

GWDG NACHRICHTEN 10-11|21

Klimafreundlicher IT-Betrieb

Kubernetes with Rancher

Data Lakes

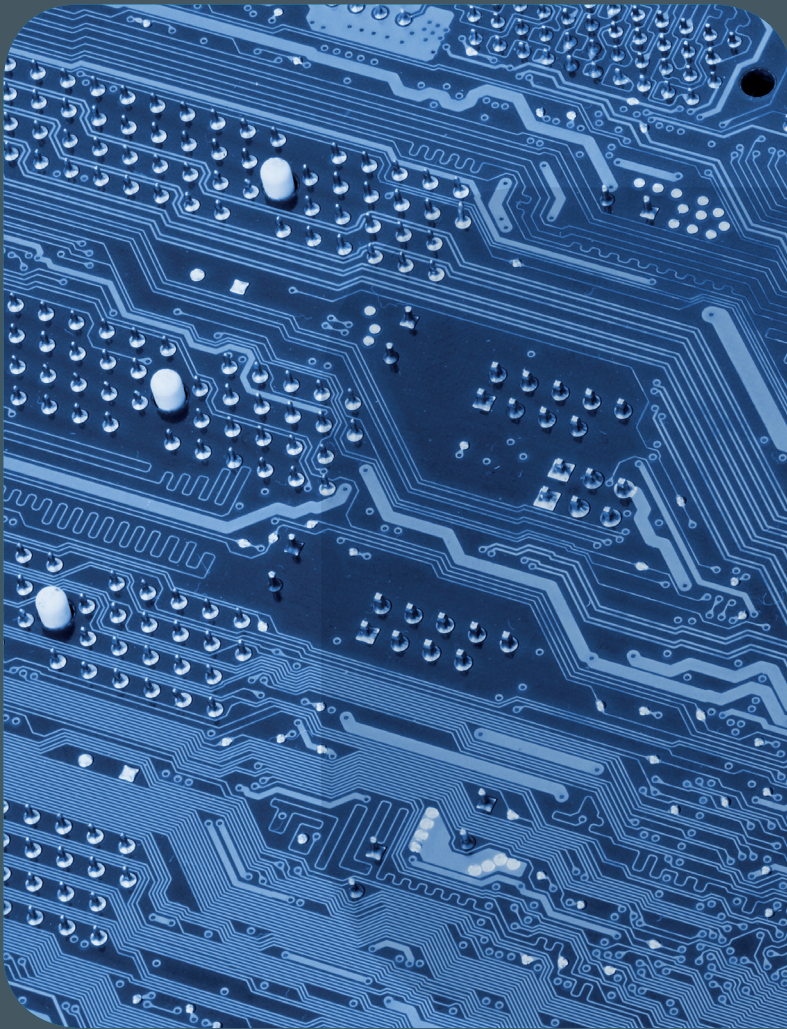
38. DV-Treffen der
Max-Planck-Institute

Treffen der
Institutsadministratoren

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDG



GWDG
Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen



GWDG NACHRICHTEN

10-11|21 Inhalt

-
- 4 GWDG setzt auf klimafreundlichen IT-Betrieb
 - 5 Kubernetes with Rancher at the GWDG –
Part 4: Setup a PostgreSQL Cluster
 - 10 Leveraging Data Lakes for Managing and
Processing of Large Data Volumes
 - 13 Nächste Ausfahrt Transformation – Blick
auf die IT der MPG im Wandel 16 Treffen der
Institutsadministratoren 17 Kurz & knapp
 - 18 Stellenangebote 21 Personalia 23 Academy

Impressum

.....
Zeitschrift für die Kunden der GWDG

ISSN 0940-4686
44. Jahrgang
Ausgabe 10-11/2021

Erscheinungsweise:
10 Ausgaben pro Jahr

www.gwdg.de/gwdg-nr

Auflage:
550

Fotos:

© peach_fotolia - stock.adobe.com (1)
© pineapple - Fotolia.com (12)
© momius - Fotolia.com (15)
© contrastwerkstatt - Fotolia.com (18-19)
© nito - Fotolia.com (20)
© Robert Kneschke - Fotolia.com (23)
© Ralf Poscher (13) © Ranga Yogeshwar (14)
© Stefan Vollmar (14)
© MPLbpc-Medienservice (3, 21, 22)
© GWDG (2)

Herausgeber:

Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen
Burckhardtweg 4
37077 Göttingen
Tel.: 0551 39-30001
Fax: 0551 39-130-30001

Redaktion:

Dr. Thomas Otto
E-Mail: thomas.otto@gwdg.de

Herstellung:

Maria Geraci
E-Mail: maria.geraci@gwdg.de

Druck:

Kreationszeit GmbH, Rosdorf



Prof. Dr. Ramin Yahyapour
ramin.yahyapour@gwdg.de
0551 201-1545

*Liebe Kund*innen und Freund*innen der GWDG,*

es ist kein Geheimnis, dass Informationstechnik signifikant Strom verbraucht und somit global einen nicht unerheblichen Anteil an den CO₂-Emissionen hat. Daher liegt es auf der Hand, sich mit dem Thema der Verbesserung der Energieeffizienz zu beschäftigen, so wie wir es bei der GWDG unter anderem mit neuen Kühl- und Wärmekonzepten in unserem neuen Rechenzentrum und bei Beschaffung und Betrieb von neuer Hardware tun. Gleichzeitig stellt sich die Frage nach der Wahl der geeigneten Energiequellen und den Möglichkeiten, den Emissions-Footprint zu verbessern. Es freut mich daher, dass wir in Zusammenarbeit mit der Universität Göttingen eine Möglichkeit gefunden haben, um künftig Naturstrom für unsere IT-Infrastruktur einsetzen zu können. So wird insbesondere ab dem 1. Januar 2022 unser Supercomputer „Emmy“ als nationale Rechnerressource mit grünem Strom betrieben, der zertifiziert CO₂-neutral ist, und wir sind damit einer der ersten unter den Betreibern der deutschen Supercomputer.

*Dies ist ein wichtiger Beitrag im Engagement der GWDG, um den Herausforderungen des Klimawandels zu begegnen. Ich bin sicher, dass unsere vielen Kund*innen, Studierende und Beschäftigte einen solchen Schritt von uns erwarten. Meines Erachtens zu Recht. Dennoch ist jeder aufgefordert, auch in seinem eigenen Nutzungsverhalten zu prüfen, welchen Beitrag er bei einer effizienten Nutzung von IT-Systemen leisten kann.*

Ramin Yahyapour

GWDG – IT in der Wissenschaft

GWDG setzt auf klimafreundlichen IT-Betrieb

Text und Kontakt:
Prof. Dr. Julian Kunkel
julian.kunkel@gwdg.de

Die GWDG ist in verschiedenen Bereichen aktiv, um den zunehmenden Herausforderungen des Klimawandels zu begegnen. Neben der Reduktion des Stromverbrauchs durch umfangreiche technische und organisatorische Maßnahmen in den Rechenzentren unternimmt sie nun mit dem Einsatz von regenerativem Strom einen weiteren wichtigen Schritt in Richtung Klimaschutz und Nachhaltigkeit.

Ökostrom für die IT! Die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) wird die Versorgung des Supercomputers „Emmy“ und ihrer IT-Infrastruktur zum 1. Januar 2022 auf regenerativen Strom umstellen

Global betrachtet ist die IT-Branche ein großer Energieverbraucher und damit einhergehend auch ein signifikanter CO₂-Produzent. Einen großen weiteren Schritt in Richtung Nachhaltigkeit geht die GWDG, indem sie ab Anfang 2022 klimaneutralen Strom nutzt, um ihre IT-Infrastruktur und den Hochleistungsrechner energieeffizient und klimaneutral zu betreiben.

Zusammen mit der Universität Göttingen wurden bereits umfangreiche technische und organisatorische Maßnahmen vorangetrieben, um den Verbrauch von Strom durch Modernisierung der technischen Ausstattung und Optimierung der Energieeffizienz der Rechenzentren zu reduzieren. In einem Modularen Rechenzentrum (MDC) und dem Ende Juni 2021 eröffneten neuen Rechenzentrum am Burckhardtweg werden verschiedene energiesparende Kühlverfahren wie Warmwasserkühlung und sowohl direkte als auch indirekte freie Kühlung realisiert. Mit der durch die Rechnerleistung anfallenden Abwärme können benachbarte Gebäude beheizt werden. In der Übergangszeit ist ein Mischbetrieb möglich, bei dem Kältemaschinen die freie Kühlung unterstützen, um die Nutzungszeit der freien Kühlung zu maximieren. Die Energieaufnahme der IT- und Kühlsysteme der GWDG kann erfasst werden, um den Stromverbrauch eingesetzter Hardware zu optimieren.

Die Energieeffizienz misst man bei Rechenzentren typischerweise mit der sogenannten Power Usage Effectiveness (PUE). Sie stellt das Verhältnis aus der verbrauchten Gesamtenergie zur Energieaufnahme der IT-Infrastruktur dar. Rechenzentren gelten bei einem Wert von unter 1,2 als sehr effizient. Mit den genannten Maßnahmen konnte die GWDG beim Betrieb von Emmy eine PUE von bis zu 1,03 erreichen, so dass nur 3 % zusätzlicher Strom für Kühlung und Gebäudetechnik benötigt wird und damit eine hervorragende Energieeffizienz erreicht wird.

Neben der Reduktion des Stromverbrauchs unternimmt die GWDG nun mit dem Einsatz von regenerativem Strom einen weiteren wichtigen Schritt, den zunehmenden Herausforderungen des Klimawandels zu begegnen. Dabei ist auch die Herkunft des Stroms ein wichtiges Kriterium: Die Naturstrommengen werden ausschließlich aus erneuerbaren Energien erzeugt. Dies wird

über das Herkunftsnachweisregister (HKNR) für Ökostrom beim Umweltbundesamt (UBA) nachgehalten, in dem für jede erzeugte Megawattstunde sowohl die Erzeugungsart als auch die Herkunft in Form von Herkunftsnachweisen hinterlegt sind. Zusätzlich fließt ein Teil des Geldes sofort in den Bau regenerativer Anlagen für die Stromerzeugung.

Nicht nur die GWDG nimmt eine Vorreiterrolle ein, auch die Universität Göttingen möchte das Thema Klimaschutz vorantreiben und plant, die Hochschule bis 2030 klimaneutral zu gestalten. „Die Beweislage für den Klimawandel ist inzwischen erdrückend. Es geht um unser aller Zukunft, deshalb müssen wir jetzt handeln“, so Universitätspräsident Prof. Dr. Metin Tolan.

Die Universität Göttingen ist mit der GWDG eines von acht Rechenzentren im Verbund Nationales Hochleistungsrechnen (NHR) und ebenso Teil der Nationalen Forschungsdateninfrastruktur (NFDI).

Der aktuell schnellste Rechner ist „Emmy“, benannt nach der Göttinger Mathematikerin Emmy Noether, und liefert eine Rechenleistung von 9,28 PFlop/s (Billiarden Rechenoperationen pro Sekunde) mit 1.569 Rechenknoten mit 3.192 Intel-Prozessoren. Daneben betreibt die GWDG weitere Hochleistungsrechner wie den Supercomputer „CARO“ für das Deutsche Zentrum für Luft- und Raumfahrt (DLR).

Weitere Informationen

- Nationales Hochleistungsrechnen (NHR):
<https://www.nhr-gs.de>
- Hochleistungsrechnen bei der GWDG:
<https://hpc.gwdg.de>

GWDG Focuses on Climate-friendly IT Operation

The GWDG is active in various areas to meet the increasing challenges of climate change. In addition to reducing electricity consumption through extensive technical and organisational measures in the data centres, it is now taking another important step towards climate protection and sustainability by using renewable electricity.

