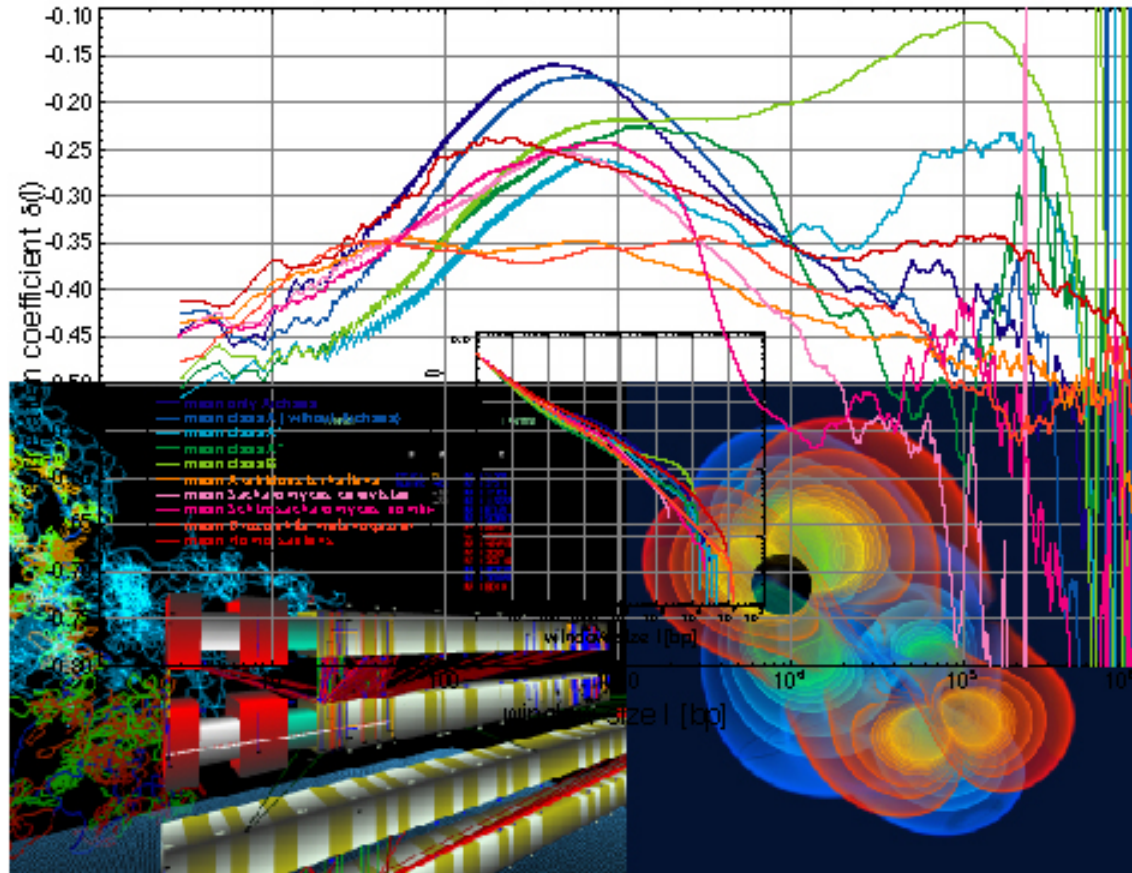


Produktions-Grid für Genom- und LISA-Forschung



Einstein@Home & LISA
Alexander Beck-Ratzka (AEI)

System Biology & Genom Analysen
Tobias A. Knoch (BioQuant/DKFZ)

Agenda

- Einstein@Home im Grid
 - Voraussetzungen und Deployment
 - Ablauf und Fehlermonitoring
 - Statistiken
- System Biologie Genom Analysen im Grid
 - Project/Job/Applications Typen
 - Sicherheit
 - Statistiken
- Auswirkungen des Produktionsbetriebs auf das D-Grid
- Produktionsbetrieb für Genom- und LISA-Grid

Agenda

- Einstein@Home im Grid
 - Voraussetzungen und Deployment
 - Ablauf und Fehlermonitoring
 - Statistiken
- System Biologie & Genom Analysen im Grid
 - Project/Job/Applikations Typen
 - Sicherheit
 - Statistiken
- Auswirkungen des Produktions-betriebs auf das D-Grid
- Produktionsbetrieb für Genom- und LISA-Grid

Einstein@Home

- Für Einstein@Home sind Executables für nahezu alle Architekturen verfügbar.
- Einstein@Home basiert auf BOINC-Framework.
 - Zu analysierender Datensatz wird von Einstein@Home-Server geholt, Ergebnisse werden zurück übertragen.
- Vor Ausführung auf execution Host muss Einstein@Home Client (BOINC-Client) installiert werden

e@h – Deployment (1)

- Deployment ist weitestgehend automatisiert.
- Zum Laufen der e@h im Grid wird benötigt:
 - Das GEO600-Framework (Perl)
 - Ein BOINC-Client für e@h
 - Ein MySQL-Client (Statistik)
 - Entsprechende Perl-Module.

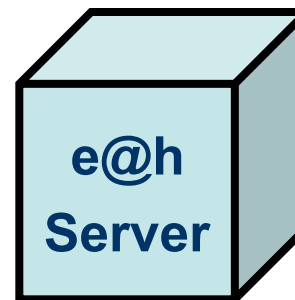
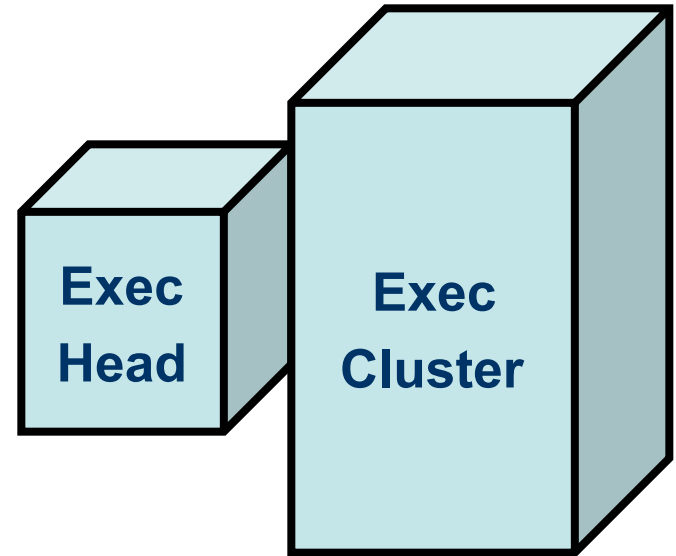
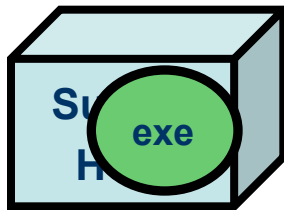
e@h – Deployment (2)

- Deployment wird über globusrun-ws gestartet.
- Manche Ressourcen erfordern unterschiedliche Softwareversionen auf Head- und Workernode -> Deployment mit Ft Fork und Ft PBS (SGE) erforderlich.

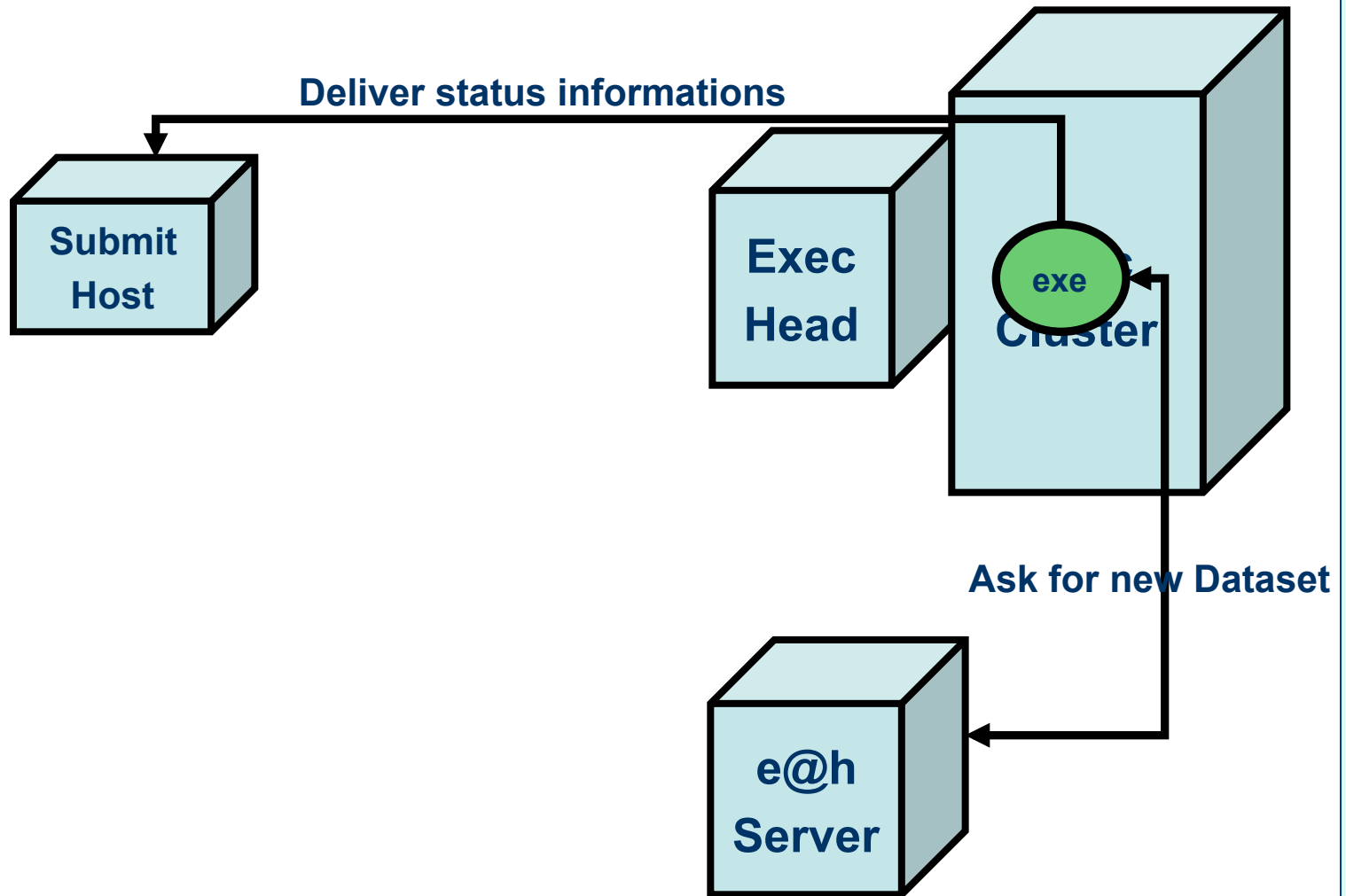
e@h – Ablauf

- Jeweils ein Einstein@Home wird auf worker nodes gestartet.
- Zur Kennung eines Einstein@Home Jobs gibt es eine eindeutige Task-ID.
- Einstein@Home Client fordert Datensatz von Einstein@Home Server an.
- Analysenergebnisse werden an Server zurück übertragen.
- Beim Erreichen des Walltime-Limits werden Checkpoints auf Execution Host gespeichert; Startpunkt für nächsten Job.

