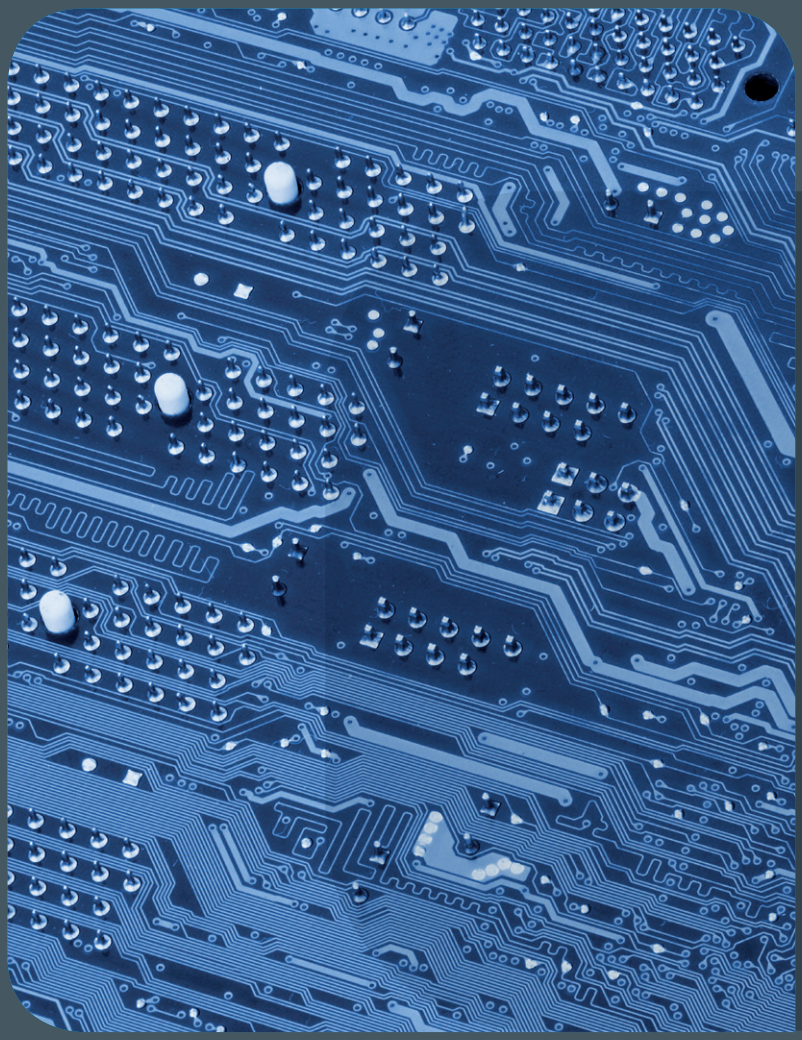
GWDG NACHRICHTEN

01-02|19

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDG

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen



GWDG
NACHRICHTEN

01-02|19

Inhalt

Impressum

Zeitschrift für die Kunden der GWDG

ISSN 0940-4686
42. Jahrgang
Ausgabe 1-2/2019

Erscheinungsweise:
monatlich

www.gwdg.de/gwdg-nr

Auflage:
550

Fotos:
© stadtratte - Fotolia.com (1)
© Brian Jackson - Fotolia.com (4)
© bluedesign - Fotolia.com (8)
© momius - Fotolia.com (9)
© SAP (28)
© vege - Fotolia.com (28)
© Contrastwerkstatt - Fotolia.com (31)
© MPIbpc-Medienservice (3, 32)
© GWDG (2, 25, 27, 33)

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:
Franziska Schimek
E-Mail: franziska.schimek@gwdg.de

Druck:
Kreationszeit GmbH, Rosdorf

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

Liebe Kunden und Freunde der GWDG,

erst Yahoo, dann Adobe, dann LinkedIn, dann Facebook. Es ist schon bemerkenswert, dass große Internet-Firmen trotz ausreichender Ressourcen immer wieder mit kleinen oder großen Daten-Hacks in der Presse landen. Zwar sind nicht immer Passwörter betroffen, aber meist sensible Account-Informationen. Zuletzt wurde eine Datenbank mit einer Sammlung von über 2 Mrd. Einträgen aus verschiedenen Leaks öffentlich.

Für Nutzer stellt sich die Frage, ob man selbst betroffen ist. In dieser Ausgabe der GWDG-Nachrichten finden Sie hierzu einen Artikel, in dem wir Hilfestellung geben wollen. Wir bieten zudem einen Dienst an, um Account-Informationen selbst prüfen zu können, ohne auf fremden Seiten sensible Informationen preisgeben zu müssen.

Es ist auch bemerkenswert, dass die Daten-Hacks für die großen Internet-Konzerne praktisch ohne Konsequenzen bleiben. Es fallen keine Geldstrafen an. Den oft beschworenen Imageverlust oder Reputationsschaden gibt es offenbar auch nicht. Zumindest ist nicht erkennbar, dass den Anbietern die Kunden weglaufen. Zu sehr haben wir uns an Meldungen zu Sicherheitslecks gewöhnt. Der nächste Datenklau kommt bestimmt.

Es ist leider ebenfalls bemerkenswert, dass einfache Grundregeln noch nicht von allen beachtet werden: gute Passwörter wählen, regelmäßig ändern und nie für verschiedene Anbieter das gleiche verwenden. So schwer dürfte das eigentlich nicht sein.

Wie gesagt – alles sehr bemerkenswert.

Ramin Yahyapour

GWDG – IT in der Wissenschaft

Passwörter zu 2,2 Mrd. E-Mail-Adressen im Internet – Fragen und Antworten

Text und Kontakt:
Dr. Holger Beck
holger.beck@gwdg.de
0551 201-1554

Nachdem Anfang Januar der Hacker-Angriff auf Politiker und andere Prominente und die damit verbundenen Veröffentlichungen teils privater und sensibler Daten die Schlagzeilen bestimmt haben, sorgen aktuell Nachrichten über riesige Sammlungen von E-Mail-Adressen und zugeordneten Passwörtern, die im Internet kursieren, für Unruhe. Die unter den Namen „Collection #1 - #5“ diskutierten Sammlungen sollen Daten zu ca. 2,2 Milliarden Online-Konten enthalten. Wir versuchen nachstehend, die wichtigsten mit diesem Datenleck verbundenen Fragen zu beantworten. Wir haben das als FAQ-Liste gestaltet und auf unseren Webseiten veröffentlicht. Diese FAQ-Liste wird bei Bedarf aktualisiert.

WAS SIND DAS FÜR DATEN ZU DEN 2,2 MRD. E-MAIL-ADRESSEN?

Im Internet kursieren immer wieder Listen mit E-Mail-Adressen und Daten, die diesen zugeordnet werden. Diese Daten können Passwörter (oder etwas, das aussieht, als ob es Passwörter sein könnten) oder auch Telefonnummern, Kreditkarten-Daten oder vieles anderes sein. Die aktuellen Listen oder Kollektionen scheinen eine Zusammenfassung solcher kleineren Datensätze zu sein.

Die Qualität dieser Daten ist unklar. Es ist durchaus möglich, dass diese Daten sehr alt sind. Die GWDG hat in der Vergangenheit immer wieder kleinere Datensätze mit im Internet kursierenden Daten zu Konten, die zu Systemen der GWDG passten, erhalten. In solchen Fällen gab es meist die Hälfte der Konten schon garnicht mehr. In Fällen, in denen Informationen zum Passwort vorlagen, hatten angesprochene Kontoinhaber fast immer mitgeteilt, dass sie ein solches Passwort für das GWDG-Konto nicht verwendet hätten. In einer einstelligen Zahl von Fällen war in den Datenlecks tatsächlich ein Passwort aufgetaucht, das (auch) für das GWDG-Konto verwendet wurde.

SIND AUCH KONTEN DER GWDG, MAX-PLANCK-GESELLSCHAFT ODER UNIVERSITÄT GÖTTINGEN BETROFFEN?

Definitiv ja. Stichproben haben das bestätigt.

FAQ about the Worldwide Dataleak “Collection #1 - #5“

While hacker attacks, targeting politians and other prominent persons, filled the headlines in the first days of January, now enormous collections of e-mail addresses and associated passwords, which circulate in the Internet, cause discussions. The so-called “collection #1 - #5“ is supposed to contain data about around 2.2 Billion online accounts. We try to explain hereafter, to answer the most important questions about these data leaks. We designed this as a FAQ list and published it on our website This FAQ will be updated as needed.

KANN ICH FESTSTELLEN, OB ICH SELBST BETROFFEN BIN?

Ja, auf mehreren Webseiten werden Prüfmöglichkeiten angeboten.

Ein Anbieter ist der australische Sicherheitsforscher Troy Hunt, der Informationen aus Datenlecks sammelt und diese auf seiner Webseite „Have I Been Pwned" (*https://haveibeenpwned.com*) zur Verfügung stellt. Dort kann in einem Webformular die eigene E-Mail-Adresse eingegeben werden, um die Sammlung zu durchsuchen. Die Webseite informiert dann, ob die E-Mail-Adresse gefunden wurde, und gibt zusätzlich Informationen, in welchen Datenquellen diese enthalten war.

Auch das Hasso-Plattner-Institut (HPI) pflegt eine solche Sammlung. Dort kann man auf einer Webseite (*https://sec.hpi.uni-potsdam.de/ilc/*) seine E-Mail-Adresse eingeben und bekommt daraufhin per E-Mail an die abgefragte Adresse eine entsprechende Auskunft (Beispiel unter *https://sec.hpi.uni-potsdam.de/ilc/publickeys*).

Die Datenbestände beider Anbieter unterscheiden sich im Detail bzgl. älterer Datenlecks, beinhalten aber beide die neuen Datenlecks.

WENN ICH BETROFFEN BIN, KANN ICH DANN FESTSTELLEN, WELCHES PASSWORT VON MIR BETROFFEN IST?

Nein (außer man verschafft sich Zugang zu den Rohdaten des Datenlecks). „Have I Been Pwned" wie auch das HPI liefern nicht zurück, welche Passwörter (oder sonstigen Informationen) zu welchem Konto in den Listen auftauchen. Das ist auch gut so, denn sonst könnte dieser Dienst böswillig missbraucht werden.

IST ES SICHER, IN „HAVE I BEEN PWNED" ODER BEIM HPI NACH MEINER E-MAIL-ADRESSE ZU SUCHEN?

Ihre E-Mail-Adresse auf einer unbekannten Webseite anzugeben, kann prinzipiell Risiken beinhalten. Vielleicht wird die Adresse dort gespeichert und in der Folge an Spam-Versender verkauft, die Sie später mit unerwünschten E-Mails überhäufen.

Troy Hunt ist allerdings ein Sicherheitsforscher mit einer guten Reputation, sodass hier nicht angenommen werden muss, dass „Have I Been Pwned" missbräuchlich genutzt wird. Das HPI ist eine renommierte Einrichtung, die akademisch an die Universität Potsdam angebunden ist. Auch hier ist ein Missbrauch nicht zu erwarten.

Ob man diese Dienste nutzen will, mag dann auch noch davon abhängen, ob man seine E-Mail-Adresse eher als eine weitverbreitete oder als eine geheime Information betrachtet.

MEINE E-MAIL-ADRESSE TAUCHT IN DER DATENSAMMLUNG AUF. WAS MACHE ICH NUN?

Das lässt sich leider nicht einfach beantworten, weil eben unklar ist, welche Informationen in der Datensammlung zu Ihrer E-Mail-Adresse stehen. Sicher ist nur, dass die E-Mail-Adresse in der Liste auftaucht. Alles andere kann falsch, harmlos, veraltet oder leider auch sensibel sein.

Wir versuchen, mit den nächsten Fragen und Antworten ein paar Hilfestellungen zu geben.

WIE GEFÄHRDET BIN ICH, WENN EINE MEINER E-MAIL-ADRESSEN IN DEN DATENSAMMLUNGEN GEFUNDEN WIRD?

Das lässt sich leider nicht einfach beantworten, da ein Treffer auf die Anfrage nach der E-Mail-Adresse zunächst nur sagt, dass die E-Mail-Adresse in den Datensammlungen auftaucht.

In der weiteren Erläuterung steht zusätzlich, in welchen Quellen die E-Mail-Adresse auftaucht. Die Erläuterungen zu den Quellen können hilfreich für die Einschätzung sein. Bei einigen der Quellen waren gar keine Passwort-Informationen enthalten, sondern z. B. nur Namen, Telefonnummern und Anschriften. Auch solche Informationen können sensibel sein (vielleicht sind es aber auch nur Informationen, die sowieso öffentlich verfügbar sind).

Zu anderen Quellen wird angegeben, dass in den Datensätzen Passwörter enthalten wären. Da aber nicht klar ist, welche Passwörter das jeweils sind, kann nicht festgestellt werden, ob das Passwort, das im Datenleck auftaucht, überhaupt ein wirklich irgendwann verwendetes Passwort ist oder ob das Passwort vielleicht vor Jahren ersetzt wurde.

WELCHES KONTO IST GEFÄHRDET, WENN EINE MEINER E-MAIL-ADRESSEN IN DEN DATENSAMMLUNGEN AUFTAUCHT?

Wenn eine Ihrer E-Mail-Adressen in den Datensammlungen auftaucht und in der Erläuterung steht, das in dem Datenleck auch Passwörter enthalten waren, dann muss das Passwort nicht zu dem Konto gehören, dass mit der E-Mail-Adresse direkt verbunden ist (z. B. zu Ihrem GWDG-Konto).

Häufig werden bei irgendwelchen Dienstleistern im Internet lokale Konten angelegt, bei denen E-Mail-Adressen des Kunden als Kontoname verwendet werden. Die meisten Internet-Nutzer dürften mehrere solcher Konten haben. Es könnte also sein, dass die Information über Sie nicht aus einem Datenleck des E-Mail-Providers stammt, sondern von einem Dienstleister, bei dem Sie ihre E-Mail-Adresse als Kontoname „wiederverwenden". Wenn Sie sich dann an die Sicherheitsempfehlung gehalten haben, für jeden Dienst ein anderes Passwort zu verwenden, wäre nur dieser eine Dienst betroffen und nicht Ihr E-Mail-Konto.

Der am wenigsten sensible Fall wäre also, dass nur ein Konto betroffen ist, das Sie bei einem Dienst bekommen haben, der mehr oder weniger belanglos ist. Es könnte aber auch ein Passwort für einen Dienst sein, bei dem sensible Daten gespeichert sind. Um das zu entscheiden, müsste man wissen, welche der eigenen Passwörter zu der enthaltenen E-Mail-Adresse aufgetaucht sind. Diese Zuordnung wird von den öffentlich verfügbaren oben beschriebenen beiden Diensten von Troy Hunt und dem HPI nicht angeboten.

KANN ICH PRÜFEN, OB MEIN PASSWORT IN DEN DATENSAMMLUNGEN ENTHALTEN IST?

Es kann geprüft werden, ob ein Passwort in den Datensammlungen enthalten ist. Diesen Dienst bietet „Have I Been Pwned"

auf einer Webseite und über eine API (Application Programing Interface) an. Die GWDG hat als Reaktion auf diesen Vorfall einen eigenen ähnlichen Dienst mit Webseite und API aufgebaut und dazu die von „Have I Been Pwned" bereitgestellte Möglichkeit zum Kopieren eines Datensatzes von SHA-1-Hashes von geleakten Passwörtern genutzt und mit Daten aus weiteren Datenquellen angereichert.

Der Dienst der GWDG ist unter *https://www.gwdg.de/pwleak/* zu erreichen. Sie können dort nach einem beliebigen Passwort suchen und erhalten als Antwort, ob dieses Passwort in einer der Datensammlungen enthalten ist. Die Antwort besagt aber nicht, in Verbindung mit welcher E-Mail-Adresse das Passwort in den Datensammlungen enthalten ist. Auch hier gilt wieder: Nur wenn etwas gar nicht gefunden wird, kann man beruhigt sein. Wenn das Passwort gefunden wird, könnte es auch zu einer ganz anderen E-Mail-Adresse in der Datenbank enthalten sein. Bei einem guten Passwort wäre das aber eher unwahrscheinlich. Man sollte daher davon ausgehen, dass man dann betroffen ist (zum weiteren Vorgehen s. u.).

IST ES NICHT RISKANT, AUF EINER UNBEKANNTEN WEBSEITE EIN GEHEIMNIS WIE EIN PASSWORT EINZUGEBEN?

Ja. Es gilt die Sicherheitsempfehlung: Geben Sie Ihr Passwort nie auf einer unbekannten Webseite ein (oder auf fremden Rechnern oder in anderen unsicheren Situationen).

Aus diesem Grund bietet die GWDG einen eigenen Dienst an. Das aktuelle Passwort eines GWDG-Kontos (oder auch früher dafür verwendete Passwörter) können Sie auf einer GWDG-Webseite ohne Bedenken eingeben. Es sind ja Passwörter, die zu diesem Dienst gehören oder gehörten. Sie sollten nur sicherstellen, dass Sie wirklich mit dem Dienst der GWDG verbunden sind! Geben Sie die Adresse selbst in der Adresszeile des Browsers ein oder prüfen Sie die Adresszeile, wenn Sie über einen Link auf einer anderen Webseite dorthin gelangt sind. Dort muss *https://www.gwdg.de/pwleak/* und vorweg auch das Schlosssymbol stehen, das die Authentizität der *https*-Webseite bestätigt.

Der Webdienst der GWDG zur Abfrage der Passwörter überträgt zudem nicht direkt Ihr Passwort zur GWDG, sondern bietet eine sogenannte k-Anonymität.

WAS BEDEUTET DIE K-ANONYMITÄT BEIM DIENST DER GWDG?

Für die technisch Interessierten soll hier kurz erläutert werden, welche Anonymisierung die GWDG bei Ihrem Passwort-Leak-Check vornimmt:

Ihr auf unserer Webseite eingegebenes Passwort verlässt nicht Ihren Browser. Zur Prüfung wird Ihr Passwort in Ihrem Browser in einen sogenannten kryptographischen Hash (einen SHA-1-Hash) überführt. Nur die ersten fünf Zeichen dieses Hashes werden an unseren Dienst übertragen. Die Antwort unseres Dienstes enthält dann eine Liste von Hashes aus den Datensammlungen, welche mit diesen fünf Zeichen beginnen. Die Prüfung, ob Ihr Hash in dieser Liste aufgeführt ist, erfolgt dann wiederum nur in Ihrem Browser. Damit ist auch die Antwort, ob Ihr Passwort bekannt ist oder nicht, nur Ihnen bekannt.

DARF ICH ÜBER DEN DIENST DER GWDG AUCH ANDERE PASSWÖRTER ALS DIE VON GWDG-KONTEN PRÜFEN?

Die GWDG bietet den Dienst zur Prüfung beliebiger Passwörter an. Streng genommen gilt natürlich auch hier die Sicherheitsempfehlung, Passwörter immer nur auf den Webseiten einzugeben, zu deren Dienst diese Passwörter gehören. Andererseits bieten die meisten Dienstanbieter einen solchen Passwort-Check-Dienst nicht an. Hier müssen Sie entscheiden, in welchen Dienstleister Sie genügend Vertrauen setzen, um diese eigentlich notwendige Prüfung vorzunehmen. Die GWDG als Dienstleister für die Universität Göttingen und die MPG ist hierfür selbstverständlich als vertrauenswürdig anzusehen.

KANN MAN SEIN PASSWORT PRÜFEN LASSEN, OHNE ES AUF EINER WEBSEITE EINZUGEBEN?

Ja, für technik-affine gibt es die oben schon erwähnte Alternative eines API-Zugriffs (bei der GWDG oder bei „Have I Been Pwned"). Diese kann verwendet werden, um nach einem Passwort zu suchen, ohne das Passwort selbst zu übermitteln. Dazu muss der SHA-1-Hash des Passworts berechnet und die ersten fünf (von 40) Zeichen des Hashes per API als Anfrage übergeben werden. Die Antwort enthält alle Hashes, die mit den fünf Zeichen beginnen. In dieser Liste kann lokal gesucht werden, ob der vollständige Hash (oder genauer die Zeichen 6-40 des Hashes) in der Antwortliste auftaucht (engl. Anleitung für den GWDG-Dienst: *https://www.gwdg.de/pwleak/api.html*). Das ist dasselbe Vorgehen, das auch beim GWDG-Webdienst über den Browser implementiert ist (nur für den Skeptiker selbst nachvollziehbar).

Damit wurde das Passwort selbst keiner externen Stelle mitgeteilt, sondern nur ein Hash und von dem auch nur ein Achtel. Die Restrisiken, die mit dem Vorgehen verbunden sind, scheinen vertretbar.

WAS SOLLTE ICH TUN, WENN ICH BETROFFEN BIN?

Wenn Anfragen nach der E-Mail-Adresse und nach einem zugehörigen Passwort einen Treffer ergeben haben, sollte das betreffende Passwort sofort geändert werden. Sie sollten auch überlegen, welche Konsequenzen der Vorfall gehabt haben könnte. Auch wenn die Antworten einer Prüfung noch nicht definitiv aussagen, dass die Kombination E-Mail-Adresse/Passwort zusammen in einem Datenleck aufgetaucht ist, muss angenommen werden, dass dies der Fall ist und ein Angreifer Zugriff auf alle Daten gehabt haben könnte, die über diese Kombination gesichert waren.

Bei Beurteilung der Konsequenzen muss berücksichtigt werden, auf welche Daten ein Angreifer Zugriff gehabt haben könnte und wie sensibel diese sind oder waren. Im Falle von personenbezogenen Daten kann es sich um einen meldepflichtigen Datenschutzvorfall handeln. Eine Information an die zuständigen Stellen muss hier umgehend erfolgen, da darüber in der Folge innerhalb von 72 Stunden eine Meldung an die zuständige Aufsichtsbehörde erfolgen muss. Forschungsdaten können ebenfalls Vertraulichkeits- oder Geheimhaltungsanforderungen unterliegen. Informieren Sie ggf. die Datenschutz- und IT- bzw. Informationssicherheits-Verantwortlichen Ihrer Einrichtung!

WAS KANN ICH VORBEUGEND TUN, UM DEN SCHADEN VON PASSWORT-LEAKS ZU MINIMIEREN?

Für den Umgang mit Passwörtern gilt die Empfehlung, für jeden Dienst ein eigenes Passwort zu verwenden. Ein Datenleck bei einem Dienstleister gefährdet dann nicht auch noch Daten bei anderen Dienstleistern. Die IT-Sicherheitsrichtlinien von GWDG, MPG und Universität fordern auch explizit, dass die dienstlichen Passwörter nicht auch noch an anderer Stelle verwendet werden.

WAS KANN ICH TUN, WEIL ICH MIR NICHT ALLE PASSWÖRTER MERKEN KANN, WENN ICH FÜR JEDEN DIENSTLEISTER EIN ANDERES PASSWORT VERWENDEN SOLL?

Das eine dienstliche Passwort, welches täglich mehrmals benötigt wird, wird man sich wahrscheinlich merken können. Für Passwörter bei Dienstleistern, die selten verwendet werden, ist dies sicherlich illusorisch. Solche Passwörter müssen notiert oder gespeichert werden. Die beste Lösung dafür sind Passwort-Manager. Das sind Programme, die in einer verschlüsselten Datei Passwörter sicher aufbewahren. Die Verschlüsselung der Datei wird dabei mit einem Master-Passwort geschützt. Man muss sich also nur noch das Master-Passwort merken. Das Master-Passwort sollte daher natürlich auch ein starkes Passwort sein (länger als einfache Passwörter, komplex und nicht von fremden Personen zu erraten).

Passwort-Manager können auch starke Passwörter erzeugen. Da der Passwort-Manager sowieso dazu dient, bei Bedarf ein Passwort aus der verschlüsselten Datei zu kopieren, um sich damit bei einem bestimmten Dienst anzumelden, kann auf diese Weise ohne großen Aufwand die Verwendung von starken Passwörtern umgesetzt werden.

WELCHE WEITEREN MASSNAHMEN ERGREIFT DIE GWDG ZUM SCHUTZ DER PASSWÖRTER DER VON IHR GEFÜHRTEN KONTEN?

Die GWDG hat die Sammlung von bekannten Passwörtern (genauer deren Hashes), die unter *https://www.gwdg.de/pwleak/* geprüft wird, genutzt, um über ihr Kunden-Portal eine höhere Passwort-Sicherheit zu erreichen, indem bei Passwort-Änderungen jetzt nicht nur die formalen Komplexitätskriterien geprüft werden, sondern sofort auch das neue Passwort (genauer sein Hash) gegen die Passwörter in der Sammlung geprüft wird. Falls dabei eine Übereinstimmung gefunden wird, erfolgt eine Warnung mit entsprechender Begründung, und das neue Passwort wird nicht akzeptiert.

Die GWDG prüft aktuell, inwieweit automatisiert SHA-1-Hashwerte aus dem eigenen Identity-Management-System (IdM) gegen die Sammlung verglichen werden können. Auf diesem Weg ließe sich ermitteln, welche Konten Passwörter verwenden, die in den Passwort-Leaks enthalten waren. Damit ist nicht belegt, dass tatsächlich ein Nutzer kompromittiert wurde. Jedoch ist das gewählte Passwort offensichtlich unsicher und sollte dringend geändert werden. Eine präventive Mitteilung an diese Personen mit einer Aufforderung zur Passwort-Änderung würde hier helfen. Die GWDG klärt zurzeit, wie dies unter Aspekten von Datenschutz und Mitbestimmungs- bzw. Personalrecht umsetzbar ist. Wir werden selbstverständlich über das weitere Vorgehen informieren.

Die GWDG erhält gelegentlich von Sicherheitseinrichtungen wie dem DFN-CERT oder dem CERT-Bund Listen von Konten, die der GWDG zugeordnet werden können, wenn diese in Datenlecks enthalten waren und diese Einrichtungen davon Kenntnis erhalten hatten. In solchen Fällen informiert die GWDG die betroffenen Personen. Die uns übermittelten Informationen enthalten teils nur die betroffenen E-Mail-Adressen, teils Teile (meist die beiden ersten Zeichen) des Passwortes, selten auch das komplette Passwort. Im letzteren Fall prüft die GWDG die E-Mail-Adresse und das Passwort gegen das aktuelle Passwort im IdM-System. Auch hier werden die Betroffenen bei einer Übereinstimmung umgehend informiert. Sollten die Betroffenen nicht kurzfristig erreichbar sein, sperrt die GWDG in diesen Fällen wegen Gefahr im Verzug das Konto, da in diesem Fall tatsächlich erwiesen ist, dass zu genau diesem Konto das Passwort im Internet öffentlich (zumindest in Hackerkreisen) verfügbar ist.

Die wichtigsten Empfehlungen auf einen Blick

- Falls Sie Ihr Passwort oder Ihre Passwörter unter *https://www.gwdg.de/pwleak/* oder alternativ *https://haveibeenpwned.com/Passwords* geprüft haben und eines Ihrer Passwörter dort gefunden wurde, ändern Sie es bitte unverzüglich.
- Falls Ihre E-Mail-Adresse auf *https://haveibeenpwned.com* oder *https://sec.hpi.uni-potsdam.de/ilc/* gefunden wird und Sie auf eine Prüfung des Passworts verzichten, empfehlen wir sicherheitshalber, Ihr entsprechendes Passwort zu ändern.
- Das Passwort sollte neu sein, also nicht schon einmal in dienstlichen oder privaten Kontexten verwendet worden sein.
- Passwort-Manager wie KeePass (*https://keepass.info*) bieten die Möglichkeit, sichere und individuelle Passwörter für beliebige IT-Dienste und Webseiten zu generieren und zu verwalten.

Das Ende ist nahe – Ablauf der Generation 1 der DFN-PKI Global

Text und Kontakt:
Thorsten Hindermann
thorsten.hindermann@gwdg.de
0551 201-1837

Am 10. Juli 2019 läuft, wie schon länger angekündigt, die Generation 1 der DFN-PKI Global ab. Spätestens an diesem Tag sind dann auch alle Benutzer-, Dienst- und Server-Zertifikate unweigerlich abgelaufen. Eine Verlängerung wird es nicht geben. Dieser Artikel gibt Hinweise, was es zu tun und zu beachten gibt.

EINLEITUNG

Die Generation 1 der DFN-PKI Global (PKI = Public Key Infrastructure) läuft am 10. Juli 2019 ab. Diese basiert auf dem Wurzelzertifikat „Deutsche Telekom Root CA 2“. Schon vor geraumer Zeit hat der DFN die Generation 2 der DFN-PKI eingeführt. Diese basiert auf dem Wurzelzertifikat „T-TeleSec Global Root Class 2“. Weiterhin hat der DFN auch schon in seinen Hinweis-E-Mails, die frühzeitig auf den Ablauf von Zertifikaten hinweisen, die URLs für das Beantragen des neuen Zertifikats auf die Generation 2 der DFN-PKI umgestellt.

WAS IST ZU TUN?

In den folgenden Abschnitten wird kurz erklärt, was jetzt am besten zu tun ist. Denn, wie schon beschrieben, das Ablaufdatum für die Generation 1 steht mit dem 10. Juli 2019 unweigerlich fest.

Für Benutzer

Warten Sie ab, bis Sie in den nächsten Tagen, Wochen oder Monaten die Hinweis-E-Mails bekommen. Diese informieren Sie über den Ablauf Ihres Zertifikats. Beantragen Sie über den URL in der E-Mail einfach Ihr neues Zertifikat. Dieser Antrag wird automatisch richtig in der Generation 2 der DFN-PKI gestellt.

Für Server-Administratoren

Sie als Administrator von Diensten oder Servern sollten am besten wissen, ob Ihr Dienst oder Server noch ein Zertifikat der Generation 1 aktiv benutzt. Falls Sie sich unsicher sind, wenden Sie sich bitte an Ihre RA-Operatoren vor Ort und fragen bei diesen nach. Sie können Ihnen Auskunft über die aktiven Zertifikate für Ihre(n) Dienst(e) oder Server geben. Falls noch Zertifikate der Generation 1 aktiv sind, sollten Sie diese sobald wie möglich austauschen, um nicht in Bedrängnis zu geraten, wenn plötzlich und

Expiration of Generation 1 of the DFN-PKI Global

As announced some time ago, Generation 1 of the DFN-PKI Global will expire on 10 July 2019. On this day at the latest, all user, service and server certificates will inevitably expire. There will be no extension. This article gives hints on what to do and what to consider.

unerwartet, trotz der Hinweis-E-Mails zum Ablauf der Zertifikate, der 10. Juli 2019 vor der Tür steht. Da sehr viele Dienst- und Server-Zertifikate aktiv sind, wird es sicherlich um diesen Termin herum viel Ansturm auf die Ausstellung neuer Zertifikate der Generation 2 geben. Wenn dann die RA-Operatoren überlastet sind und ein paar Tage lang nicht mit dem Ausstellen nachkommen, ist das unschön für Ihren Dienst bzw. Server, weil dann in den Browsern beim Zugriff Warnmeldungen erscheinen. Um diese Situation zu vermeiden, reagieren Sie rechtzeitig in den nächsten Tagen, Wochen oder Monaten.

Für RA-Operatoren

Als RA-Operatoren sind Sie ja schon seit der Einführung der Generation 2 der DFN-PKI über die entsprechenden E-Mail-Verteilerlisten informiert. Für Sie gilt der Rat, dass Sie Ihre Server-Administratoren aktiv darauf hinweisen, dass noch Dienst- und/oder Server-Zertifikate aus der Generation 1 aktiv sind. Informieren Sie Ihre betroffenen Kollegen gezielt, welche Zertifikate noch in der Generation 1 aktiv sind. Somit vermeiden Sie möglichst einen großen Ansturm und für sich selber viel Stress an oder um den 10. Juli 2019 herum, wenn dann die Generation 1 der DFN-PKI endgültig ausläuft.

OPENJDK-KOMPATIBLE GUIRA-VERSION FÜR RA-OPERATOREN

Im Zuge der Lizenzierung von Oracle Java hat der DFN reagiert und das RA-Operator-Werkzeug **GUIRA** OpenJDK-kompatibel gemacht. Die auf Oracle Java basierende GUIRA-Version wird nicht mehr aktiv weiterentwickelt. Hierfür stellt der DFN nur noch Sicherheitskorrekturen zur Verfügung, und die Unterstützung für diese Version wird nur noch bis Mitte dieses Jahres gewährleistet.

Die OpenJDK-Version ist somit die aktuelle Version und RA-Operatoren sollten bis Mitte dieses Jahres auf alle Fälle auf diese Version umgestiegen sein. Die neue Version steht unter dem URL *https://blog.pki.dfn.de/tag/guira-releases* zum Herunterladen bereit, nebst Anleitung und Start-Dateien für die drei gängigen Betriebssysteme macOS, UNIX und Windows.

ANSPRECHPARTNER BEI FRAGEN

Wenn Sie Fragen haben, wenden Sie sich bitte zuerst an die RA-Operatoren vor Ort in Ihrem Institut. Haben Sie keine RA-Operatoren vor Ort, können Sie sich auf alle Fälle gerne auch an die Service-Hotline der GWDG per E-Mail an *support@gwdg.de* oder über *https://www.gwdg.de/support* wenden. Wir können Ihnen Auskunft geben, was zu tun ist und welche Möglichkeiten Sie haben.

Verbesserungen am Backbone des Universitätsnetzes

Text und Kontakt:
Steffen Klemer
steffen.klemer@gwdg.de
0551 201-2170

Die GWDG nimmt am zentrale Rückgrat des GÖNET, dem Netzwerk der Universität Göttingen, einige wichtige Veränderungen vor, um dessen Leistungsfähigkeit zu verbessern. Während die allgemeine Struktur in den letzten fünf Jahren gut funktioniert hat und mit dem Wachstum der Anzahl der Netzwerkgeräte wie WiFi, IP-Telefone, Gebäudeleittechnik, Multimedia-Systeme und der zunehmenden Nutzerzahl gut zurechtkam, machte das aktuelle Wachstum des Netzwerkverkehrs einige Änderungen erforderlich. Wir haben die Firewall-Last nun auf beide Geräte unseres ehemaligen aktiv-passiv Firewall-Paares verteilt und die Verbindung zum DFN und damit zum Internet um 50 % auf jetzt zweimal 15 Gbit/s erweitert. In den kommenden Monaten werden wir weitere Institute auf ein neues „privates IP"-NAT-Konzept umstellen, um Administration und weiteres Wachstum zu erleichtern, und neue dezentrale Firewall-Geräte einführen.

DIE ENTWICKLUNG IM ÜBERBLICK

Das Rückgrat des Göttinger Universitätsnetzes, der GÖNET-Backbone, existiert in seiner heutigen Form seit jetzt fünf Jahren (siehe Abb. 1). Die während des letzten großen Umbaus (siehe die GWDG-Nachrichten 3/2014] eingeführten Komponenten und Strukturen haben sich nach einigen Fehlerbehebungen durch den Router- und Firewall-Hersteller und ein paar Fein-Justierungen bewährt. Die einzige fundamentale Änderung gegenüber der Einführung im Jahr 2014 betrifft den Aufbau der zentralen Firewall. Aufgrund von zuvor nicht klar durch den Hersteller kommunizierten Einschränkungen und zahlreichen Softwarefehlern im sogenannten Cluster-Betrieb wurde bereits 2015 auf einen klassischen aktiv-passiv-Redundanz-Betrieb umgestellt.

Auf dieser Basis konnten in den letzten Jahren tausende zusätzliche Endgeräte der verschiedensten Gattungen versorgt werden. Das WLAN wurde um ca. 1.500 Access Points (APs) erweitert, die Telefonie großflächig auf IP-Telefone umgestellt, die Gebäudeleittechnik (Türsteuerungen, Heizungssysteme etc.), Multimediatechnik und Videokonferenzsysteme massiv aus- und umgebaut und in den Backbone integriert. Und mehr Studierende, Mitarbeiter und Wissenschaftler brachten viele neue und immer komplexere Notebooks, Smartphones, Tablets, Uhren, PCs, eBook-Reader usw. mit.

LEISTUNGSFÄHIGERE INTERNET-ANBINDUNG

Während der gesamte, summierte Netzwerkverkehr im Backbone aufgrund von Zentralisierung und Virtualisierung im Rechenzentrum seit Anfang 2015 nur um etwa 60 % angestiegen ist, hat er sich hingegen in Richtung Internet verdoppelt (siehe Abb. 2). Um dem Rechnung zu tragen, veränderten wir bereits vor zwei Jahren die Routingpfade derart, dass dauerhaft beide 10-Gbit/s-Leitungen zum DFN (Internet) aktiv genutzt werden, und vergrößerten im Herbst des letzten Jahres die Anbindung auf zweimal 15 Gbit/s. Technischer Hintergrund: Der DFN übergibt uns je zwei sogenannte LR-Leitungen zu den DFN-Knoten Hannover und Frankfurt, auf denen er ein Traffic-Shaping betreibt, sobald mehr als 15 Gbit/s pro Knoten übertragen wird.

Recent and Upcoming Changes for the Network Backbone

GWDG is changing the central backbone of the GÖNET, the network of the University of Göttingen. While its general structure served well for the last five years and coped well with all the growth in number of network devices like WiFi, IP-telephones, central building control, multimedia systems and many more users, recent traffic growth necessitated some changes. We distributed firewall-load onto both devices of our former active-passive firewall pair and broadened the link to the DFN by 50% to now two times 15 Gbit/s. In the upcoming months we will switch more institutes to a new "private IP"-NAT concept to facilitate administration and growth and roll out new decentralized firewall devices.

FIREWALL LOAD-BALANCING

Problematischer stellte sich das zentrale Firewall-Paar dar, das mit Beginn des Wintersemesters 2018/19 aufgrund der oben erwähnten Umstellung von 2015 leider früher als geplant an seine Grenzen stieß. Da auch fünf Jahre später der dynamisch lastverteilte Cluster-Betrieb über zwei Geräte noch Einschränkungen mit sich bringt, entschieden wir uns für eine „manuelle" und statische Verteilung der virtuellen Firewalls bzw. Interfaces auf beide Hardware-Geräte. Dies wurde in drei Schritten in den letzten Wochen im Rahmen unserer periodischen Wartungsfenster vollzogen. Dabei stellten wir fest, dass das bisher einzeln aktive Gerät deutlich mehr überlastet war als vermutet. Mit den ersten zwei Umzugs-Schwüngen von Interfaces stieg der Datendurchsatz auf dem zweiten Gerät an, nahm auf dem ersten jedoch nicht ab.

Im 2er-Verbund und im Zusammenspiel mit ein paar Software-Optimierungen sind beide Geräte nun selbst in den Tageszeiten mit maximalem Durchsatz jeweils deutlich unter den Werten des „Einzelgängers" zuvor. Zudem gab es von zahlreichen Nutzern positive Rückmeldungen, dass ihre Dienste wieder die gewohnte Geschwindigkeit erreichen. Vor allem stark Latenz-behaftete (WAN-)Verbindungen waren zuvor doch deutlich eingeschränkt.

NEUES KONZEPT FÜR MEHR IP-ADRESSEN

Immer mehr elektronische Geräte brauchen immer mehr IP-Adressen und der Durchbruch von IPv6 steht zwar unmittelbar bevor, lässt aber leider (zugegebener Maßen auch im GÖNET) weiter auf sich warten. Die Folge ist ein Mangel an IPv4-Adressen in den Instituten. Um dem zu entkommen, haben wir ein allgemeines Konzept für mehr IPv4-Adressen aus dem privaten IANA-Bereich 10.0.0.0/8 im Zusammenspiel mit NAT auf eine öffentliche IP für den Internet-Zugang geplant und erproben es bereits mit einigen Instituten. Dies wollen wir in den kommenden Monaten noch einmal ausführlich vorstellen und weiter ausrollen. Und das Beste daran: IPv6 fügt sich nahtlos in das Konzept ein.

WEITERE PLANUNGEN

An der Firewall-Front sollen noch im Laufe des Jahres zusätzliche „Next-Generation"-Geräte angeschafft werden, die als dezentrale Instituts-Firewalls das zentrale Gerätepaar entlasten und das Routing und damit die tägliche Arbeit der Netz-Administratoren vereinfachen sollen. In diesem Zusammenhang werden wir auf die Institute zukommen, um den individuellen Satz an Regeln („ACLs") zu besprechen und ggf. zu optimieren.

Um einem „Verkehrskollaps" zuvorzukommen, werden wir zusätzlich die Bandbreite zwischen den GÖNET-Routern in diesem Jahr verdoppeln. Alles in allem sind wir zuversichtlich, dass damit der Backbone fit für die Jahre 2019 und 2020 ist. Die geplanten Neuerungen im Netzwerkbereich sind auch Thema beim nächsten Treffen der Netzwerkbeauftragten am 14.03.2019 (siehe auch die Ankündigung auf der nächsten Seite).

Fun Facts

Verbindung zum X-WiN

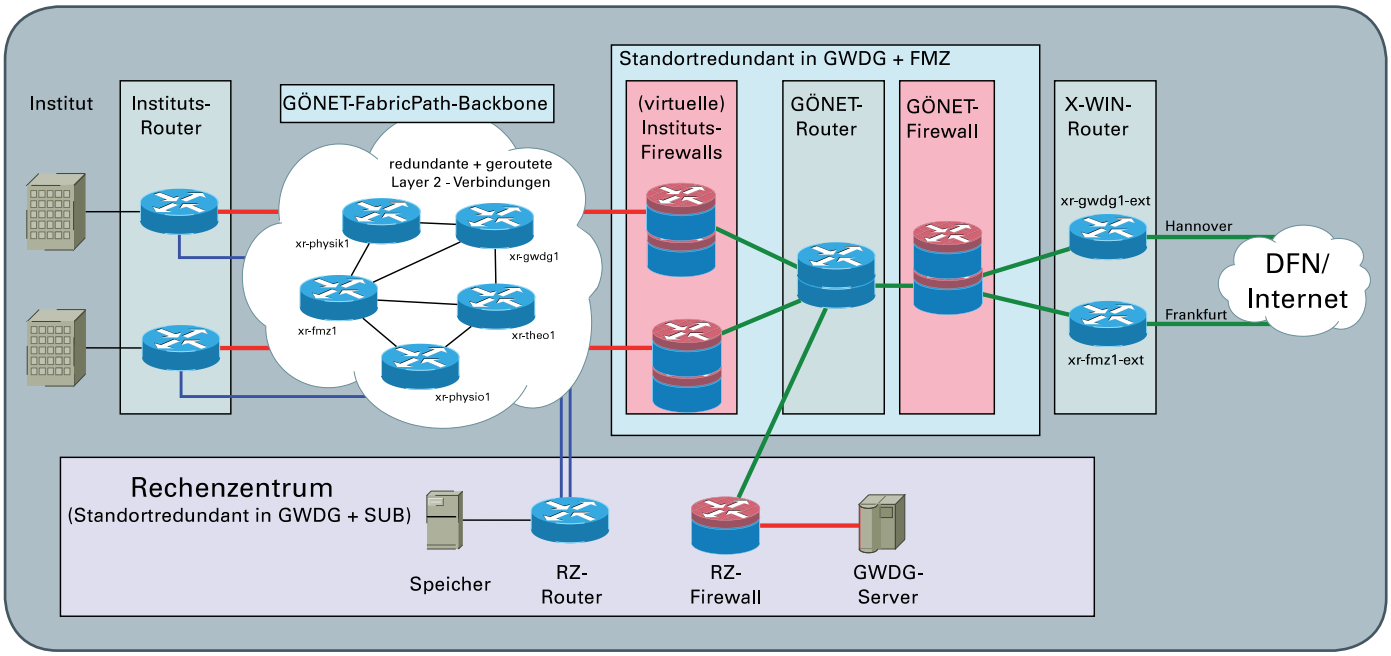
- Anbindung: 2 x 15 Gbit/s symmetrisch
- Jahresmittel: 4,1 Gbit/s down, 2,9 Gbit/s up
- 95 % des Traffics unterhalb von 6,2 Gbit/s down und 3,4 Gbit/s up
- Maximum im Januar 2019: 13,5 Gbit/s down, 11,5 Gbit/s up

Traffic zum Rechenzentrum (ohne FTP-Server)

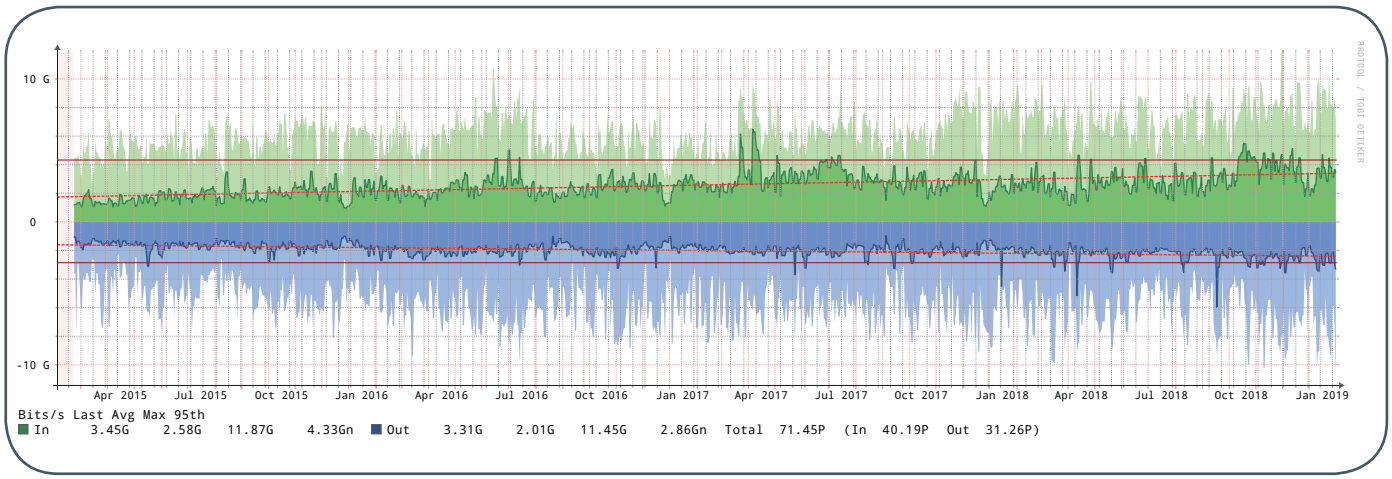
- Anbindung: 40 Gbit/s symmetrisch
- Jahresmittel: 4,8 Gbit/s down, 2,8 Gbit/s up
- 95 % des Traffics unterhalb von 8,9 Gbit/s down und 5,5 Gbit/s up

Traffic im WLAN

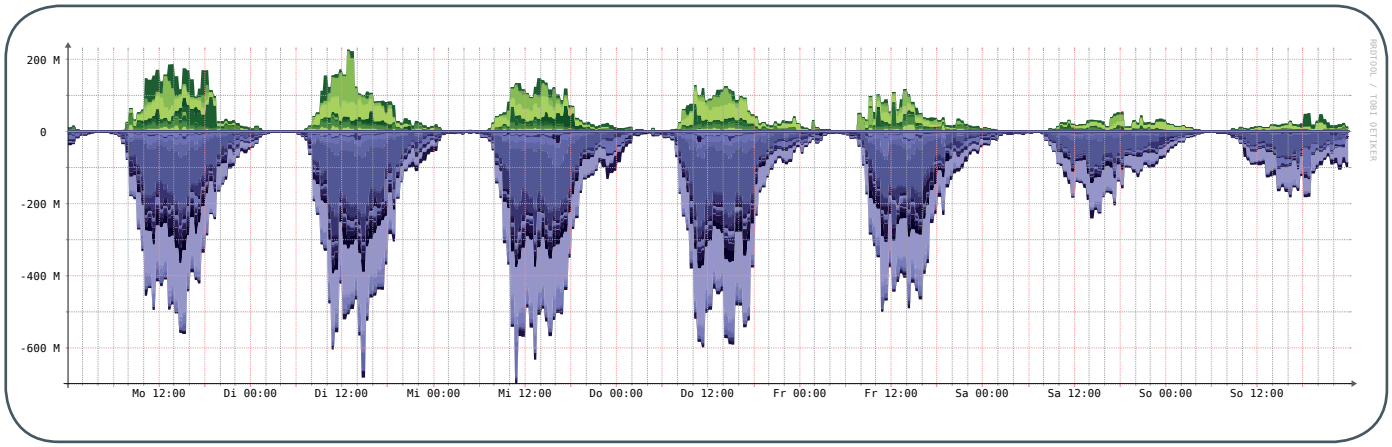
- Tages-Maxima: 1 Gbit/s down, 300 Mbit/s up (siehe auch Abb. 3)



1_Routing-Pfade des GÖNET-Backbone (siehe auch die GWDG-Nachrichten 3/2014)



2_Entwicklung des Netzwerkverkehrs vom und zum DFN über die vergangenen fünf Jahre auf der primär verwendeten Strecke zum Kernnetzknoten Hannover



3_Netzwerkverkehr vom und zum WLAN im Wochenverlauf [in Bit/s]

Treffen der Netzwerkbeauftragten am 14. März 2019

Das letzte Treffen der Netzwerkbeauftragten fand 2017 statt. In der Zwischenzeit hat sich vieles, auch im Netzwerkbereich, getan und es gibt daher in diesem Zusammenhang auch einiges zu berichten. Überdies möchten wir auch unsere gemeinsame Planung für den Netzwerkbereich vorstellen. In den nächsten zwei Jahren werden einige bedeutende Neuerungen im Netzwerk auf uns zukommen. Vor diesem Hintergrund planen wir, das Treffen der Netzwerkbeauftragten zukünftig halbjährlich stattfinden zu lassen, damit zum einen der Erfahrungs- und Informationsaustausch intensiviert wird und zum anderen, zusätzlich zu den bereits bestehenden bekannten Kommunikationskanälen, auch der für alle Beteiligten wichtige Dialog öfter stattfinden kann.

Das nächste Treffen der Netzwerkbeauftragten wird am Donnerstag, den 14.03.2019, von 9:00 – 10:45 Uhr, im Hörsaal ZHG 006 des Zentralen Hörsaalgebäudes, Platz der Göttinger Sieben 5, stattfinden (Lageplan: *https://lageplan.uni-goettingen.de/?ident=5257_1_EG_0.154*).

Vorläufige Agenda:

- Neuerungen bei eduroam
- Mehr Selfservice (IMC)
- DNS-Updates und -Änderungen
- Neues NAT-Konzept
- WLAN-Ausbau und zukünftige Planungen
- Diskussion

Eine explizite Einladung erfolgt in Kürze per E-Mail an alle Netzwerkbeauftragten. Wir freuen uns auf eine hohe Beteiligung und einen regen Austausch.

Ißleiber

Backup of Large File Systems Using ISP/TSM

Text and Contact:
Bjørn Nachtwey
bjoern.nachtwey@gwdg.de
0551 201-2181

As the topic "Looking for suggestions to deal with large backups not completing in 24 hours" was discussed on the „ADSM-L" mailing list recently [1] we decided to update and translate a text published in the GWDG News in November 2016. The growth of data mentioned two years ago is still ongoing and therefore the file administrators and backup operators have to face the challenge of coping with the backup of data within the specified time window and thus complying with the promised protection against manipulation and data loss. In this article, some approaches to speed up the backup using ISP/TSM are discussed. They will be explained briefly, but the scope is on the chances and limitations of each. The second part of the article develops an approach starting with the basic idea of parallelizing the backup towards different variants of a script including some reporting, error handling and statistics.

CURRENT SITUATION

The amount of data is growing, different analyses show a value of 20% each year [2, 3]. In addition to the challenges of storing this data sensibly and efficiently, one aspect often falls out of focus: How can this growing data be backed up?

The nominal performance of the tape systems is growing faster than the growth of the data itself [2, 3], but even this view is unfortunately incomplete, since the process of backing up, the actual backup, often represents the bottleneck.

"IBM Spectrum Protect (ISP)" (formerly known as "WDSF/VM", "DFDSM", "ADSM" or "Tivoli Storage Manager (TSM)") has long pursued the approach of backing up only the changed files since the last run only – instead of doing a full dump sometimes. As there are no planned full dumps, this approach is called "incremental forever".

The Advantage is obvious: Especially for large file systems (say > 10 TB) the amount of daily changed data is relatively small. So that even with many versions of older data (the GWDG standard backup policy allows up to 350 versions in 90 days) the additionally needed space is small [4] compared to the space needed for the secured data. But, if a full backup is done periodically, the backup capacity must be several times greater than the secured data itself. For mixed approaches, e.g. the „Grandfather-Father-Son"-principle by means of

- monthly full backups (Grandfather),
- weekly differential hedges (Father) and
- daily backups (Son)

the necessary backup capacity is larger than with "incremental forever" due to the full dumps.

However „incremental forever" does not solve an essential problem of any incremental backup: the answer to the question of which data must be backed up at all. ISP identifies the data to be backed up ("backup candidates") by comparing all directories and files on the computer (to be backed up) with those from the last backup and remembering changed files. This process usually runs at a speed of 1 – 2 million objects (files and folders) per hour.

Searching through a 100 TB file system with around 100

Backup großer Filesysteme mit ISP/TSM

Nachdem das Thema „Das Backup läuft in 24 Stunden nicht durch" (also ein Synonym für das Backup großer Datenbereiche) zuletzt auf der TSM-Mailing-Liste „ADSM-L" diskutiert wurden, haben wir uns entschlossen, den Text aus den GWDG-Nachrichten 11/2016 zu aktualisieren und vor allem auch englischsprachig für die internationalen Interessenten zu veröffentlichen. Das bereits vor zwei Jahren angesprochene permanente Wachstum von Daten stellt die Filesystem-Administratoren weiterhin vor die Herausforderung, das Backup der Daten innerhalb des vorgegebenen Zeitfensters zu bewältigen und so die zugesagte Absicherung gegen Manipulation und Datenverlust einzuhalten. Der nachfolgende Artikel beleuchtet anhand der Backup-Lösung „IBM Spectrum Protect (ISP)" (ehemals „Tivoli Storage Manager (TSM)") verschiedene Möglichkeiten und Ansätze. Es werden diese kurz erläutert und die Grenzen und Beschränkungen aufgezeigt. Der zweite Teil des Artikels entwickelt auf Basis der Grundidee „Parallelisierung mittels mehrerer Backup-Threads" verschiedene Varianten, das Backup zu beschleunigen.

million objects takes between 50 and 100 hours, **a daily backup of such a file system is therefore not possible with the common approaches**.

In addition, there is the problem that within this long search time, a considerable amount of data will be changed or even deleted. As a result, ISP throws many error messages (*ANS4037E Object <NAME>' changed during processing. Object skipped* or *ANS4005E Error processing ‚<NAME>': file not found*). How to solve this problem?

NON-WORKING SOLUTIONS

One possible solution, "bird ostrich method": just adapt the service description of the file servers by not guaranteeing daily backups but (initially) only every two days. As the amount of data grows, the backup frequency needs to be continuously adjusted. When reaching about 150 million objects the interval will only be a monthly backup.

At this point, the second "zero solution" should be considered: To give up the backup of corresponding file systems completely and not to lull the users into a (data) security that does not (no longer) exist.

Since searching for "backup candidates" is the problem with backups, one could consider biting the bullet and doing full backups, as tapes are relatively inexpensive compared to DISK storage. Unfortunately, in our experience, this is definitely **no solution**:

For a full backup of 100 TB, theoretically, only about 24 hours are required with a 10GE connection. However, actual operational experience shows that only about 2 – 4 TB per day are effectively saved and about 25 – 50 days are required for each full backup. In other words, approximately the same time as for an "incremental" backup.

ACCELERATION WITH ISP ON-BOARD TOOLS

IBM offers several on-board tools to speed up the backup process:

Simplified identification exclusively via the change date

Usually the ISP client compares numerous meta data to select objects for the new backup. Besides the date of the last modification, these are also file size, checksum, access rights/ACLs. In the interactive call *dsmc i* and/or as *Object* in the client schedule, the check can be reduced to the comparison of the change date of the object with the date of the last backup by the option *-INCRbydate* and thus considerably accelerated.

However, the option also has some problems: Especially if no snapshots are used or if the backup fails, files that are modified or created while the backup is running will be shipped by the next run with *-INCRbydate* if they have not been modified again. IBM therefore strongly recommends running a normal "incremental" regularly [5]. Similar problems can occur if client and server have different system times.

Another important point: Deleted files are not recognized, they remain in the backup and files that come into the system with an old date, e.g. due to the installation of software, are not backed up!

In summary, the *-INCRbydate* option can only be used for the daily backups together with a "normal" backup at the weekend if the normal backup lasts slightly longer than 24 hours.

Turning off ACLs and checksums

Processing ACLs (and thus previously checking) and creating checksums slows down the identification process and can be influenced by several options. **However, it should be carefully considered whether the relatively low speed gain sufficiently outweighs the loss of information.**

- *skipacl* completely disables ACL processing, but ACLs will probably not be saved either (option for UNIX and macOS only).
- *skipaclupdatecheck* also disables checksum calculation (option for UNIX, macOS and Windows).
- *skipntsecuritycrc* prevents the calculation of a CRC checksum (option for Windows only).

Parallelization of backups for multiple file spaces

If the data to be backed up is on several partitions, the backup process can be distributed to parallel streams using the *RESSOURCEUTILIZATION* option (in contrast to IBM documentation, significantly more than 10 are possible, > 100 streams are reported in practice). This makes better use of the bandwidth and considerably reduces the search time through parallelization. Since this also generates additional sessions on the ISP server side, the number of *MAXSESSIONS* may have to be increased.

This approach works only when backing up multiple file spaces. As a workaround, of course, a single file space can be split into seemingly multiple file spaces with the *VIRTUALMOUNTPOINT* option and then this approach works, of course. (See also the Excursus "Workaround *VIRTUALMOUNTPOINT* for Windows Clients".)

Explicit backup of changed files only

If information is available, which files have changed since the last backup and which files have been deleted since then, ISP can only back up these files. Instead of an "incremental backup", a "selective backup" with the explicit specification of these files is then possible:

- *dsmc sel –filelist=<File with Names of files changed>*

or

- *dsmc expire –filelist=<File with Names of files deleted>*

The basic principle of "selective backup" is also used in the following approach and in the "file systems that support fast backup", but requires two explicit lists of files that have been modified or deleted.

JournalBasedBackup / FilepathDemon

IBM has been offering the "JournalBasedBackup (JBB)" method since TSM 5. The "JBB Demon" (or "Filepath Demon") monitors the file system to be backed up and collects information on new, modified and deleted files. During backup, the TSM/ISP client uses this information in the same way as doing a "selective backup". The effort for identifying the backup candidates is eliminated and the backup is reduced to the transfer of the new/changed data.

Tests done by the GWDG with a Linux fileserver with about 150 TB capacity distributed over 22 file spaces were not successful: The resource requirements for the JBB were extensive, but the time saving, especially due to regular re-indexing, was rather limited. In other constellations, the JBB may bring clear advantages.

There is also an important limitation: JournalBasedBackup

only works with local file systems. CIFS/NFS and cluster file systems do not work.

Note: Optimizations for data transmission can be found in the Performance Tuning-Guide [5].

HYBRID APPROACH WITH SNAPSHOTS

Numerous file systems and most filers offer the possibility to create snapshots. A hybrid approach can be implemented by combining snapshots and ISP backup: Backups are done as often as possible, e.g. weekly, in between snapshots.

In addition to the considerable expansion of the backup time window, there is usually the positive side effect that the end users can access the snapshots directly and the administrators are relieved of numerous restore requests. If the backup is also based on a snapshot, the problem of the opened files is solved (error message *ANE4987E Error processing ‚<NAME>': the object is in use by another process*).

A prerequisite for this approach is, of course, that the file systems support snapshots – and in sufficient quantities.

FILE SYSTEMS THAT SUPPORT FAST BACKUP

Some file systems / filers support a fast backup using ISP by identifying the necessary backup candidates and making them available to the ISP client. This list is only a selection:

IBM Spectrum Scale (formerly GPFS)

IBM's cluster file system naturally supports backup with ISP and even offers its own script *mmbackup*. This not only uses the information about the backup candidates, but can also parallelize the data transfer over several (ISP) nodes and GPFS servers.

However *mmbackup* does not simply run "out-of-the-box": The initial creation of the configuration requires a little trial and error, but afterwards *mmbackup* runs both stable and performant.

In addition to ISP, IBM Spectrum Scale also offers close integration with HPSS as an HSM system, so that the problem can also be reduced by (partially) transferring the data to HPSS whereby ISP/ISS can also back up very large data volumes in a comparatively short time.

NetApp SnapDiff

NetApp has also been supporting backup of its own NAS filers since TSM 5 in a variety of ways. In addition to NDMP, the *SnapDiff* function also accelerates the incremental backup. *SnapDiff* transfers the changes to files and directories between two snapshots to the ISP client. The integration goes so far that the ISP client can even trigger the required snapshots on the filer and after a successful backup can delete the previous one on its own.

Since the *SnapDiff* function compares only two snapshots, but does not take into account in any way whether the last backup was successful, the same problems arise as when using the *INCRbydate* option: errors from the last backup are not compensated and a regular "normal" incremental backup is strongly recommended. *mmbackup*, in contrast, takes into account the backup status of all data and is fault-tolerant with regard to the problems mentioned above.

Basically, each cluster/scale-out file system should be able to provide a list of new, modified and deleted files, since this (meta) information is necessary for the consistency of the data (and especially the caches) on the cluster nodes. In practice, the problems are that this information is not easily accessible and there are no tools by manufacturers to access this data. Quantum has responded to customer demands and is currently examining how this information can be made available for the StorNext file system. DELL/EMC also offers ScaleOut NAS systems with the ISILON systems. In version 7 of the operating system, called OneFS, there is the possibility to log changed files, but the resource requirements are so high that there is a lasting impairment of the entire system. With OneFS 8 there should also be improvements.

TWO (SIMPLE) IDEAS FOR ALL FILE SYSTEMS

For all users who do not have an IBM Spectrum Scale in operation (*mmbackup* is the best solution for this!) and neither full backups nor NDMP this raises the question of what to do now?

As previously mentioned, identifying backup candidates takes most of the time during ISP backup. This process examines the entire file tree of the file system to be backed up – sequentially in a single thread. The solution is to turn this one process into several parallel processes.

Users can usually be divided into groups (e.g. working groups or institutes). Especially in academic environments, this classification can also be found in file systems, since there is often a folder level with faculties or institutes for easier access control, and below this level are the user and workgroup directories.

Variant 1

Parallel backup is possible by setting up a separate node for each faculty or institute instead of a single ISP node for the entire file system and performing the backup "faculty by faculty" / "in-stitute by institute". Instead of a single process, several processes search the file system in parallel (file servers are able to process even several hundred parallel processes) and the search times should be significantly reduced. In practice, this approach reveals at least two problems:

- Not all faculties/institutes have the same amount of data; usually there are one or two that use almost the entire capacity of the file system alone. Therefore, the backups of some run much faster than of others, for the "big users" the backup time is only slightly reduced in the worst case. Overall, the (time) gain is usually only marginal.
- If further faculties (probably not so often) or institutes are added, the backup administrator must adapt his configuration in time, otherwise the new ones are left out.

Using UNIX, the nodes can be separated relatively elegantly using the *VIRTUALMOUNT* option, for Windows you either have to create *exclude.dir* rules for each node, which is both complex and error-prone, or work with a trick (see info-box "Workaround for *VIRTUALMOUNTPOINTS* for Windows Clients").

Variant 2

Often, however, the users on the file systems are not organized in groups, but all directories lie flat next to each other on the entry level. Creating a separate ISP node for each user directory

Workaround for VIRTUALMOUNTPOINTS for Windows Clients

For UNIX, Linux, and macOS it is possible to configure individual directories as virtual drives in ISP/TSM. This simplifies the configuration of the backup, since the virtual drive can be specified directly as backup source instead of specifying the actual drive and excluding all directories that are not to be backed up using exclude rules.

Unfortunately, there is no comparable function for Windows. This also eliminates the possibility of parallelizing the backup via different virtual drives.

However, if only individual directories are to be backed up, but not the "remaining" root directory in parallel, numerous exclusion rules must usually be created in the form of *exclude.dir* records in the *dsm.opt*. Of course, this way is highly error-prone, additionally all directories newly created in the root directory of a drive are not automatically excluded, but are included in the backup.

The following workaround simplifies configuration and parallelization of the backup under Windows:

- Create an "advanced share" for each directory you want to back up.
- By adding a *$* to the share name, the Windows SMB service also does not list it on the network map ("hidden share").
- Access to this share is only required by the local administrators of the backup node.
- The paths to be backed up can be accessed via the loopback device:
  DOMAIN \\127.0.0.1\<Share1>
  DOMAIN \\127.0.0.1\<Share2>

From the point of view of the ISP/TSM client, the shares are independent network shares and can be backed up in parallel!

repeats the second problem mentioned above and is very time-consuming regarding the number of users.

It is therefore easier to distinguish the directories according to a pattern, for example after the first character(s): *^[a,A]*, *^[b,B]*, ...*^[z,Z]*, *^[0-9]*. (ISP even provides "regular expressions" at this point!)

You get 27 or 729 ISP nodes, which automatically include all new directories. Unfortunately, the Regular Expressions (RegEx) formulations only capture the directories that exist, not the deleted ones. Remedy is possible if you additionally back up all directories of the start path without subdirectories.

Although this variant is often better than the first, it does not meet all expectations:

- Solving the "deleted directories" problem is cumbersome.
- The configuration becomes – especially if one distinguishes between the first two letters – very extensive.
- Changes to the directory names distribute the data across several ISP nodes and the restore in particular becomes time-consuming.

In summary, there are certainly application scenarios for both approaches, but experience at the GWDG shows that the effort is quite high and there are always a few "power users" who again need special treatment with these two approaches in order to achieve a usable benefit.

ONE APPROACH FOR ALL FILE SYSTEMS

Idea and first steps using BASH

Already in the last decade the (at that time) Generali Versicherungs-AG was faced with the problem outlined at the beginning and Rudolf Wüst as backup administrator extended the aforementioned approach by a decisive idea. From this, he developed a solution that successfully parallelized the "search problem" with up to 2,000 threads. Mr. Wüst kindly shared his extension and the author took it up and developed it further within the scope of his work at the GWDG.

The goal of a practicable solution must be to capture all directories, store them in a single ISP node and still parallelize the search. This can be done by executing a script instead of a simple backup call, which in turn starts several parallel threads to back up the directories. The core of the script consists of a loop of the following form (see example 1).

```
For all (find all directories in the start path)
{
        Start a backup thread for the current directory
}
```

Example 1: Pseudo code

Instead of "one incremental backup", many partial incremental backups are performed for each directory.

The deleted directories are recorded with a subsequent backup of the start path without subdirectories – the last specification is extremely important, otherwise, a normal "incremental backup" is made on the entire file system.

As source code for the BASH this looks like in example 2.

```
startpath=<start path>;
folderlist=<path to a file containing foldernames>;
find $startpath –xdev –mindepth 1 –maxdepth 1 –type d \
  –print > $folderlist;
for $i in $(cat $folderlist)
do
        dsmc –i $i –subdir=yes &
done
dsmc –i $startpath –subdir=no;
rm $folderlist;
```

Example 2: Source code BASH

During the first tests you will find out, that the script in its present form will indeed start as many threads as existing directories. On the one hand, this forces the computer that performs the backup to its knees, and on the other hand, the *MAXSESSIONS* setting of the ISP server is probably reached almost immediately and the server refuses further connections.

The remedy is a counter that simply waits when the allowed number of threads is reached. In the bash, the split backup threads

have the "parent process ID" of the script itself, so these threads can be counted even if you run the script for several file systems simultaneously. Using BASH the loop looks like example 3.

```
pid=$$;   # parents process id
startpath=<start path>;
folderlist=< path to a file containing foldernames >;
maxthreads=<max. number of parallel threads>;
find $startpath –xdev –mindepth 1 –maxdepth 1 –type d \
  –print > $folderlist;

while [ -s $folderlist ]
do
    nthr=$(ps axo ppid,cmd | grep $ppid | grep -v grep | wc -l)
    if [ $nthr -le $maxthreads ]
    then
            # get new start path
            folder=$(head -n 1 < $folderlist);

            # backup actual folder
            dsmc i $folder/ -subdir=yes -quiet >> $ppid.log &

            # remove first line from folderlist
            sed -i '1 d' $folderlist
     else
            sleep 5; # wait to complete another thread
     fi;
done
dsmc –i $startpath –subdir=no

# Waiting for all running threads at the end
while [ $nthr -gt 1 ]
do
      >&2 echo "Waiting for $nthr threads to end"
      sleep 60;
      nthr=$(ps axo ppid,cmd | grep $ppid | grep -v grep | wc -l)
done
rm $folderlist;
```

Example 3: Extended BASH code

In the extended form, essential goals are now achieved, but one cannot be completely satisfied:

- A return value is missing for reporting for the ISP server. In the simplest case, you can add a line *return 0* at the end, then the schedule is always successful – regardless of whether errors occur or not. As already added in example 3, one should rather collect the output of the individual "partial incremental backups" and evaluate them at the end of the script, e.g. search for errors or summarize the "summaries". Depending on the type (and number if necessary) of errors and the "Files failed", the script can then give the appropriate return values. (This extension is already included in the published source code.)
- The problem mentioned with the idea that individual directories use a considerable proportion of the file system size and thus significantly influence the runtime of the backup is not solved by the script. The inequality of the data set / number of objects will certainly be smaller, but will only shift. The next step would be to create the directory list over several levels and thus increase the number of partial backups. As a result, inequality should be more evenly balanced out.
- In addition to the return code of a "client schedule", detailed error messages and an overview in the form of a summary can also be read out in the reporting of the ISP server; this is (currently) not possible with the specified script; it only provides a "traffic-light" status via the return code.

An excursion to the PowerShell

Based on the BASH code, an attempt was made to implement the algorithm with the Windows PowerShell. UNIX affinity combined with reservations about the PowerShell and above all the double effort ended this project after some work without having created an executable version.

PERL – ONE SOLUTION FOR ALL (?) WORLDS AND FURTHER DEVELOPMENT OF THE SIMPLE APPROACH

The closest solution was initially overlooked: a programming/scripting language for all operating systems, neither BASH/MinGW/WSL nor PowerShell / PowerShell Core on Linux, but PERL.

PERL offers numerous functions – also in the area of access to files and directories, which are encapsulated by the respective implementation in such a way that the actual command is independent of the operating system. File system paths can even be specified in both UNIX and Windows nomenclature (i.e. with / or \ as directory separator) and thanks to the *File::Spec->canonpath* function they are converted to the correct format. To a large extend the source code does not need be individually adapted for the respective platform. Exceptions are the paths to the binaries, i.e. *\opt\tivoli\client\ba\bin\dsmc* or *C:\Program Files\Tivoli\baclient\dsmc.exe* and (currently) only partial readout of the directory tree using *find* (Linux) and *Robocopy* (Windows).

Another reason for PERL is that it allows the use of threads in a simple way (see info-box "Threads in PERL") and also ensures that only a certain number of (sub) threads run at the same time and thus the start of further threads only takes place after completion of previous threads – and this independent of the operating system!

The steps outlined for the BASH are thus reduced to three essential steps in PERL:

1. Create a new subthread with the *fork()* function
2. Branch the source code into the paths "main script" and "script for the subthread", in the main script only the number of started threads is incremented, in the subthread the "partial incremental backup" takes place.
3. Check whether the desired number of threads has been reached and waiting for a thread to be terminated and then start a new one.

In detail, the source code is of course somewhat more complex and also takes into account, for example, if that starting a subthread was not successful.

Threads in PERL

PERL offers its own thread module and thus a much more elegant method than the complex solutions for the BASH or the PowerShell:

Using the *fork()* function, the PERL interpreter creates a second thread that starts at just this point in the script. This thread processes all of the statements below in the same way as the original script. It therefore makes sense to use an IF statement to branch the different tasks. For this, the return value query of the *fork()* routine provides: If the value is not defined, no thread could be created, if the value is "true", there is a new thread – and it is the parent routine in which this IF was executed. The value is also defined in the child thread, but "false". The query could therefore look as follows:

```
my $cpid = fork();
if (! defined $cpid)
{       # forking failed!
        exit 1; # stop script
}
if ($cpid)
{       # parent process
        # … further commands for the parent thread
}
else
{       # child process
        # … further commands for the child thread
}
```

Collecting/waiting for the started threads is also much easier: The *wait()* function waits for a child thread to end, so that for all can be waited with a simple loop:

```
While (wait() != -1 ) ;
```

Further development: Deeper dive into the directory tree and start parallel threads based on multiple directory levels

The tests with the parallelization approach directly below the base path showed exactly those effects that were already addressed during parallelization via institutes: Individual directories are (usually) larger than all other parallel-lying directories together, so that the speed gain is considerably lower than expected or desired. A better balance can only be achieved via additional directories; these can be found by searching through further levels in addition to the first, highest directory level below the start path and then allowing the backup to be made via all these directories. The first problem is that the directories are nested, i.e. a partial backup of a directory from a higher level also includes those subdirectories that are backed up in other parallel threads anyway. In this script, this problem was solved by saving all directories above the set "dive depth" with the option *-SUbdir=No*, i.e. only the contents of these directories including the names of the subdirectories, but not their contents. In a second step, the directories are backed up at the lowest level specified with their subdirectories (*-SUbdir=Yes* option). (Since backups without subdirectories are usually much faster, those directories with subdirectories are backed up first and those without are backed up second.)

Evaluation of the individual runs

Not only for profiling (see below) but also to create a summary of the backup, each sub-thread writes its output to a separate file that contains its own ID in the name in addition to the process ID of the script. Thus, even if the script is aborted, the output can be clearly assigned.

Although the overall evaluation can only take place at the end of the backup, a sufficiently deep dive into the directory tree in practice quickly leads to several thousands to hundreds of thousands of small files and thus to considerable problems. Therefore, the sub threads write the content of the output file after completion of their backup to a central log file, which is evaluated at the end of the script. Within the context of this writing, the information whether subdirectories have been processed is also stored and the runtime is already converted into seconds for profiling and saved. The return value of the backup call is also added.

In the current implementation (July 2018), the final evaluation sums up

- the number of objects inspected
- the number of objects saved
- the number of updated objects
- the number of deleted objects
- the number of orphaned objects
- the number of objects with errors
- the number of inspected bytes
- the number of bytes transferred
- the time for data transfer
- the elapsed time

If necessary, a conversion to the common size (bytes, seconds) takes place.

In addition, the number of

- of warnings (*/^AN[RS][0-9]{4}W/*)
- serious errors (*/^AN[RS][0-9]{4}S/*)
- errors due to *ANS1228E* and *ANS1820E*
- "Server-Out-of-Space"-errors (*ANS1292S*) additionally

are counted in each case and as a total.

From the sum of the elapsed times and the runtime of the loop via the directories ("wall clock time") the script calculates a parallel speedup, which shows how much faster the parallelization is compared to the sum of the individual times.

Performance optimization by profiling

The runtime of the parallel backup is essentially determined by the runtimes of the individual backup runs. Without a detailed measurement (but by comparing the time for the script call with skipping the backup itself) it is assumed that the runtime of the PERL statements is negligible in comparison. The aim of the optimization is the "best" order of the directories, so that

1. large, long-running ones run as parallel as possible,
2. large ones are started first because a mismatch balance has a less dramatic effect on the shorter runtimes of the smaller directories.

Since the runtime of the backups cannot be estimated in advance, the optimization is based on the last backup (and does assume no dramatic changes, which could only be predicted by complex and therefore time-consuming analyses). As described above, the sub threads also write the runtime in seconds to the central log file at the end of their backup, so that a list of all directories with the respective runtimes is created when they are evaluated. This list is sorted by descending runtimes and written to a profile file.

The next time the script is called, it first creates a list of all directories to be backed up. In the next step (this part does not yet work for Windows and has therefore been swapped out again) the backup script compares this list with the entries from the profiling file.

Directories that are in the profiling file but no longer exist in the directory list are ignored. New entries in the directory list for which there is no runtime in the profiling file are assigned a long runtime (10^10 seconds ≙ > 316 years) and are therefore ranked first.

If there is no profiling file, the directories are processed in the order they appear in the list from the directory tree.

At the end of the evaluation, the profiling list is overwritten.

OPEN ISSUES / OUTLOOK

There are still some questions left, for example about transferring the summary to the server log. For a good solution, error handling should be added to make the script fault-tolerant to certain situations.

It is also possible to split the work steps "identify directories" and "partial incremental backup", so that for very large file systems, the list of directories to be processed is filled up again as soon as the backup window has expired, – but probably increasing the immersion depth is the better approach.

One problem that cannot be solved is the fact that "partial incremental backups" do not change the "last backup" attributes of the nodes or file spaces and, of course, this is not done within the scope of the outlined script. You should refrain from writing to the DB2 of the ISP servers, as this affects IBM's warranty. IBM expressly prohibits direct access to the ISP-DB2 outside of corresponding instructions within the scope of support.

IN ADDITION, HOW DO YOU SPEED UP THE RESTORE?

The previously mentioned approaches with ISP on-board tools and the outlined approach for parallelization only work for backup. If many files are to be restored from the backup, this is very easy with the approaches with several nodes for a file space, since a separate restore must run for each node anyway and the processes run in parallel. For the parallel threads approach, an adjustment for the restore based on a file list is easily possible: Instead of a "folder list", a file list is used for the restore.

However, it should be noted that in an environment with a tape library as a storage backend, the number of drives usually limits the performance of the restore. Furthermore, ISP usually organizes the restore (without the *-disablenqr=yes* option) so that the tape mounts are optimized. If a file list is processed in parallel by numerous parallel threads, the server cannot optimize the tape accesses. However, if a disk-based FILE or container pool is used, the parallel restore over numerous threads is faster. If the data is stored on two servers via server replication, the restore can also be distributed over both servers and thus additionally accelerated.

Unfortunately, experience shows that "full restores" also involve enormous effort when parallelizing and can only be accelerated unsatisfactorily.

AVAILABILITY / ACCESS TO SOURCE CODE / ALTERNATIVES

It can be assumed that neither the author of the original idea nor the GWDG can claim to be the only one to have had and implemented the idea outlined. Rather, many TSM/ISP users may have faced the same problem and found similar solutions.

A commercial implementation that follows a similar approach to parallelization can be found in the product "MAGS" of General Storage [6]. In addition to realiable support, "MAGS" offers regular further development and uses several NAS nodes for parallelization with ISILON Scale Out systems. A more detailed product analysis should not take place here. You must also determine the individual benefit.

The script mentioned in this article is freely available in GitLab [7] of the GWDG under the Apache 2.0 license. The script may be used and modified without restrictions. We look forward to receiving your feedback and suggestions.

TRANSFERABILITY TO OTHER BACKUP SOLUTIONS

The approaches presented address the problem of file identification and can therefore be applied to all other questions where a file list is to be created. If you replace the call of the ISP-CLI with another CLI call, you can also find all files in parallel, filtered by all attributes supported by *find* using appropriate parameters. You can also add another loop that does arbitrary operations with all entries of a complete file list. This also allows you to optimize other backup solutions that can process a directory or file list.

ACKNOWLEDGEMENT

The author thanks Gerd Becker (Empalis GmbH), Wolfgang Hitzler (IBM) and Manuel Panea (Max Planck Computation and Data Facility) for proofreading the original article and making suggestions for changes and improvements. Special thanks to Rudolf Wüst (Generali Shared Services S.c.a.r.l.) for his generosity in sharing his ideas.

FOOTNOTES AND LINKS

[1] *https://www.mail-archive.com/adsm-l@vm.marist.edu/msg102161.html*
[2] *http://www.storageconference.us/2014/Presentations/Shimizu.pdf*
[3] *http://www.insic.org/news/2015%20roadmap/15pdfs/2015%20Technical%20Roadmap.pdf*
[4] After evaluation of the GWDG ISP servers: between 15% and 84% in addition to the active data, whereby the 84% is an outlier, the average value is 39%
[5] *https://www.ibm.com/support/knowledgecenter/SSGSG7_7.1.6/client/r_opt_incrbydate.html*
[6] *http://www.general-storage.com/PRODUCTS/dsmISI-MAGS/body_dsmisi-mags.html*
[7] *https://gitlab.gwdg.de/bnachtw/dsmci*

CloudMob – a Cloud-based Mobile Big Data Processing Platform

Text and Contact:
Bo Zhao
bo.zhao@gwdg.de
0551 201-2198

This article reports the research results on a case study of mobile big data analysis. In recent years, the increasing popularity of mobile services has lead to an explosion of mobile data. The usage of mobile phones has generated large-scale diversified mobile phone data, such as call data records (CDRs), Internet access records, and location information etc. These records contain an enormous amount of information on our real life. In order to support various researches based on mobile big data, mobile operators need to collect, store, process and analyze a massive amount of data generated from each mobile phone user efficiently. To this end, we design a cloud-based mobile big data processing platform named CloudMob. Using the framework in practice, we have realized the storage and processing of the massive mobile data from millions of telecom users every day. Based on CloudMob, we can preprocess and share the mobile data with many cooperators for many further research scenarios related to mobile big data, such as social analysis, user behavior analysis, network analysis and social economic status analysis.

INTRODUCTION

Nowadays an increasing number of people rely on mobile devices for personal life. By the end of 2016 the number of mobile phone subscribers around the world reached to 4.8 billion, which represents a vast fraction of the global population [1]. Mobile phones are ubiquitous both in developing and developed world. Besides, in the recent few years, the increasing popularity of mobile services has lead to an explosion of mobile data. The usage of mobile phones has generated large-scale diversified mobile phone data, such as call data records (CDRs), Internet access records, and location information etc. These records contain an enormous amount of information on our real life. For example, CDRs record how, when, where, and with whom we communicate.

The extremely rich and informative source of mobile data represents a promising opportunity to both industry and academia. The past few years have witnessed the rise of research based on the analysis of mobile data. Naboulsi et al. [2] provide an overview of several works that identify three research fields: social, mobility and network analyses. Mobile operators have started fundamental and applied research on mobile datasets, such as the Data for Development (D4D) Challenges by Orange (*http://www.d4d.orange.com*) and the Telecom Italia Big Data Challenges (*http://www.telecomitalia.com/bigdatachallenge/*). Meanwhile, NetMob, an international conference on the analysis of mobile phone datasets, has seen a jump in the number of publications since 2013. The research span over many domains, such as health, privacy etc. Considering all the trends, the future of mobile big data analysis is reasonably promising.

In order to support various researches based on mobile big data, mobile operators need to collect, store, process and analyze a massive amount of data generated from each mobile phone user efficiently. To this end, we design a cloud-based mobile big data processing platform named **CloudMob**. Cloud computing has become popular recently owe to several advantages, such as flexibility, scalability or energy efficiency [3]. Employing cloud computing in our platform can effectively address the issue of large-scale datasets management. Our framework is build over distributed relational database and Hadoop, which provides an open source cluster computing platform. On the basis, we can deal with the massive amounts of mobile data collected by millions of databases around the whole China in parallel and thus improve the efficiency.

The rest of this paper is organized as follows. Section "Related Work" presents an overview of the related research work corresponding to mobile big data analytics, cloud computing and social economic status analysis. Section "Logical Architecture" introduces the logical architecture of our framework and Section "Implementation and Evaluation" introduces the details of the framework implementation and evaluation. We conclude our work and introduce some possible future work directions in Section "Conclusion and Future Work".

RELATED WORK

Cloud computing has its special advantages for both storing of big data and computing of big data relative to traditional solutions such as Supercomputing, Distributed Computing, Parallel Computing, and Grid Computing [4]. When designing a data center, we mainly tend to leverage the virtualization technology to maximize the utilization of computing resources and therefore it should provide the basic components such as storage, CPUs, and network bandwidth as a commodity by specialized service providers to reduce the unit cost. Most of the research institutions and enterprises introduce virtualization into cloud architectures for big data management, among which Amazon Web Services (AWS), Eucalyptus, OpenNebula, CloudStack and OpenStack are the most popular cloud management platforms for Infrastructure as a Service (IaaS) [5].

AWS is the typical pay-as-you-go cloud provider with huge usage in elastic platform and is very popular for small companies.

Nurmi et al. [6] presented Eucalyptus: an open-source cloud-computing framework that uses computational and storage infrastructure commonly available to academic research groups to provide a platform that is modular and open to experimental instrumentation and study.

OpenNebula is an open-source project aimed at building the industry standard open-source cloud computing tool to manage the complexity and heterogeneity of large and distributed infrastructures, which can offer the richest features, flexible ways and better interoperability to build private, public or hybrid clouds [7].

CloudStack (*https://cloudstack.apache.org/docs/*) is an open-source software platform, written in Java, designed for development and management of cloud Infrastructure as a Service. CloudStack aggregates computing resources for building private, public or hybrid clouds and is a turnkey technology that brings together the "Stack" of features requested by companies and users, like data centers orchestration, management and administration of users and NaaS (Network as a Service) [8].

OpenStack is a cloud software that offers capability to control large pools of compute, storage and networking resources, which empowers users providing resources on demand [9]. OpenStack was developed by Rackspace Hosting and NASA [10] aimed to provide open-source cloud solutions to build public or private clouds. In current situation, OpenStack has a good community and ecological environment while it still have some shortcomings like incomplete functions and lack of commercial supports [4]. Considering the advantages and disadvantages as well as security issues of current cloud platforms and aiming to deal with the massive data of telecom, we proposed our own cloud platform named CloudMob for mobile data storage and processing, which will be described in Section "Logical Architecture" in detail.

Mobile big data analytics (MBD) has been gaining more and more attention recently with the remarkable evolution and explosive popularization of smartphones [11]. MBD analytics is more complex and difficult than conventional big data problems because of the portable data sources and crowd sourced data traffic. Its processing object is massive amount of data collected by millions of mobile devices distributed around the world. Figure 1 shows a typical architecture of large-scale mobile systems used to connect various smartphones supported by a telecom operator. Each mobile device encapsulates its service request and own sensory

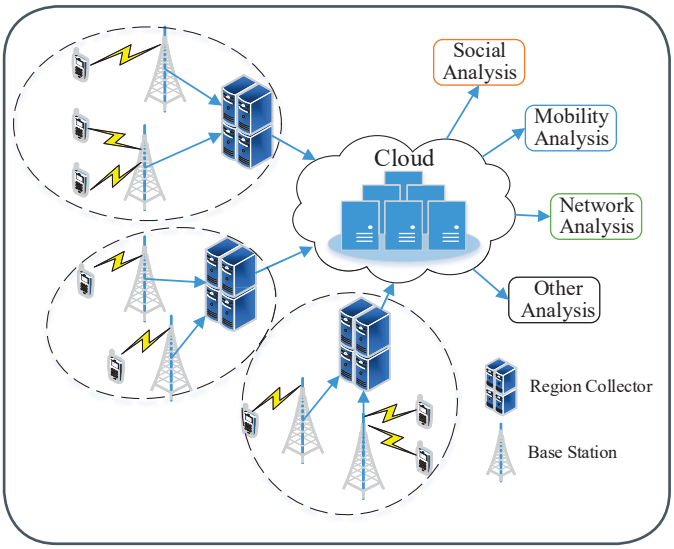


Figure 1: Typical architecture of a modern mobile network connecting smartphones

data in stateless data-interchange structure and such data can be stored by the service server temporarily of permanently. Based on the collected MBD, a service server is able to use data mining and other techniques discover hidden patterns and information, such as social analysis [12], mobility analysis [13], network analysis [14] and health care service [15] etc.

LOGICAL ARCHITECTURE

CloudMob can be divided into four layers logically, namely data collection&transmission layer, data receiver layer, data processing layer, data service layer and management layer respectively, as

CloudMob – eine cloud-basierte Plattform zur Verarbeitung großer Mengen von Mobilfunkdaten

In den letzten Jahren haben die zunehmende Popularität mobiler Dienste und die daraus resultierende verstärkte Nutzung von Mobiltelefonen auch zu einem großen Anstieg von vielfältigen Mobilfunkdaten wie z. B. Gesprächsdatensätze (CDRs), Internetzugangsdatensätze und Standortinformationen geführt. Diese Aufzeichnungen enthalten eine enorme Menge an Informationen über unser wirkliches Leben. Um verschiedene Forschungstätigkeiten auf der Grundlage solcher Daten zu unterstützen, müssen Mobilfunkbetreiber eine riesige Menge an Daten sammeln, speichern, verarbeiten und analysieren, die von jedem Mobilfunknutzer erzeugt werden. Zu diesem Zweck entwickeln wir eine cloud-basierte Plattform namens CloudMob zur Verarbeitung der riesigen Mengen an Mobilfunkdaten. Mit dem Framework haben wir in der Praxis die Speicherung und Verarbeitung der täglich anfallenden Mobilfunkdaten von Millionen von Telekommunikations-Nutzern realisiert. Basierend auf CloudMob können wir diese Daten vorverarbeiten und mit vielen Beteiligten für viele weitere Forschungsszenarien im Zusammenhang mit Mobilfunkdaten, wie z. B. Sozialanalyse, Nutzerverhaltensanalyse, Netzwerkanalyse und sozialwirtschaftliche Statusanalyse, teilen.

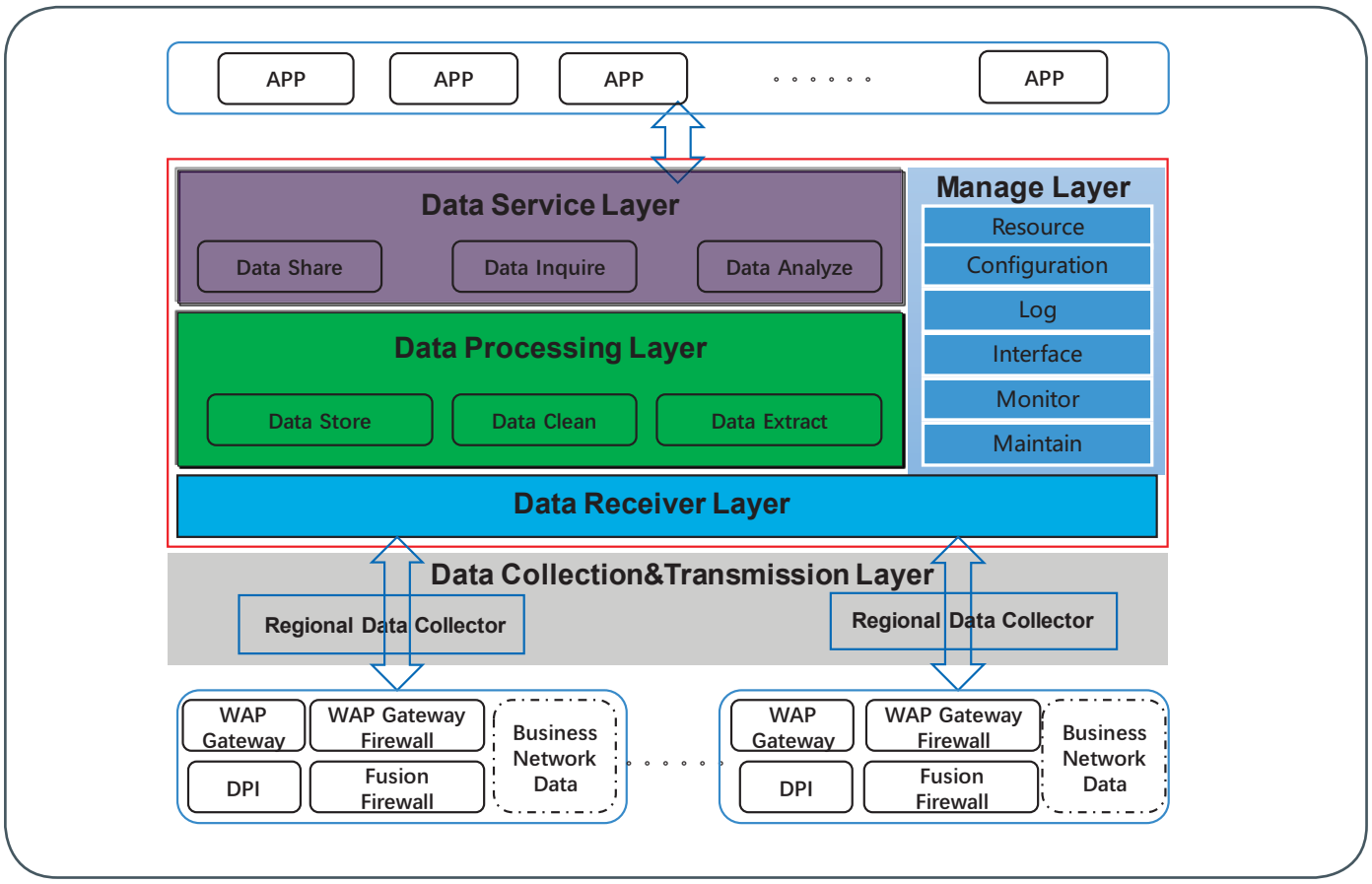


Figure 2: Logical architecture of CloudMob

shown in Figure 2. Data collection&transmission layer is designed to collect the original data from each data source and compress the uncompressed data to upload them. In addition, it also provides other functions including retransmission and cache. Data receiver layer receives data from each subcenter and check their forms, formalization, relevance and integrity. On the basis, it will process data cleaning to get rid of the redundant data and then load them into HDFS. The data collected by data receiver layer can be shared to other users. Data processing layer is mainly responsible for the data processing of original data, including data management, data statistics and summarization, relevant data storage etc. Data service layer provides uniform access and sharing service, such as data query, data sharing and business data statistic and analysis, etc. Management layer provides uniform function models, including data access management, system access management, data security management, resource monitoring and distribution, task scheduling and monitoring management, logs management etc.

IMPLEMENTATION AND EVALUATION

Implementation

Figure 3 shows the technical architecture implemented in CoudMob, including data receiver layer, data processing layer and data service layer. The details of each layer are introduced as follows.

Data receiver layer is responsible for the collection and transmission of data by FTP/SFTP and then load the data into HDFS after a series of data pre-processing, such as data checking, data splitting and merging etc.

Data processing layer provides functions of massive data

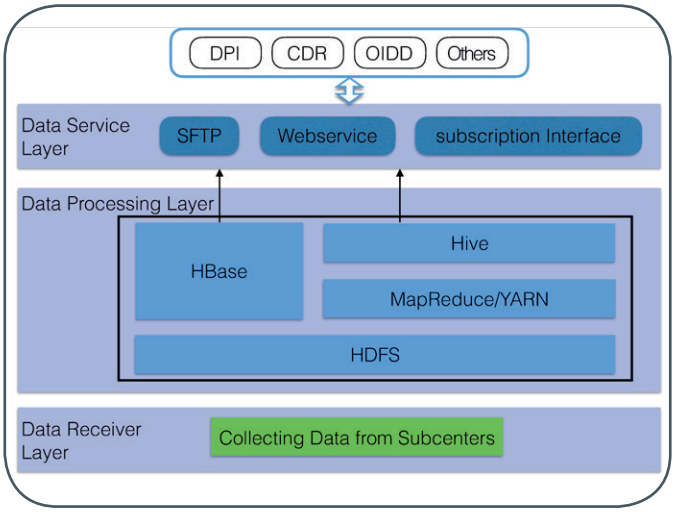


Figure 3: Technical architecture of CloudMob

storage, query and analysis etc. to support the demand scenarios and applications.

- HDFS is a distributed file storage system and is used to store the original files because of its high throughput capacity. This function model implements the original data storage, realtime business data storage, massive data cleaning and analysis (Hive) etc.
- MapReduce/YARN provides computing framework and resource management for the data processing under Hadoop. The parallel processing mechanism of MapReduce can schedule the processing of relevant tasks uniformly to take full advantage of the cluster resource for high performance processing.

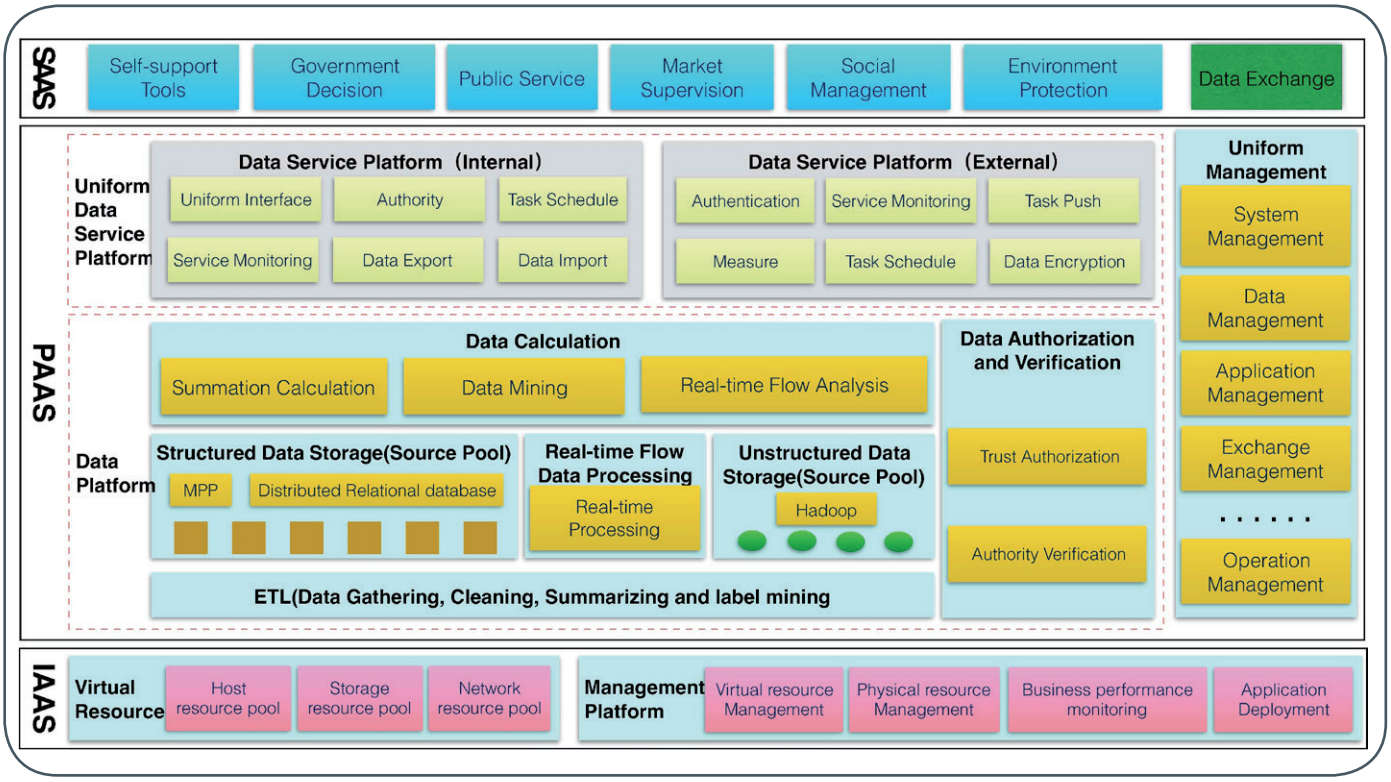


Figure 4: Cloud assisted mobile big data processing platform

- HBase is a dynamic database oriented to structured data, which is portable, reliable, efficient, distributed and column-oriented. We use BulkLoad method to transfer the data after relevance analysis into the internal format of HBase and then load them into HBase, which has a high performance and uses fewer CPU and network resources. This model is mainly used for realtime query of external data in logs storage system.
- Hive is a repository tool based on Hadoop and can map a file with structured data into a database table. It can provide some basic SQL operations, such as query, add and delete etc. In addition, it can transfer the SQL statements into MapReduce tasks for execution.

Data service layer provides the external service interface of the system, including the sharing of data files, service query and data analysis etc.

- **Support for the file sharing of batch data** through SFTP/FTP as well as the data sharing through the method of realtime query to data file in HBase.
- **Support for general message interfaces of Webservice**, used for data access, data query and data analysis etc.
- **Support for data sharing** through subscription method (Subscription interface is designed based on the characteristics of different applications).

System management layer is designed to manage the whole framework uniformly. In order to guarantee the efficiency and security of our platform, we designed the following submodules for system management layer.

- **Service management**, used for the access management aiming at service and statistics report of services etc.
- **Security management**, used to guarantee the security of users and service on data platforms, including users management and access authorization, data authorization etc.
- **Resource management**, used to manage the resource scheduling and resource usage monitoring etc.
- **Task management**, used to maintain and monitor the executive situations of tasks published by servers and support the dependent or relevant execution of tasks.
- **System management**, including the basic configuration information, log information of service and platform, monitoring and maintaining of system platform etc.
- **Data management**, providing uniform foreground data management function and managing the data lifetime of data on the platform.

The detailed function module design of the framework is shown in Figure 4. We divide the mobile big data processing platform into three parts considering the category of cloud computing services. The SAAS part is mainly corresponding to the actual application scenarios using mobile big data analysis, such as government decision, public service, environment protection and market supervision. The PAAS includes part the main work we have done to provide mobile big data processing support for many cooperators. We implemented two data service platforms for internal and external users respectively. Based on the uniform management standards, we can share the mobile data after preprocessing to our collaborators for further analysis. The IAAS part includes the virtual resources and management platforms related to our framework. This part is implemented and maintained by our internal system administrators to schedule the resource uniformly and safely.

Evaluation

Currently, we are maintaining a data center with 1000 machines and collecting data from two subcenters in the north and south China, respectively. Our datacenter is supported by tens

of province-level sub data centers and processing the massive amounts of mobile data from millions of telecom users. We evaluate our framework proposed in this paper from the following two aspects, namely performance evaluation and feasibility evaluation.

Performance evaluation: Using the cloud framework proposed in this paper, we can process 50 T original compressed mobile data on average every day currently. The time delay is controlled less than 5 minutes for end-to-end processing. Comparing with other datasets, the mobile Internet access dataset has a much larger scale. We should deal with all the mobile internet access cookies and store the data after being preprocessed temporarily for further analysis. CDR and other data with smaller scale can be stored on the platform for a longer time. We can abstract the data we need quickly with the distributed file storage strategy and parallel processing method.

Feasibility evaluation: Our cloud assisted mobile big data processing platform has been applied into practice for quite a long period and used to process the massive mobile data of China Telecom. We designed detailed standards for data collecting and processing to guarantee the security and efficiency of mobile big data processing. Based on CloudMob, we have carried out wide collaborations with many cooperators for many hot research areas, such as relationship between user mobility and society security, relationship between smartphone usage and health care, consumer behavior analysis and target marketing.

CONCLUSION AND FUTURE WORK

The increasing popularity of mobile services has lead to an explosion of mobile data. The usage of mobile phones has generated large-scale diversified mobile phone data, such as call data records (CDRs), Internet access records, and location information etc. These records contain an enormous amount of information on our real life and can be used for various researches based. Facing such large-scale mobile data, mobile operators need to collect, store, process and analyze a massive amount of data generated from each mobile phone user efficiently. To this end, we design a cloud-based mobile big data processing platform named CloudMob. We designed the logical and technical architectures of CloudMob and implemented each function model. We have been using the framework in practice and realized the storage and processing of the massive mobile data from millions of telecom users every day.

Based on CloudMob, we have carried out many collaborations with many cooperators for many further research scenarios related to mobile big data, such as social analysis, user behavior analysis, network analysis and social economic status analysis etc. For future work, we plan to deploy Spark framework to replace the current Hadoop framework in order to improve the processing efficiency. In addition, we will optimize the scalability of our architecture and deploy more machines to promote the processing capacity as whole because of the fast growing mobile data generation.

REFERENCES

[1] GSMA, "Number of global mobile subscribers to surpass five billion this year, finds new gsma study," 2017.

[2] D. Naboulsi, M. Fiore, S. Ribot, and R. Stanica, "Large-scale mobile traffic analysis: a survey," in: "IEEE Communications Surveys & Tutorials", Vol. 18, No. 1, 2016, pp. 124–161.

[3] B. Hayes, "Cloud computing," in: "Communications of the ACM", Vol. 51, No. 7, 2008, pp. 9–11.

[4] M. Bahrami and M. Singhal, "The role of cloud computing architecture in big data," in: "Information granularity, big data, and computational intelligence", Springer, 2015, pp. 275–295.

[5] C. Ji, Y. Li, W. Qiu, U. Awada, and K. Li, "Big data processing in cloud computing environments," in: "Pervasive Systems, Algorithms and Networks (ISPAN) 2012", 12th International Symposium on, IEEE, 2012, pp. 17–23.

[6] D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L. Youseff, and D. Zagorodnov, "The eucalyptus open-source cloud-computing system," in: "Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid", IEEE Computer Society, 2009, pp. 124–131.

[7] X. Wen, G. Gu, Q. Li, Y. Gao, and X. Zhang, "Comparison of opensource cloud management platforms: Openstack and opennebula," in: "Fuzzy Systems and Knowledge Discovery (FSKD) 2012", 9th International Conference on, IEEE, 2012, pp. 2457–2461.

[8] A. Barkat, A. D. dos Santos, and T. T. N. Ho, "Open stack and cloud stack: Open source solutions for building public and private clouds," in: "Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) 2014", 16th International Symposium on, IEEE, 2014, pp. 429–436.

[9] K. Jackson, OpenStack cloud computing cookbook. Packt Publishing Ltd, 2012.

[10] B. Piatt, "Openstack tutorial," 2010.

[11] J. Poushter, "Smartphone ownership and internet usage continues to climb in emerging economies," Pew Research Center: Global Attitudes&Trends, 2016.

[12] X. Zhang, Z. Yi, Z. Yan, G. Min, W. Wang, A. Elmokashfi, S. Maharjan, and Y. Zhang, "Social computing for mobile big data," in: "Computer", Vol. 49, No. 9, 2016, pp. 86–90.

[13] S. Jiang, J. Ferreira, and M. C. González, "Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore," in: "IEEE Transactions on Big Data", Vol. 3, No. 2, 2017, pp. 208–219.

[14] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5g," in: "IEEE Network", Vol. 30, No. 1, 2016, pp. 44–51.

[15] A. T. Lo'ai, R. Mehmood, E. Benkhlifa, and H. Song, "Mobile cloud computing model and big data analysis for healthcare applications," in: "IEEE Access", Vol. 4, 2016, pp. 6171–6180.

Erfolgreiche Teilnahme an der 4. Nacht des Wissens

Text und Kontakt:
Dr. Sven Bingert
sven.bingert@gwdg.de
0551 201-2164

Steffen Rörtgen
steffen.roertgen@gwdg.de
0551 201-2117

Auch die 4. Nacht des Wissens am 26. Januar 2019 in Göttingen war für den Göttingen Campus wieder eine äußerst erfolgreiche und gelungene Veranstaltung. Mit mehr als 25.000 Besuchern wurden alle Erwartungen übertroffen und ein neuer Rekord erzielt. An allen 25 Veranstaltungsorten gab es ein großes Interesse daran, Göttinger Forschung hautnah zu erleben. Die GWDG-Präsentation „Data Science: Finde Deinen Zwilling“ mit einem Experiment zur Gesichtserkennung zog viele interessierte Besucher an, deren vielfältige Fragen die Mitarbeiter am GWDG-Stand gerne beantworteten.

370 ANGEBOTE AN 25 STANDORTEN

Bei der 4. Nacht des Wissens in Göttingen, die am 26. Januar 2019 von 17:00 bis 24:00 Uhr an 25 Standorten verteilt über die gesamte Stadt stattfand, boten die Universität und die Forschungseinrichtungen des Göttingen Campus sowie weitere teilnehmende Einrichtungen mit rund 370 verschiedenen Angeboten ein abwechslungsreiches Programm an. Ziel war es, die vielfältige Forschung der beteiligten Einrichtungen einem breiten Publikum unterhaltsam zu präsentieren. Forschung wurde auf unterschiedlichste Weise für alle Alters- und Interessengruppen verständlich und erlebbar gemacht – sozusagen Forschung zum Mitmachen und Anfassen. Es gab Science Slams, Vorträge, Mitmachaktionen, interaktive Präsentationen, Führungen, Workshops, Filme und Experimente. Für jeden war etwas dabei und für manch einen hat die Zeit nicht ausgereicht, um an allen gewünschten Programmpunkten auch teilzunehmen. Rund 25.000 Besucher tauchten in die Welt der Forschung ein, um das eigene Wissen zu erweitern, und zeigten damit eindrucksvoll, dass Wissenschaft und Forschung offensichtlich ganz Göttingen und Umgebung bewegen und begeistern.

GWDG at the 4th Night of Science

Also this year the GWDG took part in the Night of Science on January 26th in Göttingen. We presented three experiments with the topic data science, especially face recognition. We offered to print out a picture taken from a photo box and overlaid with detected faces and their features. In a second experiment we showed that a simple single-board computer is sufficient to take pictures with a webcam, detect faces and present the results in live on a screen. The focus of the booth was the workflow to find the twin of a person. For that we offered two workspaces where we explained and conducted the experiment. The visitors could take a picture via the webcam. The open-source software Orange3 [1] was then used to detect the face, calculate the embeddings with the help of a neural network [3] and finally compute the nearest neighbor, the twin, from a large data set [2] of about 13.000 people.

GWDG-STAND IM ZENTRALEN HÖRSAALGEBÄUDE

Auch die GWDG war wie in den Vorjahren bei der Nacht des Wissens dabei; diesmal mit einem Experiment zum Thema Data Science und erstmalig an einem Stand im Zentralen Hörsaalgebäude der Universität Göttingen. Unter dem Motto „Finde Deinen Zwilling" wurden verschiedene Experimente zur Gesichtserkennung mit unterschiedlichen Methoden präsentiert. Diese waren eine einfache Live-Demo mit einem kleinen Einplatinen-Computer, eine Fotobox mit Druckfunktion für Erinnerungsfotos an die Nacht des Wissens und die Suche des Zwillings an zwei Arbeitsplätzen mit Erklärung der Vorgehensweise durch einen GWDG-Mitarbeiter. Da das Thema Gesichtserkennung zurzeit politisch stark diskutiert wird und es sicherlich viele Anwendungsmöglichkeiten gibt (z. B. Gebäudeeinlasskontrolle), war das Interesse an unserem Aufbau sehr groß. Wir konnten mit unseren Experimenten zeigen, wie einfach der technische Aufbau sein kann und dass mit nur wenigen Programmierkenntnissen eine solche Analyse durchgeführt werden kann.

Aufbau

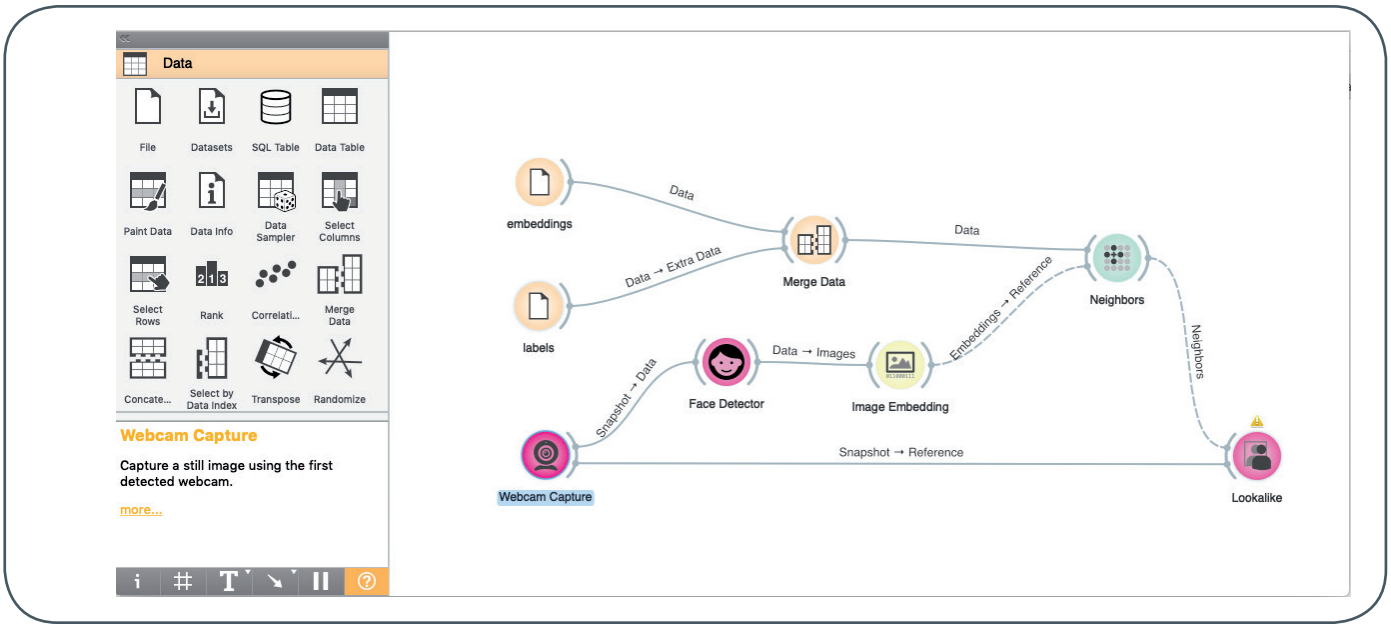
- **Fotobox:** In einer Fotobox wurde mit einer Spiegelreflexkamera ein Bild aufgenommen und an einen Einplatinen-Computer weitergeleitet. Dieser hat mit einem Algorithmus die Umrandung des Gesichts, der Augen, der Nase und des Mundes bestimmt und auf das Bild projiziert. Die Besucher konnten dann per Knopfdruck dieses Bild an dem angeschlossenen Drucker ausgeben und als Erinnerung mit nach Hause nehmen.
- **Live-Demo:** Ein Einplatinen-Computer wurde mit einer kleinen Webkamera und einem Bildschirm verbunden. Im Computer wurden die Bilder mittels eines Algorithmus mit den Umrandungen der Gesichtsmerkmale überlagert, ähnlich wie in der Fotobox. Das Experiment hat gezeigt, dass keine besondere Computer-Hardware für eine Gesichtserkennung hinter einer Web-Kamera notwendig ist. In dem Experiment konnte auch anschaulich erklärt werden, wie z. B. die Müdigkeitsüberwachung eines Autofahrers anhand der berechneten Augengröße funktionieren kann. Wenn diese unterhalb eines Schwellenwertes fällt, wird ein Alarmton ausgelöst.
- **Finde Deinen Zwilling:** An zwei Arbeitsplätzen wurde ein Workflow präsentiert, mit dessen Hilfe eine Ähnlichkeitsanalyse zwischen einem Webkamera-Bild und einem vorgegebenen Datensatz durchgeführt werden konnte. Besucher konnten ihr eigenes Bild dann mit einem Datensatz aus über 13.000 Bildern von Personen aus dem öffentlichen Leben vergleichen und herausfinden, wer ihnen am ähnlichsten ist. Für Interessierte wurde dann auch der Workflow im Detail erklärt.

WORKFLOW „FINDE DEINEN ZWILLING"

Wir verwendeten das Open-Source-Programm Orange3 [1]. Das Programm basiert auf Python und bietet die Möglichkeit, verschiedene Datenanalyse-Workflows für die unterschiedlichsten Fachrichtungen auf einfache Weise zu erstellen. Ein weiterer großer Vorteil besteht außerdem in der Konstruktion der Workflows. Durch einfaches Drag & Drop können verschiedene Operationen in den Workflow integriert werden, ohne dass dafür selbst Programmcode geschrieben werden muss. Dabei werden viele bekannte Python-Bibliotheken und bekannte Data-Science-Algorithmen, von Clustering bis Natural-Language-Processing, verwendet. Darüber hinaus besteht dank des gut dokumentierten Datenmodels die Möglichkeit, eigene Funktionalitäten einzubauen oder Python-Code zu integrieren. In Abbildung 1 ist der Workflow zur Zwillingsfindung dargestellt.

Die Analyse der Bilder erfolgt in drei Schritten:

1. Erkennung und Ausschneiden eines Gesichts in einem Bild
2. Berechnung von 128 Werten (Embeddings) mittels eines trainierten Neuronalen Netzwerkes zur numerischen Darstellung eines Gesichts
3. Berechnung des nächsten Nachbarn in diesem mehrdimensionalen Raum



1_Workflow zum Finden des Zwillings

Der erste Schritt im Workflow besteht aus dem Laden der schon berechneten Werte für den großen Datensatz pubfig83 [2]. Für das Neuronale Netzwerk verwendeten wir Openface [3], das speziell für Gesichter trainiert wurde. Diese Werte werden dann mit den Labels und den Bildnamen zu einem Datensatz zusammengefügt und für die Berechnung des nächsten Nachbarn bereitgestellt.

In einem nächsten Schritt kann dann ein Bild über die Webkamera aufgenommen werden. Bevor dieses Bild mittels des Neuronalen Netzwerks umgewandelt wird, muss das Gesicht mittels eines Haar-Classifiers [4] extrahiert werden. Der extrahierte Bereich wird dann umgewandelt und als Referenz in der Nachbarsbestimmung verwendet. In diesem nun 128 x 13.000 großen Raum wurden dann die zehn nächsten Nachbarn berechnet und im letzten Schritt des Workflows zusammen mit dem Referenzbild dargestellt.

Dieser Workflow bietet die Möglichkeit, andere Features in ähnlicher Weise zu untersuchen, indem die passenden Algorithmen zur Extraktion, das passende Neuronale Netzwerk und die passende Ähnlichkeitsbestimmung verwendet werden.

Die Einfachheit, mit der das Programm Orange3 verwendet werden kann, überzeugt immer wieder und lädt dazu ein, es in der Lehre und Forschung intensiver einzusetzen.

REFERENZEN

[1] Orange3 Version 3.19. (orange.biolab.si)
J. Demsar, T. Curk, A. Erjavec, C. Gorup, T. Hocevar, M. Milutinovic, M. Mozina, M. Polajnar, M. Toplak, A. Staric, M. Stajdohar, I. Umek, L. Zagar, J. Zbontar, M. Zitnik, and B. Zupan: „Orange: Data Mining Toolbox in Python". Journal of Machine Learning Research 14 (Aug), 2013, pp. 2349–2353.

[2] Pubfig83
Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, and Shree K. Nayar: „Attribute and Simile Classifiers for Face Verification". International Conference on Computer Vision (ICCV), 2009.

[3] Openface
B. Amos, B. Ludwiczuk, and M. Satyanarayanan: „Openface: A general-purpose face recognition library with mobile applications". CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.

[4] OpenCV Haar
Paul Viola, and Michael Jones: „Rapid Object Detection using a Boosted Cascade of Simple Features". IEEE Conf Comput Vis Pattern Recognit. 1. I-511. 10.1109/CVPR.2001.990517.

SAP Intelligent Enterprise Truck in Göttingen

In der Woche vom 18. bis 22. Februar 2019 präsentiert die SAP ihren Intelligent Enterprise Truck (*https://events.sap.com/de/truck-tour-mee/de/home*) auf dem Platz der Göttinger Sieben. Zusammen mit der GWDG, der Universitätsmedizin Göttingen und der Georg-August-Universität Göttingen wird in dieser Woche eine prototypische Umsetzung eines ausgewählten Anwendungsfalls im Innovation Lab erstellt. Im Team mit den Experten der SAP wird die minimal-funktionale Lösung basierend auf verschiedenen Technologien entwickelt, implementiert und am Ende der Woche präsentiert. Der Innenraum des Trucks verfügt über die Möglichkeit, komplexe Inhalte durch den Einsatz von 360°-Szenarios, Touch-Wänden und erlebbaren Demo-Erfahrungen einfach und greifbar zu präsentieren. Die SAP-Experten bieten deshalb verschiedene Vorträge zu spannenden Themen im Truck selbst, aber auch für ein größeres Publikum im nahegelegenen Hörsaal ZHG 004 an. Die Vorträge sind gegliedert in „Executive Tour" und „Expert-Sessions" mit jeweils unterschiedlichen Zielgruppen. Die „Executive Tour" richtet sich an Entscheider, Leiter und Verantwortliche, während sich die Vorträge der „Expert-Sessions" vorwiegend an Projektleiter, Wissenschaftler, Forscher und Studierende richten.

Termin: 18. – 22. Februar 2019
Ort: Platz der Göttinger Sieben, zwischen Bibliothek und Zentralem Hörsaalgebäude, und Hörsaal ZHG 004, Zentrales Hörsaalgebäude
Anmeldung unter: *https://s-lotus.gwdg.de/gwdgdb/SAP_Innovation_Truck.nsf/anmeldung*
Weitere Informationen: *https://www.gwdg.de/events/sap-truck*
Ansprechpartner:
Tom Weckend (SAP; E-Mail: *tom.weckend@sap.com*) und
Dr. Sven Bingert (GWDG; E-Mail: *sven.bingert@gwdg.de*)

SAP Intelligent Enterprise Truck in Göttingen

In the period from February 18th to 22nd SAP will present its Intelligent Enterprise Truck (*https://events.sap.com/de/truck-tour-mee/de/home*) on the Central Campus of the University of Göttingen. During that time a team of SAP experts together with members of the GWDG, the University Medical Center Göttingen and the University of Göttingen develop a prototype for a specific use case. As the truck offers a 360° presentation area the event will be accompanied by diverse and exciting presentations.

PROGRAMM

Montag, 18.02.2019

EXECUTIVE TOUR – INTELLIGENT ENTERPRISE TRUCK		
9:00 - 9:30	Begrüßung, Eröffnung des Intelligent Enterprise Truck Events und Start in den Hackathon	Uni Göttingen, GWDG, SAP
9:30 - 10:00	Impulsvortrag: Transformation durch Innovation	Timo Deiner
10:15 - 11:15	Live-Demo: Die digitale Universität – Integration & Innovation Die Digitale SAP Hochschulplattform	André Sadaune, Joachim Wille, Tom Weckend
EXPERT SESSIONS IM TRUCK: FOKUSTAG DATENANALYSE		
13:00 - 13:45	Optimierung der studentischen Entwicklungspfade durch prädiktive Analysen der Leistungsdaten	Luise Hogrefe, Alexander Göppel
14:00 - 14:45	Drittmittel, Leistungsstatistiken oder andere Hochschulkennzahlen analysieren, planen und steuern mit dem SAP Digital Boardroom	Luise Hogrefe, Alexander Göppel
15:00 - 15:45	Maschinelles Lernen im Drittmittelantrag: Schneller und erfolgreicher Forschungsdrittmittel erhalten durch automatisierten Abgleich Ihrer Kompetenzfelder mit den Ausschreibungen	Alexander Göppel

Mittwoch, 20.02.2019

EXPERT SESSIONS IM TRUCK: FOKUSTAG APPLIKATIONEN		
9:00 - 9:45	**SAP Ticket Intelligence:** Tickets für alle Bereiche über alle Kanäle aufgeben und über den Status der Bearbeitung informiert sein. Machine-Learning-Technologie unterstützt in der Ticket-Bearbeitung, um effizient und schnell Lösungen bereitzustellen.	
10:00 - 10:45	Optimieren Sie die Verwaltung aller Gebäude und gewinnen Sie Erkenntnisse, die man für Planung, Verwaltung und beste Flächenauslastung benötigt, mit SAP Cloud for Real-Estate	

Dienstag, 19.02.2019

EXECUTIVE TOUR – INTELLIGENT ENTERPRISE TRUCK		
9:00 - 9:30	Impulsvortrag: Transformation durch Innovation	Timo Deiner
9:45 - 10:45	Live-Demo: Die digitale Universität – Integration & Innovation Die Digitale SAP Hochschulplattform	André Sadaune, Joachim Wille, Tom Weckend
VORTRÄGE IM HÖRSAAL (ZHG 004): DIGITALE PLATTFORM		
10:00 - 10:30	Impulsvortrag: Transformation durch Innovation	Timo Deiner
10:30 - 11:15	Sammeln und integrieren Sie Daten aus all Ihren Quellen in einer vertrauenswürdigen, einheitlichen Landschaft – der SAP Data Hub	Frederic Hopt
11:15 - 12:00	Beschleunigte Anwendungsentwicklung, mobile Services, Erweiterungen und Vernetzung mit leistungsstarken APIs, innovative Lösungen für Big Data, IoT und maschinelles Lernen	Konstantin Yakubovich
12:15 - 13:00	SAP Technologie (fast) kostenfrei nutzen für Forschung & Lehre: Das University Alliance Programm	Andre Biener
EXPERT SESSIONS IM TRUCK: FOKUSTAG DIGITALE PLATTFORM		
13:00 - 14:00	Upgrade der Digitalen Plattform: Prozesse in Echtzeit, integriert, mobil, erweiterbar – mit dem SAP Data Hub in einer sicheren zukunftsfähigen HANA In-Memory-Datenplattform	Frederic Hopt, Konstantin Yakubovich
14:00 - 15:00	Diskussionsrunde & Expertengespräch, technische Details: Diskutieren Sie mit uns zu Einsatz-Szenarios, Lösungsarchitekturen und Projekterfahrungen, Q&A	Frederic Hopt, Konstantin Yakubovich, Timo Deiner, Andre Biener, Joachim Wille, Tom Weckend

Kurz & knapp

Änderungen beim TSM/ISP-Betrieb

Abschalten der Registry-Sicherung

Beim Backup vereinzelter Windows-Rechner erfolgt die Sicherung des „SystemState“ (also der Windows-Registry) durch die Option „DOMAIN ALL-LOCAL“ in der Konfigurationsdatei. Die Registry umfasst zwar nicht viele Daten, hat aber relativ viele Einträge, die sich auch regelmäßig in größerem Umfang ändern. Daher führt die Sicherung des SystemState bzw. der Registry zu überdurchschnittlich vielen Einträgen in der TSM-internen Datenbank, so dass in der Vergangenheit für diese Daten bereits von der Standard-Policy „alle Versionen in 90 Tagen“ zu „maximal 3 Versionen“ abgewichen wurde.

Während das Backup der Registry relativ einfach ist, bringt der Restore einigen Aufwand mit sich, da in einem laufenden Windows-System die Registry geändert wird. Eine teilweise Wiederherstellung ist nicht möglich. Daher werden alle Einträge überschrieben. Letztendlich ist es damit fraglich, ob das System anschließend überhaupt stabil weiterläuft. In der Praxis beobachten wir daher **keine** Restores vom SystemState bzw. der Registry, auch haben bisher **alle** angefragten Kunden verneint, ihr Windows-System aus dem Backup wieder komplett herstellen zu wollen.

Wir haben uns daher entschieden, die Server so zu konfigurieren, dass die Registry bzw. der SystemState zukünftig garnicht mehr gesichert wird. Durch diesen Ausschluss beschleunigt sich auch das Backup der entsprechenden Knoten.

Erinnerung an das Update der Clients

Wie bereits mehrfach angesprochen, unter anderem auch in den GWDG-Nachrichten 7/2018, müssen die zu sichernden Rechner auf eine aktuelle Version des TSM/ISP-Clients aktualisiert werden. Bereits im April 2016 wurde ein Bug in Clients der Versionen 5 und 6 entdeckt, der im Zusammenspiel mit TSM/ISP-Servern der Verson 7.1.4 und neuer **kein Restore** über die GUI erlaubt. Für einzelne Client-Versionen gibt es zwar Workarounds, die aber auch nur unter bestimmten Voraussetzungen funktionieren (siehe *https://www-01.ibm.com/support/docview.wss?uid=swg1IT15117*). Da wir bei der GWDG nun seit etwa drei Jahren das Update einiger Server (SM131, SM231 und SM233) ausgesetzt haben, um diese Restore-Problematik zu umgehen, laufen die Server mittlerweile auf ziemlich alten Softwareständen, die auch nicht mehr mit Bugfixes versorgt werden (obwohl sie noch offiziell supportet sind, aber der Support empfiehlt bei Fehlern zunächst das Update auf eine Version, in der ein Fehler ja bereits gefixt ist).

Diesen Zustand können wir nicht mehr lange aufrechterhalten. Daher an dieser Stelle nochmal die dringende Bitte, die Clients auf eine aktuelle Version (vorzugsweise 7.1.6.5) zu heben. Die Versionen 7.1.8 und neuer setzen per Voreinstellung auf TLS. Dies kann aber zu Problemen führen, wenn wir die Server auf diese Versionen heben: Sofern nicht zusätzlich das Server-Zertifikat installiert wurde, kommt es zu keinem Verbindungsaufbau und das Backup schlägt fehl.

Wer bereits einen neueren Client im Einsatz hat, soll nun zwar keinen Downgrade durchführen, nur möchten wir mit der Empfehlung des „letzten Clienten ohne TLS-Voreinstellung“ die potenzielle Anzahl von Umstellungsproblemen zunächst minimieren.

Verschiebung von Servern

Aufgrund der stetig wachsenden Anzahl von TSM/ISP-Servern und zusätzlichen Tape-Libraries wurde die Zuordnung der Instanzen zu den physischen Servern im Hinblick auf Optimierungspotenzial geprüft. Im Sommer diesen Jahres sollen daher verschiedene TSM/ISP-Server verschoben werden, um die (interne) Konfiguration zu vereinfachen und die vorhandene Hardware besser auszulasten. Vermutlich startet dieses „Server-Schubsen“ zu Beginn des dritten Quartals. Aktuell können wir noch nicht absehen, ob wir sukzessive einzelne Server verschieben werden oder im Rahmen einer ein- bis zweitägigen Downtime alle gleichzeitig. Details dazu werden zeitnah über die bekannte GWDG-Mailing-Liste der TSM-Nutzer (*https://listserv.gwdg.de/mailman/listinfo/gwdg-tsm*) mitgeteilt.

Nachtwey

Neuer Kurs „InDesign – Fundamentals“

Kurzfristig wurde der Kurs „InDesign – Fundamentals“ in das GWDG-Kursprogramm 2019 aufgenommen. Der zweitägige Kurs basiert auf dem bekannten InDesign-Grundlagen-Kurs und wird in Englisch gehalten. Er findet am 06./07.05.2019 im Kursraum der GWDG statt. Nähere Informationen hierzu sind unter *https://www.gwdg.de/kursprogramm* zu finden.

Otto

GWDG-Stand beim GöBiT am 23. Februar 2019

Die GWDG nimmt in diesem Jahr am Göttinger Berufsinformationstag (GöBiT) teil, der am 23. Februar 2019 von 10:00 – 15:00 Uhr in der Lokhalle, Bahnhofsallee 1, stattfindet. Sie stellt dort auf dem Stand 62 ihre beiden Ausbildungsgänge **Fachinformatiker/in – Fachrichtung Anwendungsentwicklung** und **Fachinformatiker/in – Fachrichtung Systemintegration** vor, die auch als duales Studium kombiniert mit einem Bachelorstudium in Elektrotechnik/Informationstechnik absolviert werden können. Nähere Informationen zur Ausbildung bei der GWDG sind unter dem URL *https://www.gwdg.de/ausbildung* zu finden. Der GöBit präsentiert sich unter dem URL *https://www.goebit.de*. Wir würden uns über zahlreichen Besuch freuen.

Otto

Stellenangebot

Die GWDG sucht ab sofort zur Unterstützung der Arbeitsgruppe „Verwaltung und Querschnittsaufgaben“ (AG V) eine(n)

Bilanz- oder Finanzbuchhalter(in), IHK-geprüft (m/w/d)

mit einer regelmäßigen Wochenarbeitszeit von 39 Stunden. Die Stelle ist grundsätzlich auch für Teilzeitkräfte geeignet und im Rahmen einer Elternzeitvertretung zunächst auf zwei Jahre befristet. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist je nach Qualifikation bis zur Entgeltgruppe TVöD E 9 vorgesehen.

Aufgabenbereiche

- Prüfung, Kontierung und Buchung von Ein- und Ausgangsrechnungen
- Zahlungsverkehr
- Monatliche Umsatzsteuervoranmeldung und jährliche Umsatzsteuererklärung
- Anlagenbuchhaltung
- Weitere Verwaltungstätigkeiten bei Bedarf

Anforderungen

- Abgeschlossene Ausbildung als IHK-geprüfte(r) Bilanz- oder Finanzbuchhalter(in)
- Selbstständige Arbeitsweise und mindestens drei Jahre Berufserfahrung als Bilanz- oder Finanzbuchhalter(in)
- Erfahrung im Umgang mit Personal Computern und aktueller Office-Software
- Freundliches und kompetentes Auftreten
- Sehr gute Kommunikations- und Teamfähigkeit
- Gute Sprachkenntnisse in Wort und Schrift in Deutsch und Englisch

Die GWDG strebt nach Geschlechtergerechtigkeit und Vielfalt und begrüßt daher Bewerbungen jedes Hintergrunds. Die GWDG ist bemüht, mehr schwerbehinderte Menschen zu beschäftigen. Bewerbungen Schwerbehinderter sind ausdrücklich erwünscht.
Haben wir Ihr Interesse geweckt? Dann bitten wir um eine Bewerbung bis zum **22. März 2019** über unser Online-Formular unter *https://s-lotus.gwdg.de/gwdgdb/agv/20190208.nsf/bewerbung*.

Fragen zur ausgeschriebenen Stelle beantwortet Ihnen:

Herr Dr. Paul Suren
Tel.: 0551 201-1511
E-Mail: *paul.suren@gwdg.de*

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Personalia

DUALES STUDIUM ERFOLGREICH ABGESCHLOSSEN
INES LEWANDROWSKI

Am 21. Januar 2019 hat Frau Ines Lewandrowski ihr duales Studium an der HAWK Göttingen mit dem akademischen Grad Bachelor of Engineering (B.Eng.) in Elektrotechnik/Informationstechnik erfolgreich abgeschlossen. Thema ihrer Bachelorarbeit war die Konzeption und prototypische Implementierung des Event-Sourcing-Musters am Beispiel des Event Calendars des Göttingen Campus. Den zum dualen Studium gehörenden Ausbildungsteil zur Fachinformatikerin mit Schwerpunkt Anwendungsentwicklung hatte Frau Lewandrowski bereits im Juni 2016 erfolgreich beendet und sich seitdem auf das im Praxisverbund von HAWK und GWDG durchgeführte Studium konzentriert. Mit dem Abschluss folgt zunächst eine sechsmonatige Weiterbeschäftigung, die im Bereich der Softwareentwicklung angesiedelt ist. Wir gratulieren Frau Lewandrowski sehr herzlich zum erfolgreichen Abschluss ihres Studiums und wünschen ihr viel Erfolg in ihrem neuen Aufgabenbereich bei der GWDG.

Pohl

AUSBILDUNG ERFOLGREICH ABGESCHLOSSEN
TOBIAS HEISE UND DENNIS RHODE

Herr Tobias Heise hat am 23. Januar 2019 seine auf zweieinhalb Jahre verkürzte Ausbildung zum Fachinformatiker in der Anwendungsentwicklung erfolgreich abgeschlossen. In seinem Abschlussprojekt entwickelte Herr Heise eine Lösung zur Standortermittlung von IP-Adressen für den Account-Selfservice des GWDG-Kundenportals. Dies adressiert aktuelle Herausforderungen aus den Bereichen der DSGVO (personenbeziehbare IP-Adressen müssen nicht mehr an externe Betreiber von IT-Diensten übermittelt werden) sowie der IT-Sicherheit (Erkennung von möglicherweise kompromittierten Accounts durch Login-Versuche aus unüblichen Geolokationen). Herr Heise wird bei der GWDG zunächst für sechs weitere Monate im Bereich der Kundenportal-Entwicklung weiterbeschäftigt. Wir gratulieren Herrn Heise ganz herzlich zum erfolgreichen Abschluss seiner Ausbildung und wünschen ihm einen gelungenen Start in seine neue Tätigkeit bei der GWDG.

Pohl

Herr Dennis Rhode hat am 25. Januar 2019 seine Abschlussprüfung zum Elektroniker für Geräte und Systeme bestanden und damit seine 3,5-jährige Ausbildung bei der GWDG erfolgreich beendet. Im Anschluss an seine Ausbildung wird Herr Rhode bei der GWDG weiterbeschäftigt. Er wird sich vor allem mit der Installation von Hardware im Maschinenraum und der Dokumentation der vorhandenen Verkabelung befassen. Wir gratulieren Herrn Rhode ganz herzlich zum erfolgreichen Abschluss seiner Ausbildung und wünschen ihm einen guten Start in seine neue Tätigkeit als Facharbeiter bei der GWDG.

Gutsch

INFORMATIONEN:
support@gwdg.de
0551 201-1523

Februar bis Dezember 2019

Kurse

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
GRUNDLAGEN DER BILDBEARBEITUNG MIT PHOTOSHOP	Töpfer	05.02. – 06.02.2019 9:30 – 16:00 Uhr	29.01.2019	8
STATISTIK MIT R FÜR TEILNEHMER MIT VORKENNTNISSEN – VON DER ANALYSE ZUM BERICHT	Cordes	20.02. – 21.02.2019 9:00 – 12:00 und 13:00 – 15:30 Uhr	13.02.2019	8
SHAREPOINT – EINFÜHRUNG FÜR ANWENDER	Buck, Kasper	27.02.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	20.02.2019	4
SHAREPOINT – EINFÜHRUNG IN DIE VERWALTUNG VON SITECOLLECTIONS	Buck, Kasper	28.02.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	21.02.2019	4
INDESIGN – GRUNDLAGEN	Töpfer	12.03. – 13.03.2019 9:30 – 16:00 Uhr	05.03.2019	8
PHOTOSHOP FÜR FORTGESCHRITTENE	Töpfer	26.03. – 27.03.2019 9:30 – 16:00 Uhr	19.03.2019	8
EINFÜHRUNG IN DIE STATISTISCHE DATENANALYSE MIT SPSS	Cordes	03.04. – 04.04.2019 9:00 – 12:00 und 13:00 – 15:30 Uhr	27.03.2019	8
USING THE GWDG SCIENTIFIC COMPUTE CLUSTER – AN INTRODUCTION	N.N.	08.04.2019 9:30 – 16:00 Uhr	01.04.2019	4
PARALLELRECHNERPROGRAMMIERUNG MIT MPI	Prof. Haan	09.04. – 10.04.2019 9:15 – 17:00 Uhr	02.04.2019	8
INDESIGN – FUNDAMENTALS	Töpfer	06.05. – 07.05.2019 9:30 – 16:00 Uhr	29.04.2019	8

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
PROGRAMMING WITH CUDA – AN INTRODUCTION	Prof. Haan	07.05.2019 9:15 – 17:00 Uhr	30.04.2019	4
ADMINISTRATION VON PCS IM ACTIVE DIRECTORY DER GWDG	Quentin	09.05.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	02.05.2019	4
QUICKSTARTING R: EINE ANWENDUNGSORIENTIERTE EINFÜHRUNG IN DAS STATISTIKPAKET R	Cordes	15.05. – 16.05.2019 9:00 – 12:00 und 13:00 – 15:30 Uhr	08.05.2019	8
SHAREPOINT – EINFÜHRUNG FÜR ANWENDER	Buck, Kasper	22.05.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	15.05.2019	4
SHAREPOINT – EINFÜHRUNG IN DIE VERWALTUNG VON SITECOLLECTIONS	Buck, Kasper	23.05.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	16.05.2019	4
INDESIGN – AUFBAUKURS	Töpfer	04.06. – 05.06.2019 9:30 – 16:00 Uhr	28.05.2019	8
OUTLOOK – E-MAIL UND GROUPWARE	Helmvoigt	13.06.2019 9:15 – 12:00 und 13:00 – 16:00 Uhr	06.06.2019	4
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER MIT VORKENNTNISSEN	Cordes	09.06. – 20.06.2019 9:00 – 12:00 und 13:00 – 15:30 Uhr	02.06.2019	8
STATISTIK MIT R FÜR TEILNEHMER MIT VORKENNTNISSEN – VON DER ANALYSE ZUM BERICHT	Cordes	02.07. – 03.07.2019 9:00 – 12:00 und 13:00 – 15:30 Uhr	25.06.2019	8
INDESIGN – GRUNDLAGEN	Töpfer	03.09. – 04.09.2019 9:30 – 16:00 Uhr	27.08.2019	8
SHAREPOINT – EINFÜHRUNG FÜR ANWENDER	Buck, Kasper	11.09.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	04.09.2019	4
SHAREPOINT – EINFÜHRUNG IN DIE VERWALTUNG VON SITECOLLECTIONS	Buck, Kasper	12.09.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	05.09.2019	4
AFFINITY PHOTO – GRUNDKURS	Töpfer	24.09. – 25.09.2019 9:30 – 16:00 Uhr	17.09.2019	8
ADMINISTRATION VON PCS IM ACTIVE DIRECTORY DER GWDG	Quentin	24.10.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	17.10.2019	4
INDESIGN – AUFBAUKURS	Töpfer	05.11. – 06.11.2019 9:30 – 16:00 Uhr	29.10.2019	8
EINFÜHRUNG IN DIE STATISTISCHE DATENANALYSE MIT SPSS	Cordes	13.11. – 14.11.2019 9:00 – 12:00 und 13:00 – 15:30 Uhr	06.11.2019	8
SHAREPOINT – EINFÜHRUNG FÜR ANWENDER	Buck, Kasper	20.11.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	13.11.2019	4
SHAREPOINT – EINFÜHRUNG IN DIE VERWALTUNG VON SITECOLLECTIONS	Buck, Kasper	21.11.2019 9:00 – 12:30 und 13:30 – 15:30 Uhr	14.11.2019	4

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
AFFINITY PHOTO – AUFBAU-KURS	Töpfer	26.11. – 27.11.2019 9:30 – 16:00 Uhr	19.11.2019	8
OUTLOOK – E-MAIL UND GROUPWARE	Helmvoigt	05.12.2019 9:15 – 12:00 und 13:00 – 16:00 Uhr	28.11.2019	4
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER MIT VOR-KENNTNISSEN	Cordes	11.12. – 12.12.2019 9:00 – 12:00 und 13:00 – 15:30 Uhr	04.12.2019	8

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 *https://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 *https://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 *https://www.gwdg.de/kursprogramm* zu finden.


GWDG
Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen