amazons Cloud hat man eine effiziente Lösung gefunden, die auch kostengünstig ist.

Beim CERN kam man zu dem Entschluss, das Cloud- und Grid-Computing zusammen existieren können und die beiden Systeme sich gegenseitig ergänzen. Man hat sich dazu entschieden eine Kombination aus beiden Computing Arten für ATLAS zu nutzen.

Die Daten innerhalb des Systems werden mit dem Datenmanagementsystem PanDA gemanagt. Dieses System bekommt die Daten auf den Server, wo sie in eine Globale Warteschlange gesteckt werden. Die Position in der Warteschlange hängt unter anderem von der Priorität der Aufgabe ab. Wenn die Position der Aufgabe erreicht wurde, wird die Aufgabe an einen "Job Dispatcher" weitergeleitet, wo sie dann auch bearbeitet wird. Der User bekommt dann sein Ergebnis zurück, so schafft das System bis zu 100.000 Aufgaben pro Tag. Das System ist insgesamt sehr komplex, deshalb habe ich es vereinfacht dargestellt.



Quelle: <https://twiki.cern.ch/twiki/bin/view/PanDA/PanDA>

Um ihr System aus mehreren Clouds zu beschreiben, wird der Begriff "Sky Computing" verwendet.

Der Nutzen für das Experiment ist, dass mehr Ressourcen zur Verfügung stehen und das ganze System mehr skalierbar ist.

6. Pro und Contra für die Wissenschaft

Cloud Computing hat viele Vorteile, aber natürlich auch Nachteile. Die Vor- und Nachteile haben auch mit der Größe der Forschungseinrichtung zutun, daher ist zu beachten, dass ich die komplette Wissenschaft betrachte und dazu gehören auch kleiner Wissenschaftler Gruppen.

Einer der großen Vorteile ist, dass man seine Systeme frei skalieren kann und auch nur das zahlt, was man wirklich benötigt. Dies passiert auch in kurzer Zeit, also man kann auch spontan mehr Rechenleistung buchen.

Desweiteren hat man weniger Kosten, man muss nicht in sein eigenes Rechenzentrum investieren, also hat man wenig bis keine Anschaffungskosten. Man nutzt lediglich die Ressourcen des Cloud Computing Anbieters. Außerdem kann man Personal einsparen, denn man kümmert sich nicht mehr um die Administration.

Dafür spricht auch, dass der Austausch zwischen Wissenschaftler Gruppen und Institutionen durch die Cloud erleichtert werden kann. Wissenschaftler sind über die ganze Welt verteilt, benötigte Daten werden einfach in die Cloud geladen und sind für alle verfügbar.

Hinzu kommt, dass durch z.B. Corona nicht alle ihren normalen Arbeitsplatz besuchen können. Cloud Computing ist überall verfügbar, wo Internet vorhanden ist, somit ist der Arbeitsplatz flexibel. Man ist nicht an bestimmte Orte gebunden. Dazu kommt, wie schon erwähnt die Verfügbarkeit, man kann von überall auf seine Daten zugreifen. Der Vorteil ist, dass man nicht in einem bestimmten Netzwerk sein muss um auf seine Daten zuzugreifen.

Wenn man ein eigenes Rechenzentrum hat, kommt irgendwann das Problem, dass die Technik veraltet ist. Bei Cloud Computing halten die Anbieter ihre Technik aus eigenem Interesse immer auf dem neuesten Stand. Man muss sich also nicht um irgendwelche Upgrades kümmern.

Dagegen spricht, dass der Datenschutz nicht gewährleistet ist. In Europa haben wir strenge Datenschutzrichtlinien, dies kollidiert z.B. mit Cloud Computing Anbietern aus Amerika. Also man kann bei einer Public Cloud nicht beeinflussen wie der Datenschutz dort umgesetzt wird. Bei einer Private Cloud ist das schon wieder anders, denn dort bestimmt die Institution wie der Datenschutz umgesetzt wird

Ein weiteres Argument dagegen ist, dass das Arbeitstempo von der Internetverbindung abhängt. Wenn man keine stabile Internetverbindung hat, wird das Arbeiten dadurch quasi unmöglich. Auch z.B. in armen Regionen, wo der Bandbreitenausbau nicht fortgeschritten ist, ist das Arbeiten mit der Cloud unmöglich.