Kubernetes with Rancher at the GWDG – Part 4: Setup a PostgreSQL Cluster

Text and Contact:

Samaneh Sadegh
samaneh.sadegh@gwdg.de
0551 201-2113

Rancher facilitates the deployment and the interaction with Kubernetes clusters. In this fourth part of a series of articles, we show how to use Rancher in order to deploy your application from zero to a running service. The content is based on Rancher 2.4 and could be a little different from other versions.

INTRODUCTION

At the GWDG, Rancher is the technical foundation for our managed Kubernetes service and is used to deploy and manage Kubernetes clusters for various internal and external projects. In previous parts of this series of articles, we introduced Kubernetes, explained how to utilize Rancher for providing a Kubernetes cluster, and talked about the different ways to deploy an application into the provided cluster.

In this article, we will illustrate a complete deployment of a complex application step by step from an empty cluster to a running service. The selected example is a full-fledged PostgreSQL cluster deployed using a Kubernetes operator provided by Zalando which offers a Postgres-as-a-Service capability. A Kubernetes operator is an application-specific controller that not only packages and deploys an application but also automates the entire life cycle of the software it manages (e.g. software upgrades, backups, etc.). The used operator provides support for a number of advanced PostgreSQL features like rolling updates, live volume resize, HA based on Patroni, backups to S3 and more. The operator defines a new resource type named "postgresql" and creates a PostgreSQL cluster for each yaml file whose kind is defined as "postgresql".

A user can either manually create and apply this yaml file using kubectl or work with the operator API. We use the API approach with help of another application provided by Zalando namely the "postgres-operator-ui". In this way, we will also show how a web service can be deployed on Kubernetes.

Due to space and complexity constraints this article will only cover the basic setup of a PostgreSQL cluster and won't go into additional, more advanced aspects such as backup or access control. For more information visit the very good documentation here: <https://github.com/zalando/postgres-operator>.

For the CLI commands used in this article, you need the kubeconfig file of your cluster to access it remotely. As explained in part 2 of this series, you can get its content from cluster page in the Rancher GUI. Here we use the name "prod_playground.yaml" as the kubeconfig file.

DEPLOYING THE REQUIRED SERVICES

Using application charts is the simplest way to install an application on Kubernetes. Therefore, we deploy the operator API server using its chart. To use this option, we first need to add the application's chart to the Rancher catalog. To do so, as explained in previous parts of the series, after selecting your own project from *global*, go to *tools > catalog* and then click *Add Catalog*. Specify a name for the catalog, set the scope to project and fill the other input fields similar to figure 1. The chart is published in the Zalando git repository at <https://opensource.zalando.com/postgres-operator/charts/postgres-operator/>.

To deploy the application, go to the *Apps* tab and click *Launch*. Search for the word "postgres" and select "postgres-operator" which is displayed under your newly defined catalog. In the displayed form, you can customize the following parameters:

- *Name*: We left it at the default value "postgres-operator".
- *Template version*: We selected the latest version at the time of writing this article which was 1.6.3.
- *Namespace*: We left it at the default value "postgres-operator".
- *Helm Wait*: We set it to *false*. We will monitor the process using Rancher GUI.

Click *Launch* and wait until the *Apps* state is Active. You can check the pod state from *Resources > workload*. A deployment

Kubernetes mit Rancher bei der GWDG – Teil 3: Bereitstellung eines PostgreSQL-Clusters

Rancher ermöglicht nicht nur eine einfache Installation und spätere Verwaltung von Kubernetes-Clustern, sondern auch das Deployment von Applikationen in die bereitgestellten Kubernetes-Cluster. Nachdem wir in Teil 3 dieser Artikelserie gesehen haben, welche unterschiedlichen Möglichkeiten Rancher hierzu anbietet, soll ein vollständiges Deployment beispielhaft an einer komplexen Applikation, einem HA-fähigen PostgreSQL-Cluster basierend auf dem PostgreSQL-Kubernetes-Operator von Zalando, demonstriert werden.

and a service (both named as “postgres-operator”) should be created and become active.

The “postgres-operator” listens on port 8080 and its cluster IP can be used to connect from other pods. The cluster IP is shown at *Resources > Workloads > Service Discovery* under *service name* (in our deployment it is 10.43.103.44, see figure 2).

In the next step, we need to deploy the web user interface which is going to work with the already deployed “postgres-operator” service API. As we did not make the “postgres-operator” accessible by public IP, we deploy the UI in the same K8s cluster where the API can be called by its cluster IP.

For this purpose, we first create the namespace “postgres-operator-ui” for the web UI and “postgres-cluster” for the

final PostgreSQL cluster. Then, we deploy the four components (Deployment, Service, Ingress, ServiceAccount) as explained in the Zalando git repository inside the “postgres-operator-ui” namespace. Before deployment, we customize the yaml files according to our K8s cluster parameters.

The deployment yaml file contains the main web application, the Service will provide a fixed endpoint to the deployed pods, the Ingress would provide a web URL to access the service, and the ServiceAccount is used by the application for authentication when creating a PostgreSQL cluster.

First download the four mentioned files from <https://github.com/zalando/postgres-operator/tree/master/ui/manifests> and then customize them as follows:

- In all four yaml files, change the namespace from “default” to “postgres-operator-ui” which we already created.
- In the deployment yaml file in the “env” section change the value for *OPERATOR_API_URL* to your service cluster IP, which in our case is “http://10.43.103.44:8080”. In the same section, change the value for *OPERATOR_CLUSTER_NAME_LABEL* to your K8s cluster name, which is in our case “playground” and the value for *TARGET_NAMESPACE* to “postgres-cluster”. You can optionally change the number of replicas, we kept it at the default value of “1”.
- In the ingress yaml file, change *apiVersion* to “extensions/v1beta1” as our K8s is at version v1.18. In the same file, change the host to “postgres.playground.rancher.gwdg.de” (to find the correct *apiVersion* of your cluster, run “kubectll --kubeconfig prod_playground.yaml api-resources”).
- Add the DNS “postgres.playground.rancher.gwdg.de” to your “/etc/hosts” with the IP of your load balancer (for a production environment, you should use a DNS service). You can ping the main domain (in our case “playground.rancher.gwdg.de”) to find the IP of the load balancer.
- Be sure that your cluster has a default storage class as “postgres-operator” will use the default one and no customization is provided by this article.

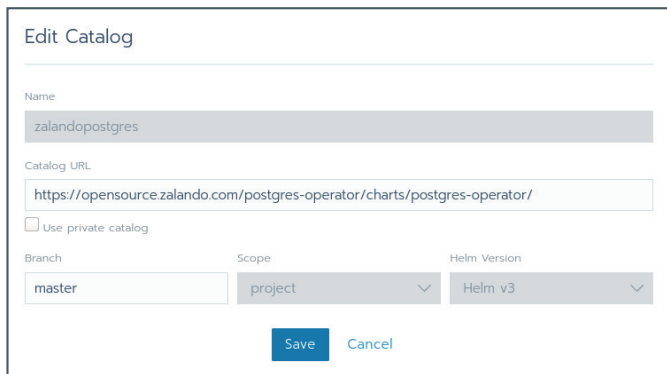


Figure 1: Define a catalog

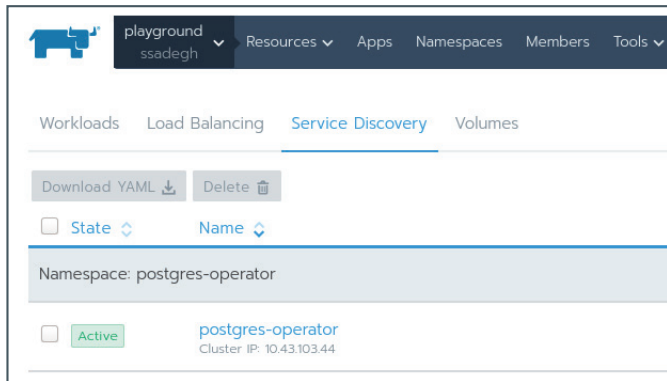


Figure 2: The “postgres-operator” service

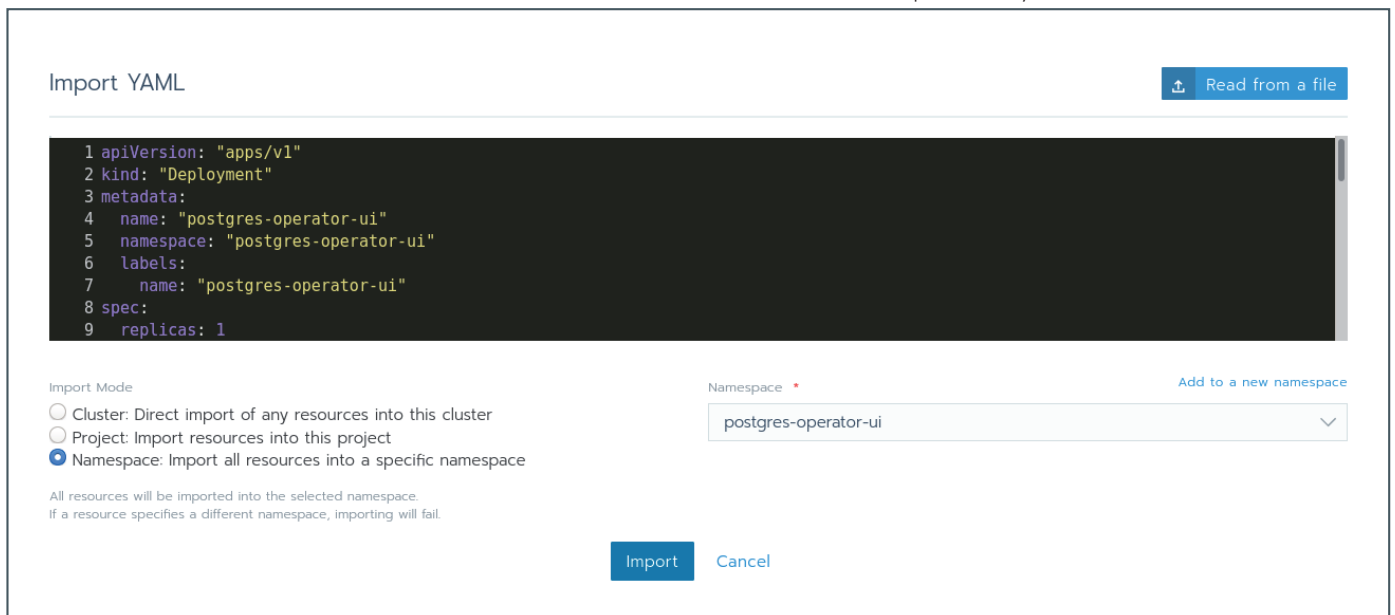


Figure 3: Rancher “ImportYAML” form

- Having the yaml files ready, there are two ways to deploy them in Kubernetes.

The first option is to use the Rancher UI. Following this approach, you should go to *Resources > Workloads* and click *Import YAML* on top-right of the page. In the displayed form, select *Namespace: Import all resources into a specific namespace* as the import mode and "postgres-operator-ui" as the namespace. Click *Read from a file* on the top-right of the page and select your deployment yaml file. It should look like figure 3. Then click *Import* to create the deployment. Repeat the process for the other three yaml files.

Alternatively, you can use the "kubectl" command to apply your yaml files. Following this approach, run the following commands to apply the files ("./manifests/" being the directory containing the four yaml files):

```
kubectl --kubeconfig prod_playground.yaml apply -f ./manifests/ -n postgres-operator-ui
```

Check the pod state using the following command and wait until the status is *Running*:

```
kubectl --kubeconfig prod_playground.yaml get pods -n postgres-operator-ui
```

After finishing this phase, the Rancher UI will look like figures 4,5, and 6.

Click the *80/http* link shown under the "postgres-operator-ui" workload to open the web UI page (see figure 7).

PROVIDE A POSTGRESQL CLUSTER USING THE WEB UI

To create a PostgreSQL cluster, select the *New Cluster* tab and fill the displayed form as follows:

- Name*: Specify your desired name for the cluster, e.g. "cluster001".
- PostgreSQL version*: Specify the *PostgreSQL* version, e.g. "13".

- Number of instances*: Specify the number of database nodes in your cluster. We left it at the default value of "1".
- Volume size*: Give a size for the volume which will be used for storing the database data. We set it to 2 "GiB" for testing purposes.
- Resources*: Specify the resources required for the application. We left it at the default value.
- Don't activate any load balancer or connection pool if your Kubernetes is running on premises (not a public cloud)

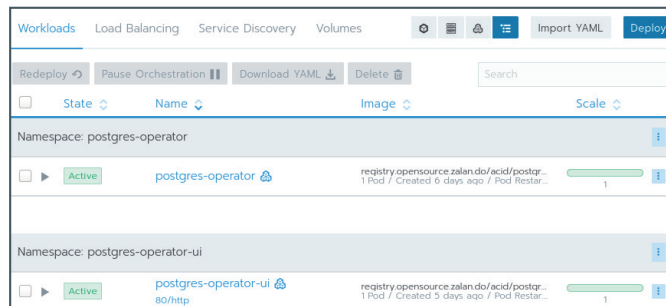


Figure 4: The workload view of "postgres-operator-ui"

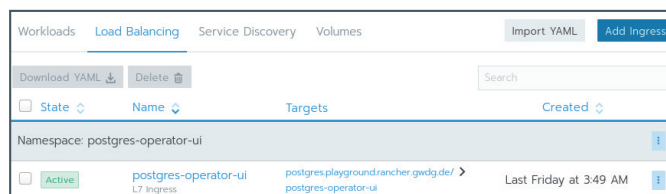


Figure 5: The ingress view of "postgres-operator-ui"

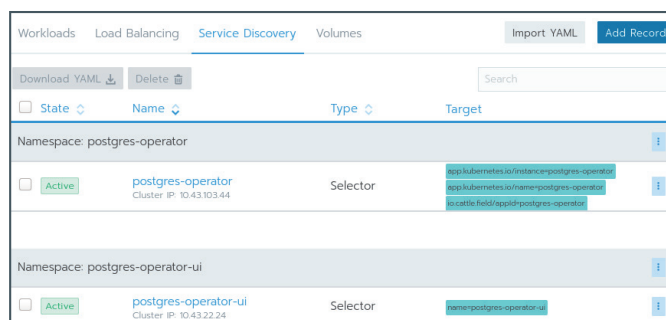


Figure 6: The service view of "postgres-operator-ui"

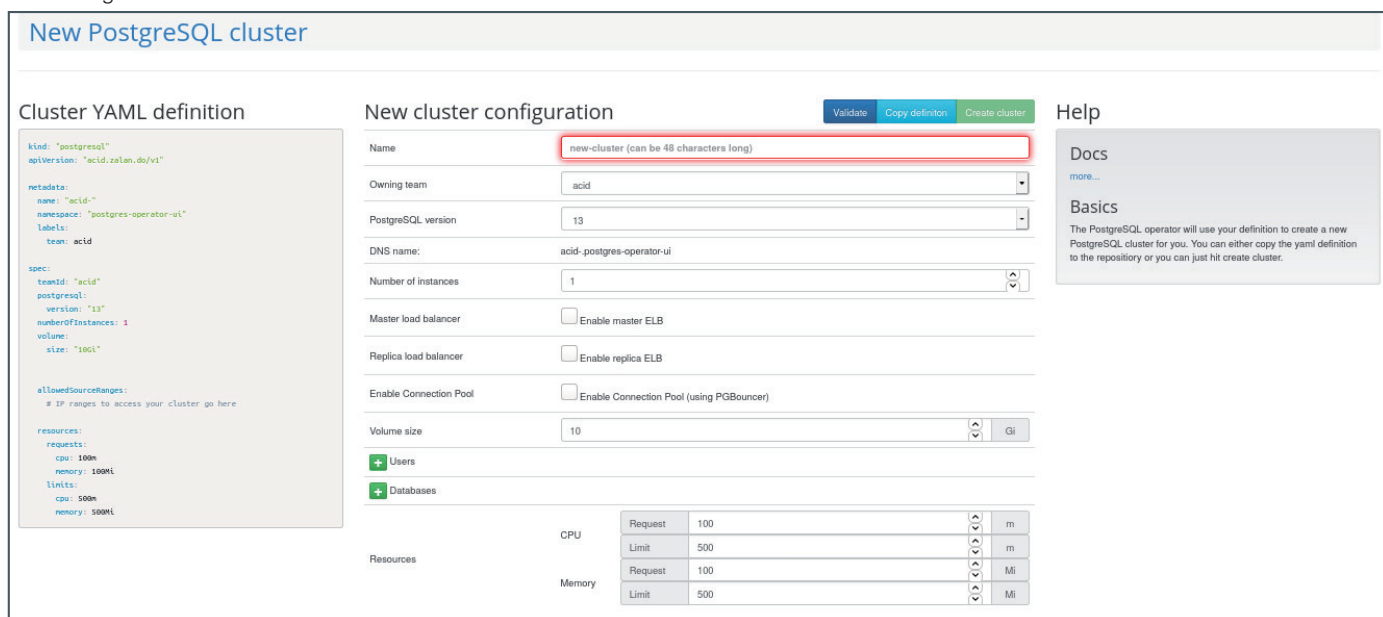


Figure 7: The web interface view for the "postgres-operator-ui"

and does not have these resources available. Although this causes your UI to not show a completed process, the cluster will be created successfully.

To create finally the cluster, first click on *Validate* to see if there is any error in the inputs, then click *Create cluster* to build your cluster (see figures 8 and 9).

In the Rancher UI, you should now be able to see a new workload, three new services and a new volume under the “postgres-cluster” namespace (see figures 10, 11 and 12).

As the final test, run the following command to check your service (“prod_playground.yaml” is the cluster kubeconfig file):

```
kubectl --kubeconfig prod_playground.yaml get all -n postgres-cluster
```

Figure 8: The filled form to create a cluster

Figure 9: The progressing state of creating a cluster

Figure 10: The workload page after deploying the PostgreSQL cluster

The output will show all created resources for delivering a PostgreSQL cluster including:

- One or more pod(s) (depending on the number you set for replica count)
- Three services
- One statefulset (which is equal to the deployment resource type but for stateful applications like databases)
- One custom resource type “postgres.acid.zalan.do”

The output for our deployment is shown in figure 13.

TEST YOUR POSTGRESQL CLUSTER

The PostgreSQL is a TCP service which behaves in the same way as the “postgres-operator”. To use it externally, you need to either configure your service as *NodePort* or deploy an ingress controller such as NGINX for your K8s cluster. In this article, as a test environment, we will use the “kubectl port-forward” command to connect to the service.

As the PostgreSQL cluster is running as a stateful application, to connect the service we need the name of the master pod (in our setup there is only one pod). The pods of a stateful application get a unique name using an ordinal index which won’t change even if the pod gets restarted. Then the IP for connecting to the service would be retrieved by the cluster DNS service. To find the name of the master pod, run the following commands:

```
kubectl --kubeconfig prod_playground.yaml get pods -o jsonpath={.items..metadata.name} -l application=spilo,cluster-name=acid-cluster001,spilo-role=master -n postgres-cluster
```

Figure 11: The service discovery page after deploying the PostgreSQL cluster

Figure 12: The created volume for the PostgreSQL cluster


```

gwdg@gwdg-pc:~/Rancher/prod$ kubectl --kubeconfig prod_playground.yaml get all -n postgres-cluster
NAME                READY   STATUS    RESTARTS   AGE
pod/acid-cluster001-0  1/1    Running   0           6d18h

NAME                TYPE          CLUSTER-IP   EXTERNAL-IP   PORT(S)    AGE
service/acid-cluster001  ClusterIP    10.43.72.203 <none>        5432/TCP    6d18h
service/acid-cluster001-config  ClusterIP    None          <none>        <none>     6d18h
service/acid-cluster001-repl  ClusterIP    10.43.215.216 <none>        5432/TCP    6d18h

NAME                READY   AGE
statefulset.apps/acid-cluster001  1/1    6d18h

NAME                TEAM   VERSION   PODS   VOLUME   CPU-REQUEST   MEMORY-REQUEST   AGE   STATUS
postgresql.acid.zalan.do/acid-cluster001  acid   13        1     2Gi     100m          100Mi           6d18h  Running

```

Figure 13: All deployed resources for the PostgreSQL cluster

In our setup, the command returns "acid-cluster001-0" as the master Pod.

Run the following command to connect your localhost port to the exposed port of your application:

```
kubectl --kubeconfig prod_playground.yaml port-forward pod/acid-cluster001-0 6432:5432 -n postgres-cluster
```

We used the port "6432", you can use any port which is unused on your local machine.

Leave your CLI terminal with this command running, then open another terminal to connect to your service.

First, install PostgreSQL client:

```
sudo apt-get install -y postgresql-client
```

Check the installation:

```
psql -version
psql --help
```

Then, complete the following steps to connect to your PostgreSQL database.

The database password is stored in a secret key and it can be found using the following command:

```
kubectl --kubeconfig prod_playground.yaml get secret postgres.acid-cluster001.credentials.postgresql.acid.zalan.do -n postgres-cluster -o 'jsonpath={.data.password}' | base64 -d
```

If you chose a different cluster name which leads to a different secret name, you can list the secret keys to find the correct name using the following command (select the one with type "Opaque" and starting with "postgres."):

```
kubectl --kubeconfig prod_playground.yaml get secret -n postgres-cluster
```

Connect to your database using the following command and enter the password you obtain from the secret key:

```
psql -U postgres -h localhost -p 6432
```

Your terminal command line now starts with "postgres=#" which shows you are successfully connected to the PostgreSQL database. To have a list of possible commands type "\?".

Some sample commands to use:

- List all databases: "\l"
- List all user accounts or roles: "\du"
- Create the database "test" (with default parameters): "create database test;"
- Connect to database "test": "\c test"
- Create the table "test_db": "create table test_db (ID INT PRIMARY KEY NOT NULL, NAME TEXT NOT NULL);"
- List the tables: "\dt"
- Insert data: "insert into test_db values (1,'test name');"
- Query the table: "select * from test_db;"

Type \q to exit from PostgreSQL database.

Type "ctrl+c" in the other terminal to terminate the port forwarding.

In the next part of these series of articles we will show how to automate the creation and management of SSL certificates inside a Kubernetes cluster by utilizing the new ACME interface available in the DFN PKI.

LINKS

- <https://github.com/zalando/postgres-operator/>
- <https://github.com/zalando/postgres-operator/blob/master/docs/administrator.md#setting-up-the-postgres-operator-ui>
- <https://www.postgresqltutorial.com/postgresql-administration/>

Leveraging Data Lakes for Managing and Processing of Large Data Volumes

Text and Contact:

Hendrik Nolte
hendrik.nolte@gwdg.de
0551 201-2119

In recent years, data lakes have become increasingly popular as central storage for unstructured data. Here, data is retained in a raw format, which ensures high reusability. Especially for use in public research institutions, the possibility of on-premise operation is just as important as the ability to perform reproducible analyses on this data on scalable compute clusters.

MOTIVATION

Due to highly specialized equipment and measurement procedures, where sometimes data is stored in rather peculiar formats, research teams often develop unique analysis workflows to process these data sets. This leads to the creation of independent data silos, i.e., isolated information systems. This means, that research teams are not easily able to exchange data and analysis workflows, but also that a cross-silo analysis is rendered impossible. This is a long-standing problem, which was first tried to be solved by introducing the concept of data warehousing in the 80's.

Here, extensive data modeling needs to be done, before disparate data from different sources can be stored in a central, physical database to yield an integrated view over all data sets. This database is usually populated by an ETL process. First, diverse data sources are accessed and the relevant data is Extracted. In order to make the different data sets uniform, they are Transformed into the before-defined and fixed schemas. Within this step, additional processing like aggregations can be done. In the end, the integrated data is Loaded into the data warehouse. Depending on the intended use of the stored data, there are drawbacks of this approach: i) During the data transformation an information loss can happen and ii) the schema-on-write approach limits the reusability of the data sets for further analysis outside of the original scope.

In 2010, James Dixon proposed a new abstract data management system, which he called a data lake, to solve these problems. In a data lake, data is not cleansed and formatted by the ETL process but data is ingested and retained in their native format within a flat hierarchy. A schema should only be inferred from the data during subsequent analysis workflows.

CONCEPT OF A DATA LAKE

Starting with the idea of storing and retaining raw data while still being able to write and execute concise queries and analysis workflows, a general concept of a data lake can be developed.

First of all, a versatile configurable but uniform ingestion interface is needed where all the different raw data can be uploaded from a range of different source systems. Thus, it should use

common protocols like HTTPS.

Thereafter, a scalable storage system is needed where all the uploaded data can be efficiently kept. Since simply dumping a vast amount of files in flat storage space will most certainly lead to an unusable state, it is crucial to maintain a data catalog. In this data catalog, all-important metadata of files within the data lake should be made available for queries.

In the next step, this raw data repository needs to be integrated with a scalable compute system to enable subsequent analysis. While serving these potentially compute-intensive workflows it is favorable to automatically reingest resulting artifacts and to transparently capture the full data lineage. The latter is easily possible since it is known, what input data were requested from the data lake, what processing has been done and what artifacts resulted from it.

Of course, uniform governance has to be enforced in all of these different areas, i.e. metadata and data storage and the integrated compute capabilities.

GENERAL WORKFLOW ON A DATA LAKE

The general workflow to work on a data lake, as shown in figure 1, always starts with the ingestion of raw data. Before storing the data, metadata extraction of the data and the subsequent indexing in the data catalog should be done. Afterwards, the data is

Nutzung von Data Lakes für die Verwaltung und Verarbeitung großer Datenmengen

In den letzten Jahren haben Data Lakes als zentraler Datenspeicher für unstrukturierte Daten und als Grundlage für eine hohe Wiederverwendbarkeit dieser Daten immer mehr an Relevanz gewonnen. Speziell für den Einsatz in öffentlichen Forschungseinrichtungen ist hierbei die Möglichkeit eines On-Premise-Betriebs ebenso wichtig wie die Möglichkeit, reproduzierbare Analysen mit diesen Daten auf skalierbaren Rechnersystemen durchführen zu können.

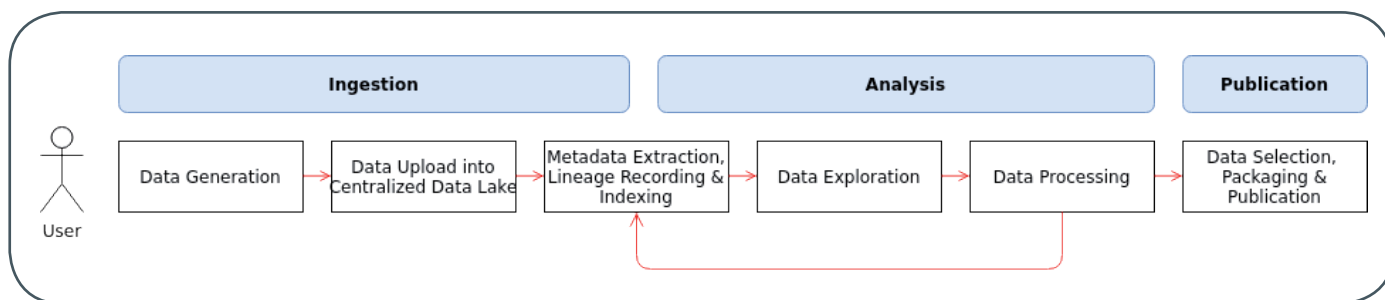


Figure 1: Workflow diagram depicting the three main stages: Ingestion, Analysis, and Publication

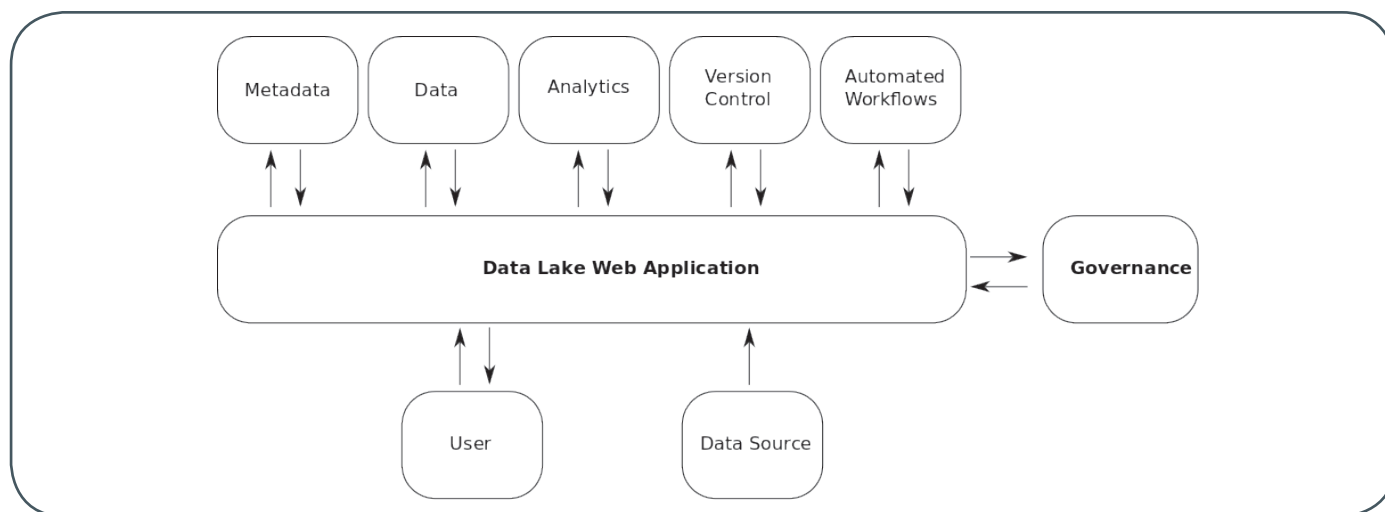


Figure 2: Sketch of the basic architectural design

available for further processing which can be triggered either automatically or manually. Particularly for a manual execution of analysis steps, it is convenient to query the data catalog in order to select the relevant data. The resulting artifacts should be ingested into the data lake, which includes the indexing of the available metadata in the data catalog.

It is also very useful to include the provenance information, which comprises all information about the exact origin. In order to share an analysis workflow with peers, it is convenient to be able to automatically retrieve all involved files in a well-structured manner, including raw and processed data, as well as execution environments, source codes, and associated data like lab notebooks.

GWDG INTERNAL DATA LAKE PROJECT

In order to gain first-hand experience with the complexity of a data lake and to further explore the possibilities particularly in research use cases, we at the GWDG started an internal data lake project. First of all, a schematic architecture had to be found which is compliant with the general requirements of a data lake:

1. A metadata catalog that enforces data quality
2. Uniform data governance across all parts of the data lake
3. Accessibility to various kinds of users
4. Integration of any type of data
5. A logical and a physical organization
6. Scalability in terms of storage and processing
7. Transparent and reliable provenance auditing

In order to be as technology agnostic as possible, a central data lake layer seemed to be the most promising option.

As depicted in figure 2, this layer is implemented as a web application, which serves as the central access point for both,

users and data sources. Within this layer the further orchestration towards the deployed services takes place. This means, that, for instance, users don't need to interact with the database directly, but that the web application will handle all necessary updates and requests, based on the higher-level operations a user performs on the data lake. This has the advantage that across all systems a consistent state is guaranteed, uniform governance can be enforced and easily scalable user management is possible.

In a prototypical implementation, this in-house developed web application was hosted in an ESX cluster and used ElasticSearch to handle the metadata, an S3 bucket in a Ceph cluster to store the data, and the local HPC system to serve to compute intensive analysis jobs.

Ingestion of Data

Before a user can ingest data into the data lake, the metadata extraction process needs to be configured. This can be done by providing a container containing an appropriate metadata extraction tool for the data which shall be uploaded. If data of a suitable type and format is now uploaded, this data will be mounted into this container and the metadata extraction tool will be executed as configured. The extracted metadata is then indexed in ElasticSearch and the data is stored in the Ceph cluster.

Performing Data Analytics

In order to perform data analytics on the data lake, users can still develop their own code as before, for instance using version control and CI/CD pipelines. Once this code should run, a container needs to be provided. Hereby, a reusable and portable execution environment is used.

Once this is done, a particular analytics task can be manually

triggered by sending a so-called job manifest to the data lake. This job manifest unambiguously describes the task which should be computed, including the input data, computational environment, compute commands, source codes, and user annotations. Based on this job manifest, the data lake creates all necessary scripts and provides all requested input files on the HPC system. The resulting artifacts are automatically ingested if specified. Since the data provenance can be solely based on these job manifests, no runtime provenance auditing is necessary, which means that there is no overhead during the actual computation. The data lake links to input and output data with the corresponding job manifest and thereby creates a provenance graph. This provenance graph allows us to easily collect all files corresponding to a certain workflow.

ACTIVITIES AT THE GWDG

At the GWDG, the development of the data lake is supported by a number of accompanying activities. As mentioned above, there is a generic data and processing infrastructure being deployed, available for all customers.

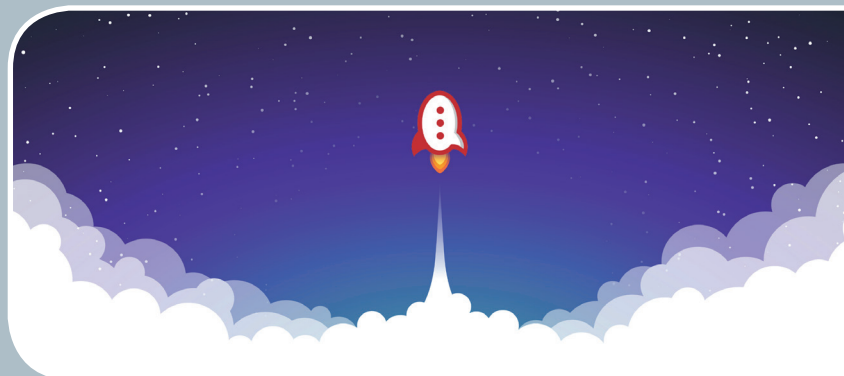
This includes a Ceph based S3 storage, virtual machines, both in an OpenStack and an ESX environment, for small-scale processing as well as a HPC system to serve even the most compute intensive tasks.

In addition, an in-house data lake solution is being developed which combines various tools within a uniform management and control layer. Here, additional software layers can be deployed, for instance to support a homogeneous governance. This is particularly important in combination with our efforts to enable secure data processing for the highest privacy levels.

During regular scheduled workshops (<https://hps.vi4io.org/events/2021/data-lake>), where everyone interested can participate, we try to bring a diverse group together to foster scientific exchange and to improve our existing solutions and steer future developments towards jointly identified problems.

SUMMARY

Data lakes are an optimal solution to prevent the creation of data silos or to integrate existing ones into a uniform data management system. In particular, keeping all data in their native format is especially for science a valuable approach since this will allow higher reusability beyond the original intended. In addition, tightly integrating a mechanism that ensures reproducible analytics by design will help researchers to be compliant with best scientific practices. ●



Rocket.Chat

Kommunikation leicht gemacht!

Ihre Anforderung

Sie benötigen einen professionellen Chat-Dienst, der eine einfache, persistente Kommunikation mit Kollegen ermöglicht – sowohl in Einzel- als auch in Gruppenunterhaltungen, die komfortabel durchsuchbar sind. Sie wollen Bilder und Dateien mit Kolleg*innen austauschen..

Unser Angebot

Wir betreiben den Messaging-Dienst „Rocket.Chat“, der es Ihnen ermöglicht, sich in Teams, Gruppen oder auch einzeln auszutauschen. Der Dienst unterstützt zusätzlich Emojis, das Versenden von Dateien, Bildern und Videos sowie die Integration von Benachrichtigungen verschiedener Dienste wie z. B. GitLab. Aufgrund einer breiten Palette von Clients, auch für mobile Geräte, sowie einer übersichtlichen Weboberfläche bieten wir komfortablen Zugriff vom Arbeitsplatz und von unterwegs..

Ihre Vorteile

- > Einfache Kommunikation im Team
- > Persistente, durchsuchbare Chat-Verläufe
- > Einfaches Teilen von Dateien und Bildern
- > Unterhaltungen mit allen Nutzer*innen, die einen Account bei der GWDG besitzen
- > Integrierte Bots und APIs für die Anbindung von GitLab oder die Einbindung von RSS-Feeds

Interessiert?

Jede*r Nutzer*in mit einem gültigen Account bei der GWDG und einem aktuellen Webbrowser oder Client kann den Dienst „Rocket.Chat“ nutzen. Für die Benutzung rufen Sie einfach <https://chat.gwdg.de> auf. Nutzer*innen ohne GWDG-Account können einen Account auf <https://www.gwdg.de/registration> registrieren.

>> www.gwdg.de/rocket.chat

Nächste Ausfahrt Transformation – Blick auf die IT der MPG im Wandel

Text und Kontakt:

Organisationsteam des DV-Treffens
dv-treffen@mpg.de

Das diesjährige 38. DV-Treffen der Max-Planck-Institute beleuchtete wesentliche IT-Themen und gab einen ersten Vorgeschmack auf die zukünftige Realität in der IT der Institute und Einrichtungen der MPG. Aktuelle IT-Themen, ein Blick in unsere Zukunft mit IT und die essenziellen Herausforderungen im Zusammenhang mit den personellen Ressourcen waren, neben vielen Aspekten mit und um IT, Kernpunkte des Treffens, das vom 28. bis 30. September 2021 coronabedingt wiederum nur virtuell stattfinden konnte.

Nicht nur gesellschaftspolitisch stehen wichtige Transformationen vor uns. Eine neue Bundesregierung, eine veränderte Verkehrspolitik und die globalen Anstrengungen zur Klimarettung sind nur einige offensichtliche Themen. Ranga Yogeshwar dachte in seiner Keynote essenzielle Themen weiter und nahm uns ein Stückchen des Weges mit.

Und wie sieht es in der MPG aus? Die Bestrebungen von Klaus Blaum (Vizepräsident CPT-Sektion) und Ariane Rauschek (Hauptabteilungsleiterin Bau/IT) zu einer stärkeren Digitalisierung stehen hier erwartungsgemäß ganz vorne. Aber auch die internen IT-Abteilungen in den Instituten und Einrichtungen der MPG stehen vor substantziellen personellen Herausforderungen. Viele neue zusätzliche Aufgaben (DSGVO, vAP, Cloud, Mobile Working, IT-Security, um nur einige zu nennen) müssen bewältigt werden und dies nicht unter den besten Voraussetzungen.

DIGITALISIERUNG

Nach der Eröffnung durch das DV-Organisationsteam stellte Ramin Yahyapour (Geschäftsführer der GWDG) aktuelle Veränderungen bei der GWDG vor. Auch hier ist durch den laufenden Umzug in den Rechenzentrums-Neubau und eine stärkere Digitalisierung eine Transformation im Gang. Neue Cloud-Dienste, IdM, Forschungsdatenmanagement und (spezifische) Projekte aus dem Umfeld der NFDI prägen das Tätigkeitsfeld der GWDG.

Auch Erwin Laure (Direktor der MPCDF) steuerte mit der Information zur Abnahme des Raven-Clusters ein Stück zum Wandel der IT bei. Ein komplett neues Betätigungsfeld wird mit AI & HPDA in der MPCDF beschritten. Auch die dargestellten neuen Entwicklungsmöglichkeiten in Martinsried zeigen klar die anstehenden Veränderungen.

Ohne eine funktionierende und fundamentale Basis innerhalb der Instituts-IT wäre dieser Wandel allerdings nicht vorstellbar. Hier hat Jörg Herrmann (IT-Sprecherkreis) den Kern der InIT-Community und deren wichtigen Beitrag zum Gelingen einer IT-Gesamtstruktur in der MPG dargestellt. Sowohl die Mitarbeit



Ralf Poscher

in den bedeutenden Gremien (BAR, Software-Kommission, KnITS etc.) und Projekten der MPG (IAM, SSO etc.), als auch die Stärkung der Forschungs-IT in Deutschland zusammen mit FhG, Leibniz und Helmholtz sind Kernpunkte unserer IT-Community.

Abseits der eher handfesten IT-Kernthemen konnte Ralf Poscher (Geschäftsführender Direktor am MPI zur Erforschung von Kriminalität, Sicherheit und Recht) einen sehr interessanten Blick auf die Frage des Grundrechts auf Selbstbestimmung

Next Exit Transformation – A Look at IT at the MPS in Transition

This year's 38th IT meeting of the Max Planck Institutes highlighted essential IT topics and gave a first taste of the future reality in IT at the institutes and facilities of the Max Planck Society (MPS). Current IT topics, a look into our future with IT and the essential challenges in the context of human resources were, among many aspects with and around IT, core issues of the meeting, which could again only take place virtually from 28 to 30 September 2021 due to Corona.

werfen. Stellt bereits die Sammlung von Daten eine Beeinträchtigung der eigenen Freiheit dar? Eigentlich nicht, da aber eine Möglichkeit des Missbrauchs besteht, stellt bereits eine Sammlung eine Gefahr dar! Beispiele zu Alumni-Daten und der namentliche Aufruf beim Arzt zeigten die formalistische Herangehensweise an das Thema, die über eine Missinterpretation dieses Grundrechts zu einer Fetischisierung des Datenschutzes mutiert.

In der Keynote von Ariane Rauschek beschrieb sie Ziele und Bedeutung der digitalen Transformation. Auf der einen Seite solle die Digitalisierung analoge Strukturen der (administrativen) Verwaltung ablösen, auf der anderen Seite könne dies nur mit vereinten Kräften, sogenannten joint forces, passieren. Eine Umfrage zur Digitalisierung der Institutsverwaltungen, die im Vorhinein stattfand, wurde genau mit diesen vereinten Kräften (Quote der Rückmeldung über 90 %) beantwortet. Allerdings blieb die Frage offen, wer und auch wie die Ergebnisse dieser Rückmeldungen analysiert und auch interpretiert werden. Eine Interpretation ist hier dringend notwendig, da die Fragen einen breiten Interpretationsspielraum geboten haben. Die notwendige Basis dieser Interpretation würde hier sicherlich durch die fachliche Expertise des IT-Sprecherkreises mitgebracht werden.

WORKSHOPS

Selbstverständlich gab es auch handfeste IT-Themen, die aktuelle Trends bei Software, Diensten und anderen Bereichen aufzeigten. Hier sind vor allem die Workshops am Vormittag des zweiten Tages zu erwähnen. Die Firma Proxmox stellte ihren neuen Backup-Server vor, der in Verbindung mit dem Virtual Environment die Virtualisierungsplattform ergänzt. Auch in Richtung Git gab es mit *DevOps mit Gitlab* eine Einführung. Und auch der Bereich der Forschungsdaten wurde mit EPICS/FAIRmat vorgestellt. Abgerundet wurden die Workshops mit der Beschreibung eines hybriden Videokonferenzsystems, das in Zeiten von Teil-Präsenz-Veranstaltungen immer mehr an Bedeutung gewinnt.

EIN BLICK IN DIE (IT-)ZUKUNFT

Einer der Höhepunkte war sicherlich die Keynote von Ranga Yogeshwar. Inhaltlich hat er einen weiten Bogen gespannt – von der aktuellen (Corona-)Lage bis hin zum nachhaltigen Wandel gesellschaftlicher Themen. Selbstverständlich kam die IT nicht zu



Ranga Yogeshwar

kurz. KI (Künstliche Intelligenz), Homeoffice, technische Komplexität und mehr flankierten seine interaktive Präsentation. Auch das Thema Fachkräftemangel kam nicht zu kurz und wurde in seiner Brisanz sogar priorisiert. Leider sieht auch er keine einfache und nachhaltige Lösung des Problems. Hier muss an „höherer Stelle“ angesetzt werden. Im Anschluss an die lebendige und erstklassige Präsentation, schloss sich noch eine knappe Stunde Frage- und Antwortrunde an. Diese lief auf „Augenhöhe“ und war sehr vielschichtig. Hier merkte man deutlich, dass er als Physiker die Wissenschaft versteht. Wir werden uns wiedersehen, Herr Yogeshwar!

IT-COMMUNITY

Aktuelle und wichtige Themen kamen natürlich auch dieses Mal von der IT-Community und deren Kontakten selbst. Von gemeinsamen Projekten (User-DB) über eine IT-Zufriedenheitsumfrage und Netzwerkthemen (Dell Networking, Campus Firewalls und DNS/DNS-over-HTTPS) bis hin zu Entwicklungsprojekten (Django, Research-Software und openBIS der ETH Zürich) war sicherlich für jeden etwas dabei.

Der Umzug der GWDG in das neue Rechenzentrum war natürlich ebenfalls Thema, daneben gab es noch Vorträge zum Organisationsmanagement, Data Lakes und der neuen PKI (TCS) von Géant. Der BAR und die Software-Grundversorgung wurden genauso vorgestellt wie die Rahmenverträge der MPG-Generalverwaltung. Zentrale Themen wie IdPortal und SSO wurden genauso wie HPC-Cloud und der Selfservice der MPCDF besprochen.

Der IT-Community-Award ging in diesem Jahr an Stefan Vollmar (MPI für Stoffwechselforschung). Hier hatten sich die Kolleg*innen im Vorhinein eine besondere Überraschung überlegt: ein Theremin, das kurz vor dem DV-Treffen persönlich (selbstverständlich unter Einhaltung aller Hygieneregeln) übergeben wurde. Ein Video dieser Übergabe zeigte diesen stimmungsvollen und bewegenden Moment.



IT-Community-Award 2021 (Stefan Vollmar – Bild Mitte)

Besonderes Augenmerk dieser Tage hat das Thema IT-Sicherheit. Neben einer gelungenen Ransomware-Übersicht wurde auch diese *Geißel der Informationstechnik* in der Tiefe betrachtet. In Zeiten von Lieferketten-Angriffen (Kaseya etc.) und Lahmlegen großer Universitäten (Uni Gießen, TU Berlin etc.) ist es nur eine Frage der Zeit, wann wir selbst die Einschläge melden. Hier sind Bestrebungen zu mehr Sicherheit von entscheidender Bedeutung. Die Frage, wie diese zusätzliche Aufgabe personell zu bewältigen ist, wurde speziell am dritten Tag thematisiert.

PERSONAL

Eine der entscheidenden Fragen dieser Tage ist die, wie die IT-Gruppen der Institute und Einrichtungen der MPG mit den stetig gestiegenen Anforderungen der letzten Jahre umgehen. Hierbei könnte man Themen wie VoIP, IoT (GLT), DSGVO, vAP, IT-Sicherheit und viele weitere nennen, die zusätzlich zum normalen IT-Alltag hinzugekommen sind. Bei diesen quantifizierbaren neuen Anforderungen ist der Fakt des reinen Technologiefortschritts noch gar nicht bedacht. Allein hier kommen jedes Jahr neue Anforderungen auf die Kolleg*innen der IT zu, die neben der operativen Arbeit (Tagesgeschäft) und dem Projektgeschäft (Migrationen, Neueinführungen, Anträge etc.) erledigt werden müssen. Die personellen Ressourcen innerhalb der IT sind andererseits über das letzte Jahrzehnt bestenfalls gleich geblieben. Die Schere zwischen Anforderungen und IT-Personal geht stetig weiter auf! Auf der anderen Seite brauchen wir Kapazitäten um die Wissenschaft adäquat zu unterstützen. Ferdi Schüth hat 2019 als damaliger Vizepräsident der MPG betont, dass Spitzenforschung Spitzen-IT bedarf, und Jochem Marotzke hat 2020 als BAR-Vorsitzender die digitale Souveränität der MPG betont. Doch wie passt das zur oben beschriebenen Realität in unseren IT-Gruppen?

Einen ersten Vorgeschmack hierauf hat uns der Vortrag von Ilko Truschkewitz von der BearingPoint GmbH gegeben. Von den

Aufgaben aktueller IT-Abteilungen über die besonderen Anforderungen im Forschungsumfeld bis hin zu den Herausforderungen in personeller Sicht – speziell in Bezug auf Konkurrenten aus der Wirtschaft – legte er die Finger in viele unserer „Wunden“. Da dieses Thema entscheidend für die IT der MPG in den nächsten Jahren sein wird, werden wir am Thema dranbleiben.

ONLINE – PRÄSENZ

Auch dieses Jahr war das 38. DV-Treffen ein Online-Erfolg! Die Aufzeichnungen der Vorträge und die Präsentationen können unter <https://dv-treffen.mpg.de/event/program/> abgerufen werden (passwortgeschützt; die Zugangsdaten sind bei den IT-Abteilungen der Max-Planck-Institute erhältlich). Im nächsten Jahr hoffen wir, dass wir uns endlich wieder in Präsenz vom 27. bis 29. September 2022 in Göttingen alle gesund wiedersehen. Bis dahin verbleiben wir mit vielen Grüßen – das DV-Organisations-Team. ●



FTP-Server

Eine ergiebige Fundgrube!

Ihre Anforderung

Sie möchten auf das weltweite Open-Source-Softwareangebot zentral und schnell zugreifen. Sie benötigen Handbücher oder Programmbeschreibungen oder Listings aus Computerzeitschriften. Sie wollen Updates Ihrer Linux- oder FreeBSD-Installation schnell durchführen.

Unser Angebot

Die GWDG betreibt seit 1992 einen der weltweit bekanntesten FTP-Server mit leistungsfähigen Ressourcen und schneller Netzanbindung. Er ist dabei Hauptmirror für viele Open-Source-Projekte.

Ihre Vorteile

- > Großer Datenbestand (65 TByte), weltweit verfügbar
- > Besonders gute Anbindung im GÖNET

- > Aktuelle Software inkl. Updates der gebräuchlichsten Linux-Distributionen
- > Unter pub befindet sich eine aktuell gehaltene locatedb für schnelles Durchsuchen des Bestandes.
- > Alle gängigen Protokolle (http, https, ftp und rsync) werden unterstützt.

Interessiert?

Wenn Sie unseren FTP-Server nutzen möchten, werfen Sie bitte einen Blick auf die u. g. Webseite. Jeder Nutzer kann den FTP-Dienst nutzen. Die Nutzer im GÖNET erreichen in der Regel durch die lokale Anbindung besseren Durchsatz als externe Nutzer.

>> www.gwdg.de/ftp-server



Treffen der Institutsadministratoren

Text und Kontakt:
Katrin Hast
katrin.hast@gwdg.de
0551 201-1808

Am 22. September 2021 fand das diesjährige, von der GWDG veranstaltete Treffen der Institutsadministratoren statt – wie schon im vergangenen Jahr coronabedingt per Videokonferenz. Bei diesen sogenannten „Admin-Treffen“ stehen Vorträge zu aktuellen Themen und der gegenseitige Informations- und Erfahrungsaustausch im Vordergrund. Im Fokus der Veranstaltung stand diesmal der Vortrag von Herrn Hans-Walter Untch von Microsoft zum Thema „Windows 11“.

Auch in diesem Jahr konnte das Admin-Treffen, das vom Active-Directory-Team der GWDG organisiert wird, coronabedingt leider nicht in Präsenz stattfinden, sondern musste per Videokonferenz abgehalten werden. Zu der jährlich stattfindenden Veranstaltung waren rund 70 Personen zugeschaltet. Wie in den Jahren zuvor gab es Vorträge zu aktuellen Themen wie z. B. Windows 11 und Public Cloud Services bei der GWDG. Es wurden aber auch Themen aufgegriffen, die durch häufige Supportanfragen „auffällig“ geworden sind, wie z. B. servergespeicherte Profile und das Speichern von Daten in den verschiedenen Speichersystemen der GWDG.

Begonnen hat das Admin-Treffen mit dem Vortrag von Frau Katrin Hast vom Active-Directory-Team zum Thema „Speicherbereiche im Vergleich“. Wichtigste Aussage war hier die Bedeutung der Kategorisierung der Daten nach Dateigröße und Häufigkeit des Zugriffs und die sich aus diesen Eigenschaften resultierende Speicherung der Daten in den verschiedenen dafür vorgesehenen Speichersystemen. In diesem Zusammenhang wurde das Angebot gemacht, den Instituten ein Protokoll für das gemeinsame Laufwerk zur Verfügung zu stellen, in dem erkennbar ist, welche Daten länger nicht mehr verwendet wurden und somit archiviert werden könnten.

Herr Hans-Walter Untch von Microsoft berichtete anschließend über die Neuerungen des Windows-11-Betriebssystems, das seit dem 05.10.2021 zur Verfügung steht. Die wichtigsten Neuerungen umfassen ein überarbeitetes Userinterface und eine verbesserte Sicherheit, deren Funktionen es zwar bereits schon in Windows 10 gab, die aber mit Windows 11 nun standardmäßig aktiviert sind. Das ist auch der Grund für die erhöhten Hardwareanforderungen. Wie erwartet wird auch die Unterstützung der Microsoft-Cloud-Dienste weiter verbessert, so z. B. das kollaborative Arbeiten mit Teams.

Microsoft verspricht, mit Windows 11 auch eine hohe Kompatibilität zu Windows 10 zu gewährleisten. Künftig soll es auch nur noch ein Update pro Jahr (Herbst-Release) geben und der Support-Zeitraum von 30 auf 36 Monate erweitert werden.

Herr Benedict Spermoser von der Abteilung IT der Philosophischen Fakultät berichtete über die Herausforderungen und Lösungen in 1,5 Jahren Homeoffice in der Pandemie. Anschließend brachte uns Herr Patrick Becker von der GWDG noch mal

das Thema „Clientmanagement mit baramundi“ näher. Die Software ermöglicht auch die Verwaltung der Windows-Clientsysteme im Homeoffice. Damit hat sie den Bordmitteln des Active Directory etwas voraus.

Der geplante Vortrag „Gemeinsame Dokumentationsplattform Confluence“ musste leider entfallen. Dafür hat uns Frau Martina Willmann (GWDG) über die aktuellen Möglichkeiten in Bezug auf servergespeicherte Profile informiert und die Teilnehmer*innen zu einer Diskussion über den künftigen Einsatz der Profile animiert. Im Bereich servergespeicherte Profile hat und wird es in Zukunft Anpassungen geben, die zu einer Verschlankung des Profils führen. Diese Änderungen werden dann selbstverständlich zeitnah über die Mailingliste GWDG-AD (gwdg-ad@gwdg.de) bekannt gemacht.

Abschließend berichtete Herr Nikolaj Kopp (GWDG) über den Status der Public-Cloud-Umgebungen und den Ausbau der Microsoft-Cloud-Dienste bei der GWDG.

Insgesamt war das Admin-Treffen eine gelungene Veranstaltung, in der aktuelle Themen aufgegriffen und zum Teil intensiv wurden. Besonders gefreut haben uns die Beiträge von Microsoft von Herrn Untch und von Herrn Spermoser von der Abteilung IT der Philosophischen Fakultät.

Wie jedes Jahr werden wir die PowerPoint-Folien der Vorträge in der GWDG ownCloud sammeln und den Link über die GWDG-AD-Mailingliste bekannt geben. Als Austauschplattform über das Admin-Treffen hinaus empfehlen wir unseren GWDG-AD-Kanal im Rocket.Chat (<https://chat.gwdg.de/channel/gwdg-ad>). ■

Meeting of the Institute Administrators

On September 22, 2021, this year's meeting of the institute administrators, organized by the GWDG, took place – as last year, due to corona, via video conference. These so-called "admin meetings" focus on presentations on current topics and the mutual exchange of information and experience. This time, the focus of the event was the lecture by Mr. Hans-Walter Untch from Microsoft on the topic of „Windows 11“.

Kurz & knapp

Prof. Dr. Julian Kunkel neuer Vorsitzender der OSSG

Prof. Dr. Julian Kunkel ist seit dem 21. Oktober 2021 Vorsitzender der Open Source Specialist Group (OSSG) in der British Computer Society (BCS).

Digitale Souveränität ist für die GWDG und ihre Kund*innen ein grundlegendes Thema. Wir sind daher bemüht, Alternativen zu kommerziellen Lösungen zu unterstützen, um Herstellerabhängigkeit zu vermeiden. Open-Source-Lösungen sind ein Eckpfeiler in dieser Strategie. Die GWDG bietet viele Open-Source-Produkte als Services für ihre Nutzer*innen an und setzt sich für die Weiterentwicklung dieser Produkte ein, um die Benutzererfahrung stetig zu verbessern.

Mit mehr als 1.500 Mitgliedern beschäftigt sich die OSSG mit allen Facetten von Open-Source-Software und Hardware und treibt mit ihren Mitgliedern verschiedenste Aktivitäten voran. Neben der Organisation themenspezifischer Abende im Headquarter in London werden Initiativen wie RISC-V und RUST unterstützt. Alle Treffen werden als hybrides Ereignis internationalen Gästen kostenlos per Videokonferenz mittels BigBlueButton (BBB) der GWDG zur Verfügung gestellt. Die OSSG nutzte bereits zuvor BBB auf einem eigenen Server bei AWS, ist mit der bisherigen Erfahrung bei der Nutzung von BBB der GWDG aber höchst zufrieden.

Die BCS ist die Vereinigung von IT-Experten in England und kann als Analogon zur Gesellschaft für Informatik (e. V.) verstanden werden. Mit mehr als 60.000 Mitgliedern in 150 Ländern ist die Mission der BCS, die technischen Standards anzuheben, um die Potenziale der IT-Industrie zu erhöhen.

Links

OSSG: <https://ossg.bcs.org/>

BCS: <https://www.bcs.org/about-us/>

Otto

Außerbetriebnahme des WINS-Dienstes der GWDG zum 01.02.2022

Seit über 20 Jahren stellt die GWDG für die Computernamensauflösung von älteren Windows-Betriebssystemen den Dienst „Windows Internet Name Service“ (WINS) bereit. Dieser Dienst wird zum 01.02.2022 außer Betrieb genommen und abgeschaltet.

Der WINS-Dienst ist ein dynamischer Registrierungs- und Auflösungsdienst, der NetBIOS-Computernamen zu deren registrierten TCP/IPv4 Adressen zuordnet.

WINS war die Haupttechnologie zur Netzwerkadressenauflösung seit Windows NT. Spätestens ab Windows 7 wird jedoch, primär aus Sicherheits- und Kompatibilitätsgründen, von Microsoft das hierarchische „Domain Name System“-Protokoll (DNS) empfohlen.

macOS und *NIX-Systeme verwenden für die Namensauflösung im GÖNET und im World Wide Web den Dienst DNS und sind von dieser Änderung nicht betroffen.