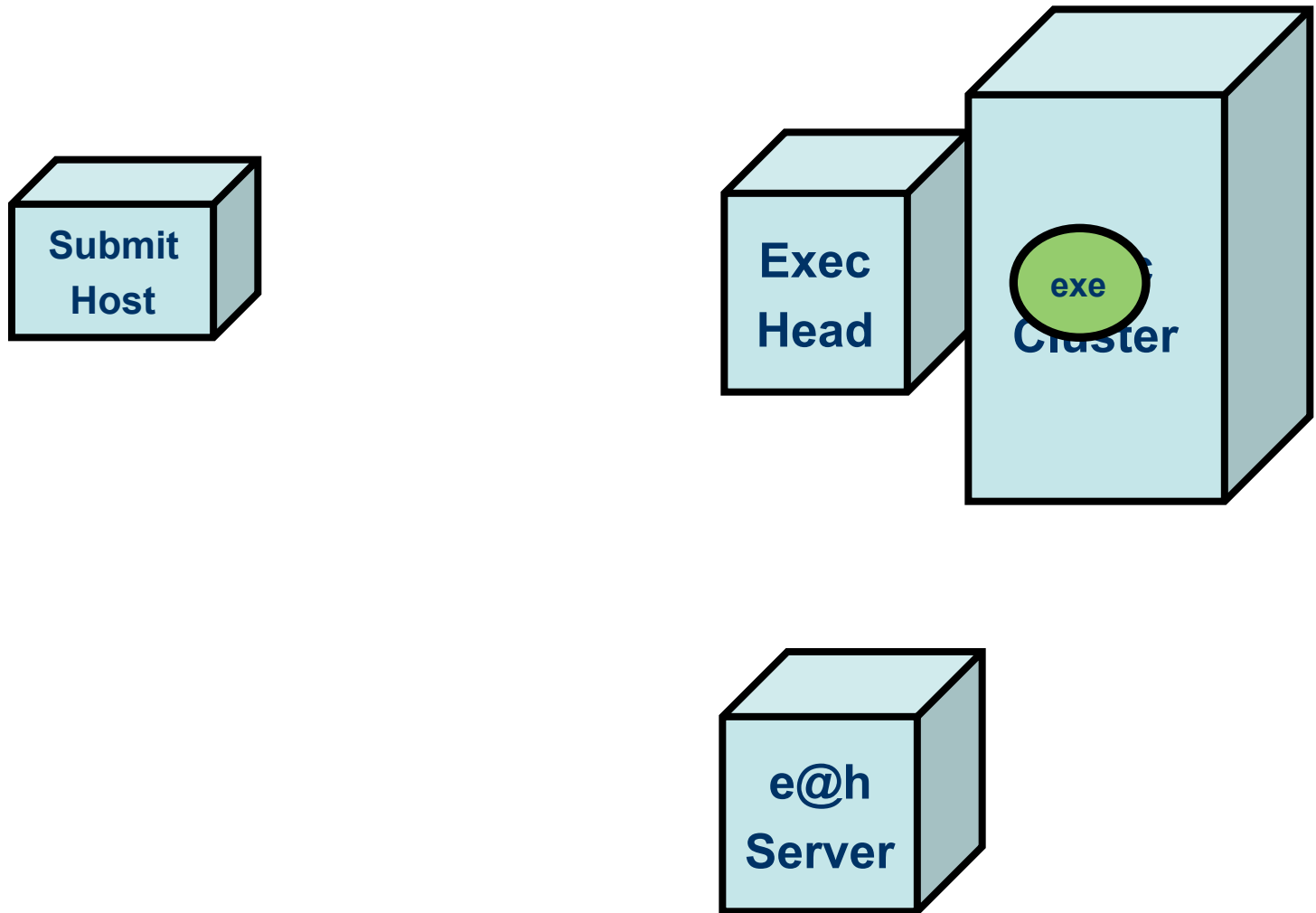
e@h Ablauf



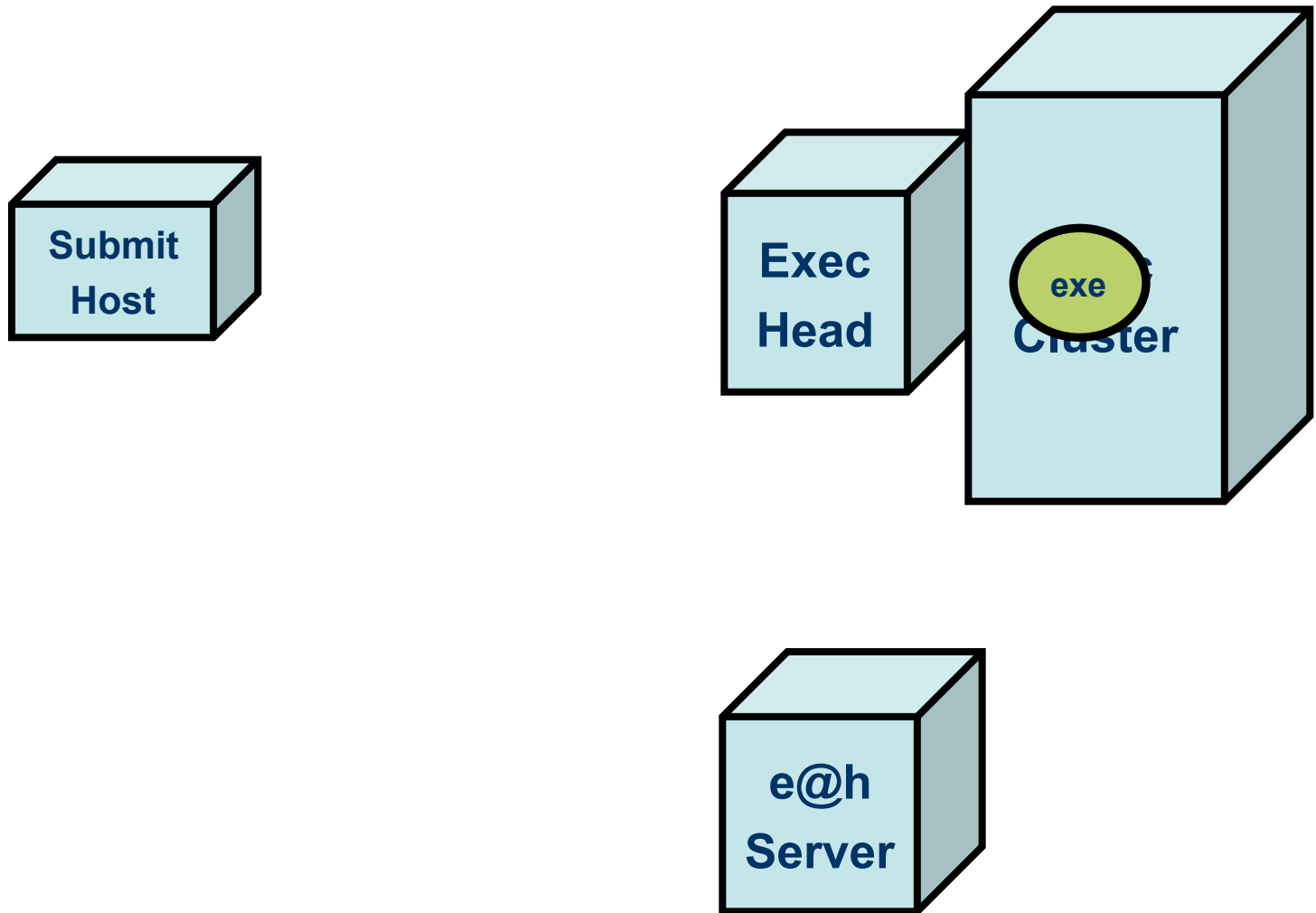
e@h Ablauf



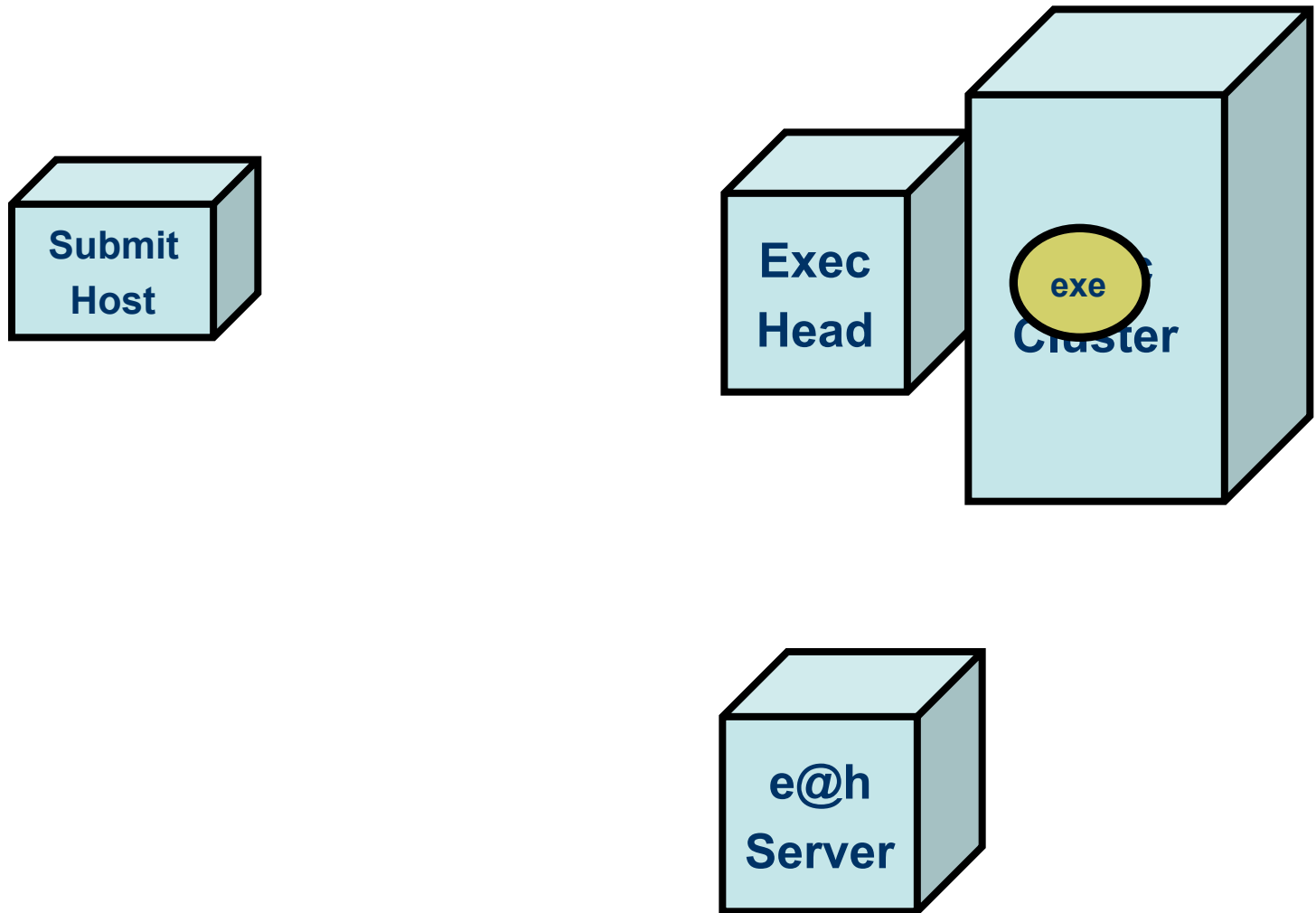
e@h Ablauf



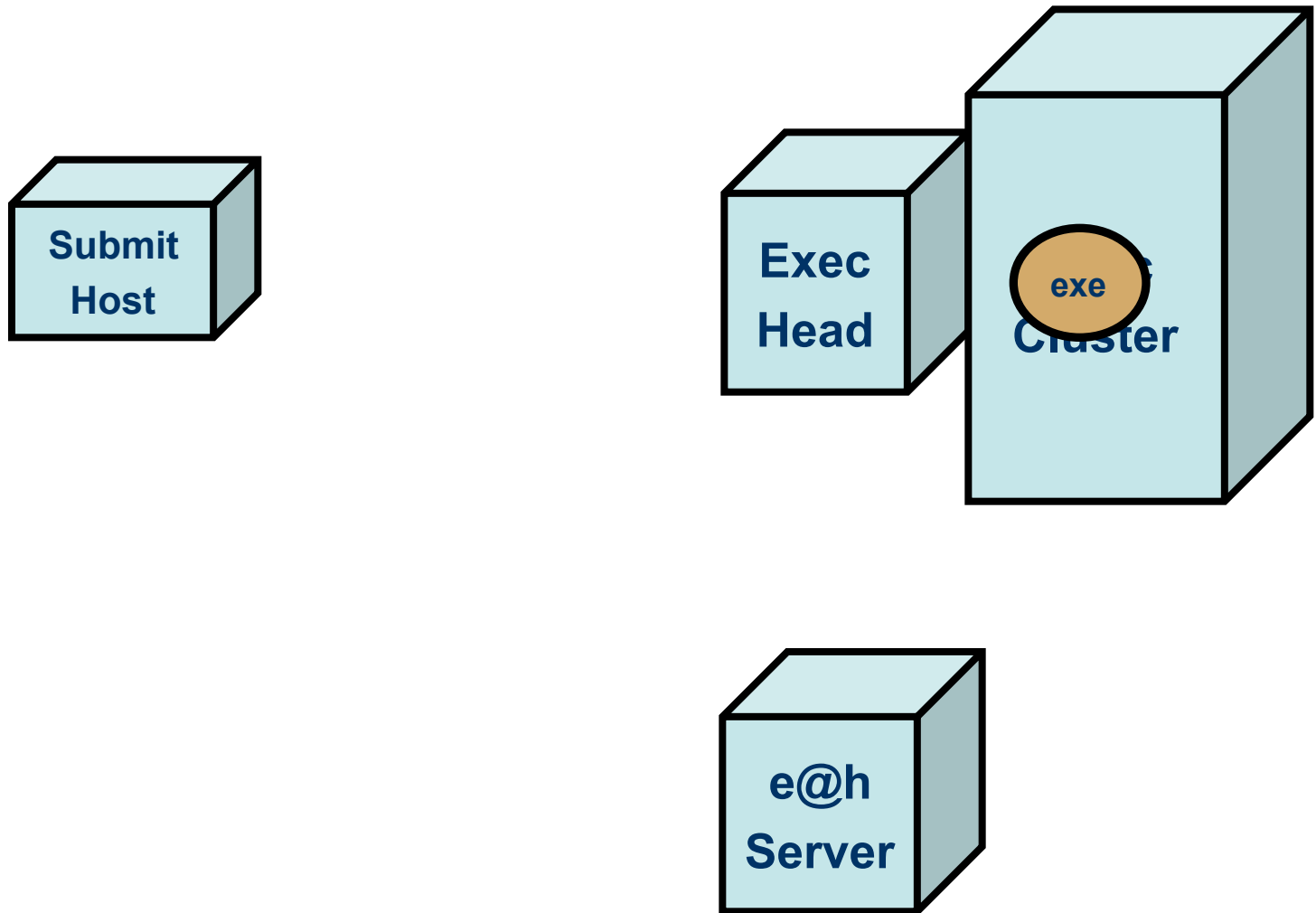
e@h Ablauf



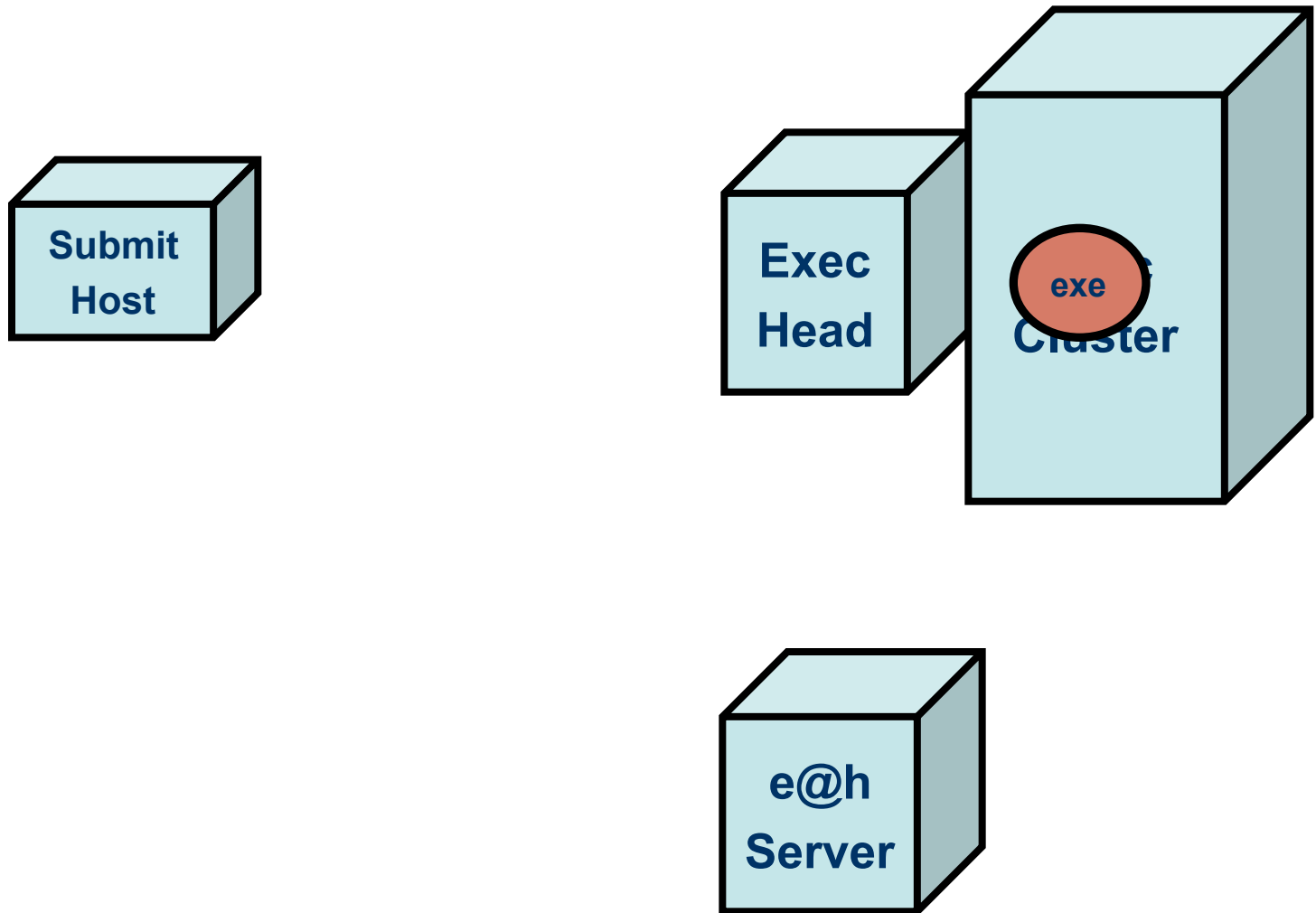
e@h Ablauf



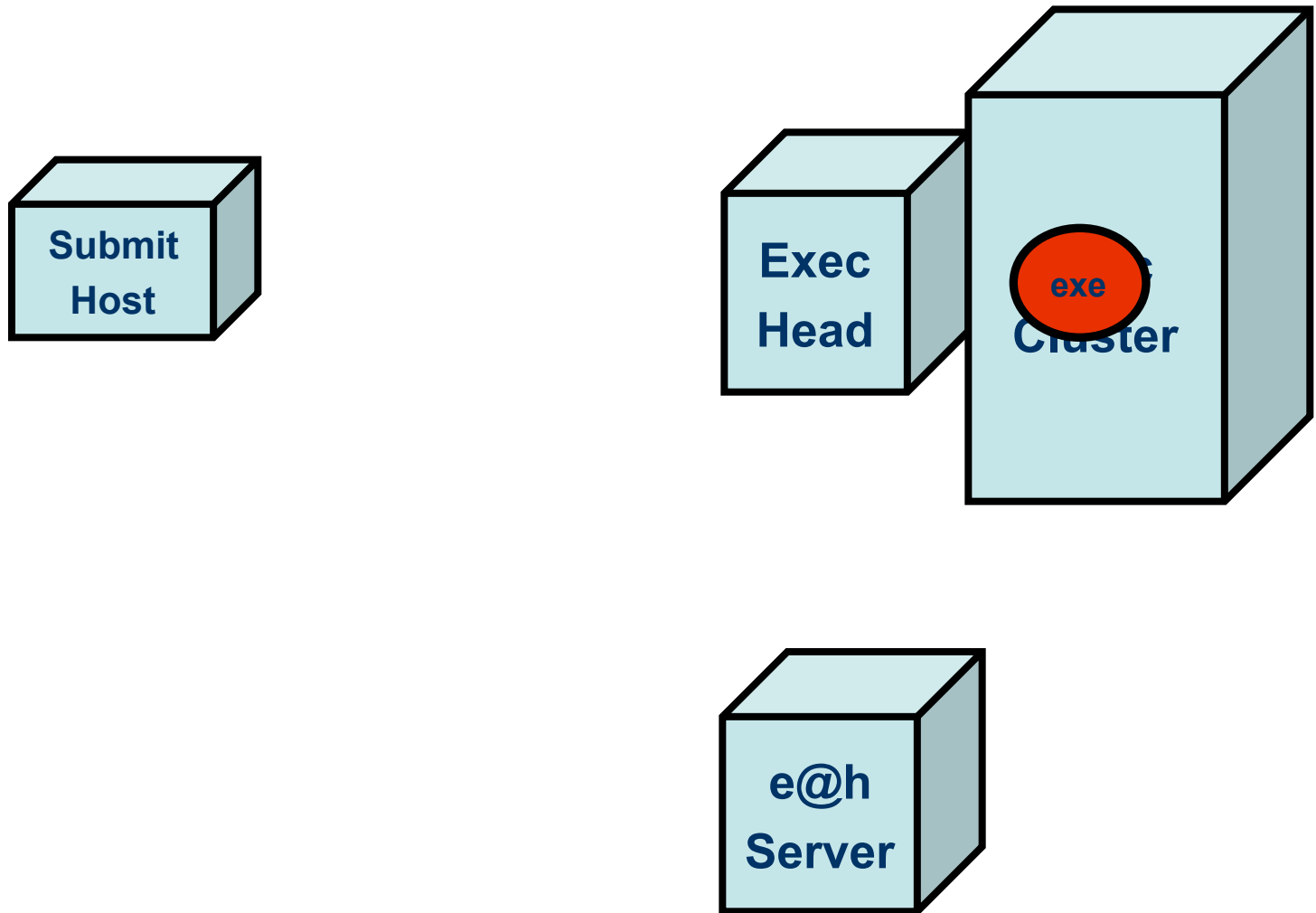
e@h Ablauf



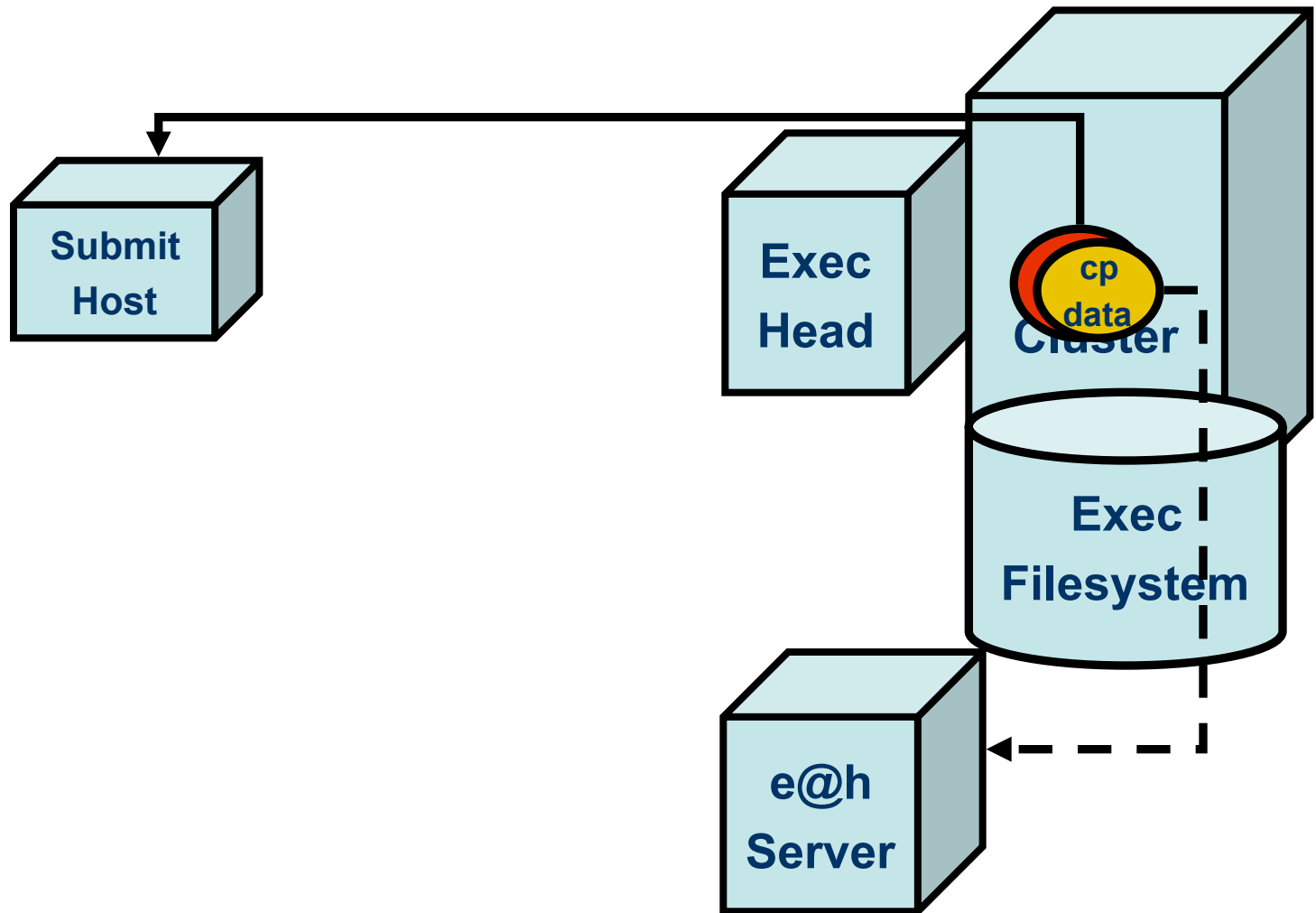
e@h Ablauf



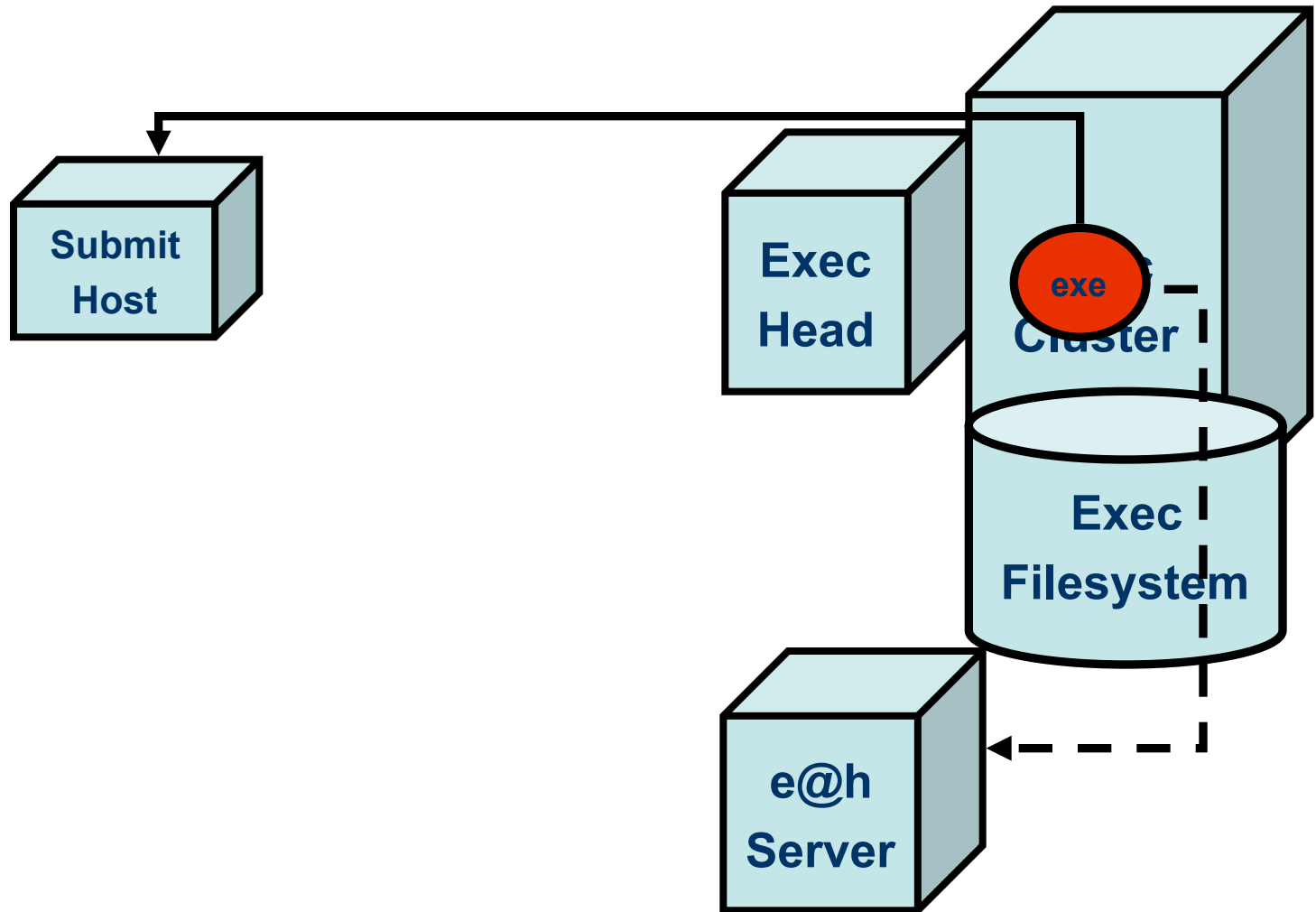
e@h Ablauf



e@h Ablauf



e@h Ablauf



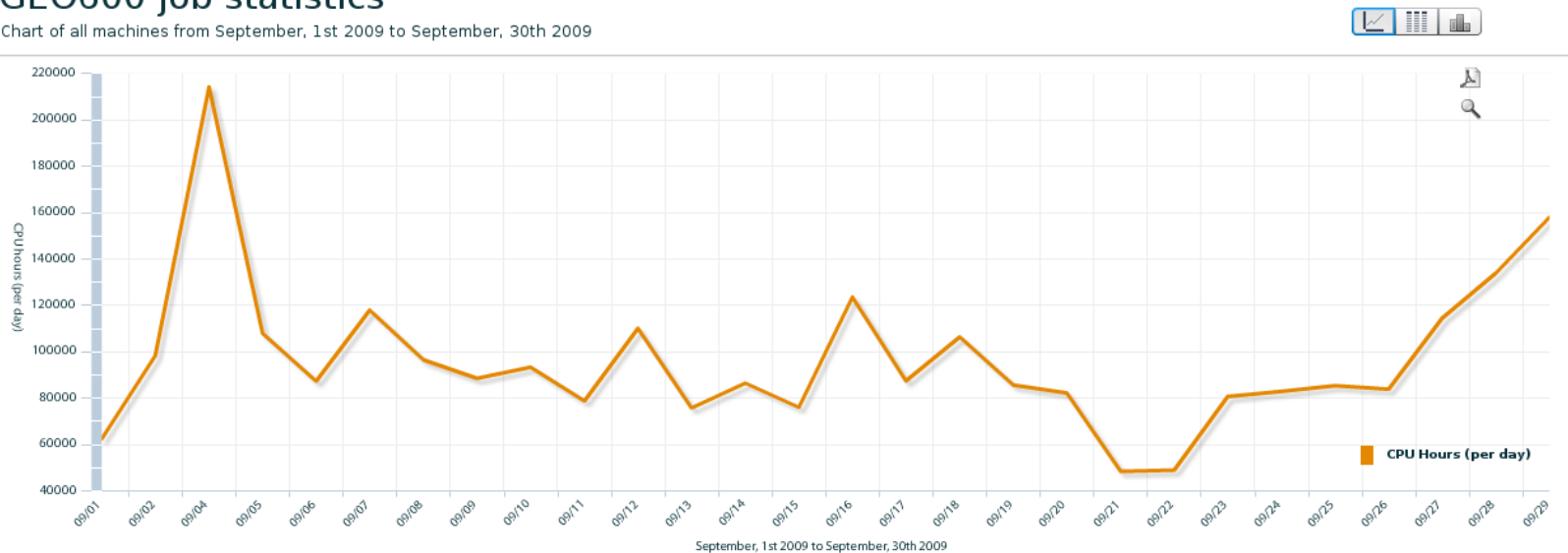
- Parallel zu den Start-Skripten werden “Inspect”-Skripte durchlaufen.
- Überprüfung der Status aller Einstein@Home Tasks auf einem Gridhost.
- Bei behebbaren Fehlern werden bestimmte Aktionen durchgeführt, so dass neue Jobs starten können.

- Regelmäßiger Neustart von e@h-Jobs und Inspect-Jobs via cron.
- Konfigurationsdateien legen Walltime, Anzahl der Jobs auf einer Ressource, Speicherort Checkpoints fest.
- Statistikportal ermöglicht schnellen Überblick, ob größere Probleme aufgetreten sind; Details über Emails mit Output der Inspect-Jobs (cron-Feature).

Rechenzeit GEO600 September 2009

GEO600 job statistics

Chart of all machines from September, 1st 2009 to September, 30th 2009



Rechenzeit GEO600 September 2009

GEO600

Chart of

CPU hours (per day)

**Einstein@Home:
Die erste von zwei
Erfolgsgeschichten
im D-Grid!**

09/20

Agenda

- Einstein@Home im Grid
 - Voraussetzungen und Deployment
 - Ablauf und Fehlermonitoring
 - Statistiken
- System Biologie Genom Analysen im Grid
 - Project/Job/Applikations Typen
 - Sicherheit
 - Statistiken
- Auswirkungen des Produktionsbetriebs auf das D-Grid
- Produktionsbetrieb für Genom- und LISA-Grid

- Projekt Typen:

- Online Visualisierung
- Pipelines im (Near-)Online Bereich
- Backfilling Super-Large andauerndes Data-Mining
- Large-scale parallele Analyse und Simulation

<= alles über mehrere Gridinfrastrukturen hinweg =>

- Performance Requirements:

- Instantan höchste Leistung.
- Höchste Zuverlässigkeit der Abläufe.
- Höchste Trackability, Stabilität und Fehlermanagement.
- Höchste instantane Verfügbarkeit und Zuverlässigkeit.

Extreme Bedingungen unter Real Life Conditions !

- Sicherheitsarchitekturen:

- Datentransfer immer verschlüsselt und komprimiert.
- Produktionsumgebung immer abgesichert.
- Im Fehlerfall immer Rückstandsfrei.
- Controlling und Fehlermanagement immer notwendig.
- Grid-Prozesse, Accounting, Billing immer „TOP SECRET“ !

<= Erfolg durch nachhaltige Produktion =>

- Statistiken:

- Weltweit ca. ~100 Nutzer (Korrelationen, Vergleiche, Association Studies, Struktursimulationen).
- Gridübergreifend >300000 CPUh per **DAY**.
- Pro Jahr > 20 Publikationen.
- Wissenschaftlicher & Industrieller Kooperationsbedarf wächst stark

+++ ... +++ ... +++ ... +++ Rauch et al., *Differentiation* 76(1), 66-82, 2008 +++ Jhunjhunwala et al., *Cell* 133(2), 265-279, 2008
+++ Knoch et al., *Stud. Health. Technol. Inform.* 147, 105-116, 2009 +++ Knoch et al., *Stud. Health. Technol. Inform.* 147, 269-276, 2009 +++ Knoch et al., *Eur. Biophys. J.* 38(6), 757-779, 2009 +++ Estrada et al., *Bioinformatics* 25(20), 2750-2752, 2009 +++ de Wit et al., *Mol. Cell. Biol.* 30(1), 2009 +++ Kepper et al., *Stud. Health. Technol. Inform.* 159, 171-180, 2010 +++ Kepper et al., *Stud. Health. Technol. Inform.* 159, 264-271, 2010 +++ ... +++ ... +++

- Sicherheitsarchitekturen:
 - Datentransfer immer verschlüsselt und komprimiert.
 - Produktionsumgebung immer abgesichert.
 - Im Fehlerfall immer Rückstandsfrei.

Controlling und Fehlermanagement immer notwendig

Genom & System Biologie:
Die zweite von zwei
Erfolgsgeschichten
im D-Grid!

26
2750-2752, 2009 +++ de Wit et al., *Mol. Cell. Biol.* 30(1), 2009 +++ Kepper et al., *Stud. Health. Technol. Inform.* 159, 171-180, 2010 +++ Kepper et al., *Stud. Health. Technol. Inform.* 159, 264-271, 2010 +++ ... +++ ... +++

9, 2008

form. 147,

atics 25(20),



Entdeckte D-Grid Probleme (1)

- Automatisches Löschen von gLite (Sammelaccounts) können Nutzerdaten von Globus-Usern löschen
- Queue Overflow: Queues vertragen hohe Jobanzahl nicht
- Verwendung zentraler Speicherressourcen ist ein Performancekiller; hier besteht Nachbesserungsbedarf



Entdeckte D-Grid Probleme (1)

- Automatisches (Sammelaccount von Globus-Usern)
 - Limits einstellbarer Jobs zwischen 1000 und 2000. Bei gLite nur 20000 bis 30000 pro Tag
- Queue Overflow: Queues vertragen hohe Jobanzahl nicht
- Verwendung zentraler Speicherressourcen ist ein Performancekiller; hier besteht Nachbesserungsbedarf

- Dokumentation der D-Grid-Ressourcen ist nicht einheitlich; man erhält nicht immer die gewünschten Informationen
- Die Behebung von Fehlern im D-Grid teilweise zu langsam.
- Übergeordnete Sicherheit und Monitoring fehlt D-Grid-weit oder ist unzureichend

Agenda

- Einstein@Home im Grid
 - Voraussetzungen und Deployment
 - Ablauf und Fehlermonitoring
 - Statistiken
- System Biologie Genom Analysen im Grid
 - Project/Job/Applikations Typen
 - Sicherheit
 - Statistiken
- Auswirkungen des Produktionsbetriebs auf das D-Grid
- Produktionsbetrieb für Genom- und LISA-Grid

- Wir haben entscheidend zur Erhöhung der Zuverlässigkeit der Globus-Ressourcen im D-Grid beigetragen
- Die Erfahrungen von Einstein@Home & den System Biologie bzw. Genom Analysen sind in das D-Grid-Betriebskonzept eingeflossen.
- Wir stärken das Vertrauen ins Grid, weil wir es erfolgreich nutzen, und sind dadurch Ansprechpartner für Interessenten geworden.

- Portierung der Job-Kontrolle für einstein@home und genom auf andere Grid-Middlewares (gLite, Unicore und Condor)
 - Abwicklung der Job-Kontrolle über das GAT / SAGA
 - Erstellung einer generischen Job-Kontrolle für Taskfarming-Applikationen
- Weitergabe unserer Erfahrung an neue Grid-Nutzer

**BA für Arbeit stellt uns
evt. idlende
Ressourcen zur
Verfügung**

- Port und **Einstein@home**
und **lewares (gLite,**
Unio
– Abwicklung der Job-Kontrolle über das GAT / SAGA
– Erstellung einer generischen Job-Kontrolle für
Taskfarming-Applikationen
- Weitergabe unserer Erfahrung an neue Grid-
Nutzer

Agenda

- Einstein@Home im Grid
 - Voraussetzungen und Deployment
 - Ablauf und Fehlermonitoring
 - Statistiken
- System Biologie Genom Analysen im Grid
 - Project/Job/Applikations Typen
 - Sicherheit
 - Statistiken
- Auswirkungen des Produktionsbetriebs auf das D-Grid
- Produktionsbetrieb für Genom- und LISA-Grid

- Erste Ausbaustufe: Zugang für System Biologie bzw. Genom- und LISA-Forschungsgruppen
- Portal als „Entry Point“ für Forscher/innen
- Modularer Ansatz, um später auch anderen Forschungsgruppen Einrichtung eines virtuellen Forschungsnetz zu ermöglichen
- Zusammenführung von CPUs und Speicherressourcen in ein Grid (Nutzung D-Grid-Technologie)

Virtuelles Forschungsnetz: Angebot spezieller Dienste über Portlets

- Portlet zur generischen Submission und Monitoring (nicht) nur von Taskfarming-Jobs
 - Anschluss an das Grid über GATLET (www.gatlet.de)
 - Auswahl der Ressource(n), Executable, Anzahl in Queue, Anzahl im Executing, Speicherort Checkpoints, etc...
- Portlet zum Suchen und Anschauen von Daten
- Portlets als virtuelle Informationsbörsen für LISA- und Genomforschungs-Informationen
- Portlet zur Generierung von Grid-Workflows
- Portlets zum Streaming von Videos
- Portlet für Video-Conferencing und virtuelles White Board (Nutzung der DFN-Dienste)

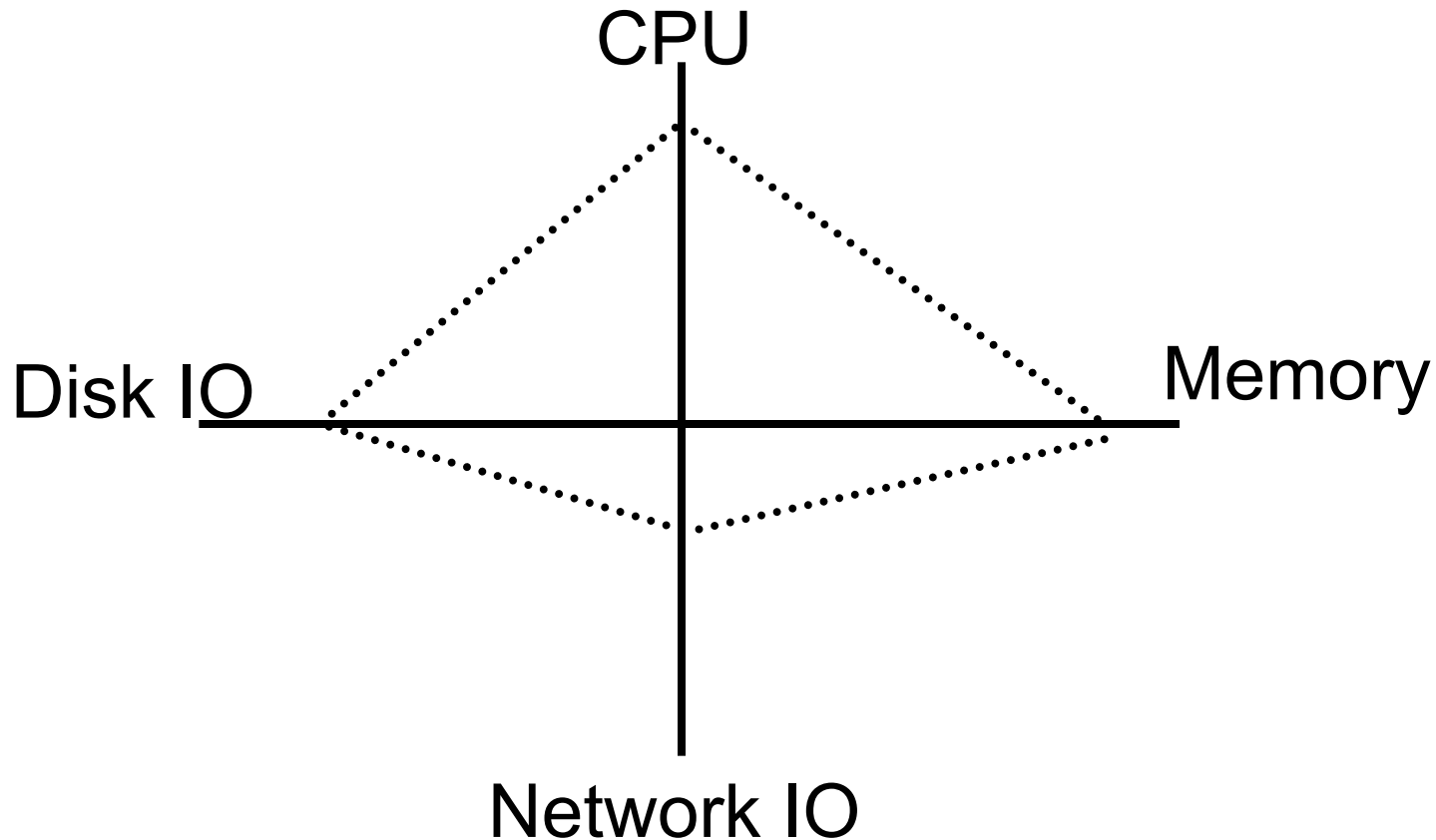
**Aus unserer Sicht
„Nice to have“**

**Aus unserer Sicht
„Nice to have”**

(aber in diesem Rahmen nicht zu leisten)

- Anwender soll für seine Applikation Benchmarking-Informationen erhalten
 - Welche Ressource ist für die Applikation am besten geeignet?
 - Informationen zur CPU-Ausnutzung (inkl. Speed), Memory, Network-IO, Disk-IO
- Entwicklung erfordert guten Kontakt zu den Grid-Admins; wir haben diesen.

Benchmarking-Graphen



- Umstellung auf andere Middlewares bringt Produktionsbetrieb für diese Ressourcen
- Wir zeigen, dass auch ein Unicore- und gLite-Grid stabil sein kann, und schaffen damit Vertrauen für das Grid
 - Nur ein stabiles Grid schafft einen realen Mehrwert auch über das D-Grid hinaus und weckt das Interesse der Industrie (Nachhaltigkeit)
- Ein generisches Portlet (nicht) für Taskfarming-Applikationen kann auch von anderen Nutzern verwendet werden

- Unterstützung neuer Nutzer bei „Gridifizierung“ ihrer Anwendungen
 - Wir stellen Kontakt zu Grid-Admins her; Grid-Admins kennen uns schon
 - Grid-Beratung basierend auf unseren umfangreichen Erfahrungen
 - Zur Verfügungstellung des halbautomatischen Deploymentmechanismus der Einstein@Home-Jobs

- 16 EraSysBio + Konsortien
 - 80 Arbeitsgruppen weltweit
 - ca. 1000 beteiligte Wissenschaftler
- LISA-Konsortium in Deutschland und Europa.
 - 14 Arbeitsgruppen europaweit
 - ca. 50 beteiligte Wissenschaftler

„Grid-Sekretariat“

- Anlaufstelle für neue Grid-Interessenten
- Muss kompetent besetzt sein
- Zwei „Standorte“:
 - Heidelberg,
 - Berlin/Potsdam
- Unterstützung für Grid-Nutzung

- Aktiv Informationen an potenzielle Interessenten weitergeben
- Die TMF als übergeordnete Anlaufstelle neben der D-Grid-GmbH etablieren
- TMF bietet die Möglichkeit aktiv auf Startups und Forschungsinstitute zuzugehen
- Erstellung einer Bibliothek von Fallbeispielen und Entwurf möglicher Szenarien
- “Forum Grid“-Veranstaltung des TMF erweitern
- Wenn eine Arbeitsgruppe sich für Grid-Computing entscheidet, TMF als Anlaufstelle

Warum schon erfolgreiche UseCases ins WissGrid?

- Unsere Erfahrungen kommen WissGrid direkt zu Gute
- Wir haben bewiesen, dass es geht. Man vertraut uns, und damit WissGrid, wenn es ums Grid geht
- Leuchtturm-Projekte sind immer gut für die Außenwirkung eines Projektes

Budget / Laufzeit

	2010 (k€)	2011 (k€)	2012 (k€)
Berlin/ Potsdam	55	56	57
Heidelberg	55	56	57
TMF	45	46	47
Gesamt	155	158	161

Budget / Laufzeit

	2010 k€	2011 k€	2012 k€	2013 k€	2014 k€
Berlin/ Potsdam	55	56	57	58	59
Heidel- berg	55	<u>Gesamtkosten</u> 805 k€			59
TMF	45				49
Gesamt	155	158	161	164	167

A Production-Grid for Genome and LISA Research

Knoch, T. A. & Beck-Ratzka, A.

TMF Forum Grid, Geschäftsstelle TMF e.V., Berlin, Germany, 13th July, 2010.

Abstract

Die D-Grid Initiative stellt Bundesweit derzeit >30.000 Rechencorens insgesamt 36 Community-Projekten zur Verfügung. Insgesamt wurden bisher in 3 „Calls“ und mehreren Infrastruktursonderinvestitionen ca. 134 Millionen Euro investiert. Auf den Ressourcen laufen mehrere sogenannte Middlewares. Die Auslastung der Ressourcen die mithilfe der Middleware Globus benutzt werden (~70% der total installierten Leistung) beträgt praktisch mittlerweile 100%. Im wesentlichen nutzen zwei Usecases diese Ressourcen zu ca. 98%: Im astrophysikalischen Bereich sind dies Gravitationswellenanalysen mittels der einstein@home-Jobs vom Max-Planck-Institut für Gravitationsphysik Albert-Einstein-Institut (AEI) und sowie Genomanalysen der Biophysikalischen Genomik, BioQuant/DKFZ. Beide nehmen derzeit mehr als 150.000 CPU-Stunden auf den Globus-Ressourcen im D-Grid auf. Vom wissenschaftlichen Standpunkt sind diese Rechnungen und Analysen von großer Bedeutung wie sich an der ansteigenden Zahl an Publikationen zeigt. Hinter den zwei hierbei verwendeten Nutzerkennungen verbergen sich dabei mehrere Nutzer die durch A. Beck-Ratzka und T. A. Knoch in Kooperationsprojekten gebündelt sind, die entweder einzelne Applikationen und/oder ganze Pipelinesysteme benutzen. Dazu gehören auch Nutzer die u.a. in nationalen, europäischen bzw. internationalen Konsortien für den Informatikpart zuständig sind. Im astrophysikalischen Fall handelt es sich dabei um 10 Nutzer/Projekte, im biomedizinischen Fall sind dies international mittlerweile ca. 200 Nutzer/Projekte (180 hierbei durch die Nutzung einer sog. Assoziation Pipeline, die von äußerst wichtiger diagnostischer Relevanz ist). Die Stärkung des Standorts Deutschland durch die erfolgreiche Nutzung im Produktionsbetrieb hat einerseits die Machbarkeit einer funktionierenden Grid-Infrastruktur für die Globus Ressourcen-Betreiber gezeigt und andererseits durch die Schaffung wissenschaftlich hoch-relevanter Ergebnisse zu einer damit verbundenen forschungspolitischen Stärkung geführt. Beides kann nicht hoch genug eingeschätzt werden, da beide Usecases vor allem auch im internationalen Vergleich eine herausragende Stellung einnehmen – faktisch gehören sie mittlerweile zu den größten Nutzern von Rechenzeit weltweit.

Der große Erfolg der zwei Usecases Gravitationswellenanalyse und Genomanalyse hat die Ressourcen bezüglich der Globus basierten Infrastruktur zu nahezu 100% ausgeschöpft. Er basiert auf unabhängig voneinander entwickelten Komponenten, welche die Middleware in eine produktive Umgebung integrieren, und damit ein Produktionsgrid erst ermöglichen. Dies ist einzigartig im D-Grid-Umfeld. Die Personalkapazität von A. Beck-Ratzka und T. A. Knoch sind aufgrund der Unterstützungsanforderungen von weiteren/neuen Nutzern zu 100% ausgereizt. Folglich stoßen wir in Bezug auf i) Rechenzeit für unsere eigenen Projekte sowie die neuer Nutzer, ii) die Unterstützung und Anwerbung neuer Nutzer, sowie iii) entsprechenden Unterstützung beim Management in Hinblick auf Daten Sicherheits/Vertraulichkeit, SLAs und Coaching, an die Grenze unserer bisher existierenden Möglichkeiten. Die entsprechenden Projektanforderungen bezüglich Rechenleistung sind klar durch die entsprechenden Projekterfolge gegeben, d.h. dass über die Laufzeit der D-Grid Infrastruktur bis 2014 massive Anforderungen über den bisher für uns zugänglichen Rahmen hinaus bestehen. Anfragen von neuen Nutzer gibt es von 10 weiteren im astrophysikalischen (Berliner Raum) und 50-60 weiteren im biomedizinischen Bereich (Raum Heidelberg), die uns als verlässliche bzw. erfahrene Garanten für erfolgreiche Hoch-Durchsatz Gridnutzung ansehen, aber entsprechende Unterstützung brauchen. Zusätzlich gibt es auch Anfragen von mehreren nationalen und internationalen Verbänden, die genau solche Komponenten wie die unsrigen sofort benutzen würden. Darüberhinaus besteht in diesem Zusammenhang der Bedarf an weiteren Lösungen im Datensicherheits- und Vertraulichkeits-Bereich, die durch diese Bereiche konkret angefordert werden, was mit entsprechendem Coaching und Management einhergehen muss und immer wieder von uns eingefordert wird. Wir sind fest davon überzeugt, dass mit dem im folgenden beschriebenen Vorschlag nicht nur diese Flaschenhälse beseitigt werden können, sondern auch die Nachhaltigkeit und damit der Erfolg des gesamten D-

Grid Projektes massiv gestärkt werden kann!

Ein weiterer Aspekt der produktiven Usecases ist, dass erst mit diesen die Grid-Ressourcen richtig für den Produktionsbetrieb getestet werden können. Die Problematik im DGUS-Ticket 853 beispielsweise ist ein Problem, dass erst durch produktive Usecases aufgedeckt wird. Die Globus-basierten Grid-Ressourcen im D-Grid sind durch die beiden produktiven Usecases Gravitationswellenanalyse und Genomanalyse stabiler geworden. Durch die in im Rahmen dieses Projektes geplante Erweiterung der Usecases auf gLite- und Unicore-Ressourcen würden auch die Betreiber dieser Ressourcen enorm profitieren, weil es damit auf diesen Ressourcen zu einem Produktionsbetrieb kommen würde.

Corresponding author email contact: TA.Knoch@taknoch.org

Keywords:

Genome, genomics, genome organization, genome architecture, structural sequencing, architectural sequencing, systems genomics, coevolution, holistic genetics, genome mechanics, genome statistical mechanics, genomic uncertainty principle, genome function, genetics, gene regulation, replication, transcription, repair, homologous recombination, simultaneous co-transfection, cell division, mitosis, metaphase, interphase, cell nucleus, nuclear structure, nuclear organization, chromatin density distribution, nuclear morphology, chromosome territories, subchromosomal domains, chromatin loop aggregates, chromatin rosettes, chromatin loops, chromatin fibre, chromatin density, persistence length, spatial distance measurement, histones, H1.0, H2A, H2B, H3, H4, mH2A1.2, DNA sequence, complete sequenced genomes, molecular transport, obstructed diffusion, anomalous diffusion, percolation, long-range correlations, fractal analysis, scaling analysis, exact yard-stick dimension, box-counting dimension, lacunarity dimension, local nuclear dimension, nuclear diffuseness, parallel super computing, grid computing, volunteer computing, Brownian Dynamics, Monte Carlo, fluorescence in situ hybridization, chromatin cross-linking, chromosome conformation capture (3C), selective high-resolution high-throughput chromosome interaction capture (T2C), confocal laser scanning microscopy, fluorescence correlation spectroscopy, super resolution microscopy, spatial precision distance microscopy, auto-fluorescent proteins, CFP, GFP, YFP, DsRed, fusion protein, in vivo labelling, information browser, visual data base access, holistic viewing system, integrative data management, extreme visualization, three-dimensional virtual environment, virtual power tool.

Literature References

- Knoch, T. A.** Dreidimensionale Organisation von Chromosomen-Domänen in Simulation und Experiment. (Three-dimensional organization of chromosome domains in simulation and experiment.) *Diploma Thesis*, Faculty for Physics and Astronomy, Ruperto-Carola University, Heidelberg, Germany, 1998, and TAK Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-010685-5 and ISBN 978-3-00-010685-9 (soft cover, 2nd ed.), ISBN 3-00-035857-9 and ISBN 978-3-00-0358857-0 (hard cover, 2nd ed.), ISBN 3-00-035858-7, and ISBN 978-3-00-035858-6 (DVD, 2nd ed.), 1998.
- Knoch, T. A., Münkler, C. & Langowski, J.** Three-dimensional organization of chromosome territories and the human cell nucleus - about the structure of a self replicating nano fabrication site. *Foresight Institute - Article Archive*, Foresight Institute, Palo Alto, CA, USA, <http://www.foresight.org>, 1- 6, 1998.
- Knoch, T. A., Münkler, C. & Langowski, J.** Three-Dimensional Organization of Chromosome Territories and the Human Interphase Nucleus. *High Performance Scientific Supercomputing*, editor Wilfried Jüling, Scientific Supercomputing Center (SSC) Karlsruhe, University of Karlsruhe (TH), 27- 29, 1999.
- Knoch, T. A., Münkler, C. & Langowski, J.** Three-dimensional organization of chromosome territories in the human interphase nucleus. *High Performance Computing in Science and Engineering 1999*, editors Krause,

- E. & Jäger, W., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3-540-66504-8, 229-238, 2000.
- Bestvater, F., **Knoch, T. A.**, Langowski, J. & Spiess, E. GFP-Walking: Artificial construct conversions caused by simultaneous cotransfection. *BioTechniques* 32(4), 844-854, 2002.
- Knoch, T. A. (editor)**, Backes, M., Baumgärtner, V., Eysel, G., Fehrenbach, H., Göker, M., Hampl, J., Hampl, U., Hartmann, D., Hitzelberger, H., Nambena, J., Rehberg, U., Schmidt, S., Weber, A., & Weidemann, T. Humanökologische Perspektiven Wechsel - Festschrift zu Ehren des 70. Geburtstags von Prof. Dr. Kurt Egger. Human Ecology Working Group, Ruperto-Carola University of Heidelberg, Heidelberg, Germany, 2002.
- Knoch, T. A.** Approaching the three-dimensional organization of the human genome: structural-, scaling- and dynamic properties in the simulation of interphase chromosomes and cell nuclei, long- range correlations in complete genomes, *in vivo* quantification of the chromatin distribution, construct conversions in simultaneous co-transfections. *Dissertation*, Ruperto-Carola University, Heidelberg, Germany, and TAK†Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-009959-X and ISBN 978-3-00-009959-5 (soft cover, 3rd ed.), ISBN 3-00-009960-3 and ISBN 978-3-00-009960-1 (hard cover, 3rd ed.), ISBN 3-00-035856-9 and ISBN 978-3-00-010685-9 (DVD, 3rd ed.) 2002.
- Knoch, T. A.** Towards a holistic understanding of the human genome by determination and integration of its sequential and three-dimensional organization. *High Performance Computing in Science and Engineering 2003*, editors Krause, E., Jäger, W. & Resch, M., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3- 540-40850-9, 421-440, 2003.
- Wachsmuth, M., Weidemann, T., Müller, G., Urs W. Hoffmann-Rohrer, **Knoch, T. A.**, Waldeck, W. & Langowski, J. Analyzing intracellular binding and diffusion with continuous fluorescence photobleaching. *Biophys. J.* 84(5), 3353-3363, 2003.
- Weidemann, T., Wachsmuth, M., **Knoch, T. A.**, Müller, G., Waldeck, W. & Langowski, J. Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal imaging. *J. Mol. Biol.* 334(2), 229-240, 2003.
- Fejes Tóth, K., **Knoch, T. A.**, Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C. P., Müller, G. & Rippe, K. Trichostatin A induced histone acetylation causes decondensation of interphase chromatin. *J. Cell Science* 117, 4277-4287, 2004.
- Ermiler, S., Krunić, D., **Knoch, T. A.**, Moshir, S., Mai, S., Greulich-Bode, K. M. & Boukamp, P. Cell cycle-dependent 3D distribution of telomeres and telomere repeat-binding factor 2 (TRF2) in HaCaT and HaCaT-myc cells. *Europ. J. Cell Biol.* 83(11-12), 681-690, 2004.
- Kost, C., Gama de Oliveira, E., **Knoch, T. A.** & Wirth, R. Spatio-temporal permanence and plasticity of foraging trails in young and mature leaf-cutting ant colonies (*Atta spp.*). *J. Trop. Ecol.* 21(6), 677- 688, 2005.
- Winnefeld, M., Grewenig, A., Schnölzer, M., Spring, H., **Knoch, T. A.**, Gan, E. C., Rommelaere, J. & Cziepluch, C. Human SGT interacts with BAG-6/Bat-3/Scythe and cells with reduced levels of either protein display persistence of few misaligned chromosomes and mitotic arrest. *Exp. Cell Res.* 312, 2500-2514, 2006.
- Sax, U., Weisbecker, A., Falkner, J., Viezens, F., Yassene, M., Hartung, M., Bart, J., Krefting, D., **Knoch, T. A.** & Semler, S. Grid-basierte Services für die elektronische Patientenakte der Zukunft. *E- HEALTH-COM - Magazin für Gesundheitstelematik und Telemedizin*, 4(2), 61-63, 2007.
- de Zeeuw, L. V., **Knoch, T. A.**, van den Berg, J. & Grosveld, F. G. Erasmus Computing Grid - Het bouwen van een 20 TeraFLOP virtuele supercomputer. *NIOC proceedings 2007 - het perspectief van lange termijn*. editor Frederik, H. NIOC, Amsterdam, The Netherlands, 52-59, 2007.
- Rauch, J., **Knoch, T. A.**, Solovei, I., Teller, K. Stein, S., Buiting, K., Horsthemke, B., Langowski, J., Cremer, T., Hausmann, M. & Cremer, C. Lightoptical precision measurements of the Prader- Willi/Angelman Syndrome imprinting locus in human cell nuclei indicate maximum condensation changes in the few hundred nanometer range. *Differentiation* 76(1), 66-82, 2008.

- Sax, U., Weisbecker, A., Falkner, J., Viezens, F., Mohammed, Y., Hartung, M., Bart, J., Krefting, D., **Knoch, T. A.** & Semler, S. C. Auf dem Weg zur individualisierten Medizin - Grid-basierte Services für die EPA der Zukunft. *Telemedizinführer Deutschland 2008*, editor Jäckel, A. Deutsches Medizinforum, Minerva KG, Darmstadt, ISBN 3-937948-06-6, ISBN-13 9783937948065, 47-51, 2008.
- Drägestein, K. A., van Capellen, W. A., van Haren, J. Tsibidis, G. D., Akhmanova, A., **Knoch, T. A.**, Grosveld, F. G. & Galjart, N. Dynamic behavior of GFP-CLIP-170 reveals fast protein turnover on microtubule plus ends. *J. Cell Biol.* 180(4), 729-737, 2008.
- Jhunjhunwala, S., van Zelm, M. C., Peak, M. M., Cutchin, S., Riblet, R., van Dongen, J. J. M., Grosveld, F. G., **Knoch, T. A.**⁺ & Murre, C.⁺ The 3D-structure of the Immunoglobulin Heavy Chain Locus: implications for long-range genomic interactions. *Cell* 133(2), 265-279, 2008.
- Krefting, D., Bart, J., Beronov, K., Dzhimova, O., Falkner, J., Hartung, M., Hoheisel, A., **Knoch, T. A.**, Lingner, T., Mohammed, Y., Peter, K., Rahm, E., Sax, U., Sommerfeld, D., Steinke, T., Tolxdorff, T., Vossberg, M., Viezens, F. & Weisbecker, A. MediGRID - Towards a user friendly secured grid infrastructure. *Future Generation Computer Systems* 25(3), 326-336, 2008.
- Knoch, T. A.**, Lesnussa, M., Kepper, F. N., Eussen, H. B., & Grosveld, F. G. The GLOBE 3D Genome Platform - Towards a novel system-biological paper tool to integrate the huge complexity of genome organization and function. *Stud. Health. Technol. Inform.* 147, 105-116, 2009.
- Knoch, T. A.**, Baumgärtner, V., de Zeeuw, L. V., Grosveld, F. G., & Egger, K. e-Human Grid Ecology: Understanding and approaching the Inverse Tragedy of the Commons in the e-Grid Society. *Stud. Health. Technol. Inform.* 147, 269-276, 2009.
- Dickmann, F., Kaspar, M., Löhnhardt, B., **Knoch, T. A.**, & Sax, U. Perspectives of MediGRID. *Stud. Health. Technol. Inform.* 147, 173-182, 2009.
- Knoch, T. A.**, Göcker, M., Lohner, R., Abuseiris, A. & Grosveld, F. G. Fine-structured multi-scaling long-range correlations in completely sequenced genomes - features, origin and classification. *Eur. Biophys. J.* 38(6), 757-779, 2009.
- Dickmann, F., Kaspar, M., Löhnhardt, B., Kepper, N., Viezens, F., Hertel, F., Lesnussa, M., Mohammed, Y., Thiel, A., Steinke, T., Bernarding, J., Krefting, D., **Knoch, T. A.** & Sax, U. Visualization in health-grid environments: a novel service and business approach. *LNCS 5745*, 150-159, 2009.
- Dickmann, F., Kaspar, M., Löhnhardt, B., Kepper, N., Viezens, F., Hertel, F., Lesnussa, M., Mohammed, Y., Thiel, A., Steinke, T., Bernarding, J., Krefting, D., **Knoch, T. A.** & Sax, U. Visualization in health-grid environments: a novel service and business approach. *Grid economics and business models - GECON 2009 Proceedings, 6th international workshop, Delft, The Netherlands*. editors Altmann, J., Buyya, R. & Rana, O. F., GECON 2009, LNCS 5745, Springer-Verlag Berlin Heidelberg, ISBN 978-3-642-03863-1, 150-159, 2009.
- Estrada, K.* , Abuseiris, A.* , Grosveld, F. G., Uitterlinden, A. G., **Knoch, T. A.**⁺ & Rivadeneira, F.⁺ GRIMP: A web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics* 25(20), 2750-2752, 2009.
- Kepper, N., Schmitt, E., Lesnussa, M., Weiland, Y., Eussen, H. B., Grosveld, F. G., Hausmann, M. & **Knoch T. A.**, Visualization, Analysis, and Design of COMBO-FISH Probes in the Grid-Based GLOBE 3D Genome Platform. *Stud. Health Technol. Inform.* 159, 171-180, 2010.
- Kepper, N., Ettig, R., Dickmann, F., Stehr, R., Grosveld, F. G., Wedemann, G. & **Knoch, T. A.** Parallel high-performance grid computing: capabilities and opportunities of a novel demanding service and business class allowing highest resource efficiency. *Stud. Health Technol. Inform.* 159, 264-271, 2010.
- Skrowny, D., Dickmann, F., Löhnhardt, B., **Knoch, T. A.** & Sax, U. Development of an information platform for new grid users in the biomedical field. *Stud. Health Technol. Inform.* 159, 277-282, 2010.
- Knoch, T. A.**, Baumgärtner, V., Grosveld, F. G. & Egger, K. Approaching the internalization challenge of grid technologies into e-Society by e-Human "Grid" Ecology. *Economics of Grids, Clouds, Systems, and Services – GECON 2010 Proceedings, 7th International Workshop, Ischia, Italy*, editors Altman, J., & Rana, O. F., Lecture Notes in Computer Science (LNCS) 6296, Springer Berlin Heidelberg New York, ISSN 0302-9743, ISBN-10 3-642-15680-0, ISBN-13 978-3-642-15680-9, 116-128, 2010.
- Dickmann, F., Brodhun, M., Falkner, J., **Knoch, T. A.** & Sax, U. Technology transfer of dynamic IT outsourcing requires security measures in SLAs. *Economics of Grids, Clouds, Systems, and Services – GECON 2010*

Proceedings, 7th International Workshop, Ischia, Italy, editors Altman, J., & Rana, O. F., Lecture Notes in Computer Science (LNCS) 6296, Springer Berlin Heidelberg New York, ISSN 0302-9743, ISBN-10 3-642-15680-0, ISBN-13 978-3-642-15680-9, 1-115, 2010.

Knoch, T. A. Sustained Renewability: approached by systems theory and human ecology. *Renewable Energy* 2, editors M. Nayeripour & M. Keshti, Intech, ISBN 978-953-307-573-0, 21-48, 2011.

Kolovos, P., **Knoch, T. A.**, F. G. Grosveld, P. R. Cook, & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics and Chromatin* 5(1), 1-8, 2012.

Dickmann, F., Falkner, J., Gunia, W., Hampe, J., Hausmann, M., Herrmann, A., Kepper, N., **Knoch, T. A.**, Lauterbach, S., Lippert, J., Peter, K., Schmitt, E., Schwardmann, U., Solodenko, J., Sommerfeld, D., Steinke, T., Weisbecker, A. & Sax, U. Solutions for Biomedical Grid Computing - Case Studies from the D-Grid Project Services@MediGRID. *JOCS* 3(5), 280-297, 2012.

Estrada, K.^{*}, Abuseiris, A.^{*}, Grosveld, F. G., Uitterlinden, A. G., **Knoch, T. A.**⁺ & Rivadeneira, F.⁺ GRIMP: A web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Dissection of the complex genetic architecture of human stature and osteoporosis*. cumulative dissertation, editor Estrada K., Erasmus Medical Center, Erasmus University Rotterdam, Rotterdam, The Netherlands, ISBN 978-94-6169-246-7, 25-30, 1st June 2012.

van de Corput, M. P. C., de Boer, E., **Knoch, T. A.**, van Cappellen, W. A., Quintanilla, A., Ferrand, L., & Grosveld, F. G. Super-resolution imaging reveals 3D folding dynamics of the b-globin locus upon gene activation. *J. Cell Sci.* 125, 4630-4639, 2012.