Nicht zu übersehen ist, dass man sich auf eine Abhängigkeit einlässt. Man ist von dem Anbieter stark abhängig, die Daten werden dort gespeichert und man bezieht seine Rechenleistung von dort. Wenn also dort etwas nicht funktionieren sollte, könnte man nicht arbeiten oder im schlimmsten Fall sind alle Daten weg.

Wie im vorherigen Punkt erwähnt, ist die Sicherheit der Daten, in einer Public Cloud, nicht unbedingt gewährleistet. Sei es durch ein schlechten Schutz nach außen oder ein schlechtes Backup System, die Daten können immer irgendwie verschwinden. Auch Hacker-Angriffe sind nicht ausgeschlossen bei großen Cloud Anbietern.

Ein weiterer Nachteil könnte sein, dass man Arbeitsprozesse an die Cloud Lösung anpassen muss. Durch nicht vorhandene Software muss man im schlimmsten Fall Prozesse an die vorhandene Cloud anpassen, damit alles funktioniert

7. Fazit

Mein Fazit zum Thema Cloud Computing in der Wissenschaft ist, dass Cloud Computing auf jeden fall eine Zukunft hat. Es wird bestimmt noch Rechenzentren und so weiter geben, aber mit der Zeit setzten große Institute immer mehr auf die Cloud. Es wird immer mehr Kombinationen aus verschiedenen Systemen geben, um die Leistung maximal auszunutzen und die Systeme sollen sich gegenseitig ergänzen Bei kleineren Wissenschaftler

Gruppen ist Cloud Computing sehr sinnvoll, denn man ist flexibel und hat keine Anschaffungskosten, man zahlt nur das was man benötigt.

Außerdem hat Corona gezeigt wie wichtig Cloud-Dienste sind. Nutzerzahlen haben sich mehr als verdoppelt. Mit der Cloud kann man flexibel von überall arbeiten, man benötigt nur einen Computer und Internet.

Natürlich hat die Cloud auch Nachteile, wie schon erwähnt z.B. der Datenschutz bei Public Clouds, aber meiner Meinung nach überwiegen die Vorteile. Public Clouds werden generell noch erforscht, z.B. vom CERN, ob sie in das vorhandene System integriert werden können. Im allgemeinen wird eher auf die Private Cloud gesetzt, denn dort kann man Dinge wie Datenschutz selber regulieren.

Auch das Thema "Sky Computing" wird wichtiger, also die Kombination aus mehreren Cloud Lösung. Wie man am CERN sieht, kann man mit "Sky Computing" maximale Leistung erzielen.

Insgesamt ist Cloud Computing ein wichtiges Thema, es wird überall relevanter, nicht nur in der Wissenschaft.

8. Quellen

<https://home.cern/>

<https://www.helix-nebula.eu/helix-nebula-initiative-europe%E2%80%99s-leading-public-private-partnership-cloud>

<https://www.exoscale.com/syslog/cern-on-exoscale/>

<https://www.cloud-mag.com/was-ist-openstack/>

<https://www.stackscale.com/blog/cloud-service-models/>

Ian Foster, Dennis B. Gannon: "Cloud Computing for Science and Engineering", 2017

<http://cds.cern.ch/record/1621892/files/ATL-SOFT-PROC-2013-034.pdf>

<https://cds.cern.ch/record/1609060/files/ATL-SOFT-SLIDE-2013-827.pdf>

https://twiki.cern.ch/twiki/bin/view/PanDA/PanDA#Architecture_and_workflow

<https://novadex.com/de/glossar-artikel/definition-cloud-computing-was-ist-cloud-computing/>

<https://www.cloudcomputing-insider.de/was-ist-cloud-computing-a-563624/>

https://ec.europa.eu/germany/news/cloud20181123_de

<https://www.cloud-mag.com/was-ist-openstack/>