Es können nach der Abschaltung von WINS in der Namensauflösung Probleme auftreten, wenn alle folgenden Punkte zutreffen:

- Bisherige Netzwerkzugriffe haben über WINS stattgefunden (z. B. bei Netzwerkdruckern oder Netzlaufwerken)
- Statt des „Fully Qualified Domain Name“ (FQDN) wurde lediglich der Servername genutzt (z. B. `\\gwdg-print` statt `\\gwdg.print.gwdg.de`).
- Der anfragende Windows-Rechner befindet sich nicht in derselben DNS-Domäne und der DNS-Suffix wurde nicht auf Windows-Ebene eingepflegt.

Ob Sie von Namensauflösungsproblemen in der Zukunft betroffen sein werden, lässt sich einfach testen. Hierfür einfach im Vorfeld auf den einzelnen Windows-Geräten in den Netzwerkkarteneinstellungen den WINS-Server entfernen. Sollte das Entfernen zu Namensproblemen führen, kann der WINS-Server wieder hinzugefügt werden und der Zugriffspfad als FQDN an den benötigten Stellen eingepflegt werden.

Quentin

Stellenangebot

Nr. 20211122

Die GWDG sucht zum nächstmöglichen Zeitpunkt zur Verstärkung des High-Performance-Computing-Teams der Arbeitsgruppe „eScience“ (AG E) drei

Expert*innen (m/w/d) für Deep Learning

mit einer regelmäßigen Wochenarbeitszeit von 39 Stunden. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist je nach Qualifikation bis zur Entgeltgruppe TVöD E 13 vorgesehen. Die Stellen sind teilzeitgeeignet und zunächst auf zwei Jahre befristet. Die GWDG strebt eine langfristige Zusammenarbeit an. Bei Interesse besteht die Möglichkeit zur Promotion

In 2020 wurde die Universität Göttingen mit der GWDG als eines von acht Rechenzentren in den Verbund Nationales Hochleistungsrechnen (NHR) aufgenommen und betreibt mit dem HLRN-IV-System „Emmy“ einen der leistungsstärksten Rechner der Welt. Des Weiteren wurde in Göttingen das Campus-Institut Data Science (CIDAS) gegründet, mit dem fakultätsübergreifend am gesamten Campus Forschung und Lehre im Bereich Data Science gefördert werden.

Die GWDG betreibt Cluster mit GPU-Systemen und stellt diese ihren Nutzer*innen zur Verfügung. In diesem Kontext ist die Weiterentwicklung der Expertise zu skalierbarer KI auf HPC-Systemen am Standort elementar.

Für den Ausbau des KI- und Deep-Learning-Teams suchen wir drei engagierte Mitarbeiter*innen mit einem nachgewiesenen Interesse in den Bereichen Deep Learning, maschinelles Lernen und KI. Sie möchten an der Verknüpfung von Hochleistungsrechnen und Deep Learning mitwirken, interdisziplinär arbeiten, neue Möglichkeiten zur Parallelisierung entwickeln oder Modelle auf maximale Performance im HPC-Bereich optimieren? Dann bewerben Sie sich!

Aufgabenbereiche

- Untersuchung der Skalierbarkeit verschiedener Algorithmen im maschinellen Lernen auf HPC-Systemen und Optimierung
- Entwicklung eigener Forschungsprojekte, Kooperationsprojekte und Services im Bereich maschinelles Lernen und Deep Learning auf HPC-Systemen
- Verbesserung von bestehenden HPC-Workflows und Simulationen durch Integration von KI
- Parallelisierung von Modellen mit Hilfe von GPUs und anderen Beschleunigern
- Unterstützung von Forschenden am Standort Göttingen im Bereich HPC und Deep Learning
- Entwicklung von Workshops an der Schnittstelle

zwischen HPC und maschinellem Lernen

- Unterstützung bei der Beratung der Nutzer*innen zum Thema KI / maschinelles Lernen

Anforderungen

- Abgeschlossenes Hochschulstudium oder vergleichbare Qualifikation mit einschlägiger Berufserfahrung
- Erfahrungen bei der Anwendung von KI-Methoden in der Wissenschaft, bspw. Medizin, Forstwirtschaft, Life Sciences oder Digital Humanities
- Theoretische Kenntnisse im Bereich maschinelles Lernen / Deep Learning / KI
- Erfahrungen bei der Nutzung von HPC-Systemen
- Wünschenswert sind Kenntnisse aus dem Performance Engineering und Erfahrungen in der GPU-Programmierung
- Gute Programmierkenntnisse in Python und anderen relevanten Sprachen
- Gutes analytisches Denkvermögen
- Selbstständige, strukturierte und systematische Arbeitsweise
- Ausgeprägte Team- und Kommunikationsfähigkeit
- Sehr gute Deutsch- und Englischkenntnisse in Wort und Schrift

Unser Angebot

- Flexible Arbeitszeiten und die Möglichkeit zu mobilem Arbeiten
- Ein modernes, vielfältiges und außergewöhnliches Arbeitsumfeld mit großer Nähe zu Wissenschaft und Forschung an der Schnittstelle mehrerer innovativer Technologiesektoren
- Mitarbeit in einem kompetenten und engagierten Team
- Unterstützung bei der Qualifizierung und Weiterentwicklung Ihrer Fähigkeiten
- Sozialleistungen des öffentlichen Dienstes

Die GWDG strebt nach Geschlechtergerechtigkeit und Vielfalt und begrüßt daher Bewerbungen jedes Hintergrunds. Die GWDG ist bemüht, mehr schwerbehinderte Menschen zu beschäftigen. Bewerbungen Schwerbehinderter sind ausdrücklich erwünscht.

Haben wir Ihr Interesse geweckt? Dann bitten wir um eine Bewerbung über unser Online-Formular unter <https://s-lotus.gwdg.de/gwdgdb/age/20211122.nsf/bewerbung>.

Das Auswahlverfahren wird in zwei getrennten Verfahren durchgeführt. Beim ersten Auswahlverfahren werden die Bewerbungen berücksichtigt, die **bis zum 03.12.2021** eingehen, beim zweiten Auswahlverfahren dann die Bewerbungen, die nach dem 03.12.2021 und **bis zum 10.01.2022** eingehen.

Fragen zu den ausgeschriebenen Stellen beantwortet Ihnen:

Herr Prof. Dr. Julian Kunkel

E-Mail: julian.kunkel@gwdg.de



Stellenangebot

Nr. 20211104

Die GWDG sucht zum nächstmöglichen Zeitpunkt zur Verstärkung des High-Performance-Computing-Teams der Arbeitsgruppe „eScience“ (AG E) eine

Fachkraft (m/w/d) für Bioinformatik- Services

mit einer regelmäßigen Wochenarbeitszeit von 39 Stunden. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist je nach Qualifikation in den Entgeltgruppen TVöD E 11 bis TVöD E 13 vorgesehen. Die Stelle ist teilzeitgeeignet und zunächst auf zwei Jahre befristet. Die GWDG strebt eine langfristige Zusammenarbeit an.

Die HPC-Systeme der Universität Göttingen / GWDG werden durch diverse Forschungsgruppen aus dem Bereich der Lebenswissenschaften genutzt. Dabei ist es wichtiger Bestandteil unserer Arbeit, diese Forschenden optimal zu unterstützen und ihnen HPC-Hardware und Software auf dem neuesten Stand bereitzustellen. In vielen Bereichen der computergestützten Lebenswissenschaften nehmen Machine Learning und die Analyse großer Datensätze einen stetig wachsenden Stellenwert ein. Zur Unterstützung dieses Feldes stellt die GWDG den Forschenden eine Reihe von Softwarepaketen und Hardware zur Verfügung. Dies geschieht sowohl auf den HPC-Systemen als auch über spezielle, z. T. mit Web-Schnittstellen ausgestattete Bioinformatik-Server.

Zur Verstärkung unseres High-Performance-Computing-Teams suchen wir eine*n engagierte*n Mitarbeiter*in mit Interesse an der Pflege und dem Ausbau unseres Bioinformatik-Angebots und an der Zusammenarbeit mit Forschenden im Bereich Bioinformatik sowie allgemeinem Interesse an High-Performance-Computing.

Aufgabenbereiche

Ihre Aufgaben in unserem Team werden sich u. a. folgendermaßen zusammensetzen:

- Bereitstellung, Dokumentation und Pflege verschiedener Bioinformatik-Module im
- HPC-System und von Web-Tools.
- Pflege der Lizenzserver
- Unterstützung der Forschenden bei der Nutzung der Tools und der Entwicklung ihrer Workflows
- Administration der Linux-Server und des Windows-Dialogservers im Bereich Bioinformatik

Anforderungen

- Abgeschlossene Ausbildung oder vergleichbare Qualifikation mit Erfahrung im Bereich Administration von Linux-Systemen
- Sicherer Umgang mit der Kommandozeile
- Erfahrung mit Skriptsprachen, z. B. Bash, R oder Python
- Kenntnis von relevanten Softwarepaketen aus dem Bereich der Bioinformatik, bspw. Galaxy, GeneXplain, MASCOT oder RStudio, sind vorteilhaft.
- Gutes analytisches Denkvermögen
- Selbstständige, strukturierte und systematische Arbeitsweise
- Ausgeprägte Team- und Kommunikationsfähigkeit
- Gute Englischkenntnisse in Wort und Schrift, Grundkenntnisse in Deutsch

Unser Angebot

- Flexible Arbeitszeiten und die Möglichkeit zu mobilem Arbeiten
- Ein modernes, vielfältiges und außergewöhnliches Arbeitsumfeld mit großer Nähe zu Wissenschaft und Forschung an der Schnittstelle mehrerer innovativer Technologiesektoren
- Eine interessante, vielseitige Tätigkeit in einem großen, international agierenden
- IT-Kompetenzzentrum
- Mitarbeit in einem kompetenten und engagierten Team
- Unterstützung bei der Qualifizierung und Weiterentwicklung Ihrer Fähigkeiten
- Sozialleistungen des öffentlichen Dienstes

Die GWDG strebt nach Geschlechtergerechtigkeit und Vielfalt und begrüßt daher Bewerbungen jedes Hintergrunds. Die GWDG ist bemüht, mehr schwerbehinderte Menschen zu beschäftigen. Bewerbungen Schwerbehinderter sind ausdrücklich erwünscht.

Haben wir Ihr Interesse geweckt? Dann bitten wir um eine Bewerbung bis zum **05.12.2021** über unser Online-Formular unter <https://s-lotus.gwdg.de/gwdgdb/age/20211104.nsf/bewerbung>.

Fragen zur ausgeschriebenen Stelle beantwortet Ihnen:

Herr Prof. Dr. Julian Kunkel

E-Mail: julian.kunkel@gwdg.de oder

Herr Dr. Christian Boehme

E-Mail: christian.boehme@gwdg.de

Stellenangebot

Nr. 20211117

Die GWDG sucht ab sofort zur Unterstützung des High-Performance-Computing-Teams der Arbeitsgruppe „eScience“ (AG E) eine

Studentische Hilfskraft (m/w/d)

mit einer Beschäftigungszeit von bis zu 80 Stunden im Monat. Die Vergütung erfolgt entsprechend den Regelungen für Studentische/Wissenschaftliche Hilfskräfte. Die Stelle ist zunächst auf ein Jahr befristet mit der Option auf Verlängerung.

In 2020 wurde die Universität Göttingen mit der GWDG als eines von acht Rechenzentren in den Verbund Nationales Hochleistungsrechnen (NHR) aufgenommen und betreibt mit dem HLRN-IV-System „Emmy“ einen der leistungsstärksten Rechner der Welt.

Sie möchten uns bei der Entwicklung eines speziell für Forschungsanwendungen des Max-Planck-Instituts für chemische Energiekonversion (MPI CEC) entworfenen Data Lakes unterstützen und dabei eigenständig wissenschaftliche Workflows implementieren? Dann bewerben Sie sich!

Aufgabenbereiche

- Entwurf, Implementierung, Test und Dokumentation von Funktionalitäten des Data Lakes
- Erstellung automatisierter Deployments für den Data Lake
- Enge Zusammenarbeit mit den involvierten Wissenschaftler*innen für
 - › Kundensupport
 - › Requirement Engineering
- Datenmodellierung diverser Datensets
- Entwurf, Implementierung, Test und Dokumentation kanonischer wissenschaftlicher Workflows

Anforderungen

- Gute Linux-Kenntnisse
- Sehr gute Python-Kenntnisse
- Erfahrung im Umgang mit und der Entwicklung von RestAPIs
- Gutes analytisches Denkvermögen
- Selbstständige, strukturierte und systematische Arbeitsweise
- Ausgeprägte Team- und Kommunikationsfähigkeit
- Gute Englischkenntnisse in Wort und Schrift

Unser Angebot

- Flexible Arbeitszeiten und die Möglichkeit zu mobilem Arbeiten
- Ein modernes, vielfältiges und außergewöhnliches Arbeitsumfeld mit großer Nähe zu Wissenschaft und Forschung an der Schnittstelle mehrerer innovativer Technologiesektoren
- Eine interessante, vielseitige Tätigkeit in einem großen, international agierenden IT-Kompetenzzentrum
- Mitarbeit in einem kompetenten und engagierten Team
- Unterstützung bei der Qualifizierung und Weiterentwicklung Ihrer Fähigkeiten

Die GWDG strebt nach Geschlechtergerechtigkeit und Vielfalt und begrüßt daher Bewerbungen jedes Hintergrunds. Die GWDG ist bemüht, mehr schwerbehinderte Menschen zu beschäftigen. Bewerbungen Schwerbehinderter sind ausdrücklich erwünscht. Haben wir Ihr Interesse geweckt? Dann bitten wir um eine Bewerbung **bis zum 01.12.2021** über unser Online-Formular unter <https://s-lotus.gwdg.de/gwdgdb/age/20211117.nsf/bewerbung>.

Fragen zur ausgeschriebenen Stelle beantwortet Ihnen:

HPC-Team

E-Mail: hpc-team@gwdg.de oder

Herr Prof. Dr. Julian Kunkel

E-Mail: julian.kunkel@gwdg.de



NEUER MITARBEITER CHRISTIAN LORENZEN

Seit dem 1. August 2021 unterstützt Herr Christian Lorenzen in der Arbeitsgruppe „Basisdienste und Organisation“ (AG O) die Arbeiten in den Bereichen Authentifizierungs- und Autorisierungs-Infrastrukturen (AAI) und Single Sign-On (SSO). Sein Tätigkeitsschwerpunkt liegt in der Weiterentwicklung und Verbesserung von bestehenden, zumeist auf der Programmiersprache PHP basierenden SSO-Lösungen für die MPG, Universität Göttingen, GWGD und Academic Cloud. Herr Lorenzen schloss 2015 sein Studium der Elektrotechnik/Informationstechnik an der Hochschule für Angewandte Wissenschaft und Kunst (HAWK) Hildesheim/Holzminde/Göttingen mit einem Bachelor of Engineering ab und konnte in den vergangenen Jahren berufliche Erfahrungen in der Softwareentwicklung sammeln. Herr Lorenzen ist per E-Mail unter christian.lorenzen@gwdg.de zu erreichen.



Pohl



NEUER MITARBEITER EMMANUEL TCHOUMKEU-NGATAT

Seit dem 1. August 2021 ist Herr Emmanuel Tchoumkeu-Ngatat als studentische Hilfskraft in der Arbeitsgruppe „eScience“ (AG E) tätig. Dort unterstützt er das HPC-Team im Bereich Anwendungssoftware und Statistik-Tools. Im Rahmen seines Bachelor-Studiums in den Fächern Mathematik und Volkswirtschaftslehre an der Georg-August-Universität Göttingen hat er bereits Erfahrungen mit statistischen Methoden (insbesondere über das Tobit-Modell) sowie u. a. RStudio und Python gesammelt. Herr Tchoumkeu-Ngatat ist per E-Mail unter emmanuel.tchoumkeu-ngatat@gwdg.de zu erreichen.

Köhler

NEUER MITARBEITER ALEXANDER GOLDMANN

Seit dem 6. Oktober 2021 ist Herr Alexander Goldmann als Community Manager für den NHR@Göttingen in der Arbeitsgruppe „eScience“ (AG E) tätig. Seine Aufgaben liegen im HPC-Team unter anderem beim Aufbau und der Unterstützung von lokalen und nationalen HPC-Community und der Entwicklung und Umsetzung von PR-Strategien. Herr Goldmann hat an der Ostfalia Hochschule für angewandte Wissenschaften am Standort Salzgitter in der Fachrichtung Medienmanagement (B.A.) studiert. Im Anschluss hat er ein Volontariat als PR-Berater in der Berliner Werbeagentur „Zum goldenen Hirschen“ absolviert. Darauf folgend war er im Berliner Coworking Space des St. Oberholz als Community Manager zum einen für den Aufbau und die Pflege der Community und zum anderen für das komplette Marketing und die PR zuständig. Im Rahmen dieser Tätigkeit sammelte er zudem Erfahrungen in der Betreuung verschiedener Community-Mitglieder. Neben der Arbeit als Community Manager war er auch für die Planung und Umsetzung von Events, Schulungen, Workshops und anderweitigen Veranstaltungen sowohl für interne als auch externe Teilnehmer*innen zuständig. Herr Goldmann ist per E-Mail unter alexander.goldmann@gwdg.de zu erreichen.



Boehme

NEUE MITARBEITERIN GESA SALOGA

Seit dem 15. August 2021 verstärkt Frau Gesa Saloga das Support-Team im Helpdesk der GWDC als wissenschaftliche Hilfskraft. Ihre Haupttätigkeiten sind der First-Level-Support, wo sie Anfragen ratsuchender Nutzer*innen entgegennimmt, diesen direkt hilft oder die Anfragen an Kolleg*innen im Second-Level-Support zur weiteren Bearbeitung übergibt. Frau Saloga studiert zurzeit an der Georg-August-Universität Göttingen im Fach Kulturanthropologie/Soziologie.

Helmvoigt



NEUER MITARBEITER MOHAMMAD HOSSEIN BINIAZ

Seit dem 1. September 2021 ist Herr Mohammad Hossein Biniaz als studentische Hilfskraft in der Arbeitsgruppe „eScience“ (AG E) tätig. Dort unterstützt er das HPC-Team bei der Softwareentwicklung zur Integration der HPC-Systeme in andere Dienste. Im Rahmen seines Master-Studiums hat er bereits Erfahrungen mit Machine Learning und Paralleler Programmierung gemacht und mit Python, C und R gearbeitet. Herr Biniaz ist per E-Mail unter mohammad-hossein.biniaz@gwdg.de zu erreichen.

Köhler

NEUER MITARBEITER LARS QUENTIN

Seit dem 1. Oktober 2021 ist Herr Lars Quentin als studentische Hilfskraft in der Arbeitsgruppe „eScience“ (AG E) tätig. Sein Aufgabenbereich liegt dort in der Unterstützung des HPC-Teams im Hinblick auf proaktive Forschung und Entwicklung des Angebots. Als erste Aufgabe forscht er an der Weiterentwicklung eines Blackbox-Ansatzes zur automatischen Klassifizierung von I/O-Anfragen. Hierbei soll automatisch durch statische Informationsaggregation sowie aktives Mikro-Benchmarking und darauffolgender statistischer Analyse ein Modell zur Leistungscharakterisierung einzelner Rechenknoten erfolgen. Anhand dieses Modells soll es zukünftig möglich sein, suboptimale I/O-Zugriffe zu klassifizieren und mittels Anomalieerkennung präventive Wartung der Rechenknoten zu ermöglichen. Herr Quentin studiert zurzeit Angewandte Informatik sowie Mathematik an der Georg-August-Universität Göttingen. Während seines Studiums hat er bereits als Systemadministrator und Entwickler gearbeitet. Zu seinen vorherigen Arbeitgebern zählen das Max-Planck-Institut für biophysikalische Chemie, das Max-Planck-Institut für Dynamik und Selbstorganisation sowie die Universität Göttingen. Seine Forschungsinteressen umfassen High Performance Computing mit Fokus auf Speichersysteme sowie Architekturen für verteilte Systeme. Herr Quentin ist per E-Mail unter lars.quentin@gwdg.de zu erreichen.

Kunkel



INFORMATIONEN:
support@gwdg.de
0551 201-1523



Dezember 2021

Academy

KURS	DOZENT*IN	TERMIN	ANMELDEN BIS	AE
OUTLOOK – E-MAIL UND GROUPWARE	Helmvoigt	02.12.2021 9:15 – 12:00 und 13:00 – 16:00 Uhr	25.11.2021	4
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER MIT VORKENNTNISSEN	Cordes	08.12. – 09.12.2021 9:00 – 12:00 und 13:00 – 15:30 Uhr	01.12.2021	8
WORKING WITH GRO.DATA	Király	14.12.2021 10:00 - 11:30 Uhr	13.12.2021	0

Teilnehmerkreis

Das Angebot der GWDG Academy richtet sich an die Beschäftigten aller Einrichtungen der Universität Göttingen, der Max-Planck-Gesellschaft sowie aus wissenschaftlichen Einrichtungen, die zum erweiterten Kreis der Nutzer*innen der GWDG gehören. Studierende am Göttingen Campus zählen ebenfalls hierzu. Für manche Kurse werden spezielle Kenntnisse vorausgesetzt, die in den jeweiligen Kursbeschreibungen genannt werden.

Anmeldung

Für die Anmeldung zu einem Kurs müssen Sie sich zunächst mit Ihrem Benutzernamen und Passwort im Kundenportal der GWDG (<https://www.gwdg.de>) einloggen. Wenn Sie zum Kreis der berechtigten Nutzer*innen der GWDG gehören und noch keinen GWDG-Account besitzen, können Sie sich im Kundenportal unter dem URL <https://www.gwdg.de/registration> registrieren. Bei Online-Kursen kann das Anmeldeverfahren abweichen. Genauere Informationen dazu finden Sie in der jeweiligen Kursbeschreibung. Einige Online-Angebote stehen Ihnen jederzeit und ohne Anmeldung zur Verfügung.

Absage

Absagen können bis zu sieben Tagen vor Kursbeginn erfolgen. Bei kurzfristigeren Absagen werden allerdings die für den Kurs angesetzten Arbeitseinheiten (AE) vom AE-Kontingent der jeweiligen Einrichtung abgezogen.

Kursorte

Aufgrund der aktuellen Corona-Situation finden zurzeit nahezu alle Kurse in einem geeigneten Online-Format und nicht als Präsenzkurse statt. Nähere Informationen dazu finden Sie bei den jeweiligen Kursen. Auf Wunsch und bei ausreichendem Interesse führen wir auch Kurse vor Ort in einem Institut durch, sofern dort ein geeigneter Raum mit entsprechender Ausstattung zur Verfügung gestellt wird.

Kosten bzw. Gebühren

Die Academy-Kurse sind – wie die meisten anderen Leistungen der GWDG – in das interne Kosten- und Leistungsrechnungssystem der GWDG einbezogen. Die den Kursen zugrundeliegenden AE werden vom AE-Kontingent der jeweiligen Einrichtung abgezogen. Für alle Einrichtungen der Universität Göttingen und der Max-Planck-Gesellschaft sowie die meisten der wissenschaftlichen Einrichtungen, die zum erweiterten Kreis der Nutzer*innen der GWDG gehören, erfolgt keine Abrechnung in EUR. Dies gilt auch für die Studierenden am Göttingen Campus.

Kontakt und Information

Wenn Sie Fragen zum aktuellen Academy-Kursangebot, zur Kursplanung oder Wünsche nach weiteren Kursthemen haben, schicken Sie bitte eine E-Mail an support@gwdg.de. Falls bei einer ausreichend großen Gruppe Interesse besteht, könnten u. U. auch Kurse angeboten werden, die nicht im aktuellen Kursprogramm enthalten sind.



Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen