

GWDDG NACHRICHTEN 09-10|15

DavMail

Windows 10

TSM-Backup bei FreeBSD

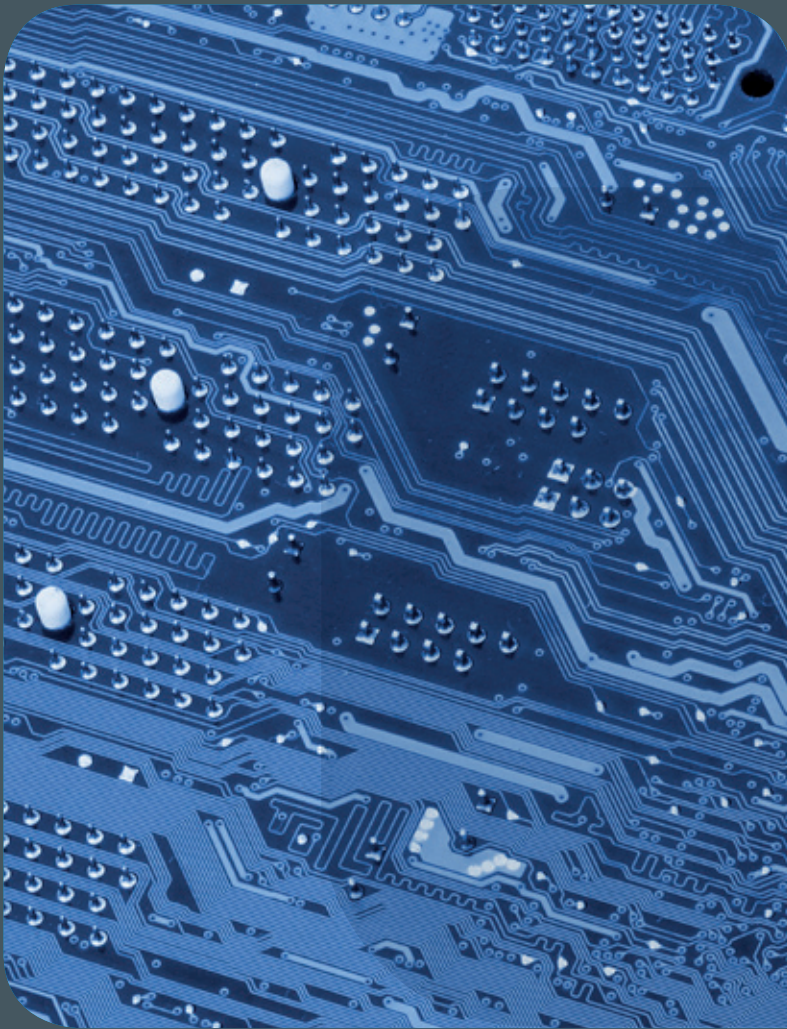
Digitale
Langzeitarchivierung

Parallelverarbeitung mit
numerischen Bibliotheken

EU-Projekt „PaaSage“

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDDG





GWDG NACHRICHTEN

09-10|15 Inhalt

-
- 4 Exchange-Server der GWDG und DavMail
 - 6 Effizientes mobiles Arbeiten per CalDAV
 - 9 muttprint 10 Erste Eindrücke zu Windows 10
 - 12 Backing Up FreeBSD Systems using TSM
 - 14 Auftragsdatenverarbeitung personen-
bezogener Daten 16 Digitale Langzeitarchi-
vierung für die DNB 19 Parallelverarbeitung
mit numerischen Bibliotheken 26 Interaktive
Visualisierung wissenschaftlicher Daten im
Web 30 Kurz & knapp 31 EU Project "PaaSage"
 - 36 Personalia 38 Kurse

Impressum

Zeitschrift für die Kunden der GWDG

ISSN 0940-4686
38. Jahrgang
Ausgabe 9-10/2015

Erscheinungsweise:
monatlich

www.gwdg.de/gwdg-nr

Auflage:
550

Fotos:
© bluebay2014 - Fotolia.com (1)
© storm - Fotolia.com (4)
© Robert Kneschke - Fotolia.com (6)
© MK-Photo - Fotolia.com (14)
© Sashkin - Fotolia.com (25)
© momius - Fotolia.com (35)
© MPLbpc-Medienservice (3, 36, 37)
© GWDG (2, 38)

Herausgeber:

Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen
Am Faßberg 11
37077 Göttingen
Tel.: 0551 201-1510
Fax: 0551 201-2150

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 Kunden und Freunde der GWGD,

aus den vielen Artikeln dieser umfangreichen Doppelausgabe der GWGD-Nachrichten möchte ich besonders auf das Thema Windows 10 hinweisen. Die neue Version des Microsoft Betriebssystems macht vieles anders und verbindet doch bekannte Funktionen aus Windows 7 und 8 miteinander. Diese neue Version ist umstritten, da in der Standardkonfiguration zahlreiche Nutzerinformationen gesammelt und an den Hersteller übermittelt werden. Diese Sammelwut kann durch Anpassung der Einstellungen zu Datenschutz und Privatsphäre reduziert werden.

Dennoch ist der Wandel zu einer stärkeren Verschmelzung von Desktop-Anwendung zu Cloud-Lösung und der damit einhergehenden Weitergabe von Daten kaum zu verhindern. Entweder man sucht nach alternativen Software-Lösungen, was nicht immer einfach ist, oder man lässt sich zunehmend auf diese Betriebskonzepte ein, wobei man den Firmen weitreichend ausgeliefert ist. Es benötigt einen sensiblen Umgang mit den eigenen Daten und ein kritisches Abwägen, ob Nutzen und Risiken in einem sinnvollen Verhältnis stehen. Ein anderes, meines Erachtens ungelöstes Problem sind die permanenten, kontinuierlichen Software-Updates, wie sie bei Apple, Microsoft und diversen anderen Software-Anbietern Einzug gehalten haben. Auf der einen Seite ist es aus Sicherheitsgesichtspunkten löblich, wenn Fehler kurzfristig beseitigt werden. Auf der anderen Seite wird es zunehmend schwierig sicherzustellen, ob nach einer neuen Version von iOS oder Windows noch die Kompatibilität mit anderen Anwendungen besteht. Für größere Organisationen ist dies eine echte Herausforderung.

Ich wünsche Ihnen viel Spaß beim Lesen dieser Ausgabe.

Ramin Yahyapour

GWGD – IT in der Wissenschaft

Nutzung des Exchange-Servers der GWDG mit DavMail

Text und Kontakt:

Claudius Laumanns
claudius.laumanns@gwdg.de
0551 201-1881

Sie möchten mit Ihrem Lieblings-Mailclient die volle Funktionalität des Exchange-Servers der GWDG nutzen. Sie sind Linux- bzw. BSD-Nutzer und möchten gemeinsame Termine mit Ihren Kollegen anlegen und auch das Exchange-Adressbuch nutzen. Die GWDG bietet Ihnen ab sofort mit DavMail die Möglichkeit dazu: DavMail ist eine Gateway-Software, greift auf die Outlook-Weboberfläche zu und bietet Mailclients die standardkonformen Protokolle CalDAV für Kalender sowie CardDAV für Adressbücher.

Basis des GWDG-Mailservice ist eine hochverfügbare Exchange-2010-Umgebung, die Anwendern mit Microsoft Outlook als Mailclient-Programm den vollen Funktionsumfang bietet. Mittlerweile kann auch Apple Mail oder Gnome Evolution mit Exchange Web Services kommunizieren, und iOS sowie Android-Smartphones verstehen Exchange Server ActiveSync.

Insbesondere Nutzer mit Open-Source-Betriebssystemen und/oder alternativen populären Mailprogrammen waren jedoch bislang leider nur auf POP- oder IMAP-Zugang beschränkt und von der Kalender- und Adressbuchfunktionalität des Exchange-Servers ausgeschlossen. Eine Ausnahme bildet Thunderbird, für den über spezielle Erweiterungen in gewissem Umfang Nutzungsmöglichkeiten bestehen (siehe auch die GWDG-Nachrichten 12/2014). Im Rahmen einer erweiterten Testumgebung bietet die GWDG seit Kurzem mit einer DavMail-Schnittstelle einen erweiterten Funktionsumfang als generelle Lösung an.

DavMail greift auf die Outlook-Weboberfläche zu und bietet Mailclients ein Gateway zum Exchange-Server der GWDG über die standardkonformen Protokolle CalDAV für Kalender sowie CardDAV für Adressbücher. Auch der Zugriff auf das globale Adressbuch des Exchange-Servers kann über LDAP vermittelt werden. Es handelt sich bei DavMail um Open-Source-Software, die unter der General Public License (GPLv2) veröffentlicht worden ist. Abb. 1 zeigt die grundlegende Funktionsweise von DavMail.

Ein Mailclient stellt per WebDAV eine Anfrage an DavMail, das die Authentifizierung an die Outlook Web App durchleitet und von dort die Kalender und Adressbücher abfragt, um diese wiederum als CalDAV, CardDAV oder LDAP an den Mailclient zurückzugeben. Die Kommunikation zwischen Exchange und DavMail sowie zwischen DavMail und dem Mailclient wird mit SSL verschlüsselt.

Unterstützt werden Thunderbird sowie iOS und Mac OS X. Auf davmail.sourceforge.net finden Sie Anleitungen für verschiedene Mailclients. Darüber hinaus haben wir DavMail mit der KDE Groupware, Vdirsyncer sowie Khal und Khard getestet. Da strikte

Standardkonformität ein wichtiges Ziel der DavMail-Entwickler ist, ist zu erwarten, dass auch weitere Mailclients von DavMail profitieren werden.

Um DavMail bei der GWDG zu nutzen, müssen folgende Adressen in der jeweiligen Konfiguration des Mailclients verwendet werden (*userid@gwdg.de* ist Ihre primäre GWDG-E-Mail-Adresse):

- <https://davmail.gwdg.de:1080/users/userid@gwdg.de/calendar> (für CalDAV)
- <https://davmail.gwdg.de:1080/users/userid@gwdg.de/contacts> (für CardDAV)

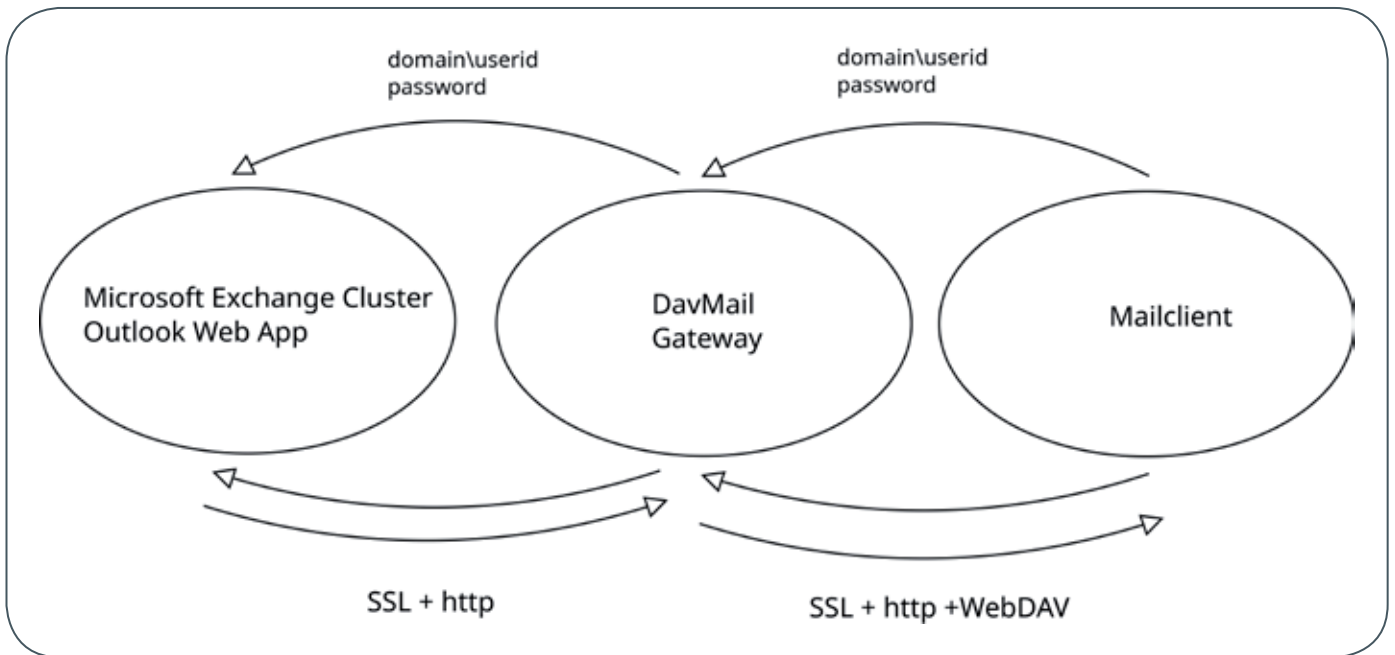
Zur Nutzung des Exchange-Adressbuches per LDAP müssen der URL <https://davmail.gwdg.de:1389> sowie der LDAP Base Context *ou=people* eingetragen werden.

Als Besonderheit ist bei iOS und Mac OS X beim Adressbuchzugriff <https://davmail.gwdg.de:1080/principals/users/userid@gwdg.de/> zu konfigurieren.

Unter Mac OS X wird das Directory *Utility* aufgerufen. (Seit

Using GWDG's Exchange Server with DavMail

GWDG's mail service is based on a highly available Exchange 2010 cluster. Using Microsoft Outlook as a mail client program provides full functionality to you. Unfortunately users with open source operating systems and/or popular alternate mail programs are excluded from calendar and addressbook functionality. Recently, within an enhanced test environment, a DavMail gateway has been put into operation to overcome this shortcomings. DavMail is a Open Source Gateway to the proprietary Microsoft Exchange Server. It provides clients with standard compliant CalDAV for calendar and CardDAV for personal contacts or LDAP for Exchange's global addressbook.



1_Funktionsweise von DavMail

Yosemite ist das unter `/System/Library/CoreServices/Applications` zu finden.) Hier wird der `LDAPv3` ausgewählt und `davmail.gwdg.de` mit SSL und dem Search Base Suffix `o=od` angegeben.

DavMail ist nicht Teil des Exchange-Clusters der GWGD, sondern greift auf dieses zu. Die für das Exchange-Cluster abgeschlossenen Service Level Agreements (SLAs) sind deshalb hierfür nicht anwendbar. Im Unterschied zum redundant ausgelegten Exchange-Cluster läuft DavMail in der Testphase nur auf einem einzigen virtuellen Server. Wir können deshalb verständlicherweise nicht das gleiche hohe Maß an Zuverlässigkeit garantieren.

Um Bedenken zu prüfen, dass DavMail starkem Nutzeraufkommen nicht gewachsen sein könnte, wurden ausführliche Stresstests mit Tsung, DavMail und einer

Exchange-Testumgebung durchgeführt. Tsung ist ein Stress Testing Tool, welches hier die Rolle der Mailclients übernahm. Simuliert wurden CalDAV- und CardDAV-http-requests (also Kalender und Adressbuch). Es hat sich gezeigt, dass DavMail ausreichend viele Anfragen gleichzeitig behandeln kann, wobei für jede Anfrage genau eine Verbindung aufgebaut wird.

Wir hoffen, mit dem neuen DavMail-Service ein interessantes Zusatzangebot für unsere Kunden bereitzustellen, auch wenn keine dem Exchange-Cluster selbst vergleichbaren hohen Verfügbarkeiten zugesagt werden können. Die während der Testphase gesammelten Erfahrungen werden zeigen, ob der Übergang in den Produktivbetrieb sinnvoll ist. ●





Effizientes mobiles Arbeiten mit Exchange-Kalendern per CalDAV

Text und Kontakt:

Eric Helmvoigt
eric.helmvoigt@gwdg.de
0551 201-1845

Olga Smirnova
olga.smirnova@gwdg.de
0551 201-1892

Mit der Einführung des DavMail-Services bietet die GWDG Ihnen die Möglichkeit, auch mit mobilen Geräten auf freigegebene Kalender und Kontakte eines Exchange-Postfachs zuzugreifen, diese zu lesen, zu beschreiben und zu synchronisieren.

Im Alltag gibt es öfters Situationen, wo der Einsatz mehrerer Kalender als Organisationsmittel vorkommt. Die Handhabung mehrerer freigegebener Kalender in einer Exchange-Umgebung ist jedoch nur aus Outlook heraus bequem und unkompliziert möglich. Auf Smartphones beschränkt sich die Nutzung der Exchange-Kalenderfunktionen rein auf den primären Kalender.

Die GWDG bietet seit Kurzem einen DavMail-Service an, der eine alternative Möglichkeit für den Zugriff auf z. B. Exchange-Kalender bietet. DavMail ist eine Gateway-Software, die auf OWA (Outlook Web App) zugreift. Es bietet so auch alternativen E-Mail-Clients den Zugriff auf Exchange-Kalender und -Adressbücher, wenn der Client die Protokolle für CalDAV sowie CardDAV beherrscht. Im Vergleich zur Standardlösung kann über CalDAV auch auf freigegebene Kalender zugegriffen werden.

Auf Android-Geräten kann man die Vorteile des DavMail-Service unabhängig von der verwendeten Kalenderapplikation nutzen, da in Android nicht die einzelnen Apps, sondern ein Hintergrunddienst für das Synchronisieren und Bereitstellen der Kalenderdaten in den lokalen Speicher zuständig ist. Hierfür ist der Android-Dienst „Kalenderspeicher“ („Calendar Storage“) zuständig.

Für diesen Artikel wurden die Tests mit dem **Business Kalender** und der **Android Standard Kalender App** durchgeführt. Es wurde ebenso mit verschiedenen Android-Versionen getestet (4.2.1, 4.2.2 und 5.0.2). Natürlich können wir nicht sicherstellen, dass diese Anwendungen in jedem Fall problemlos funktionieren.

Im Folgenden beschreiben wir die Schritte, die bei der

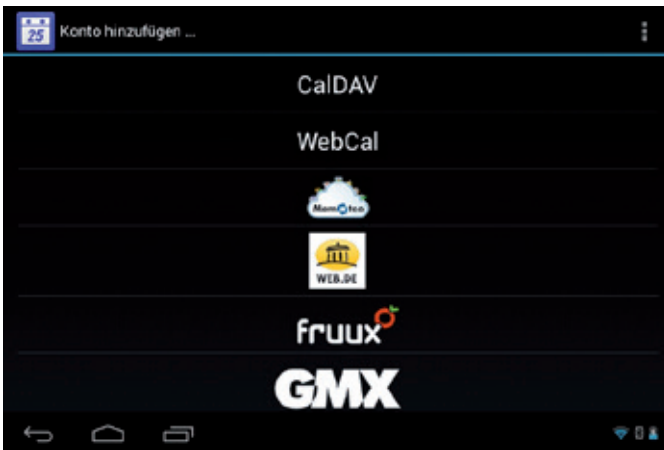
Installation und Konfiguration erforderlich sind. Die Wahl der App fiel auf die kostenpflichtige App **CalDAV-Sync** [1]. Die App hat bei den internen Tests die Synchronisation mit dem Kalenderspeicher mit hoher Zuverlässigkeit durchgeführt und bietet einen großen Komfort bei der Einrichtung sowie Konfiguration, den freie Apps nicht bieten.

Als erstes muss die App z. B. aus dem Google Store [2] oder von Amazon [3] bezogen und installiert werden. Nachdem Sie die **CalDAV-Sync App** gestartet haben, können Sie direkt den GWDG-Maildienst konfigurieren. Voraussetzung dafür ist, dass Sie die Freigabeadresse des einzubindenden Kalenders kennen. Diese bezieht sich normalerweise auf den Primärkalender der Benutzer und sieht wie folgt aus: <https://davmail.gwdg.de:1080/users/userid@gwdg.de/calendar>.

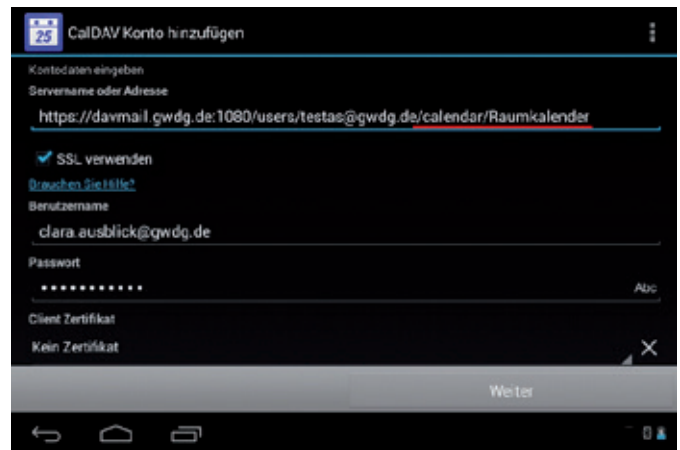
Bei der Freigabeadresse müssen zwei Parameter angepasst werden:

Efficient mobile working with Exchange calendars via CalDAV

The GWDG has newly introduced DavMail-Service, which offers you an extended access to shared calendar and contacts of an exchange account. This possibility allows viewing, editing and synchronizing shared context from several sources on your mobile device.



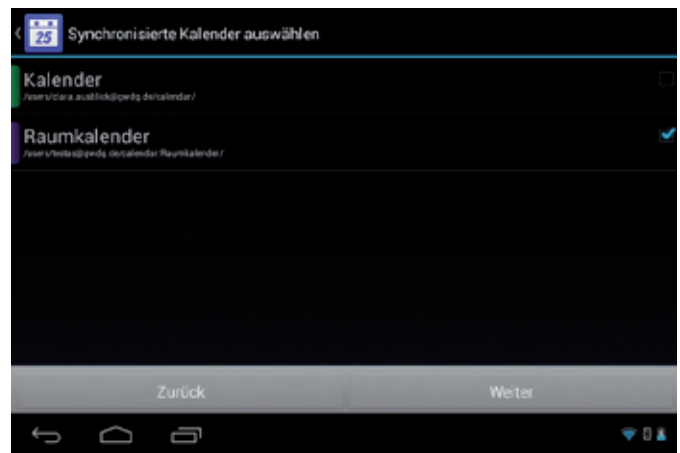
1_Auswahl des CalDAV-Dienstes in CalDAV-Sync



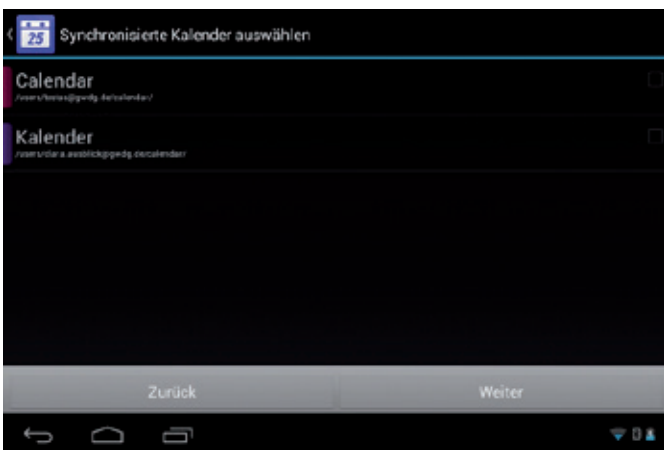
4_Das Hinzufügen eines zusätzlichen Kalenders (als Beispiel der „Raumkalender“; Synchronisation ist gesetzt)



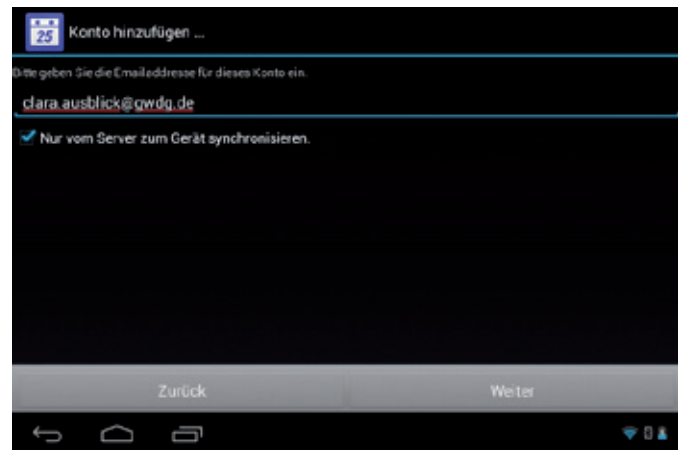
2_Das Hinzufügen des freigegebenen Primärkalenders (im Beispiel Synchronisation nicht gesetzt)



5_Das Hinzufügen eines zusätzlichen Kalenders (als Beispiel der „Raumkalender“; Synchronisation ist gesetzt)



3_Das Hinzufügen des freigegebenen Primärkalenders (im Beispiel Synchronisation nicht gesetzt)



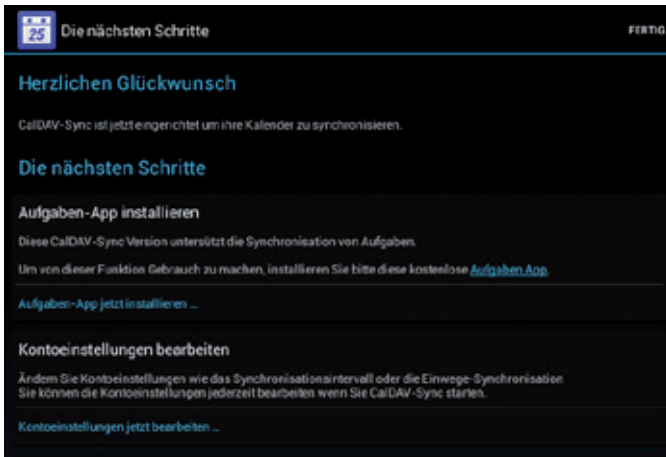
6_Unidirektionale Synchronisation ist eingestellt

1. *userid@gwdg.de*; die primäre E-Mail-Adresse des Postfachs, dessen Kalender freigegeben wurde (s. Abb. 2 und Abb. 3)
2. *calendar*; bezieht sich auf den primären Kalender, kann aber durch zusätzliche Kalender erweitert werden. Bei der Freigabe im Outlook oder im OWA (Outlook Web App) ist darauf zu achten, wo der Kalender in der Ordnerstruktur (s. Abb. 9) angelegt wird, da diese bei der Gestaltung der Freigabeadresse berücksichtigt werden muss. Ist der freigegebene Kalender parallel zum primären Kalender

angelegt, so sieht die Freigabeadresse wie folgt aus:
<https://davmail.gwdg.de:1080/users/userid@gwdg.de/freigegebener-kalender>

Im folgenden Beispiel wurden die untergeordneten Raum- und Konferenzkalender (s. Abb. 9) des Benutzers *testas@gwdg.de* für den Benutzer *clara.ausblick@gwdg.de* freigegeben. Dadurch ändert sich die Freigabeadresse wie unten erklärt.

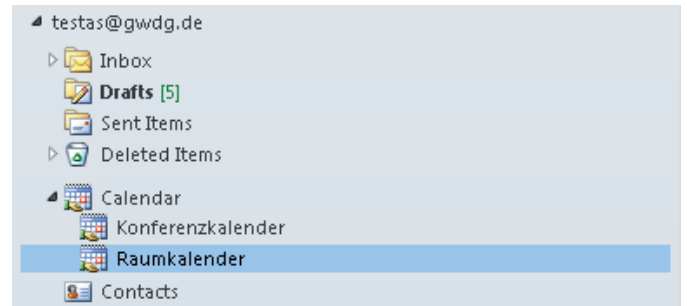
Nach dem Start von CalDAV-Sync werden Sie aufgefordert, den Kontotyp für die Verbindung zum Server auszuwählen. Für unser Beispiel wählen Sie die Option *CalDAV* aus, was in der



7_Erfolgreiche CalDAV-Einrichtung



8_Die Ansicht des freigegebenen Kalenders (Raum- und Konferenzkalender) in der Android Standard Kalender App



9_Kalenderstruktur der primären und zusätzlichen freigegebenen Kalender

derzeitigen Version ganz oben in der Auswahlliste angeboten wird (s. Abb. 1).

1. Unter „Servername oder Adresse“ (s. Abb. 4) muss die Freigabeadresse des Kalenders, in diesem Beispiel ist es der Raumkalender des Benutzers *testas@gwdg.de*, eingegeben werden. Hierbei ist auf die Kalenderstruktur (s. Abb.9) zu achten. Es ergibt sich daraus die Freigabeadresse: <https://davmail.gwdg.de:1080/users/testas@gwdg.de/calendar/Raumkalender>. Bei „Benutzername“ und „Passwort“ geben Sie bitte Ihre E-Mail-Adresse und Passwort an.
2. CalDAV-Sync bietet Ihnen danach den freigegebenen Kalender neben Ihrem eigenen Kalender in einem Auswahl-Menü an (s. Abb. 5).
3. Während der Konfiguration kann eine bidirektionale oder unidirektionale Synchronisation ausgewählt werden (s. Abb. 6), die später noch änderbar ist.
4. Als nächstes bekommen Sie eine Meldung über die erfolgte Einrichtung. Die Konfiguration wird mit einem Tippen auf die Schaltfläche „Fertig“ abgeschlossen (s. Abb. 7).
5. Den Kalender können Sie nun mit Ihrer bevorzugten Kalender-App angezeigt lassen und bearbeiten (s. Abb. 8). Die Kalender-Apps bieten hierzu normalerweise einen eigenen Punkt innerhalb der Einstellungen an.

Falls Sie weitere Kalender benötigen, so lassen sich diese auf gleichem Weg per CalDAV-Sync in Ihr System einbinden (s. Abb.3).

REFERENZEN

- [1] <http://dmfs.org/caldav/>
- [2] <https://play.google.com/store/apps/details?id=org.dmfs.caldav.lib>
- [3] <https://www.amazon.com/gp/mas/get/android/>

Formatierter Ausdruck von E-Mails mit muttprint

Text und Kontakt:

Dr. Wilfried Grieger
wilfried.grieger@gwdg.de
0551 201-1512

Dr. Konrad Heuer
konrad.heuer@gwdg.de
0551 201-1540

Textbasierte E-Mail-Programme, beispielsweise unter UNIX-Derivaten, haben in der Regel das Problem, dass der Ausdruck einzelner E-Mails doch recht spartanisch aussieht. Das lässt sich mit muttprint deutlich verbessern. muttprint ist in eine Vielzahl von E-Mail-Programmen integrierbar.

EINLEITUNG

In den GWDG-Nachrichten 12/2013 haben wir das textbasierte E-Mail-Programm **mutt** vorgestellt, das auch die Anforderungen an einen modernen und zukunftssträchtigen E-Mail-Verkehr erfüllen kann. Aber mutt hat den Nachteil, dass der Ausdruck einzelner E-Mails in der Regel nicht mit der Qualität der Drucke aus beispielsweise Thunderbird oder Outlook mithalten kann. Es ist zwar eine Umwandlung des bloßen Textes z. B. mit **a2ps** in ein PostScript-Format möglich, aber damit sind auch eigentlich alle Möglichkeiten erschöpft. Abhilfe kann nun das Programmpaket **muttprint** schaffen. Es wurde von Bernhard Walle geschrieben. Es handelt sich dabei um freie Software, die gemäß der GNU General Public License genutzt werden darf.

EIGENSCHAFTEN VON MUTTPRINT

muttprint ist in der Programmiersprache Perl geschrieben und nutzt eine LaTeX-Distribution zur Umwandlung des reinen E-Mail-Textes in eine ansprechende formatierte Druckausgabe. Damit eröffnet sich eine Fülle von Möglichkeiten, die auch bei Thunderbird oder Outlook nicht zu finden sind.

Alle Angaben zur Formatierung müssen im File `~/muttprintrc` abgelegt, können aber auch als Option beim Zeilenaufruf von muttprint angegeben werden.

Es kann beispielsweise dort festgelegt werden, welche Header-Zeilen in welcher Reihenfolge gedruckt werden sollen und welcher Font und welches Format für den gesamten Header zu nutzen sind. Auch die teilweise recht beliebten X-Face-Bilder können mit ausgedruckt werden.

Für den eigentlichen E-Mail-Text kann ein eigener Font festgelegt werden. Dabei kann auch die Signatur anders als der Text formatiert oder auch ganz weggelassen werden.

Weitere Eigenschaften entnehme man dem Handbuch.

INTEGRATION VON MUTTPRINT IN ANDERE E-MAIL-PROGRAMME

muttprint lässt sich einfach in andere E-Mail-Programme integrieren:

- Für **mutt** muss im File `~/muttrc` die folgende Zeile geändert oder eingefügt werden: `set print_command="muttprint"`.
- Für **pine** oder **alpine** muss im Printer-Setup als Druckkommando `muttprint` eingetragen werden.

Weitere Beispiele können dem Handbuch entnommen werden.

SOFTWARE

Die Software ist bei der GWDG auf dem Rechner `gwdu60.gwdg.de` in der Version 0.72d installiert. Dort finden sich in der Directory `/usr/local/share/doc/muttprint` sowohl das Handbuch als auch Beispiele für die Konfigurationsdatei `~/muttprintrc`.

Für eine lokale Installation kann die Software von <http://muttprint.sourceforge.net> heruntergeladen werden. ■

Formatted e-mail printing with muttprint

Text-based e-mail clients, i. e. under UNIX derivatives, typically have the problem that the printed single e-mail looks quite spartan. This can significantly be improved with muttprint. muttprint is integrable in a plurality of e-mail clients.

Erste Eindrücke zu Windows 10

Text und Kontakt:

Patrick Becker
patrick.becker@gwdg.de
0551 201-1809

Das neue Windows 10 von Microsoft zieht große Aufmerksamkeit auf sich: Es soll die Anwender, die den traditionellen Windows-Desktop gewöhnt waren und durch Windows 8 erschreckt wurden, mit der neuen universellen Oberfläche für alle Plattformen versöhnen, es ist kostenlos als Upgrade für Systeme mit Windows 7 und 8 erhältlich, und es soll die finale Windows-Version sein, die dann über Jahre hinweg permanent von Microsoft aktuell gehalten wird. Wir beschreiben in diesem Artikel die ersten Erfahrungen mit Windows 10 bei der GWDG auch im Hinblick auf das Active Directory.

Microsoft hat Windows 10 am 29. Juli 2015 veröffentlicht. Es ist der Nachfolger von Windows 8.1 und wurde erstmals am 30. September 2014 vorgestellt. Das Betriebssystem wird Benutzern der Vorgängerversionen Windows 7 und Windows 8 für ein Jahr nach Veröffentlichung als kostenloses Update angeboten.

Windows 10 stellt eine einheitliche Plattform für verschiedene Geräteklassen wie Desktop-Rechner, Smartphones, Notebooks und Tabletcomputer dar. Je nach verwendetem Gerät passt sich die Oberfläche automatisch an die Eigenschaften des Systems an. Wegen des fehlenden Startmenüs und der Einführung des Kachelbildschirms bei Windows 8 wurde Microsoft von einem Großteil der Nutzer kritisiert. Mit Windows 10 erhält das Startmenü in einer überarbeiteten Form wieder Einzug (siehe Abb. 1).

Weitere Neuerungen sind der Webbrowser „Microsoft Edge“, der Personal Assistant „Cortana“, Virtuelle Desktops (siehe Abb. 2), Copy and Paste per Tastatur in der Kommandozeile, der nicht mehr zwingend vorgeschriebene Vollbildmodus für Apps und ein für alle Anwendungen zur Verfügung stehender PDF-Drucker. Das als „Action Center“ benannte neue Benachrichtigungs-Center informiert über Meldungen wie E-Mail, System- und App-Benachrichtigungen. Die zunehmende Bedeutung der Cloud-Dienste von Microsoft wird durch die noch stärkere Integration von „OneDrive“, als noch in Windows 8 deutlich. Neben dem App-Store erfordern auch einige weitere Apps ein Microsoft-Konto.

Eine weitere interessante, nach Meinung des Autors zu wenig beachtete, Neuerung ist der Powershell-basierte Paketmanager „PackageManagement“. Damit lassen sich wie unter Linux Programme, wie beispielsweise Sysinternals, über die Kommandozeile finden, installieren, aktualisieren und deinstallieren.

Auch das Update-Intervall und das Einspielen von Updates haben sich in Windows 10 verändert. Konnte man in vorherigen Versionen noch einzelne Windows-Updates von der Installation ausschließen, kann man jetzt nur noch den Zeitpunkt des Neustarts nach einem Update festlegen, nicht aber, ob ein bestimmtes Update überhaupt installiert werden soll. Durch das Microsoft-Tool

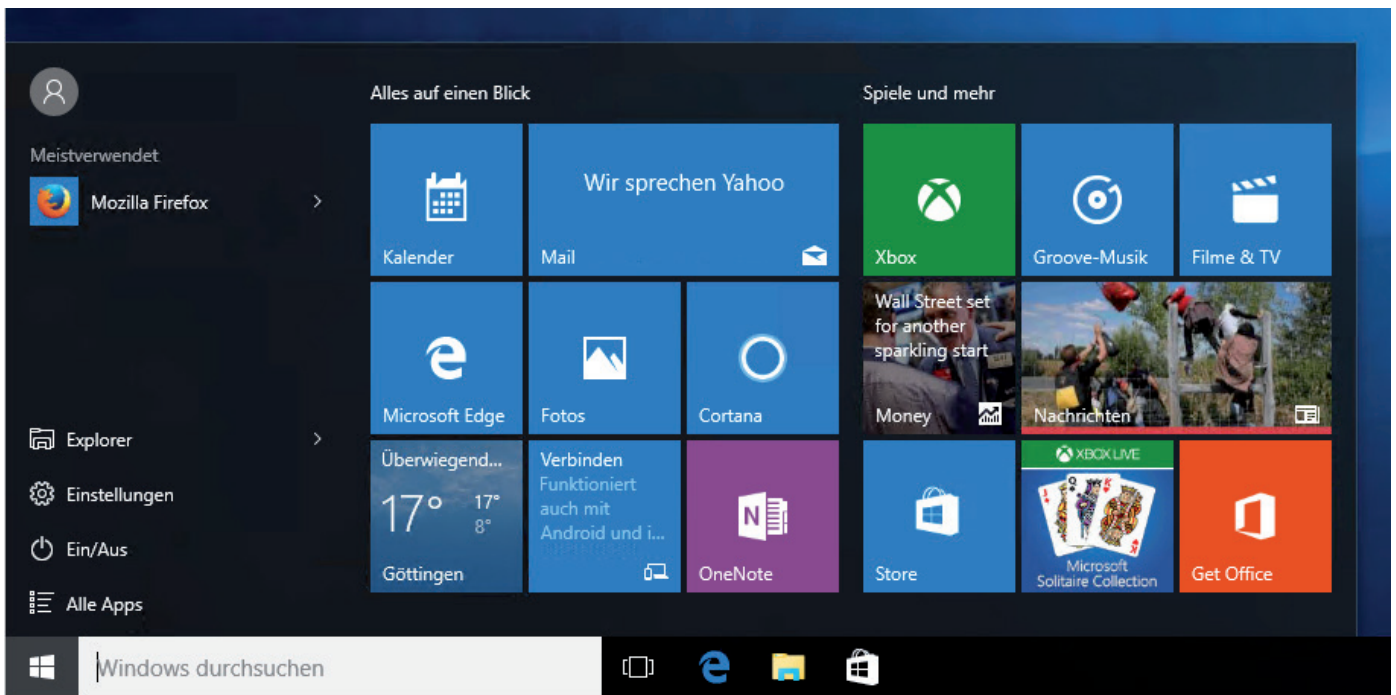
KB3073930 kann man verhindern, dass Updates nach einem Fehlerfall erneut installiert werden. Es bleibt abzuwarten, ob Microsoft diese Funktion durch ein zukünftiges Upgrade direkt in Windows integriert oder ob man weiterhin den Umweg über das Zusatztool nehmen muss.

Den bisher monatlichen Patchday führt Microsoft bei Windows 10 nicht weiter fort, sondern stellt die Patches nach Fertigstellung und internen Tests direkt zur Verfügung. In einem Active Directory kann der Installationszeitpunkt über Richtlinien festgelegt werden. Die GWDG wird in ihrem Active Directory das von Microsoft eingeführte Update-Intervall von einmal im Monat auch für Windows 10 übernehmen.

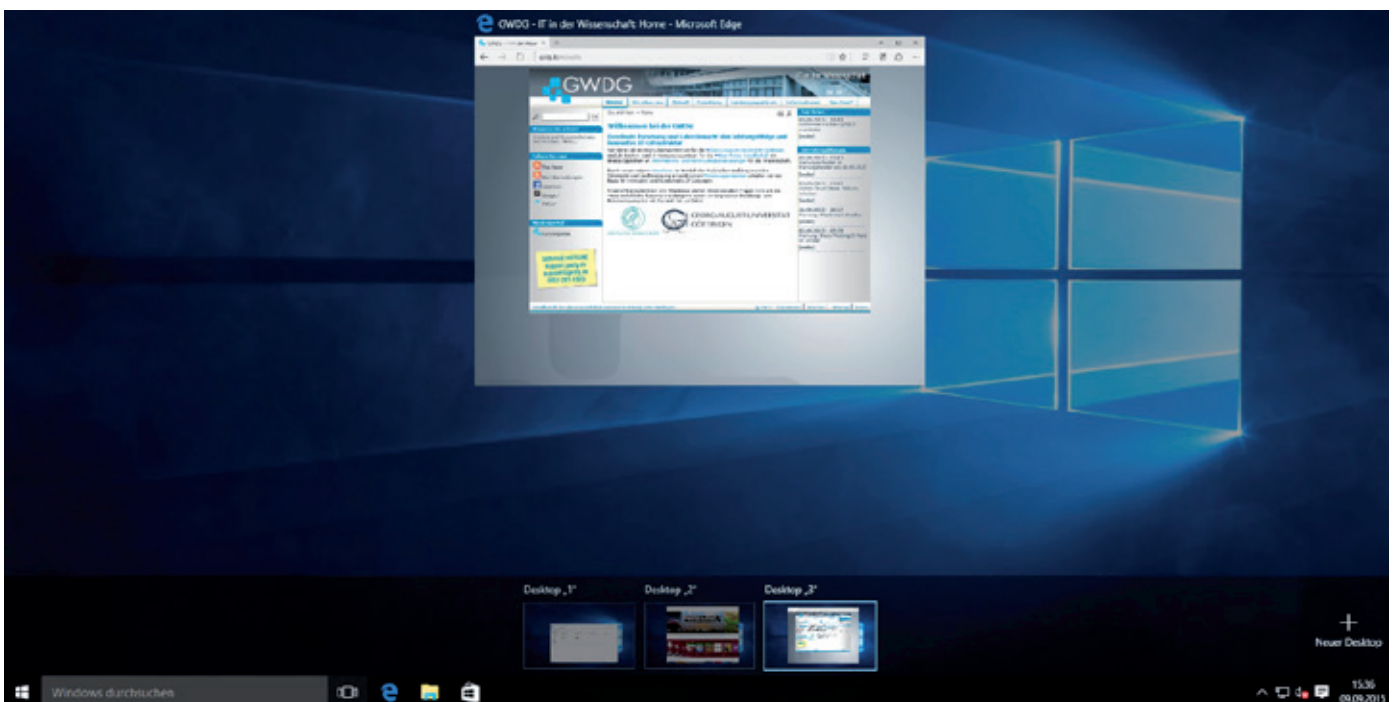
Datenschutz ist auch in Windows 10 ein wichtiges Thema. Kritisiert wird hier im Speziellen, neben der fehlenden Möglichkeit, die Übermittlung persönlicher Daten an Microsoft komplett abzustellen, dass Microsoft das Opt-Out-Verfahren anwendet. Also, dass bei einer Standardinstallation erst einmal alles eingeschaltet ist und man selbst dafür sorgen muss, dass möglichst keine privaten Daten an Microsoft gesendet werden. Die Überprüfung dieser Optionen gestaltet sich etwas zeitaufwändig, da die Einstellungen an verschiedenen Stellen in Windows 10 zu finden sind. Aus diesem Grund werden diverse Tools angeboten, die die Einstellungen

First impressions of Windows 10

Microsoft Windows 10 draws much attention: It shall reconcile users who were accustomed to the traditional Windows desktop, and who were frightened by Windows 8, with the new universal interface for all platforms. Furthermore, it is a free upgrade for systems running with Windows 7 and 8, and it is announced to be the final version of Windows which will be maintained permanently for years to come. We describe our first experiences with Windows 10 and its compatibility with our Active Directory here at GWDG.



1_Das neue Startmenü in Windows 10



2_Virtuelle Desktops jetzt auch für Windows

unter einer Arbeitsoberfläche zusammentragen. Nach Meinung des Autors ist z. B. „W10Privacy“ ganz gut gelungen, das von der Seite www.winprivacy.de kostenlos heruntergeladen werden kann. Wie immer gilt bei der Anwendung spezieller Tools, dass unbedarftes An- bzw. hier Abschalten auch Folgefehler in Windows mit sich bringen kann und daher Vorsicht geboten ist. Die Aufnahme von Windows 10-PCs in das Active Directory verhält sich äquivalent zu Windows 8, weswegen keine Schwierigkeiten zu erwarten sind. Microsoft lässt Windows 10 allerdings bei der Anmeldung

eines Benutzers ein neues Benutzerprofil im Active Directory anlegen. Das hat den Vorteil, dass Unstimmigkeiten durch ein früheres System gar nicht erst auftreten, aber auch den Nachteil, dass Einstellungen und Dateien, die im bisherigen Profil liegen, nicht übernommen werden.

Abschließend bleibt zu sagen, dass bei einer Neuinstallation nichts gegen die Verwendung von Windows 10 spricht. Ob man jetzt sein altes, noch funktionierendes System auf Windows 10 updaten soll, muss Jeder für sich selbst entscheiden. ●

Backing Up FreeBSD Systems using Tivoli Storage Manager (TSM) and File System Snapshots

Text und Kontakt:

Dr. Konrad Heuer
konrad.heuer@gwdg.de
0551 201-1540

Thomas Körmer
thomas.koermer@gwdg.de
0551 201-1555

System backups are essential in professional IT environments. Although FreeBSD is an interesting and powerful server operating system there is no native TSM client for FreeBSD. The Linux client can be used, but configuration is a little bit tricky.

TSM is widely used within the Max Planck Society, the University of Göttingen and the GWDG itself as a central software solution for backing up and restoring data to protect from severe data losses. The TSM client software is available for a couple of operating systems, and the Linux client also runs on FreeBSD as many other Linux applications do since FreeBSD provides a Linux run-time environment.

FreeBSD is a well-known open source operating system with several interesting features especially on the server side. To mention some examples, there is an outstanding implementation of different firewall solutions, an excellent integration of the ZFS file system, and in general FreeBSD is a pleasant mix of keeping UNIX traditions while nevertheless constantly introducing modern features.

However, there is basic problem when backing up FreeBSD systems with Linux applications like TSM.

The Linux run-time environment is composed of two parts joined together: a FreeBSD kernel module providing Linux system calls, and a typically small Fedora Linux userland containing

a bundle of Linux binaries, shared libraries and configuration files. This userland is installed downwards starting from the directory `/compat/linux` with its own typical UNIX/Linux directory tree.

When starting a Linux application, the FreeBSD loader identifies the associated binary file format and initializes Linux mode which by some means defines `/compat/linux` to be the preferred root directory of the process (as illustrated in figure 1). The process is not imprisoned here and can access any file at any location on the system as long as granted by the POSIX access rights, but in case of identical pathnames the Linux userland will always win. This is a must; for example, a Linux application might try to map a shared library `/lib/libc.so.6` into the virtual address space of the process, and this library could exist as a FreeBSD version in `/lib` as well as a Linux version in `/compat/linux/lib`. Any Linux application will have to use the second one, otherwise it would die.

Inescapable, this leads to consequences when backing up a FreeBSD system by using the Linux TSM client. Imagine a server with three file systems mounted on `/`, `/usr` and `/var` and a file `dsm.opt` containing the lines:

```
DOMAIN /
DOMAIN /usr
DOMAIN /var
```

TSM-Backup bei FreeBSD-Systemen

System Sicherungen sind in professionellen IT-Umgebungen unverzichtbar. Obwohl FreeBSD ein interessantes und leistungsstarkes Server-Betriebssystem ist, gibt es keinen nativen TSM-Clienten für FreeBSD. Der Linux-Client kann stattdessen verwendet werden, die Konfiguration ist aber ein wenig trickreich.

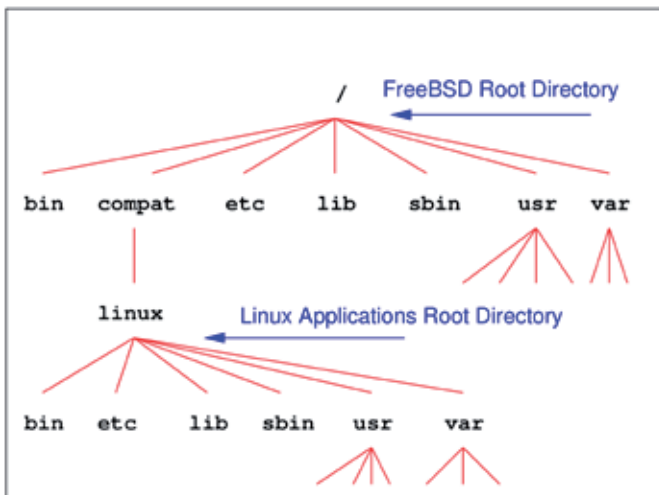


Figure 1: Part of the FreeBSD directory tree

The TSM client will never traverse the real file systems to save them; instead, it will never leave the directory `/compat/linux` since all directory names can be found there, too.

One trick to make things work is to mount file systems additionally on different directories by „nullfs“ mount commands (corresponding to loop-back mount processes in the Linux world). As a consequence, these file systems will be accessible by additional path names. An excerpt of `/etc/fstab` might be:

```
/dev/da0s1a / ufs rw 1 1
/dev/da0s1e /usr ufs rw 2 2
/dev/da0s1f /var ufs rw 2 2
/usr /tsm/usr nullfs ro 0 0
/var /tsm/var nullfs ro 0 0
```

Changing `dsm.opt` to

```
DOMAIN /tsm
DOMAIN /tsm/usr
DOMAIN /tsm/var
```

will help for `/usr` and `/var` since pathnames with a leading `/tsm` do not exist in `/compat/linux`. But this is not a solution for the root file system itself because `/` can in principle not be mounted anywhere else by a nullfs mount command.

Backing up the root file system is possible by making use of file system snapshots. Snapshots are useful for backup purposes anyway since files that currently change while being read cannot be saved, and snapshots are static by definition. To create and mount a snapshot, only few commands are necessary:

```
mksnap_ffs /snapshot
mdconfig -a -t vnode -o readonly -f /snapshot -u 0
mount -r /dev/md0 /tsm
```

Snapshots and nullfs-mounted file systems can be combined, of course, if several file systems are to be saved on a single server. But please keep in mind that FreeBSD does not generate `/etc/mstab` automatically, thus it is necessary to create `/compat/linux/etc/mstab` manually (`ext2` is just a useful fake here):

```
/dev/md0 /tsm ext2 ro 0 0
/dev/da0s1e /tsm/usr ext2 rw 0 0
/dev/da0s1f /tsm/var ext2 rw 0 0
```

For a periodic backup schedule it is of course essential to refresh the file system snapshot periodically as well. This can easily be done by daily cron job calling a small script like this:

```
#!/bin/sh

MDU=0
MP=/tsm
PATH=/bin:/sbin:/usr/bin:/usr/sbin
SSN=/snapshot

DEV=`df $MP | tail -1 | awk '{ print $1 }'`

if [ "$DEV" = /dev/md$MDU ]; then
    umount -f $MP
    mdconfig -d -u $MDU
    rm -f $SSN
fi

mksnap_ffs $SSN
mdconfig -a -t vnode -o readonly -f $SSN -u $MDU
mount -r /dev/md$MDU $MP

exit $?
```

The latest versions of FreeBSD automatically configure UFS filesystems to run with soft-updates and journaling to ensure fast reboots also after unclean shutdowns. Unfortunately, currently file system snapshots are not possible if journaling is turned on. To disable journaling the file system has to be unmounted e.g. by rebooting into single-user mode and by entering:

```
tunefs -j disable /
```

Soft-updates itself will still be in function and will allow fast reboots, too. An essential difference is that journaling eliminates any need for file system checks whereas soft-updates still require them but let them run in the background when the system has already gone to multi-user mode. This is a small disadvantage one has to accept.

So what is the final conclusion? File system snapshots and nullfs-mounted file systems help to make use of the Linux TSM client on FreeBSD systems in many different scenarios. For virtual machines, it is often favorable to work with only one virtual disk and one file system which makes things easier for the hypervisor in general and when extending virtual disk and file system. Here file system snapshots are the only and a „stylish“ solution. Nullfs-mounted file systems are advantageous for large file systems where files frequently change and where file system snapshots would grow rapidly in size (evidently, file systems snapshots need disk space).

In case of any remarks or questions, please feel free to mail to support@gwdg.de.



Auftragsdatenverarbeitung personenbezogener Daten

Text und Kontakt:
Dr. Wilfried Grieger
wilfried.grieger@gwdg.de
0551 201-1512

Sollen auf den IT-Systemen der GWDG personenbezogene Daten verarbeitet werden, so sind dafür selbstverständlich die gesetzlichen Vorschriften einzuhalten. Dazu gehören insbesondere der § 11 des Bundesdatenschutzgesetzes sowie der § 6 des Niedersächsischen Datenschutzgesetzes. Beide befassen sich generell mit der Verarbeitung personenbezogener Daten im Auftrag, den beispielsweise die GWDG entgegennehmen kann. Ein spezieller Vertrag ist erforderlich.

EINORDNUNG

Sehr viele Paragraphen befassen sich in deutschen Gesetzestexten mit der Erhebung, Verarbeitung oder Nutzung personenbezogener Daten. Zu diesen Texten gehören das Informationsfreiheitsgesetz, die Sozialgesetzbücher, das Telekommunikationsgesetz, das Strafgesetzbuch, das Urheberrechtsgesetz, das berühmte Volkszählungsgesetz von 1987, das Bürgerliche Gesetzbuch und viele andere Gesetze, die hier aus Platzgründen gar nicht alle genannt werden können.

Am bekanntesten sind in Bezug auf personenbezogene Daten natürlich die Datenschutzgesetze. An erster Stelle ist da das **Bundesdatenschutzgesetz** (BDSG) zu nennen, das von allen öffentlichen Stellen des Bundes sowie von allen nicht-öffentlichen Stellen (§ 1 (2) BDSG), wie beispielsweise die GWDG als GmbH oder die Max-Planck-Gesellschaft als eingetragener Verein (e. V.), anzuwenden ist. Allerdings ist das BDSG lediglich ein sogenanntes Auffanggesetz, das erst dann anzuwenden ist, wenn andere Rechtsvorschriften (z. B. aus den oben aufgeführten Gesetzen) nicht greifen (§ 1 (3) BDSG).

Analog verhält es sich mit den Landesdatenschutzgesetzen, von denen hier das **Niedersächsische Datenschutzgesetz** (NDSG) als Beispiel herangezogen werden soll. Das NDSG ist von den öffentlichen Stellen „des Landes, der Gemeinden und

Landkreise“ sowie von den der „Aufsicht des Landes unterstehenden ... Stiftungen des öffentlichen Rechts“ (§ 2 (1) NDSG) anzuwenden. Damit unterliegt insbesondere auch die Georg-August-Universität Göttingen zum größten Teil dem NDSG, wenn nicht besondere Rechtsvorschriften den Bestimmungen dieses Gesetzes vorgehen (§ 2 (6) NDSG).

Sowohl das BDSG als auch das NDSG regeln in ihren Paragraphen das Erheben, das Verarbeiten und das Nutzen von personenbezogenen Daten. Diejenige Person, die personenbezogene Daten für sich selbst erhebt, verarbeitet oder nutzt oder dies durch andere im Auftrag vornehmen lässt, wird als **Verantwortliche Stelle** bezeichnet (§ 3 (7) BDSG). Wenn sich beispielsweise die von einem Institut erhobenen und gespeicherten personenbezogenen

Processing of personal data on behalf

If personal data should be processed on the IT systems of the GWDG, then legislation must be observed. These include in particular the § 11 of the Federal Data Protection Act and the § 6 of the Lower Saxony Data Protection Act. Both deal with the processing of personal data on behalf, for example, of the GWDG. A special agreement is required.

Daten anderer Personen (**Betroffene**) auf den IT-Systemen der GWDG befinden, dann lässt die Verantwortliche Stelle, das Institut oder die zuständigen Wissenschaftler, die Verarbeitung der personenbezogenen Daten durch die GWDG im Auftrag vornehmen.

Das ist mit **Auftragsdatenverarbeitung personenbezogener** Daten gemeint. Den Begriff „Auftragsdatenverarbeitung“ gibt es rein rechtlich nur im Zusammenhang mit personenbezogenen Daten.

§ 11 BDSG UND § 6 NDSG

Im § 11 BDSG und im § 6 NDSG ist aufgeführt, welche Pflichten eine Verantwortliche Stelle im Fall einer Auftragsdatenverarbeitung personenbezogener Daten übernehmen muss. Grundsätzlich bleibt die Verantwortliche Stelle weiterhin für die personenbezogenen Daten verantwortlich! Aus diesem Grund ist es erforderlich, dass die Verantwortliche Stelle als Auftraggeber mit dem Auftragnehmer, beispielsweise der GWDG, einen schriftlichen Vertrag über den Auftrag abschließt.

Der § 11 (2) BDSG sagt dazu:

„... Der Auftrag ist schriftlich zu erteilen, wobei insbesondere im Einzelnen festzulegen sind:

1. *der Gegenstand und die Dauer des Auftrags,*
2. *der Umfang, die Art und der Zweck der vorgesehenen Erhebung, Verarbeitung oder Nutzung von Daten, die Art der Daten und der Kreis der Betroffenen,*
3. *die nach § 9 zu treffenden technischen und organisatorischen Maßnahmen,*
4. *die Berichtigung, Löschung und Sperrung von Daten,*
5. *die nach Absatz 4 bestehenden Pflichten des Auftragnehmers, insbesondere die von ihm vorzunehmenden Kontrollen,*
6. *die etwaige Berechtigung zur Begründung von Unterauftragsverhältnissen,*
7. *die Kontrollrechte des Auftraggebers und die entsprechenden Duldungs- und Mitwirkungspflichten des Auftragnehmers,*
8. *mitzuteilende Verstöße des Auftragnehmers oder der bei ihm beschäftigten Personen gegen Vorschriften zum Schutz personenbezogener Daten oder gegen die im Auftrag getroffenen Festlegungen,*

9. *der Umfang der Weisungsbefugnisse, die sich der Auftraggeber gegenüber dem Auftragnehmer vorbehält,*
10. *die Rückgabe überlassener Datenträger und die Löschung beim Auftragnehmer gespeicherter Daten nach Beendigung des Auftrags.“*

Die Pflichten des Auftragnehmers sind in § 11 (4) BDSG aufgeführt. § 9 BDSG enthält die zu treffenden technischen und organisatorischen Maßnahmen, um den Schutz der personenbezogenen Daten geeignet zu gewährleisten.

Entsprechende Vorschriften finden sich im § 6 NDSG, die technischen und organisatorischen Maßnahmen zum Schutz der personenbezogenen Daten sind im § 7 NDSG aufgeführt.

Zusammengefasst: Der Auftraggeber (Verantwortliche Stelle) ist verpflichtet, mit dem Auftragnehmer (z. B. GWDG) einen entsprechenden Vertrag abzuschließen, nicht umgekehrt!

RAHMENVEREINBARUNGEN DER MAX-PLANCK-GESELLSCHAFT UND DER UNIVERSITÄT GÖTTINGEN MIT DER GWDG

Da solche Verträge zur Auftragsdatenverarbeitung personenbezogener Daten durchaus komplex werden und für Nichtjuristen eine nahezu unüberwindliche Hürde darstellen können, haben sowohl die Generalverwaltung der Max-Planck-Gesellschaft als auch die Georg-August-Universität Göttingen jeweils mit der GWDG eine Rahmenvereinbarung abgeschlossen, die alle oben geforderten Punkte enthält. Für die einzelnen Verantwortlichen Stellen ist dann lediglich eine Zusatzvereinbarung erforderlich, die auf die jeweilige Rahmenvereinbarung verweist und den speziellen Zweck der Datenverarbeitung und die vereinbarten technischen und organisatorischen Maßnahmen enthält. Selbstverständlich liegen auch für die Zusatzvereinbarung entsprechende Muster vor, die lediglich angepasst werden müssen.

Sowohl die Rahmenvereinbarungen als auch die zugehörigen Muster können bei Bedarf über support@gwdg.de bei der GWDG angefordert werden. Für Institutionen, die nicht der Max-Planck-Gesellschaft und der Universität Göttingen angehören und bei der GWDG personenbezogene Daten im Auftrag verarbeiten lassen wollen, müssen jedoch Einzelverträge abgeschlossen werden. ■

Digitale Langzeitarchivierung für die Deutsche Nationalbibliothek

Text und Kontakt:

Björn Braunschweig
bjoern.braunschweig@gwdg.de
0551 201-2133

Seit mehr als zehn Jahren ist die GWDG Kooperationspartner der Deutschen Nationalbibliothek (DNB) im Bereich der digitalen Langzeitarchivierung (LZA). Ziel ist die langfristige Erhaltung des deutschen schriftlichen Erbes. Dieser Artikel zeigt die Aktivitäten der GWDG und der DNB im Bereich LZA, die mit dem Projekt kopal begründet wurden. Es wird im Folgenden sowohl auf die Bedeutung der LZA als auch auf die technische Realisierung und aktuelle Aktivitäten der GWDG in diesem Bereich eingegangen.

BEDEUTUNG DER LZA

Die meisten der heute produzierten Daten liegen in elektronischer Form vor. Um diese Daten langfristig zur Verfügung stellen zu können, müssen ganz andere Herausforderungen gemeistert werden als beispielsweise die physische Archivierung alter Schriften.

Als Grundbaustein der digitalen Archivierung muss sichergestellt werden, dass ein in digitaler Form vorliegendes Objekt mit seinen „Nullen“ und „Einsen“ auch in 100 Jahren noch genauso vorliegt wie am Tag der Archivierung. Dies wird als „Bitstream Preservation“ bezeichnet.

Doch Nullen und Einsen alleine reichen nicht aus, um den Inhalt zu erfassen. Es sind weitere beschreibende Metadaten erforderlich. Bei einer Doktorarbeit z. B. der Autor und Titel, aber auch technische Daten wie z. B., in welcher Formatversion das zugrunde liegende PDF abgelegt wurde.

Aufgrund des kontinuierlichen technologischen Wandels kommt es immer wieder vor, dass Dateiformate nicht mehr unterstützt werden. So kann nach einer längeren Zeit der Fall eintreten, dass archivierte Daten nicht mehr darstellbar sind. Um diesem Informationsverlust vorzubeugen, ist es nötig, Formatmigrationen durchzuführen. Hierbei wird von einem Ausgangsformat in ein neueres Zielformat konvertiert.

Die zuvor genannten Punkte sind nur die wichtigsten einer Reihe von Überlegungen, die beachtet werden müssen, wenn eine Lösung für die Frage der digitalen Langzeitarchivierung gesucht wird.

AUFGABEN DER DNB

Die DNB mit ihren Standorten Leipzig und Frankfurt am Main ist die zentrale Archivbibliothek und das nationalbibliografische Zentrum der Bundesrepublik Deutschland. Die DNB hat durch das „Gesetz über die Deutsche Nationalbibliothek“ (DNBG) [1] vom 22. Juni 2006 den Auftrag zur Sammlung, Erschließung,

Verzeichnung und Archivierung von Medienwerken in körperlicher und unkörperlicher Form. Medienwerke in unkörperlicher Form sind digitale Objekte, sogenannte Netzpublikationen.

Die Pflichtablieferungsverordnung [2] vom 23.10.2008 und die Sammelrichtlinien [3] regeln u. a., dass Veröffentlichungen der DNB unaufgefordert und kostenfrei zur Verfügung gestellt werden müssen. Als Ablieferungspflichtige gelten jede Art von Körperschaften, die berechtigt sind, die Medienwerke zu verbreiten oder öffentlich zugänglich zu machen, und den Sitz, eine Betriebsstätte oder den Hauptwohnsitz in Deutschland haben.

KOPAL UND DIAS

Im Zuge dieser Pflicht, digitale Objekte dauerhaft zu erhalten und nutzbar zu machen, beteiligte sich die DNB in den Jahren 2004 bis 2007 an dem Projekt „Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen“ – kopal [4], welches vom Bundesministerium für Bildung und Forschung (BMBF) gefördert wurde. Gemeinsam mit der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB), IBM und der GWDG erfolgte im Projekt die Entwicklung und der Aufbau einer Langzeitarchivierungslösung. Dieser liegt zum einen das „Digital Information Archiving System“ (DIAS) der Firma IBM zugrunde und zum anderen das im Projekt entwickelte Softwarepaket „kopal Library for Retrieval

GWDG's long-term archiving activities for the German National Library

For more than ten years the GWDG is a cooperation partner of the "Deutsche Nationalbibliothek" (German National Library) in the field of digital preservation. The objective is the long-term preservation of the German cultural property. This article is intended to show the current long-term archiving activities in the GWDG.

and Ingest“ (koLibRI). koLibRI stellt u. a. die Generierung technischer Metadaten für die zu archivierenden Netzpublikationen und auch Import- und Exportschnittstellen zur Kommunikation mit dem DIAS-System zur Verfügung.

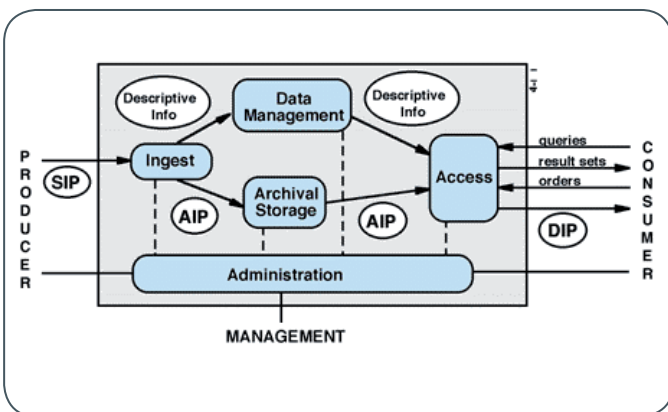
DIE ROLLE DER GWDG

Die GWDG als Rechen- und IT-Kompetenzzentrum hostet das DIAS-System, berät die DNB zum langfristigen Weiterbetrieb und arbeitet an der Verbesserung des Betriebs unter den Gesichtspunkten der Datensicherheit und des Datenmanagements. Außerdem arbeitet die GWDG intern an mehreren Forschungsprojekten im Bereich Datenmanagement, woraus ein vorteilhafter Austausch zwischen Wissenschaft und Anwendung für beide Seiten entsteht.

ARCHITEKTUR

Das Open Archival Information System (OAIS) [5] ist ein Referenzmodell für ein Archivinformationssystem und als ISO-Standard veröffentlicht. Es wurde durch die NASA initiiert und mit der Raumfahrtorganisation ESA und weiteren Institutionen weiter spezifiziert. Durch das OAIS ist u. a. ein allgemeingebräuchlicher Sprachschatz definiert. Beispielsweise „Ingest“ für die Einspeisung und „Access“ für das Abholen von Daten aus dem Archivsystem. Abb. 1 zeigt die funktionalen Bestandteile des OAIS-Referenzmodells.

Die beteiligten Systeme auf Seiten der DNB und der GWDG bilden eine „service-orientierte Architektur“ (SOA), welche in der Gesamtheit als ein OAIS-konformes Langzeitarchiv verstanden wird. Das DIAS-System ist ein Bestandteil davon, der vor allem die Bereiche Data Management und Archival Storage abdeckt.



1_OAIS Functional Entities

ABLIEFERUNGS-WORKFLOW

Im Zuge des Ablieferungs-Workflows werden die zu archivierenden digitalen Objekte zunächst vom Ablieferer an die DNB gesendet. Anschließend erfolgen u. a. eine automatische Extraktion von technischen Metadaten sowie eine Bewertung des Archivguts mit einem sogenannten Ingest-Level. Der Ingest-Level gewährleistet anhand von Kriterien wie z. B. Dateintegrität, Beschränkungsfreiheit, Formatvalidität, dass eine definierte Qualität eingehalten wird. Beispiel: Die Checksumme eines PDF-Dokuments konnte erfolgreich validiert werden; dadurch hält das Dokument das Ingest-Level 0. Darauf folgend wird geprüft, ob das Dateiformat erkennbar ist, was zum Ingest-Level 1 führt usw.

Jede erfolgreich durchlaufene Stufe bescheinigt dem Dokument somit eine höhere technische Qualität. DNB-seitig ist definiert, bis zu welchem Ingest-Level welche Dateiformate zur Archivierung akzeptiert werden. Die Objekte und zugehörigen Metadaten werden anschließend an DIAS übertragen. DIAS validiert nochmals die Konformität, indem es u. a. die Metadaten gegen definierte Schemata validiert und Checksummen prüft. Als nächstes wird das Archivgut mit weiteren technisch notwendigen Daten angereichert und im Tivoli Storage Manager (TSM) [6] gesichert. Die Daten werden durch TSM-Mechanismen redundant in der Bandbibliothek der GWDG im Rechenzentrum „Am Faßberg“ und in der Universitätsmedizin Göttingen (UMG) gespeichert. Aktuell sichert das System die täglichen Neueingänge der digitalen Pflichtablieferung bei der DNB. Die Archivierung weiterer Datenbestände aus anderen Digitalisierungsprojekten ist in Vorbereitung. Abb. 2 zeigt die Architektur von der Ablieferung (Ingest) bis zur Bereitstellung (Access).

DIAS – TECHNISCHE REALISIERUNG

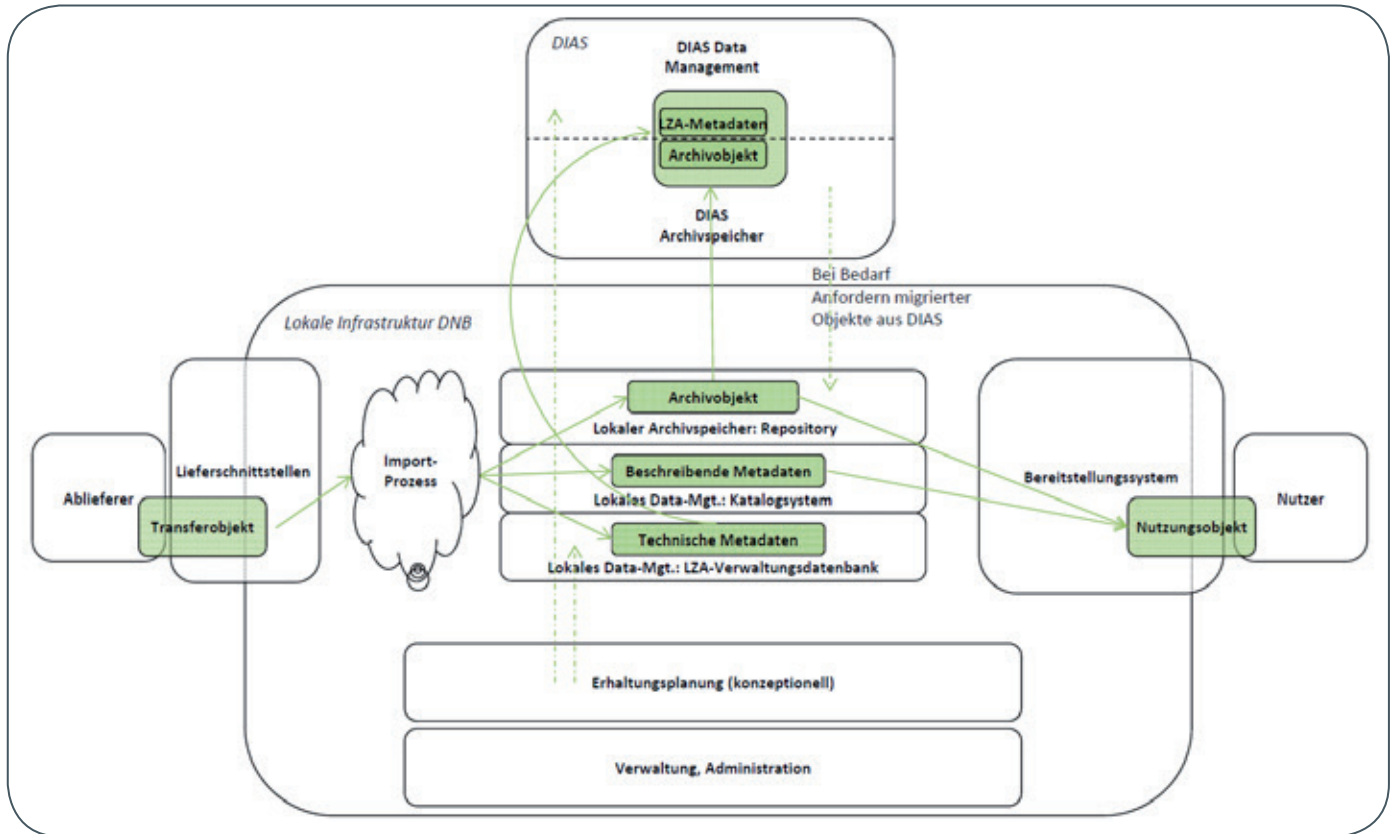
Die DIAS-Installation der GWDG nutzt eine IBM Power7 Hardware mit acht CPUs und 64 GByte RAM. Mit der Power Hardware steht eine eigene Virtualisierungsplattform zur Verfügung, in der aktuell zehn virtuelle Maschinen (sogenannte LPARs) für das Produktiv- und Testsystem genutzt werden. Mit dem Host verbunden ist ein IBM DS3500 Disk Storage mit ca. 11 TByte Brutto-Speicherkapazität, der den virtuellen Maschinen im RAID-6-Modus zur Verfügung gestellt wird. Durch gängige Maßnahmen sind die DIAS-Systeme gegen den Zugriff von außen geschützt. U. a. wird die Verbindung zu DIAS durch TLS verschlüsselt und durch Firewallbeschränkungen ausschließlich der DNB zur Verfügung gestellt. Als schematischer Aufbau der Produktivumgebung soll Abb. 3 dienen.

AKTUELLE ENTWICKLUNGEN UND ZUKUNFT

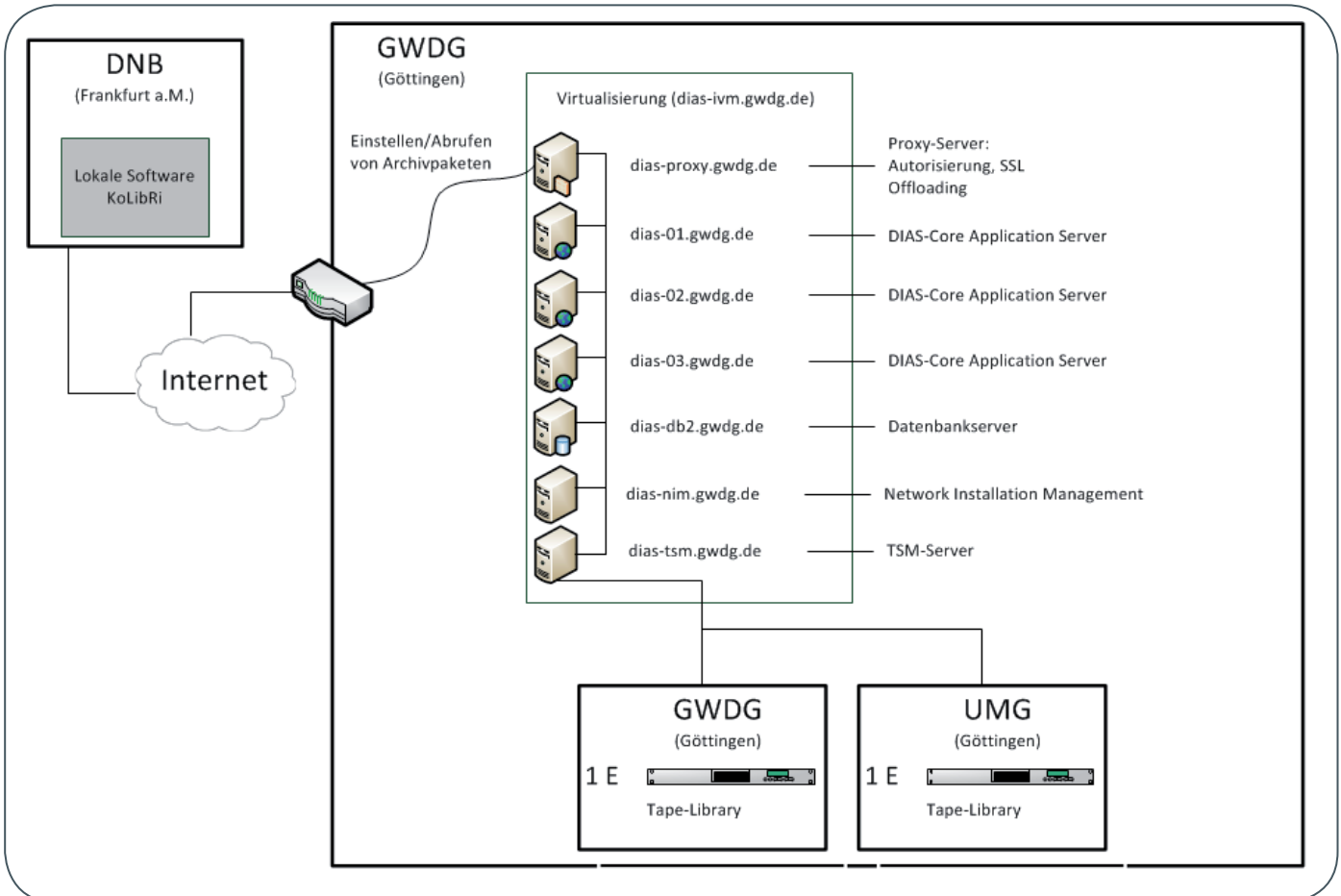
2014 – 2015 wurde das Archivsystem der DNB nach den Richtlinien des Data Seal of Approval [7] als vertrauenswürdigen Langzeitarchiv zertifiziert. Als nächste Stufe wird aktuell eine Zertifizierung nach dem nestor-Siegel [8] für vertrauenswürdige digitale Langzeitarchive angestrebt. Nestor ist ein Kompetenznetzwerk zur digitalen Langzeitarchivierung mit Partnern aus verschiedenen Bereichen wie Bibliotheken, Archive und Museen. Die erfolgreiche Zertifizierung der GWDG nach ISO 9001 und die geplante Zertifizierung nach ISO 27001 werden auch in diesem Bereich für eine ständige Qualitätssicherung und kontinuierliche Qualitätsverbesserung sorgen.

Im Jahr 2014 führte die DNB den Dienst „Archivierung und dauerhafte Erhaltung digitaler Objekte“ – AREDO [9] ein. Mit AREDO bietet die DNB ihre Erfahrungen und Fähigkeiten auf dem Gebiet der digitalen Langzeitarchivierung Kultur- und öffentlichen Einrichtungen in Form einer Dienstleistung mit dem Ziel an, gemeinsam und partnerschaftlich ein kooperatives und spartenübergreifendes Langzeitarchivierungssystem aufzubauen. Mit AREDO archivierte Daten setzen auf die gleiche wie in diesem Artikel vorgestellte Infrastruktur auf.

Durch die bereits langjährige Erfahrung der GWDG und der DNB mit dem DIAS-System ergaben sich einige neue Anforderungen. Daher stehen momentan Überlegungen im Raum, DIAS in der jetzigen Version auf lange Sicht abzulösen bzw. neu aufzulegen. Hierzu müssen allerdings noch weitere Überlegungen wie Funktionsumfang, Kosten und technische Realisierbarkeit mit



2_DNB-Langzeitarchivierungsinfrastruktur



3_Produktive Systemumgebung von DIAS

einbezogen werden. Es ist geplant, im Laufe des nächsten Jahres ein Konzept zu erstellen, in dem diese Bedingungen näher erörtert werden.

REFERENZEN

- [1] Gesetz über die Deutsche Nationalbibliothek (DNBG): <http://www.gesetze-im-internet.de/dnbg/BJNR133800006.html>
- [2] Verordnung über die Pflichtablieferung von Medienwerken an die Deutsche Nationalbibliothek: <http://www.gesetze-im-internet.de/bundesrecht/pflav/gesamt.pdf>
- [3] Sammelrichtlinien: <http://nbn-resolving.de/urn:nbn:de:101-2012022707>
- [4] kopal: <http://kopal.langzeitarchivierung.de>
- [5] Open Archival Information System (OAIS): <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- [6] Siehe auch die GWDG Nachrichten 8/2014, 11/2014 und 8/2015.
- [7] Data Seal of Approval: <http://datasealofapproval.org/en/assessment/>
- [8] Nestor: http://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html
- [9] AREDO: <http://www.dnb.de/aredo.html> ■

Parallelverarbeitung mit numerischen Bibliotheken – Leistungsmessungen für ScaLAPACK- und PBLAS-Routinen

Text und Kontakt:
Prof. Dr. Oswald Haan
oswald.haan@gwdg.de

Der Einsatz parallelisierter Softwarebibliotheken kann die effiziente Nutzung der Parallelverarbeitung auf dem GWDG Scientific Compute Cluster unterstützen. In diesem und einem Folgeartikel werden die Rechengeschwindigkeiten und das Skalierungsverhalten von Programmen mit Algorithmen der linearen Algebra aus den Paketen PBLAS und ScaLAPACK untersucht. Im vorliegenden Beitrag geht es um die parallele Matrizenmultiplikation, in dem in einer der nächsten Ausgaben der GWDG-Nachrichten erscheinenden Beitrag um die parallele Lösung von Gleichungssystemen und die Lösung von Eigenwertproblemen.

Die Parallelisierung numerischer Anwendungen zur effizienten Nutzung moderner Rechencluster ist aufwändig und der erzielte Erfolg hängt oft davon ab, in welchem Maße die Eigenschaften der numerischen Algorithmen, der verwendeten Parallelisierungssoftware und der in den Clustern eingesetzten Hardware ausgeschöpft werden. Für häufig verwendete numerische Verfahren stehen sehr effiziente Softwarebibliotheken zur Verfügung, mit denen auch ohne großen eigenen Programmieraufwand die hohe Rechenleistung der Parallelverarbeitung genutzt werden kann. Für Verfahren der linearen Algebra, wie Matrizenmultiplikation, das Lösen von linearen Gleichungssystemen oder der Bestimmung von Eigenwerten und -vektoren gibt es die Bibliotheken PBLAS [1] (Parallel

Parallel processing with numerical libraries

The efficient use of parallel processing on the GWDG Scientific Compute Cluster can be supported by employing available parallelized software libraries. In this and in a following second article performance and scaling behavior of linear algebra programs from PBLAS and ScaLAPACK will be analyzed. The first part deals with parallel matrix multiplication, whereas the second part covers solver for systems of linear equations and the solution of eigenvalue problems.

Basic Linear Algebra Subprograms) und ScaLAPACK [2] (Scalable Linear Algebra PACKage), die weit verbreitet sind und in den mathematischen Bibliotheken von Software- und Hardwareherstellern integriert sind. Die auf den Rechenclustern der GWDG bereitgestellte Intel MKL [3] (Intel® Math Kernel Library)-Bibliothek enthält alle PBLAS- und ScaLAPACK-Routinen.

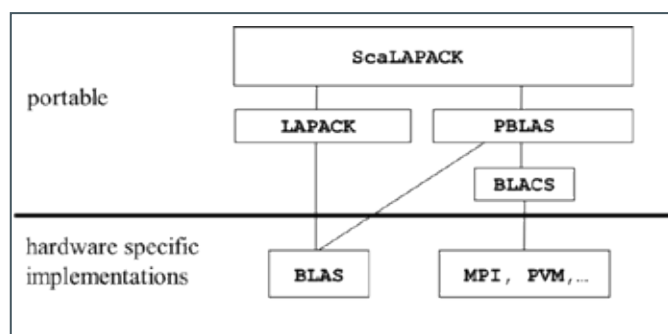
Vor der Parallelisierung von Anwendungen sollte stets eine Abschätzung des durch Parallelverarbeitung zu erwartenden Rechenzeitgewinns stehen, um zu entscheiden, ob der Aufwand zur Parallelisierung lohnenswert ist. Können in der Anwendung Routinen aus numerischen Bibliotheken verwendet werden, erfordert diese Abschätzung eine Kenntnis von deren Rechenzeiten auf Parallelrechnern. Die Analyse der Performanz einiger Routinen aus den PBLAS- und ScaLAPACK-Bibliotheken auf den Rechenclustern der GWDG kann für die Rechenzeitabschätzung für solche Anwendungen herangezogen werden, in denen diese Bibliotheksroutinen verwendet werden können. In einem ersten Teil soll hier die Matrizenmultiplikation aus PBLAS untersucht werden, in dem zweiten Teil dann Gleichungslösungen und Eigenwertprobleme aus ScaLAPACK.

PBLAS / SCALAPACK: PARALLELISIERUNGSKONZEPT UND DATENVERTEILUNG

Parallelverarbeitung in ScaLAPACK beruht auf der Verteilung von Operationen auf unabhängige Prozesse, die durch Nachrichtenaustausch miteinander kommunizieren. Diese Kommunikation erfolgt über die zu ScaLAPACK gehörige Bibliothek BLACS (Basic Linear Algebra Communication Subprograms), die ihrerseits auf die jeweils verfügbare Kommunikationsschnittstelle des Parallelrechnersystems zugreift. Auf den Rechenclustern der GWDG ist dies das standardisierte Message Passing Interface (MPI). ScaLAPACK kann so die Parallelität von vernetzten Einzelprozessoren, von einzelnen Multicore-Systemen und von Rechenclustern mit vernetzten Multicore-Systemen nutzen.

Die in ScaLAPACK implementierten Algorithmen der linearen Algebra verwenden die in den PBLAS (Parallel Basic Linear Algebra Subprograms) bereitgestellten parallelisierten einfachen Matrizenoperationen. Die Verteilung der Rechenlast erfolgt dabei so, dass alle beteiligten Prozesse Operationen auf Teilmatrizen durchführen, die jeweils unterschiedliche Blöcke der ursprünglichen Gesamtmatrix enthalten. Diese lokalen Matrizenoperationen werden durch Aufrufe der sequentiellen Softwarepakete BLAS (Basic Linear Algebra Subprograms) und LAPACK (Linear Algebra PACKage) realisiert.

Abb. 1 verdeutlicht diese Hierarchie der verschiedenen Softwarepakete. Sie zeigt auch, dass Programme durch die



1_Softwarehierarchie von ScaLAPACK

a11	a12	a13	a14	a15	a16	a17
a21	a22	a23	a24	a25	a26	a27
a31	a32	a33	a34	a35	a36	a37
a41	a42	a43	a44	a45	a46	a47
a51	a52	a53	a54	a55	a56	a57

2_Blockzyklische Aufteilung einer Matrix mit $m=5 \times n=7$ Elementen mit einer Blockgröße $mb=2 \times nb=2$

a11	a12	a15	a16	a13	a14	a17
a21	a22	a25	a26	a23	a24	a27
a51	a52	a55	a56	a53	a54	a57
a31	a32	a33	a34	a35	a36	a37
a41	a42	a45	a46	a43	a44	a47

3_Blockzyklische Verteilung der Elemente der globalen Matrix aus Abb. 2 auf ein Prozessgitter der Größe $nrow=2 \times ncol=2$

Verwendung von ScaLAPACK-Routinen auf allen Parallelrechnersystemen lauffähig sind, auf denen das BLAS-Paket und eine der von den BLACS unterstützte Kommunikationssoftware vorhanden sind und dass die Effizienz der Routinen davon abhängt, wie gut das BLAS-Paket und die Kommunikationsschicht auf der verwendeten Hardware implementiert sind.

PBLAS/ScaLAPACK legt der Parallelisierung der Verfahren der linearen Algebra eine blockzyklische Verteilung der beteiligten globalen Felder auf lokale Felder in einem zweidimensionalen Prozessgitter zugrunde. Die Abb. 2 und 3 illustrieren das an dem Beispiel einer Matrix mit $m=5$ Zeilen und $n=7$ Spalten, die blockzyklisch mit einer Blockgröße von $mb=2 \times nb=2$ auf die vier Prozesse eines Prozessgitters der Größe $nrow=2 \times ncol=2$ verteilt sind.

Die blockzyklische Datenverteilung dient einer guten Lastverteilung zwischen allen beteiligten Prozessen für die Parallelisierung der Algorithmen der linearen Algebra. Diese Algorithmen bearbeiten sukzessive immer kleinere Submatrizen der ursprünglichen Matrix. Die blockzyklische Verteilung ermöglicht, dass sich auch in den späteren Phasen der Algorithmen noch möglichst viele Prozesse an der Bearbeitung der Submatrizen beteiligen können.

Eine ausführlichere Beschreibung der blockzyklischen Datenverteilung und Beispielprogramme zur Realisierung solcher Verteilungen sind im GWDG-Wiki in der Kategorie „Scientific Computing:“ [4] unter dem Stichwort „ScaLAPACK“ zu finden.

LEISTUNGSERWARTUNG FÜR PARALLELE PROGRAMME

Parallelverarbeitung verspricht eine Verkürzung der Rechenzeit für die Programmausführung, im besten Fall um den Faktor $1/np$, wenn np Prozesse gleichzeitig für die Bearbeitung des Programms eingesetzt werden. Mit einer Leistungsmessung für ein gegebenes Programm kann ermittelt werden, inwieweit diese

Erwartung erfüllt wird.

Zur genauen Analyse der Leistung haben sich folgende Größen eingebürgert:

ANZAHL VON OPERATIONEN	N_{OP}
Ausführungszeit mit np Prozessen	T_{np}
Ausführungsgeschwindigkeit	$r_{np} = N_{op}/T_{np}$
Beschleunigung	$s_{np} = T_1/T_{np}$
Effizienz	$e_{np} = s_{np}/np$

Die zwei wichtigsten Faktoren, die die Effizienz der Parallelverarbeitung auf einen Wert unter 1 senken, sind:

1. Kommunikation zwischen den Prozessen
2. Ungleiche Verteilung der Rechenlast auf die Prozesse

Der Aufbau eines Kommunikationskanals zwischen den Prozessen nimmt Zeit in Anspruch, während der keine Operationen bearbeitet werden können. Diese Latenzzeit beträgt in den heute gängigen Kommunikationsnetzen ca. eine Mikrosekunde. In dieser Zeit könnte ein Prozessor mit einer Rechengeschwindigkeit von 10 GigaFlop/s (also 10 Milliarden Fließkommaoperationen pro Sekunde) 10.000 Operationen durchführen. Der eigentliche Austausch von Daten nimmt ebenfalls Zeit in Anspruch. Die Bandbreite innerhalb heutiger Netze liegt typischerweise bei 40 GBit/s. In der Zeit, um eine Fließkommazahl (Länge 64 Bit) zu senden, könnten 16 Fließkommaoperationen bearbeitet werden. Die zu erreichende Effizienz hängt also in starkem Maße ab von Anzahl und Inhalt der bei der Parallelverarbeitung anfallenden Kommunikation im Verhältnis zur Menge der von dem einzelnen Prozess durchzuführenden Operationen. Für viele Anwendungen wächst die Zahl der Rechenoperationen mit der Problemgröße schneller als der Kommunikationsaufwand, so dass die Effizienz bei gleichbleibender Prozesszahl mit der Problemgröße ansteigt. Bei konstanter Problemgröße und wachsender Prozesszahl jedoch sinkt die Rechenlast pro Prozess, der Kommunikationsaufwand jedoch bleibt gleich oder steigt sogar, so dass in diesem Fall die Effizienz kleiner wird.

Die Verminderung der Effizienz durch ungleiche Lastverteilung wird besonders deutlich, wenn Teile des Programms nicht parallelisiert werden können und von nur einem Prozess bearbeitet werden müssen. Dann gilt das Amdahlsche Gesetz [5], das bereits 1967 formuliert wurde: Bei einem Anteil s an nicht parallelisierbaren Operationen sinkt die Effizienz auf $e = 1/(1+(np-1)s)$. Bei konstanter Problemgröße sinkt damit unweigerlich die Effizienz mit steigender Zahl der eingesetzten Prozesse. Bei wachsender Problemgröße wächst in vielen Anwendungen jedoch nur der Anteil der parallelisierbaren Operationen, so dass dann bei gleichbleibender Prozesszahl die Effizienz steigt.

DER RECHENCLUSTER DER GWDG

Die absolute Größe der von parallelen Programmen erzielbaren Rechengeschwindigkeit hängt natürlich von der Leistung der verwendeten Hardware ab: von der Anzahl der Recheneinheiten der einzelnen CPU und deren Taktung, von Größe und Geschwindigkeit des Speichersystems der einzelnen Rechenknoten und von der Leistungsfähigkeit des Kommunikationsnetzwerks, das die Rechenknoten miteinander verbindet.

Der Rechencluster der GWDG für das wissenschaftliche Rechnen (GWDG Scientific Compute Cluster) besteht aus mehreren Gruppen von Rechenknoten mit jeweils unterschiedlichen

Hardwareausstattungen, die durch ein gemeinsames Infiniband-Netzwerk zusammengeschaltet sind. Die Finanzierung des Clusters erfolgt zum einen durch Investitionsmittel, die der GWDG durch ihre beiden Gesellschafter Universität Göttingen und Max-Planck-Gesellschaft zur Verfügung gestellt werden, zum anderen auch durch Projektmittel einzelner Forschungsgruppen der Universität und der Max-Planck-Gesellschaft. Entsprechend den Finanzierungsanteilen ist die Nutzung von Teilen des Clusters auf die Mitglieder der finanzierenden Gruppen beschränkt. Die Leistungsfähigkeit des GWDG Scientific Compute Clusters wird regelmäßig erhöht durch Hinzufügen von Rechenknoten mit der aktuell verfügbaren Technologie und Abschalten veralteter Systeme. Für die allgemeine Nutzung durch alle Institute der Universität Göttingen und der Max-Planck-Gesellschaft stehen zurzeit über 500 Rechenknoten mit insgesamt fast 15.000 Cores und einer maximalen Rechenleistung von 200 TeraFlop/s zur Verfügung. Eine detaillierte Übersicht über den Rechencluster ist auf den Webseiten zum Thema High Performance Computing in Göttingen (GöHPC) [6] und im GWDG-Wiki in der Kategorie „Scientific Computing:“ unter dem Stichwort „Scientific Compute Cluster“ zu finden

Die Messung der Rechenleistung für die PBLAS- und ScaLAPACK-Routinen wurde auf den neuesten Systemen des Clusters durchgeführt, die mit jeweils zwei CPUs vom Typ Intel E5-2640 v3 bestückt sind. Diese Haswell-Knoten wurden durch Projektmittel einzelner Institute beschafft und sind deshalb nicht für die allgemeine Nutzung freigegeben. Die auf diesen Systemen ermittelten Leistungswerte können aber auch zur Abschätzung der auf den älteren Systemen zu erwartenden Leistung verwendet werden. Das Skalierungsverhalten bzw. die Effizienz der Parallelisierung sollte auf den älteren Systemen besser sein als auf den Haswell-Systemen, da die geringere Leistung des einzelnen älteren Systems ein günstigeres Verhältnis von Kommunikations- zu Rechenleistung erzeugt und somit die Effizienzverluste bei der Parallelisierung vermindert. Die absolute Größe der Rechenleistung der parallelen Anwendungen wird auf den älteren Systemen natürlich um einen Faktor vermindert, der die verminderte Leistung des älteren Einzelsystems im Vergleich zum Haswell-System berücksichtigt. Ebenso können die jetzt ermittelten Leistungswerte später auch zur Abschätzung der Leistung auf zukünftig im Cluster eingesetzten neueren Systemen herangezogen werden.

Die wichtigsten Kenndaten dieser Haswell-Rechenknoten sind:

CPUs pro Knoten	2
Cores pro CPU	8
Taktfrequenz [GHz]	2,6
Flops pro Takt und Core	16
Hauptspeicher pro Knoten [GByte]	256

Aus diesen Kenndaten allein kann die maximal erreichbare Rechenleistung der einzelnen Cores und des gesamten Rechenknotens allerdings nicht ermittelt werden. Dies liegt zum einen an der Turbo-Funktion der Intel-Prozessoren, die bei genügender Kühlleistung die Prozessoren mit einer höheren Taktfrequenz arbeiten lässt. Zum anderen werden die angegebenen 16 Flops pro Takt nur bei Verwendung der AVX2-Befehle erreicht, alle anderen Befehle können maximal 8 Flops pro Takt generieren. Die AVX2-Befehle wiederum werden mit einer geringeren Grundfrequenz abgearbeitet, die allerdings auch durch die Turbo-Funktion erhöht werden kann.

Die im Rechenknoten insgesamt erreichbare Leistung hängt dann davon ab, welche Taktfrequenz bei Aktivität aller Cores gefahren wird. Ein weiterer Faktor, der die Gesamtleistung beeinflusst, ist der allen Cores gemeinsame Hauptspeicher und Level3-Cache. Wenn alle Cores gleichzeitig einen hohen Datenverkehr erzeugen, kann dies zu einer Reduktion der Zugriffsgeschwindigkeit für den einzelnen Core führen.

Die im einzelnen Rechenknoten erreichbare Rechenleistung hängt deshalb von der jeweiligen Anwendung ab und kann nur empirisch bestimmt werden. Als Vergleichswert für die zu bestimmenden Rechenleistungen für die parallelen PBLAS- und ScaLAPACK-Anwendungen wird die Geschwindigkeit der sequentiellen BLAS-Routine DGEMM zur Matrizenmultiplikation zweier quadratischer Matrizen der Ordnung n herangezogen. Die Anzahl der Fließkommaoperationen für diese Matrizenmultiplikation ist $N_{op} = 2n^3$, die Rechengeschwindigkeit wird aus der Laufzeit T für die Multiplikation als $r = N_{op}/T$ bestimmt und in der Einheit GFlop/s (Milliarden Fließkommaoperationen pro Sekunde) gemessen. Abb. 4 zeigt die erreichten GFlop/s-Raten bei der Ausführung der Multiplikation für eine unterschiedliche Anzahl gleichzeitig laufender Berechnungen der sequentiellen DGEMM-Routine in Abhängigkeit von der Matrixordnung n , Abb. 5 stellt die Abhängigkeit der DGEMM-Rechengeschwindigkeit von der Zahl gleichzeitig laufender Prozesse für die Matrixordnung $n=4000$ dar. Die Geschwindigkeit sinkt von 50 GFlop/s für die exklusive Nutzung nur eines Cores auf 34,5 GFlop/s, wenn alle 16 Cores des Knoten aktiv sind. Die Rechengeschwindigkeit von 34,5 GFlop/s pro Prozess kann als maximal erreichbare Rechenleistung für parallele Anwendungen auf diesen Systemen angesehen werden

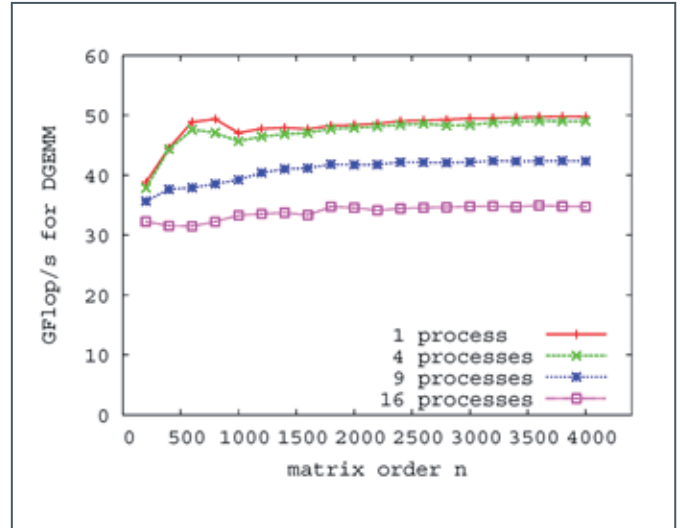
RECHENGESCHWINDIGKEIT FÜR PARALLELE MATRIZENMULTIPLIKATION MIT DER PBLAS-ROUTINE PDGEMM

Die Ausführungsdauer der PDGEMM-Routine zur Berechnung von $C=AB$ hängt von einer Vielzahl von Faktoren ab:

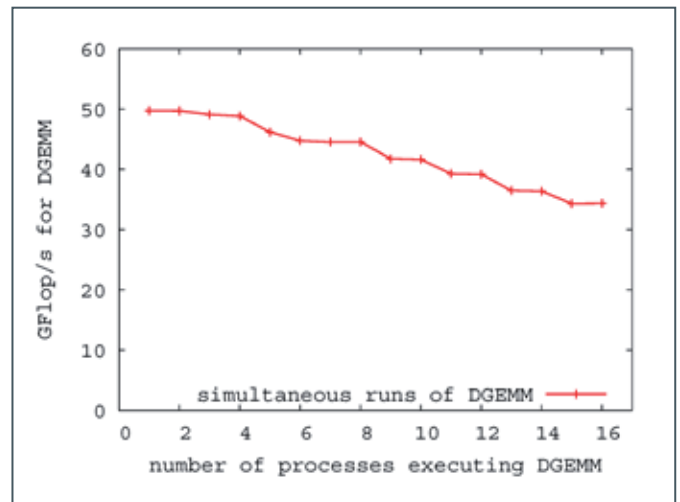
- die Dimensionen m, k, n bei der Multiplikation einer Matrix A mit m Zeilen und k Spalten mit einer Matrix B mit k Zeilen und n Spalten,
- die Dimensionen $nrow$ und $ncol$ des zweidimensionalen Prozessgitters, auf dem die Multiplikation parallel bearbeitet wird,
- die Blockgrößen $mb_A, nb_A, mb_B, nb_B, mb_C, nb_C$, die die blockzyklische Verteilung der drei Matrizen A, B und C auf die Prozesse des Prozessgitters festlegen.

Die mit PDGEMM erreichbaren Rechenleistungen wurden in einem kleinen Teil dieses großen Parameterraums untersucht: Quadratische Matrizen mit n Zeilen und n Spalten, quadratische Prozessgitter mit $nq \times nq$ Prozessen, gleiche quadratische Blockgrößen $nb \times nb$ für alle Matrizen. PDGEMM parallelisiert die Matrizenmultiplikation durch Aufteilung in viele Matrizenmultiplikationen reduzierter Matrizen der Größe $n/nq \times n/nq$, die auf die beteiligten Prozesse verteilt werden. Zur besseren Vergleichbarkeit wird deshalb die Rechenleistung pro Prozess angegeben und jeweils in Abhängigkeit der skalierten Matrixdimension n/nq dargestellt.

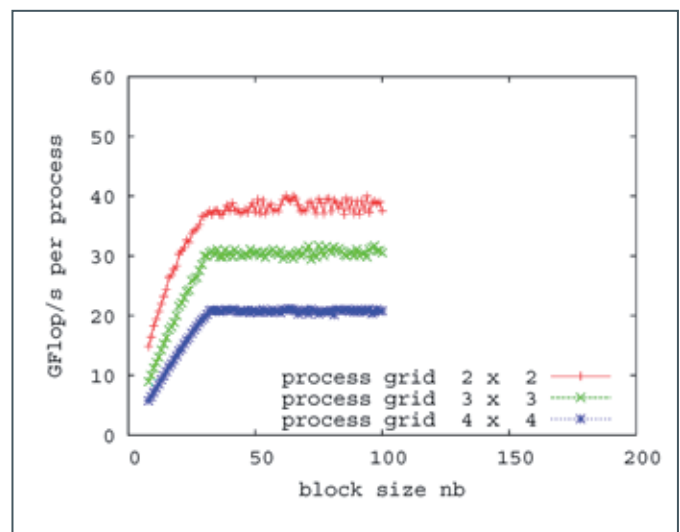
Abb. 6 zeigt die Rechenleistung für eine skalierte Matrixdimension von $n/nq=4000$ für verschiedene Größen des Prozessgitters in Abhängigkeit der Blockgröße nb .



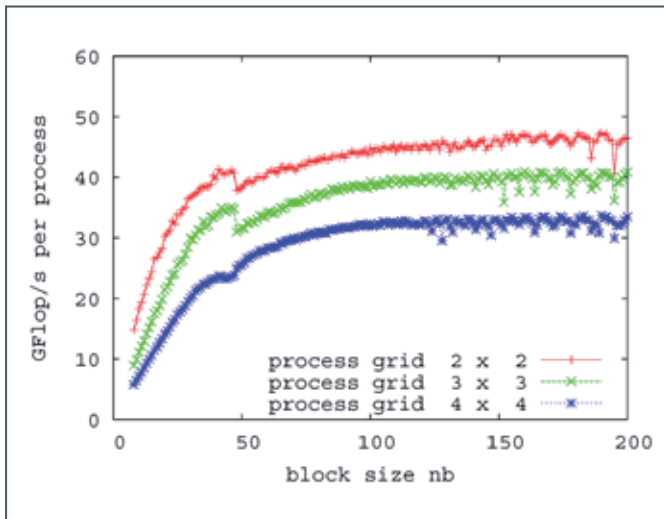
4_Rechenleistung der BLAS-Routine DGEMM in Abhängigkeit der Matrixordnung n



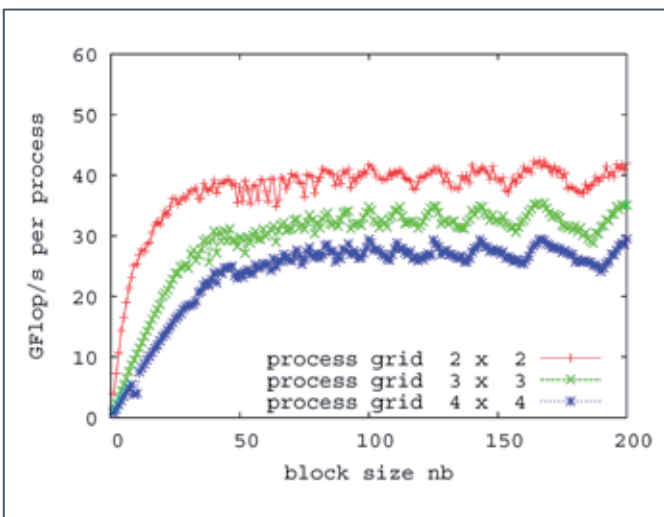
5_Rechenleistung der BLAS-Routine DGEMM in Abhängigkeit der Anzahl gleichzeitig laufender Prozesse, die Matrixordnung ist n=4000.



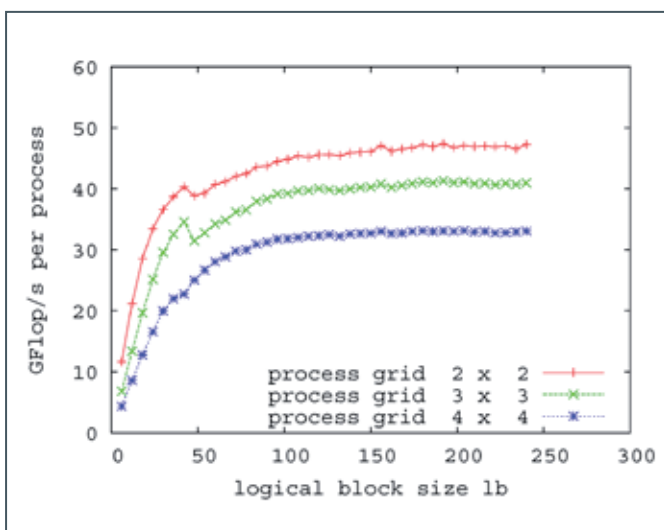
6_Rechenleistung pro Prozess für die Matrizenmultiplikation zweier quadratischer Matrizen mit PDGEMM mit einer logischen Blockgröße von $lb=32$. Die verwendete Matrixdimension ist $n=4000nq$ bei der parallelen Ausführung auf einem Prozessgitter der Größe $nq \times nq$.



7_Rechenleistung pro Prozess für die Matrizenmultiplikation zweier quadratischer Matrizen mit PDGEMM mit einer logischen Blockgröße von $lb=156$. Die verwendete Matrixdimension ist $n=4000nq$ bei der parallelen Ausführung auf einem Prozessgitter der Größe $nq \times nq$.



8_Rechenleistung pro Prozess für die Matrizenmultiplikation zweier quadratischer Matrizen mit PDGEMM mit einer logischen Blockgröße von $lb=156$. Die verwendete Matrixdimension ist $n=1000nq$ bei der parallelen Ausführung auf einem Prozessgitter der Größe $nq \times nq$. Die Maxima treten bei Werten von $nb=1.000/5, 1.000/6, 100/7, \dots$ auf.



9_Rechenleistung pro Prozess für die Matrizenmultiplikation zweier quadratischer Matrizen mit PDGEMM mit einer Blockgröße von $nb=250$. Die verwendete Matrixdimension ist $n=4000nq$ bei der parallelen Ausführung auf einem Prozessgitter der Größe $nq \times nq$.

Offensichtlich steigt die Rechenleistung pro Prozess mit wachsender Blockgröße, allerdings nur bis zu einem Wert von $nb=32$ und bleibt dann mit Schwankungen konstant unterhalb eines Wertes von etwa zwei Dritteln der für die simultane Ausführung einer äquivalenten Zahl von DGEMM-Prozessen gemessenen Leistung (vgl. Abb. 4).

Die Ursache für diese unbefriedigende Leistungsausbeute liegt darin, dass in den PBLAS- und ScaLAPACK-Routinen die lokalen Matrizenoperationen durch DGEMM-Aufrufe mit einer intern festgelegten Matrixgröße ausgeführt werden, die durch eine „logischen Blockgröße“ lb bestimmt wird [7]. Die logische Blockgröße wird in den PBLAS- und ScaLAPACK-Routinen durch die intern definierte Funktion PILAENV festgelegt. In der derzeitigen Version von ScaLAPACK wird in PILAENV der Wert $lb=32$ gesetzt, unabhängig von Größe und Verteilung der beteiligten Matrizen. Die für PBLAS und ScaLAPACK bereitgestellte Nutzerdokumentation erwähnt die Verwendung der internen logischen Blockgröße mit dem festgelegten Wert nicht. In der Literatur und in Internet-Foren wird die logische Blockgröße angesprochen [8] und auch die Möglichkeiten, sie vom Anwender zu modifizieren [9].

Das in Abb. 6 gezeigte Verhalten der Rechenleistung erklärt sich aus der Tatsache, dass unabhängig von der Blockgröße der Matrizenaufteilung die Matrixdimensionen der lokalen Matrizenmultiplikationen mit DGEMM durch die logische Blockgröße $lb=32$ auf einen suboptimalen Wert festgelegt sind.

Für die weitere Analyse der PDGEMM-Performanz wurde die Festlegung von lb aufgehoben durch Einbindung einer eigenen Funktion PILAENV mit der in der ScaLAPACK-Dokumentation beschriebenen Schnittstelle [10]. Damit kann lb auf einen beliebigen Wert gesetzt werden.

Abb. 7 zeigt die Rechenleistung in Abhängigkeit der Blockgröße nb bei einer größeren logischen Blockgröße $lb=156$. Jetzt steigt die Leistung für Blockgrößen über $nb=32$ weiter an bis zu Werten, die weniger als 10 % unter denen der in Abb. 4 mit DGEMM erreichten liegen.

Es zeigt sich auch, dass die Leistungen ab etwa $nb=120$ größeren Schwankungen unterworfen sind. Diese Schwankungen sind noch ausgeprägter für kleinere Werte der skalierten Matrixdimension, wie in der Abb. 8 für $n/nq=1000$ zu sehen ist. Die Maxima werden erreicht für Werte von nb , die nahe bei $1000/5, 1000/6, 1000/7, \dots$ liegen. Die blockzyklische Verteilung der globalen Matrizen ergibt für diese Werte von nb nahezu gleich große lokale Anteile der Matrizen in allen beteiligten Prozessen, was zu einer besonders homogenen Lastverteilung für Kommunikation und Rechenoperationen bei der Parallelisierung führt.

Diese Untersuchungen zeigen also, dass auf einem Haswell-Rechenknoten im GWDG-Cluster die Rechenleistung für die parallelisierte Matrizenmultiplikation PDGEMM auf einem $nq \times nq$ Prozessgitter maximiert wird, wenn eine Blockgröße von über $nb=100$, die in der Nähe eines Teilers der skalierten Matrixdimension n/nq liegt, sowie eine logische Blockgröße von über $lb=150$ gewählt werden. Die damit erzielte Rechenleistung für $n=16000$ auf 16 Prozessen mit $nq=4$ liegt bei 33 GFlop/s und damit weniger als 5 % unter der von der sequentiellen Matrizenmultiplikation DGEMM auf 16 Prozessen erzielten Leistung.

Es bleibt noch zu analysieren, wie sich PDGEMM bei Verwendung mehrerer Rechenknoten, die über das QDR-Infiniband-Netz kommunizieren, verhält. Die in Abb. 10 dargestellte Gesamtrechenleistung zeigt, dass eine Sättigung der Rechenleistung für

wachsende Prozessgitter bei immer größeren Matrixdimensionen eintritt. Für das größte untersuchte Prozessgitter mit $16^2=256$ Prozessen wächst die Rechenleistung noch deutlich bis zu bei einer skalierten Matrixdimension von 8.000. Der Speichergröße der Haswell-Rechenknoten mit 16 Cores ist 256 GByte, so dass die drei für die Matrizenmultiplikation benötigten Matrizen bis zu einer skalierten Matrixdimension $n/\sqrt{nproc}=18000$ im Speicher Platz hätten. Eine Bestimmung der Rechenzeit für größere Matrixdimensionen wurde jedoch nicht durchgeführt, weil dies zu einer unverhältnismäßigen Belastung des GWDG-Rechenclusters geführt hätte. Bei der Verwendung von 256 Prozessen benötigt eine parallele Matrizenmultiplikation mit einer skalierten Dimension von 8000 bereits über 10 Min., bei einer skalierten Dimension von 18000 mehr als zwei Stunden

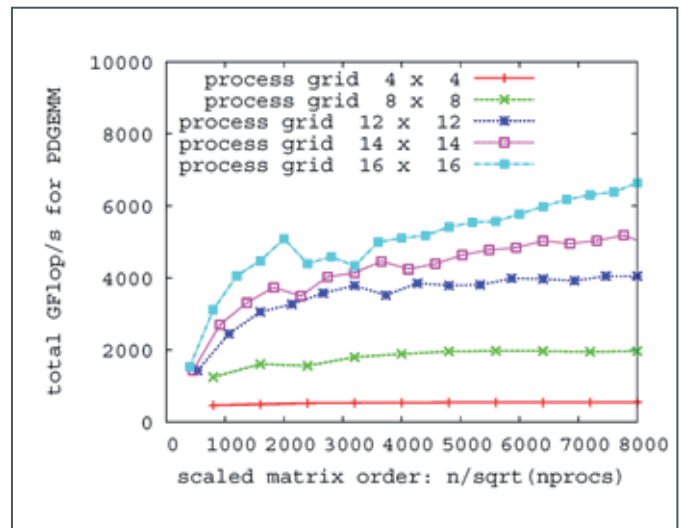
Die in der Abb. 10 sichtbaren Abweichungen des Verlaufs der Rechenzeiten von einer glatten Kurve sind auf zwei Effekte zurückzuführen: Zum einen werden in PDGEMM je nach Matrixgröße unterschiedliche Strategien für die Kommunikation der Matrixblöcke zwischen den Prozessen verwendet. Der Einbruch der Leistung zwischen den skalierten Dimensionen von 2000 und 3000 könnte auf einen Wechsel in der Kommunikationsstrategie zurückzuführen sein. Zum anderen steht das Infiniband-Netz nicht exklusiv für eine Anwendung zur Verfügung. Die Belastung des Netzes durch andere Rechenknoten kann die nutzbare Netzleistung vermindern und dadurch unkontrollierbare Schwankungen in der Ausführungszeit der untersuchten Anwendung verursachen.

Die in der Abb. 10 enthaltene Information ist in den Abb. 11 und 12 für die Darstellung der wichtigen Leistungsgrößen Beschleunigung und Effizienz umgesetzt worden. Es wird deutlich sichtbar, dass auch für die größten untersuchten Matrizen die Beschleunigung sublinear ist und die Effizienz deshalb unter 1 liegt.

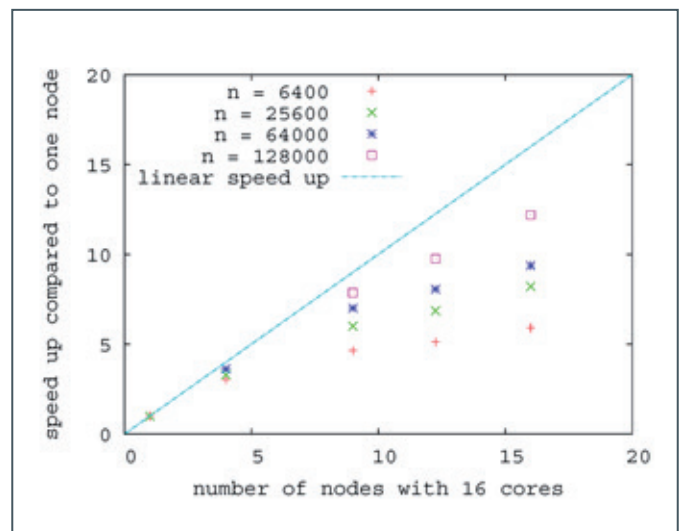
Die beiden Abbildungen zeigen, dass die ScaLAPACK-Routine PDGEMM für parallele Matrizenmultiplikation bei 256 Prozessen bei einer Matrixgröße von $n=128000$ mit einer Rechenleistung von mehr als 6,6 TeraFlop/s eine Effizienz von über 75 % erreicht und dass eine Steigerung dieser Leistung für noch größere Matrizen zu erwarten ist.

REFERENZEN

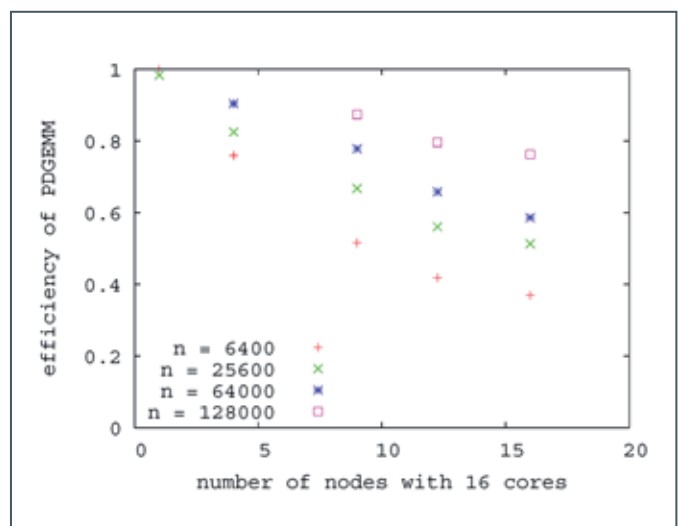
- [1] http://www.netlib.org/scalapack/pblas_qref.html
- [2] <http://www.netlib.org/scalapack/>
- [3] <https://software.intel.com/en-us/intel-mkl>
- [4] http://wiki.gwdg.de/index.php/Kategorie:Scientific_Computing
- [5] Gene Amdahl: Validity of the Single Processor Approach to Achieving Large-Scale Computing Capabilities. In: AFIPS Conference Proceedings. 30, 1967, S. 483–485
- [6] <http://hpc.gwdg.de/systems.html>
- [7] Sascha Hunold; Thomas Rauber: Automatic Tuning of PDGEMM Towards Optimal Performance 2005, Springer, in: Proc. of the Euro-Par Conference 2005
- [8] <https://icl.cs.utk.edu/lapack-forum/viewtopic.php?f=2&t=4046>
- [9] http://doxygen.cp2k.org/db/d7f/pilaenv__hack_8f90_source.html
- [10] http://www.netlib.org/scalapack/explore-html/d7/d8f/pilaenv_8f_source.html



10_Gesamtrechenleistung für die Matrizenmultiplikation zweier quadratischer Matrizen mit PDGEMM mit einer Blockgröße von $nb=200$ und logischer Blockgröße von $lb=156$ bei der parallelen Ausführung auf einem Prozessgitter der Größe $n \times n$



11_Beschleunigung der Matrizenmultiplikation mit PDGEMM bei Erhöhung der Anzahl beteiligter Rechenknoten. Als Referenz ist die Ausführungszeit auf einem Knoten mit 16 Cores genommen.



12_Effizienz der Matrizenmultiplikation mit PDGEMM bei Erhöhung der Anzahl beteiligter Rechenknoten. Als Referenz ist die Ausführungszeit auf einem Knoten mit 16 Cores genommen.



Using the Parallel Processing Power of the GWDG Scientific Compute Cluster

Upcoming Introductory and Parallel Programming Courses

GWDG operates a scientific compute cluster with currently 14,800 cores and a total compute power of 200 Teraflops ($2.0 \cdot 10^{14}$ floating point operations per second), which can be used by all scientists of the institutes of GWDG's supporting organisations, University of Göttingen and Max Planck Society.

In order to facilitate the access to and the efficient use of these computing resources, GWDG offers introductory and parallel programming courses, held at GWDG's site 'Am Faßberg'.

The next courses in 2015 are

> October 12th, 9:30 am - 4:00 pm

Using the GWDG Scientific Compute Clusters – an Introduction

This course explains all steps for accessing GWDG's clusters, to compile and install software, and to work with the batch system for the execution of application jobs. The course is intended for new or inexperienced users of the clusters.

> October 13th - 14th, 9:15 am - 5:00 pm

Parallel Programming with MPI (Including MPI for Python)

This course introduces the message passing interface (MPI) for programming parallel applica-

tions in FORTRAN, C, and in Python. All concepts will be illustrated with hands on exercises. Examples of parallel applications will be presented and analysed.

> December 3th, 9:15 am - 4:30 pm

High-level, High-performance Technical Computing with Julia

Julia is a modern programming language combining high-level dynamic programming with high performance. The course covers the basics of Julia including numerical computing, parallel computing, and statistical methods.

These three courses are repeated regularly. Other courses on parallel computing, dealing with more specialized topics can be arranged on demand. The possible subjects include parallel programming for shared memory systems and for graphics processors, and using extensions of C or Fortran with high level parallel constructs.

More Information about the courses held regularly or on demand at www.gwdg.de/scientific-computing-courses.

Information for registering for the courses at www.gwdg.de/courses.

If you have any further questions please contact support@gwdg.de.

Interaktive Visualisierung wissenschaftlicher Daten im Web

Text und Kontakt:

Oliver Schmitt
oliver.schmitt@gwdg.de
0551 201-2176

Alexander Wildschütz
alexander.wildschuetz@gwdg.de
0551 39-20442

Die Visualisierung von Forschungsdaten hat einen festen Platz im Werkzeugkasten der Wissenschaftler aller Fachdisziplinen in Göttingen und der Max-Planck-Gesellschaft. Viele hochwertige Visualisierungen sind in Publikationen zugänglich für die Fachwelt, können aber auch Interesse bei einem Publikum über den Fachkreis hinaus erzeugen. Hierbei ergibt sich die Herausforderung, wie man Visualisierungen einem breiten Publikum zugänglich macht, um wissenschaftliche Erkenntnisse begreifbar zu vermitteln. Der technische Fortschritt im Bereich Webtechnologien erlaubt es heute, auch komplexere Visualisierungen ohne proprietäre Technologien wie Adobe Flash oder schwerfällige Browser-Plugins umzusetzen und dabei auch Nutzer mobiler Endgeräte zu erreichen. In diesem Artikel stellen wir Beispiele vor, bei denen Informatiker der GWDG mitwirkten, wissenschaftliche Daten und Visualisierungen unter Einbeziehungen dieser Technologien für eine interaktive Darstellung im Internet vorzubereiten.

VISUALISIERUNG IM WEB IST INTERAKTIV

Möchte man bestehende Visualisierungen von Forschungsdaten für eine Nutzung im Internet aufbereiten, so ist zunächst zu klären, mit welchem Konzept man die Daten präsentieren möchte. Statische Visualisierungsformen wie Diagramme oder Bilder lassen sich ohne größere Herausforderungen ins Web bringen. Das stellt seit fast 20 Jahren kein Problem dar. Interessant wird es bei Animationen, Filmen, 3D-Modellen und Forschungsdaten, die einen mehrdimensionalen Bezug aufweisen. Hier sollten die Daten nicht nur statisch im Web dargestellt werden, sondern es muss in der Regel eine ganze Benutzeroberfläche geschaffen werden, die es den Nutzern erlaubt, Animationen zu steuern, Landkarten zu vergrößern oder einen Zeitstrahl mit geschichtlichen Ereignissen abzuschreiten. D. h., neben der Frage, wie man die Visualisierung technisch für das Web vorbereitet, stellt sich auch die Frage, wie der Benutzer mit der Visualisierung in Interaktion treten kann. In den Beispielen, die in diesem Artikel vorgestellt werden, liefern Wissenschaftler die Visualisierungen in eigenen Dateiformaten, und in Zusammenarbeit mit Informatikern der GWDG wurden diese in webbrowserskompatible Formate überführt und gemeinsam Konzepte zur Präsentation der Daten im Web entwickelt. Besonders bei der korrekten Darstellung der Visualisierung ist das Feedback der Wissenschaftler unerlässlich, um beispielsweise virtuelle Lichtquellen bei 3D-Objekten korrekt auszurichten und

Raum-Zeit-Bezüge exakt im wissenschaftlichen Kontext wiederzugeben. Das Feedback der Informatiker ist für die Wissenschaftler hierbei genauso wichtig, um Visualisierungen an die Anforderungen von Webbrowsern anzupassen und eine verständliche Benutzeroberfläche zu entwickeln.

Interactive visualization of scientific data

The visualization of research data are well-established methods in the toolbox of researchers from all disciplines in Göttingen and the Max Planck Society. Many high-quality visualizations are available in publications for professionals, but can be also interesting for a broader audience. This results in the challenge of how to offer scientific visualizations to a wide audience in order to convey scientific information understandable. The technical progress in the field of web technologies allow today to implement even more complex visualizations without proprietary technologies such as Adobe Flash or cumbersome browser plug-ins to reach even users of tablets. In this article we present examples in which computer scientists of the GWDG supported to bring scientific data and visualizations using state of the art web technology to a broader audience on the web.

VON DER IDEE ZUR INTERAKTIVEN VISUALISIERUNG

Die Erstellung einer Visualisierung für das Web ist ein Prozess, der ein hohes Maß an Arbeitsteilung zwischen Wissenschaftler und Informatiker mit sich bringt. Im Fall der Informatik erfordert dies ein gewisses Grundverständnis für die wissenschaftlichen Sachverhalte, die im Web dargestellt werden sollen, um Algorithmen und Methoden für eine flüssige und detailgetreue Wiedergabe ideal auszuwählen. Auf Seiten der Wissenschaftler ist eine Affinität zu digitalen Methoden hilfreich. Idealerweise setzt man auf bereits bestehende Visualisierungen auf, die es bereits in einer Desktop-Applikation gibt oder auf konkrete Vorstellungen des Wissenschaftlers, die sich z. B. durch Zeichnungen oder Papiermodelle ausdrücken lassen. Bei der Umsetzung werden hierbei verschiedene Schritte durchlaufen, die hier in Kürze zusammengefasst werden.

Im ersten Schritt werden zunächst Funktionsmodelle, sog. Mockups, erstellt, die es erlauben, einen Eindruck von der späteren interaktiven Visualisierung zu gewinnen und den Gesamteindruck der Bildschirmabläufe festzuhalten. Ein hilfreiches Open Source-Werkzeug für die Erstellung von Mockups ist Pencil [1]. Zusammen mit den Wissenschaftlern können hier beispielsweise die Funktionen von Schaltflächen und auch interaktive Abläufe rasch durchgespielt werden, ohne zeitaufwändige Programmierung durchzuführen.

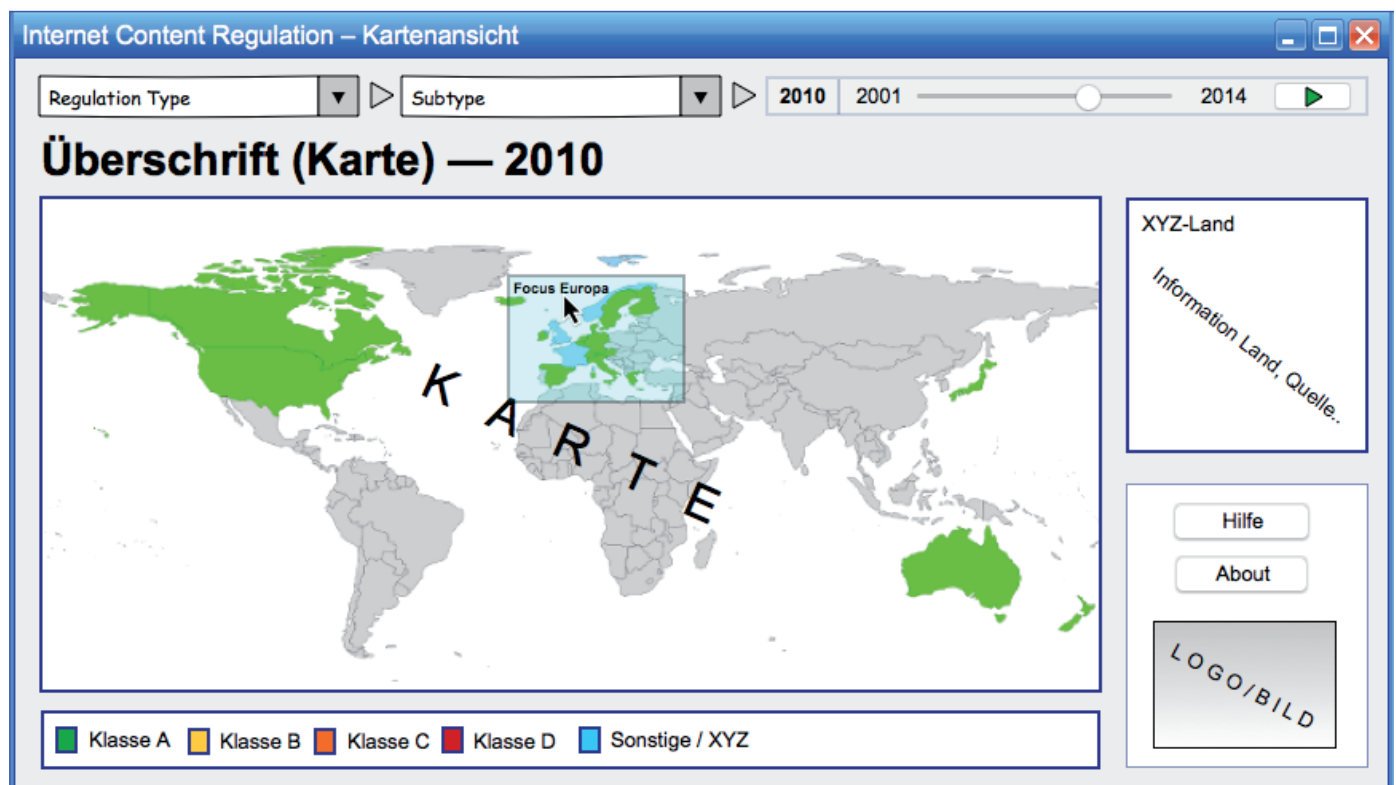
Im zweiten Schritt müssen die Forschungsdaten bzw. die fertigen Visualisierungen in ein für das Internet kompatibles Format gebracht werden. In der Regel steht hier eine Verdichtung / Reduktion von Informationen an. Beispielsweise müssen für 3D-Visualisierungen der Detailgrad der 3D-Modelle reduziert und Bilder sowie Texturen komprimiert werden, um lange Wartezeiten beim Laden der Webseite zu vermeiden und auch leistungsärmeren

Computern eine flüssige Drehung der Modelle im virtuellen Raum zu erlauben. Hierzu müssen die Werkzeuge zur Erzeugung von Grafiken und Modellen der Wissenschaftler angepasst werden, um die Daten Internet-gerecht zu erzeugen. Als Mittel der Wahl kommen hier Open Source-Werkzeuge zum Einsatz, die massenhaft Grafiken komprimieren können, wie beispielsweise ImageMagick [2] oder Blender [3] zum Anzeigen und Prüfen von 3D-Modellen.

Im letzten Schritt wird das Funktionsmodell mit JavaScript und HTML5 durch Programmierung der Informatiker (und auch durch Wissenschaftler selbst) in eine fertige Webanwendung überführt. Dabei kommen je nach Aufgabenstellung unterschiedliche Open Source-Komponenten zum Einsatz. Bei generischen Visualisierungen kann beispielsweise auf fertige Lösungen, wie etwa den DARIAH Geobrowser [4], zurückgegriffen werden. In der Mehrzahl der Fälle müssen allerdings neue Lösungen implementiert werden, da viele Visualisierungen sehr fachspezifisch sind.

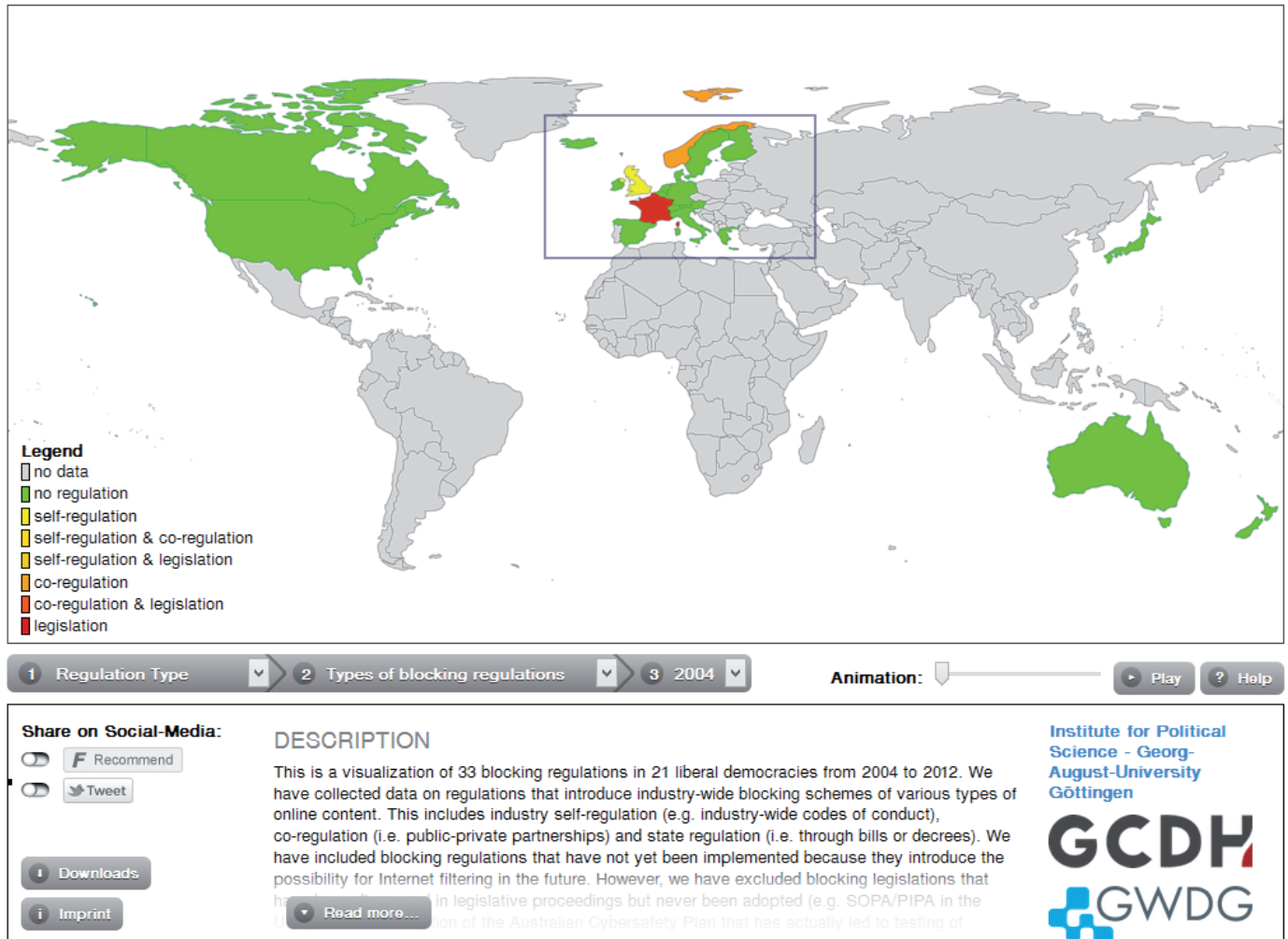
BEISPIEL 1: VISUALISIERUNG VON DATEN ZUR REGULIERUNG VON INTERNETINHALTEN IN LIBERALEN DEMOKRATIEN

Für das Forschungsprojekt „Internetinhaltsregulierungen in liberalen Demokratien“, das am Göttinger Center for Digital Humanities (GCDH) im Rahmen des Digital Humanities Forschungsverbands (DHFV) durchgeführt wurde, wurde eine geeignete Form gesucht, die Forschungsergebnisse interaktiv im Internet zu präsentieren. Hierbei sollte die Perspektive auf Internetinhalte und deren Regulierung in 23 liberalen Demokratien über eine Zeitschneise von mehreren Jahren gegenübergestellt werden. Unter der Leitung von Dr. Yana Breindl vom Institut für Politikwissenschaften der Universität Göttingen wurden hierzu Forschungsfragestellungen formuliert und beantwortet. Als ein Ergebnis entstand hierbei



1_Mit Pencil erstelltes Mockup zur Visualisierung von Inhaltsregulierung im Internet

Types of blocking regulations in liberal democracies — 2004



2_Die fertige Webanwendung zur Visualisierung von Inhaltsregulierung im Internet [6]

im Team von Frau Dr. Breindl eine Datenbank mit Forschungsdaten zur Inhaltsregulierung im Internet in verschiedenen Bereichen wie etwa Glücksspiel, Urheberrechtsverletzung oder jugendgefährdende Inhalte. In gemeinsamer Arbeit mit den Informatikern der GWDG wurde ein Konzept zur Präsentation der Daten im Internet entworfen (siehe Abb. 1). Zunächst mussten die Datenbankinhalte in Kartendarstellungen umgewandelt werden, sodass die Unterschiede zwischen den Ländern farblich deutlich wurden. Hierzu wurde die Statistiksoftware R mit dem Paket rworldmaps [5] eingesetzt, um aus Einträgen der Datenbank automatisch Länderkarten zu generieren. Mit IimageMagick [2] wurden die Grafiken anschließend automatisch verkleinert und komprimiert, um Ladezeiten im Browser zu verkürzen. Im Folgenden wurden die zeitlichen Abläufe in einem Datenmodell abgebildet und eine entsprechende Webanwendung von der GWDG unter der Leitung von Alexander Wildschütz programmiert und getestet (siehe Abb. 2). Sie können die fertige Visualisierung unter [6] in Ihrem Browser ansehen und unterschiedliche Staaten in Bezug auf die Regulierung von Internetinhalten vergleichen. Durch die Nennung der Internetadresse auf einem bekannten Blog zum Thema Politik wurden die wissenschaftlichen Ergebnisse von Dr. Breindl als Visualisierung bis zu 50.000 Mal am Tag aufgerufen [7].

BEISPIEL 2: INTERAKTIVE ANZEIGE VON 3D-MODELLEN IN DER ARCHÄOLOGIE

Ebenfalls im Rahmen des Digital Humanities Forschungsverbunds am GCDH wurde nach einer Möglichkeit gesucht, 3D-Modelle ohne den Einsatz von proprietären Technologien wie Java oder Flash im Browser wiederzugeben und Benutzern die Installation von Software auf dem Computer zu ersparen. Der konkrete Anwendungsfall besteht darin, 3D-Modelle aus archäologischen Sammlungen, die von Prof. Dr. Martin Langner, Professor für Klassische Archäologie und ihre digitale Methodik der Universität Göttingen, und seinem Team mit photogrammetrischen Methoden erstellt werden, für Internetnutzer im Webbrowser zugänglich zu machen. Prof. Dr. Langner kam hierbei auf die GWDG zu, um seine digitale Werkzeugsammlung zur Erstellung und Bearbeitung von 3D-Modellen um eine Plattform zur Anzeige der 3D-Modelle im Web zu erweitern. Hier konnte nach einer kurzen Anforderungsanalyse direkt mit der Auswahl der Softwarekomponenten und dem Bau eines 3D-Viewers auf Basis von JSC3D [8] [9] begonnen werden. Der 3D-Viewer ist hierbei als eine Webanwendung mit modernen Technologien aufgebaut. Die benutzerseitige Schnittstelle ist als Single Page Application mit HTML5 ausgeführt. Alle durch den Benutzer abrufbaren Inhalte

werden bei Bedarf dynamisch nachgeladen, um Wartezeiten zu verkürzen oder gänzlich einzusparen. Durch den Einbau einer Suchfunktionalität können die Beschreibungen der 3D-Modelle mit Filtern durchsucht werden. Serverseitig werden die Erfassung und Bereitstellung von 3D-Modellen sowie der zugehörigen Metadaten von mehreren PHP-Skripten umgesetzt. Es wird für jede Ressource wie 3D-Modell, Textur und Vorschaubild ein eindeutiger Code generiert und den Metadaten zugeordnet. Fordert ein Benutzer die Anzeige eines Modells an, werden dem Code zugehörige Daten an die Webapplikation übertragen und Querreferenzen (z. B. auf Texturen) während der Übertragung codiert.

Um ein flüssiges Drehen und Zoomen der 3D-Modelle umzusetzen, wird auf die WebGL-Schnittstelle des Browsers zugegriffen, die die Modelle, wie bei modernen 3D-Computerspielen üblich, direkt auf der Grafikkarte rendert. Somit können die 3D-Scans der archäologischen Objekte mit einer Mausgeste rasch gedreht werden, obwohl die Originale in der Realität mit einem Gewicht von über 100 Kilogramm nur mit einem Hubwagen oder Gabelstapler bewegt werden können.

Eine möglichst realistische Darstellung der Objekte stand bei

der Entwicklung im Vordergrund. Im Besonderen war die korrekte Texturierung, d. h. das Abbilden der echten Fotoaufnahmen auf dem 3D-Gittermodell wichtig. Ein neutraler Lichteinfall (Shading) und ein Verzicht auf besondere Interpolationsmethoden zeigten dabei das beste Ergebnis, da so die ursprünglichen Farbwerte auf dem 3D-Modell abgebildet werden. Insbesondere an der Texturkante, dem Schnitt zwischen verschiedenen Fototeilen, werden keine Farben benachbarter Pixel zur Interpolation herangezogen, welche die Farben verfälschen können. Dieses Feedback von Prof. Dr. Langner an die Informatiker der GWDG war wichtig, um eine Abweichung der Abbildungen durch die Rendering-Verfahren zu verhindern, um beispielsweise matte Oberflächen nicht glänzend erscheinen zu lassen. Nach Fertigstellung der ersten Serien an 3D-Scans und Abschluss der Testphase wird der 3D-Viewer in einigen Wochen auf den Servern der GWDG Fachwissenschaftlern und der Öffentlichkeit zugänglich gemacht werden. Somit kann ein breites Publikum archäologische Sammlungsgegenstände von jedem Ort der Welt aus betrachten. Ein Screenshot der aktuellen Version ist in Abb. 3 zu sehen.

[Home](#) • [Info](#) • [Impressum](#)

1111 Relief: Thronende Kybele



Modell Schließen

Beschreibung

- **Standort:**
Wien, KHM Inv. I 1111
- **Material/Maße:**
Marmor, H. max. 37 B. max. 23,5 T. max. 14 cm. Statuette ohne Kopf: H. 37 B. 23,5 T. 14 cm
- **Datierung:**
antik, 4. Jh. v. Chr.
- **Letzte Änderung:**
10.07.2015 12:46:50

Steuerung

- **Drehen:** Mit der Maus klicken und ziehen.
- **Zoomen:** Scrollrad drehen, oder **shift**-Taste halten und mit der Maus klicken und ziehen
- **Verschieben:** **ctrl**-Taste halten und mit der Maus klicken und ziehen.

Neustes Modell:
10.07.2015 12:46:50



3_Photogrammetrisch erstelltes 3D-Model eines Reliefs im HTML 5 3D Viewer

ZUSAMMENFASSUNG UND AUSBLICK

Das Erstellen individueller Visualisierungen im Internet, die es den Besuchern erlauben, in Interaktion mit den wissenschaftlichen Inhalten zu treten, ist heute mit aktuellen Webbrowsern technologisch einfacher denn je. Neben den technischen Komponenten ist es wichtig, dass Informatiker und Wissenschaftler hierbei ein gegenseitiges Grundverständnis ihrer Methoden entwickeln und eine einfache Benutzung der Anwendung im Vordergrund steht. Liegen Inhalte einmal als Webanwendung vor, so können diese in einer weiteren Stufe auch für mobile Endgeräte wie Tablets und Smartphones als App bereitgestellt werden. Moderne Web-Frameworks wie Meteor [10] bieten hierzu bereits erste Ansätze für eine App-Konvertierung.

Sollten Sie Beratung bei der technischen oder konzeptionellen Umsetzung von Visualisierungen im Internet benötigen, so sprechen Sie die zuständigen Mitarbeiter der GWDG an. Gerne teilen

wir mit unseren Kunden und Wissenschaftlern unsere Projekterfahrung und loten mit Ihnen gemeinsam aus, wie Sie Ihre wissenschaftlichen Ergebnisse im Internet anspruchsvoll präsentieren können.

REFERENZEN

- [1] <http://pencil.evolus.vn/>
- [2] <http://www.imagemagick.org/>
- [3] <https://www.blender.org/>
- [4] <http://geobrowser.de.dariah.eu/>
- [5] <https://cran.r-project.org/web/packages/rworldmap/>
- [6] <http://internet-content-regulation.uni-goettingen.de/>
- [7] <http://blog.fefe.de/?ts=abc12aa6>
- [8] <https://code.google.com/p/jsc3d/>
- [9] <https://github.com/humu2009/jsc3d>
- [10] <https://www.meteor.com/> ■

Kurz & knapp

Informationsveranstaltung „Security & Privacy: Lösungen für Backup und Collaboration von Code42 und VNC am 14.10.2015

Die GWDG lädt in Zusammenarbeit mit Code42 und VNC alle Interessierten ganz herzlich zu einer Informationsveranstaltung zum Thema „Security & Privacy: Lösungen für Backup und Collaboration von Code42 und VNC“ am 14.10.2015 ein. Die Veranstaltung findet von 11:00 – 17:00 Uhr im Großen Seminarraum des Max-Planck-Instituts für biophysikalische Chemie, Am Faßberg 11, 37077 Göttingen statt und ist speziell für Institute aus dem Bereich Forschung & Lehre gedacht. Der Fokus liegt dabei auf CrashPlan PROe, einer leistungsfähigen Backuplösung für Endgeräte, die von der GWDG seit Frühjahr 2015 als Dienst für ihre Kunden angeboten wird (siehe auch die GWDG-Nachrichten 3/2015). Weitere Informationen zum Programm und zur Anmeldung finden Sie unter <http://www.eventbrite.co.uk/e/security-privacy-losungen-fur-backup-und-collaboration-von-code42-und-vnc-bei-gwdg-in-gottingen-tickets-18392859538>.

Heider

Vier neue Dienstkataloge

Soeben sind die vier bekannten Dienstkataloge der GWDG in aktualisierter Fassung 09/2015 erschienen und stehen allen Interessierten sowohl in gedruckter als auch in elektronischer Form zur Verfügung:

- Dienste für die Max-Planck-Institute – Grundversorgung
- Services for the Max Planck Institutes – Basic Services
- Dienste für die Universität Göttingen – Grundversorgung
- Services for the University of Göttingen – Basic Services

Auf Wunsch senden wir Ihnen gerne ein oder mehrere Exemplare dieser Broschüren zu. Schicken Sie dazu bitte eine E-Mail mit den entsprechenden Angaben an gwdg@gwdg.de. Auf den Webseiten der GWDG sind die Broschüren unter <http://www.gwdg.de/index.php?id=mpg-dienstkatalog> bzw. <http://www.gwdg.de/index.php?id=uni-dienstkatalog> zu finden.

Otto



EU Project "PaaSAGE" – Model based Cloud Platform Upperware

Text und Kontakt:
Edwin Yaqub
edwin.yaqub@gwdg.de
0551-3920427

PaaSAGE is a large-scale integrated research project funded by the FP7 framework programme of the European Union. The project consortium constitutes of 19 partners from leading industrial and academic institutions, which aim at developing model-based solution for quality and policy aware deployment of complex applications on multi-Cloud infrastructures.

SUMMARY

PaaSAGE is researching and developing an open and integrated platform to support design and deployment of next generation Cloud applications. This is accompanied by a methodology that allows model-based development, configuration, optimisation, and deployment of existing and new applications independent of the underlying Cloud infrastructures. Furthermore, application lifecycle control is also provided because the quality, non-functional or cost constraints change over time. The architecture builds upon components that coherently work together to provide an advanced platform for managing next generation Cloud applications while leveraging state-of-the-art in technology and extending available standards.

THE CHALLENGE

The PaaSAGE perspective is to provide a tool for an application developer to master the complex and often error-prone task of deploying an application to the Cloud. The application developer could learn the interface and features of one Cloud provider, but it will be very costly to learn the development for many providers. It is a challenge to orchestrate the simultaneous deployment to many different Clouds at the same time as would be the case if some parts of the application can only run on private Cloud resources for confidentiality reasons while one would like to use public Cloud providers for application scalability.

THE PAASAGE SOLUTION

The main objective of PaaSAGE is to assist the developer with difficult deployment scenarios through autonomic cloud

deployment. Thus, PaaSAGE contemplates an application model that annotates platform extensions, user's goals and preferences; all of which are structured into a Cloud Application Modelling and Execution Language (CAMEL) model. The PaaSAGE platform transforms a CAMEL model into a deployed application on one or more Clouds. To support the developer, PaaSAGE needs not only a model of the application to be deployed, but models of the features of the available Cloud platforms, and goals and preferences to be

EU-Projekt „PaaSAGE“

PaaSAGE ist ein großangelegtes Forschungsprojekt, das von dem FP7-Rahmenprogramm der Europäischen Union gefördert wird. Das Projektkonsortium besteht aus 19 Partnern der führenden Industrie- und Hochschulinstitutionen. PaaSAGE ermöglicht eine modellbasierte Lösung für eine qualitäts- und richtlinienorientierte Verteilung von komplexen Applikationen auf Multi-Cloud-Infrastrukturen. Das Projekt ist im Oktober 2012 gestartet und auf eine Dauer von 48 Monaten angelegt. PaaSAGE betrachtet ein Applikationsmodell, das Plattformerweiterungen, Benutzerziele und Prioritäten adressiert; all dies ist in dem Cloud Application Modelling and Execution Language (CAMEL)-Modell strukturiert. Die PaaSAGE-Plattform überführt ein CAMEL-Modell in eine verteilte Applikation auf einer oder mehreren Clouds. Dadurch unterstützt PaaSAGE den Entwickler durch die Bereitstellung einer automatischen Verteilungslösung für komplexe Anwendungen und ermöglicht ständige Optimierung während des Applikations-Lebenszyklus. Die Projektergebnisse wurden erfolgreich bei der jährlichen Berichterstattung an die Europäische Kommission vorgestellt.

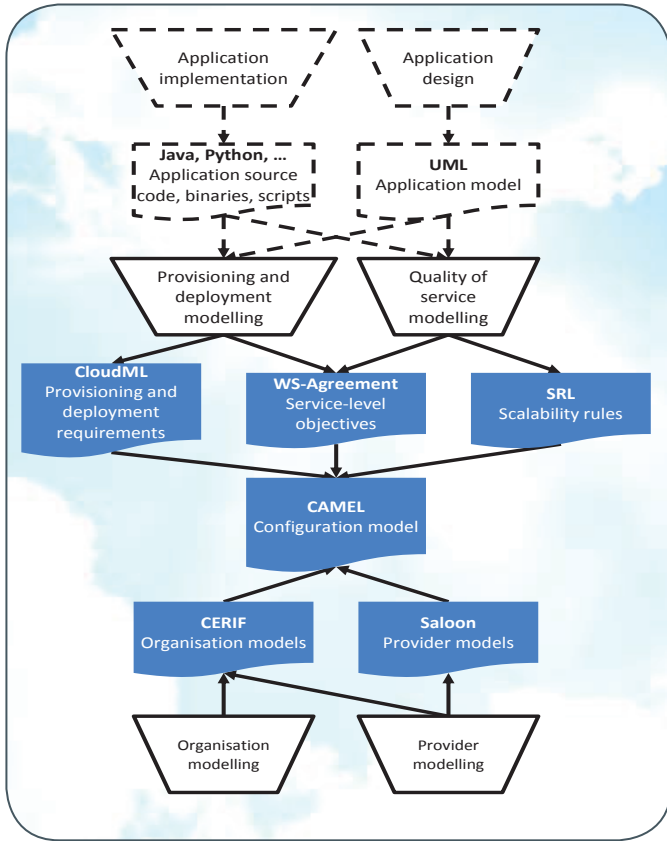


Figure 1_CAMEL Modelling Overview

satisfied by the deployment like response times or deployment cost budgets. These different models are specified in domain specific languages (DSLs) such as CloudML [1], WS Agreement [2], Scalability Rules Language (SRL) [3], CERIF [4] and Saloon [5]. CAMEL integrates the various DSLs using the Eclipse Modelling Framework (EMF) on top of the Connected Data Objects (CDO) persistence solution to maintain model information between different application deployments. A top-down view of CAMEL modelling is illustrated in Figure 1, which is progressively developed considering an application design in UML, implementation, provisioning, scalability, quality of service requirements along with provider and organizations to be targeted for deployment. The CDO repository is referred to as the metadata database, and the intention is that different PaaSage users will be able to form a social network and exchange models. Thus, a parameterised model developed by one user for a particular Cloud provider can be shared with other PaaSage users. This allows an immediate deployment solution for particular Cloud providers.

ARCHITECTURE OVERVIEW

The CAMEL model of an application is first transformed by what is referred to as the Upperware. It consists of a set of components that are used to derive a specific configuration satisfying all constraints and goals for the deployment as set by the user in the CAMEL model. This implies selecting one or more Cloud providers and generating the necessary deployment scripts. These scripts are then passed to the PaaSage Executionware – the component responsible for i) instantiating different parts of the application on selected Cloud providers, and ii) monitoring a set of defined metrics in order to make autonomous scalability decisions within

the boundaries of the deployment configuration. If the application execution triggers conditions that cannot be satisfied by the current deployment configuration, the Executionware passes control back to the Upperware to find a better configuration for the current application context. The monitored metrics are subsequently consolidated as statistical knowledge in the metadata database to guide the decisions on future deployment configurations.

Figure 2 shows the overall PaaSage architecture and workflow. The yellow parts are concerned with modelling, the blue parts are the Upperware, the orange parts represent the persistent metadata structure, while the green parts are the Executionware.

The feedback loop controlling the application deployment is also depicted in Figure 2. This illustrates the PaaSage workflow and the different logical parts of the Upperware and the Executionware. The Upperware consists of a Profiler, which combines information from the different CAMEL DSLs and produces a consistent model-based profile of the deployment scenarios. This model is passed on to the Reasoner, which finds a deployment configuration that satisfies all constraints and optimises the goals set for the deployment by the developer. The Adapter produces the platform-specific deployment scripts. It also maintains a connection between the running application and the model, and it reacts to context changes by issuing commands to keep the deployment within the current deployment configuration found by the Reasoner. The Executionware has one part taking care of the actual mapping of the deployment configuration onto the Cloud platforms chosen by the Reasoner. This includes setting up the necessary resources and installing both the monitoring infrastructure and the application components. The execution is then monitored. Based on the observed values, the execution control will then make autonomous scalability decisions within the boundary conditions given by the deployment configuration, or pass control back to the Upperware to produce a new deployment configuration that better suits the current execution context.

PROJECT STATUS

The focus in the first year of PaaSage was on defining the architecture for an automated deployment system as well as synthesizing requirements from project use cases and state of the art on models and standards. This was followed in the second year by the definition of the interfaces, integration testing of components and their capability to manipulate the involved CAMEL DSLs. In 2014, the project was highly successful at the second annual

Key Facts

- Duration:** 01.10.2012 – 30.09.2016 (48 months)
- Programme:** European Union's 7th Framework Programme (FP7)
- Partners:** 19 Partners across Europe
- EC Contribution:** 8.4 million EUR



PaaSage is a project funded in part by the European Union

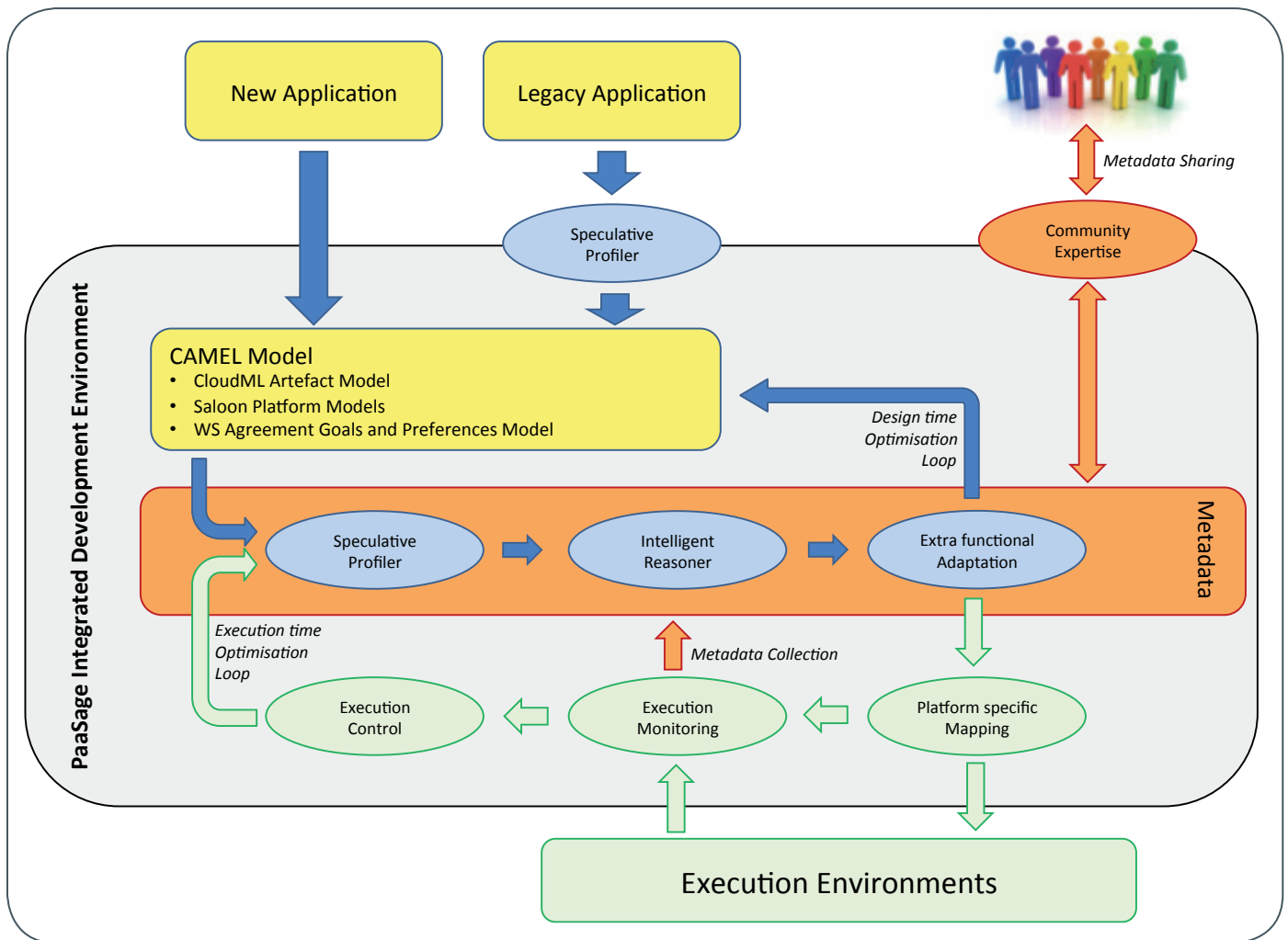


Figure 2_PaaSage Architecture and Workflow

review held at the European Commission, where PaaSage was exclusively demonstrated on GWDG’s OpenStack Cloud. The current focus is to harden the PaaSage platform, assist the use case partners to implement their applications the PaaSage way, and ensure a timely preparation for the third annual review at the European Commission.

THE TECHNOLOGY

The main parts of the prototype and the components they encapsulate are shown in Figure 3 and briefly explained below. The components are integrated around the CDO server.

Profiler

The Profiler part is supported by a **Constraint Program (CP) Generator** and a **Rule Processor** component. The former reads the CAMEL model and converts it into a constraint programming model by defining the variables of the model, their domains, and the constraints that must be satisfied by the deployment. The Rule Processor checks all constraints of the CAMEL model and sets the domains of the variables accordingly. It also removes redundant variables and constraints from the model, e.g., if the model defines that only virtual machines of a given size should be used, only providers offering such virtual machines can be selected and all other providers must be removed from the domain for the ‘provider’ variable.

Reasoner

The Reasoner contains four components: The **Meta Solver**, the **MILP Solver**, the **LA Solver**, and the **Solver to Deployment** component. The Meta Solver analyses the model and selects the solver most suited for the problem. If the problem is linear in its constraints and utility function, the MILP Solver will be used. Otherwise the LA Solver will be used. The two solvers are therefore interchangeable but with different characteristics. It is envisioned for PaaSage to also support other solvers in addition to these two. It could also happen that the Meta Solver will be able to decouple the problem into a linear and a non-linear part, and use the solvers in parallel. The ‘utility function’ can be any way of evaluating a candidate deployment, for instance a Cloud simulator or even a real world test deployment. The Meta Solver consequently informs the solvers how deployment candidates should be evaluated. Once a solution has been found, the Solver to Deployment component converts the solver output to **Cloud Provider Specific Models (CPSM)** for the providers involved in the proposed deployment.

Adapter

The Adapter part is currently supported by the **Adaptation Manager** component. It takes the CPSMs, produces and validates a configuration plan, and sends this plan to the Executionware.

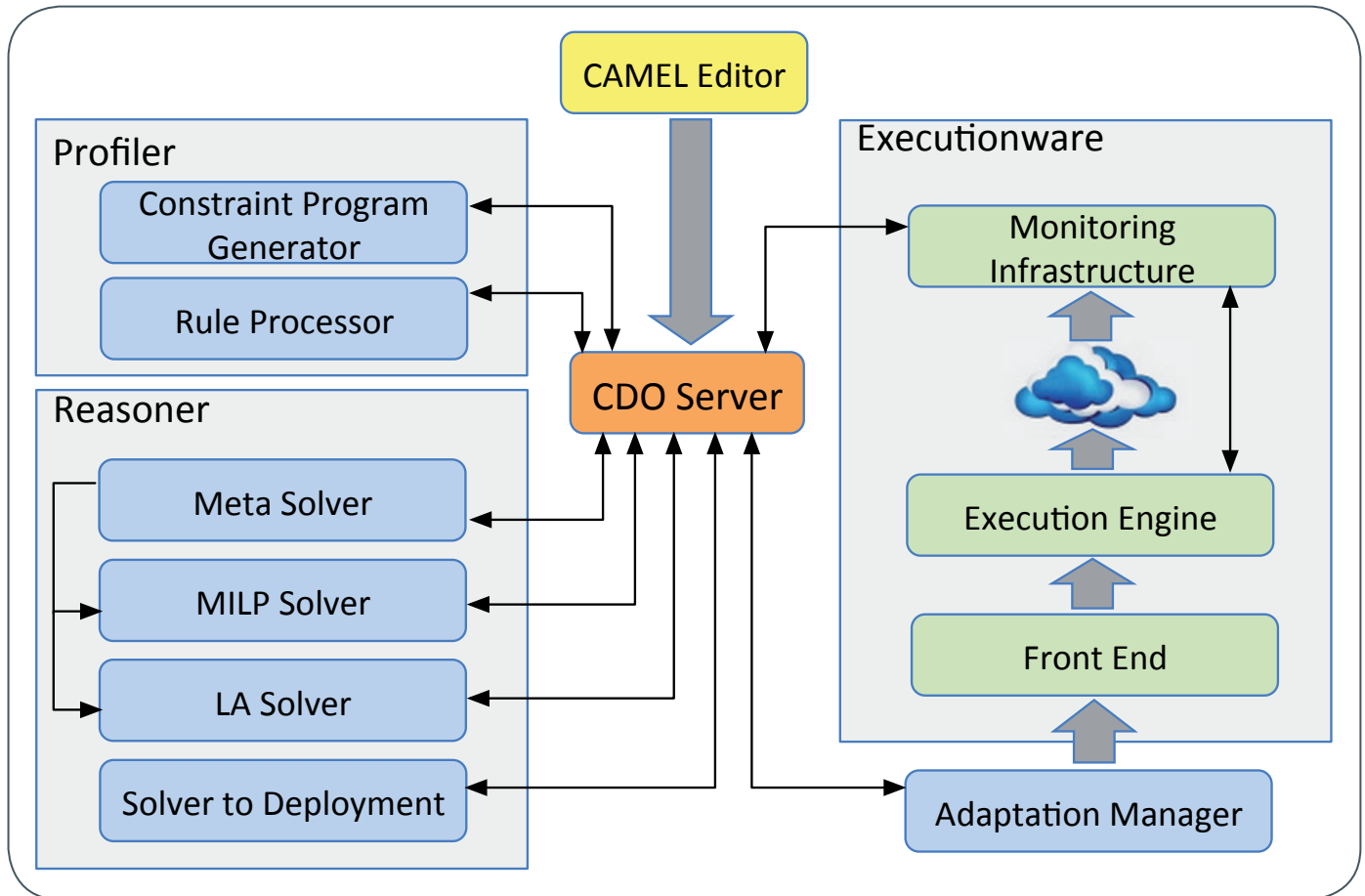


Figure 3_Platform Components

Executionware

The Executionware uses three components: the **Front End**, an **Execution Engine**, and the **Monitoring Infrastructure**. The Front End receives the deployment plan from the Adapter and enacts the deployment of the application on the selected providers. The control is then passed to the Engine that interacts with the Cloud providers, acquires the virtual machines, configures them and launches the user application on the set of virtual machines; instantiating Docker containers if the application model specifies use of containers. Once the machines are running, the Monitoring Infrastructure takes over and collects sensor data for the running application, triggering re-configurations if necessary.

EXPLOITATION, DISSEMINATION AND COMMUNITY BUILDING

As part of exploitation strategy and fruitful dissemination of results, PaaSage members decided to build a sustainable community around PaaSage. For this purpose, the project entered into an agreement with the OW2 consortium [6]. The OW2 is an independent open source community committed to making available to everyone the best and most reliable enterprise computing infrastructure software, including middleware, application platforms and Cloud Computing technologies. The OW2 consortium assists PaaSage by hosting its code base, promoting it to broad audiences and fostering the community and business ecosystem. OW2 will also provide dissemination and community building services to PaaSage for up to five years after the completion of the project.

GWDG AND PAASAGE

PaaSage showcases a variety of industrial and academic use cases spanning diverse domains and computational requirements. GWDG provides scalable Cloud services to PaaSage consortium and is actively involved in supporting the deployment of various use cases/components as listed below:

- Flight Scheduling System of Lufthansa Systems, Germany
- Scalarm Data Farming System of AGH University of Science and Technology, Poland
- HyperFlow WorkFlow Execution Engine of AGH University of Science and Technology, Poland
- Computational Fluid Dynamics System of the Automotive Simulation Center Stuttgart (ASCS), Germany
- Cloud test-bed for Executionware component by University of Ulm, Germany
- Cloud test-bed for Foundation for Research & Technology – Hellas (FORTH), Greece

Additionally, GWDG has contributed seven research publications conducted under the auspices of the PaaSage project.


ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement number 317715 (PaaSage).

DISCLAIMER

The article reflects the views of the author. The European Commission is not responsible for any use made of the provided information.

REFERENCES

- [1] <http://cloudml.org/>
- [2] <http://cloudindustryforum.org/downloads/standards/GFD.192.pdf>
- [3] <http://dl.acm.org/citation.cfm?id=2758780.2759588>
- [4] http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf
- [5] <https://hal.inria.fr/hal-01103560/document>
- [6] <http://www.ow2.org/bin/view/Main/> 



FTP-Server

Eine ergiebige Fundgrube!

Ihre Anforderung

Sie möchten auf das weltweite OpenSource-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 GWGD betreibt seit 1992 einen der weltweit bekanntesten FTP-Server, seit sieben Jahren mit leistungsfähigen Ressourcen für schnellen Service

Ihre Vorteile

- > Großer Datenbestand (40 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, ftp, rsync und nfs) 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

NEUER MITARBEITER MARCEL HELLKAMP

Seit dem 1. August 2015 unterstützt Herr Marcel Hellkamp die Arbeitsgruppe „eScience“ (AG E) im Rahmen des Projekts PERICLES. Schon während seines Studiums der Angewandten Informatik an der Georg-August-Universität Göttingen war er als studentische Hilfskraft erst für das Max-Planck-Institut für biophysikalische Chemie und dann bei der GWDG als Software- und Systementwickler tätig. Dort war er insbesondere an der Entwicklung von CDSTAR (siehe auch den GWDG-Bericht 78) beteiligt. Herr Hellkamp ist per E-Mail unter marcel.hellkamp@gwdg.de und telefonisch unter 0551 39-21381 zu erreichen.



Wieder



NEUER MITARBEITER DR. SONG YANG

Seit Anfang August 2015 ist Herr Dr. Song Yang als neues Mitglied der AG „eScience“ (AG E) im EU-Projekt CleanSky tätig. Er bringt einen Bachelor-Abschluss in Software-Engineering (2008) und einen Master-Abschluss in Informatik (2010) mit, die er an der Dalian University of Technology in der Provinz Liaoning in China erworben hat. Im Sommer 2015 hat er sein PhD-Studium an der Technischen Universität Delft in den Niederlanden erfolgreich abgeschlossen. Seine Forschungsinteressen liegen vor allem im Bereich Netzwerk-Optimierungsalgorithmen, insbesondere Routing, Überlebensfähigkeit und Datenmanagement in optischen Netzen, stochastischen Netzwerken und Netzwerken in Rechenzentren. Als Postdoc im Projekt CleanSky arbeitet er im Bereich Langzeitdatenspeicherung von Gensequenzierungsanwendungen. Ziel dieses Projektes ist die Entwicklung forschungsdomänenspezifischer Lösungen für die Speicherung von Gensequenzierungsdaten in Cloud-Infrastrukturen unter Berücksichtigung von Ausfallwahrscheinlichkeiten, Verzögerungen, Einschränkungen und anderer Quality of Service (QoS)-Faktoren. Herr Dr. Yang ist per E-Mail unter song.yang@gwdg.de und telefonisch unter 0551 39-20427 zu erreichen.

Wieder

AUSBILDUNG ERFOLGREICH ABGESCHLOSSEN JANNIS THOMAE

Mit der Präsentation seines Abschlussprojektes „Integration der GWDG-Dienstkataloge in das Kundenportal der GWDG“ bei der IHK hat Herr Jannis Thomae am 14. Juli 2015 seine dreijährige Ausbildung zum Fachinformatiker Anwendungsentwicklung erfolgreich abgeschlossen. Wie üblich wird er zunächst für sechs Monate bei der GWDG weiterbeschäftigt und während dieser Zeit hauptsächlich an der Weiterentwicklung des Kundenportals mitwirken. Wir gratulieren Herrn Thomae ganz herzlich zur bestandenen Prüfung und wünschen ihm alles Gute für seinen Einstieg ins Berufsleben.



Pohl

NEUE AUSZUBILDENDE CARSTEN BIEKER UND DENNIS RHODE

Am 1. September 2015 hat Herr Carsten Bieker seine 2½-jährige Ausbildung zum „Elektroniker für Geräte und Systeme“ in der Arbeitsgruppe „IT-Infrastruktur“ (AG I) bei der GWGD begonnen. Nach dem Abitur an der Stiftsschule St. Johann in Amöneburg hat er an der Philipps-Universität Marburg Chemie mit dem Bachelor-Abschluss studiert. Sein Interesse an den chemisch-technischen Laborapparaturen hat ihn zu der beruflichen Umorientierung bewogen. Herr Bieker ist per E-Mail unter carsten.bieker@gwdg.de und telefonisch unter 0551 201-1533 zu erreichen.



Ebenfalls am 1. September 2015 hat Herr Dennis Rhode seine 3½-jährige Ausbildung zum „Elektroniker für Geräte und Systeme“ in der AG I bei der GWGD begonnen. Nach dem Erreichen des Sekundarabschlusses I an der Personn-Realschule Weende hat er die Fachoberschule Informatik an der BBS II in Göttingen besucht. Sein Praktikum absolvierte er in einem Göttinger Computergeschäft. Herr Rhode ist per E-Mail unter dennis.rhode@gwdg.de und telefonisch unter 0551 201-1533 zu erreichen. Gutsch



ABSCHIED VON KAI UWE MATHER

Herr Kai Uwe Mather hat die GWGD Mitte August 2015 nach fast 9,5-jähriger Tätigkeit verlassen. Herr Mather war in dieser Zeit hauptsächlich im Bereich der Installation und Pflege der über 1.000 Layer-2-Komponenten des Netzwerkes der GWGD und der Universität Göttingen tätig. Darüber hinaus hat er die Installation der Wireless-Infrastruktur mit sieben Controllern und daran angebotenen etwa 950 Accesspoints wesentlich vorangetrieben. Wir danken Herrn Mather für seine langjährige, erfolgreiche Arbeit und sein großes Engagement und wünschen ihm für seinen weiteren Berufs- und Lebensweg alles Gute und weiterhin viel Erfolg.

Handke



INFORMATIONEN:
support@gwdg.de
0551 201-1523

Oktober bis
Dezember 2015

Kurse

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
INSTALLATION UND ADMINISTRATION VON WINDOWS 8	Buck	07.10.2015 9:00 – 12:30 und 13:30 – 15:30 Uhr	30.09.2015	4
USING THE GWDG SCIENTIFIC COMPUTE CLUSTER - AN INTRODUCTION	Dr. Boehme, Ehlers	12.10.2015 9:30 – 16:00 Uhr	05.10.2015	4
PARALLELRECHNERPROGRAMMIERUNG MIT MPI	Prof. Haan	13.10. – 14.10.2015 9:15 – 17:00 Uhr	06.10.2015	8
MAC OS X IM WISSENSCHAFTLICHEN ALLTAG	Bartels	28.10. – 29.10.2015 9:30 – 16:30 Uhr	21.10.2015	8
PHOTOSHOP FÜR FORTGESCHRITTENE	Töpfer	02.11. – 03.11.2015 9:30 – 16:00 Uhr	26.10.2015	8
OUTLOOK – E-MAIL UND GROUPWARE	Helmvoigt	05.11.2015 9:15 – 12:00 und 13:00 – 16:00 Uhr	29.10.2015	4
ADMINISTRATION VON PCS IM ACTIVE DIRECTORY DER GWDG	Buck	09.11.2015 9:00 – 12:30 und 13:30 – 15:30 Uhr	02.11.2015	4
EINFÜHRUNG IN DIE STATISTISCHE DATENANALYSE MIT SPSS	Cordes	11.11. – 12.11.2015 9:00 – 12:00 und 13:00 – 15:30 Uhr	04.11.2015	8
INDESIGN – AUFBAUKURS	Töpfer	16.11. – 17.11.2015 9:30 – 16:00 Uhr	09.11.2015	8
EINFÜHRUNG IN DAS IP-ADRESSMANAGEMENTSYSTEM DER GWDG FÜR NETZWERKBEAUFTRAGTE	Dr. Beck	18.11.2015 10:00 – 12:00 Uhr	11.11.2015	2

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
QUICKSTARTING R: EINE ANWENDUNGSORIENTIERTE EINFÜHRUNG IN DAS STATISTIKPAKET R	Cordes	25.11. – 26.11.2015 9:00 – 12:00 und 13:00 – 15:30 Uhr	18.11.2015	8
UNIX FÜR FORTGESCHRITTENE	Dr. Sippel	30.11. – 02.12.2015 9:15 – 12:00 und 13:15 – 15:30 Uhr	23.11.2015	12
HIGH-LEVEL, HIGH-PERFORMANCE TECHNICAL COMPUTING WITH JULIA	Chronz	03.12.2015 9:15 – 16:30 Uhr	26.11.2015	4
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER MIT VORKENNTNISSEN	Cordes	09.12. – 10.12.2015 9:00 – 12:00 und 13:00 – 15:30 Uhr	02.12.2015	8
DIE SHAREPOINT-UMGEBUNG DER GWDG	Buck	16.12.2015 9:00 – 12:30 und 13:30 – 15:30 Uhr	09.12.2015	4

Teilnehmerkreis

Das Kursangebot der GWDG richtet sich an alle Mitarbeiterinnen und Mitarbeiter aus den Instituten der Universität Göttingen und der Max-Planck-Gesellschaft sowie aus einigen anderen wissenschaftlichen Einrichtungen.

Anmeldung

Anmeldungen können schriftlich per Brief oder per Fax unter der Nummer 0551 201-2150 an die GWDG, Postfach 2841, 37018 Göttingen oder per E-Mail an die Adresse support@gwdg.de erfolgen. Für die schriftliche Anmeldung steht unter <http://www.gwdg.de/antragsformulare> ein Formular zur Verfügung. Telefonische Anmeldungen können leider nicht angenommen werden.

Kosten bzw. Gebühren

Unsere Kurse werden wie die meisten anderen Leistungen der GWDG in Arbeitseinheiten (AE) vom jeweiligen Institutskontingent abgerechnet. Für die Institute der Universität Göttingen und

der Max-Planck-Gesellschaft erfolgt keine Abrechnung in EUR.

Absage

Sie können bis zu acht Tagen vor Kursbeginn per E-Mail an support@gwdg.de oder telefonisch unter 0551 201-1523 absagen. Bei späteren Absagen werden allerdings die für die Kurse berechneten AE vom jeweiligen Institutskontingent abgebucht.

Kursorte

Alle Kurse finden im Kursraum oder Vortragsraum der GWDG statt. Die Wegbeschreibung zur GWDG sowie der Lageplan sind unter <http://www.gwdg.de/lageplan> zu finden.

Kurstermine

Die genauen Kurstermine und -zeiten sowie aktuelle kurzfristige Informationen zu den Kursen, insbesondere zu freien Plätzen, sind unter <http://www.gwdg.de/kurse> zu finden.



Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen