

GSI+FAIR Computing Kick-Off-Meeting am 12. Mai 2011

Protokoll

Raum: Seminarraum Theorie, GSI

Beginn: 12:06 Uhr

Persönlich anwesend: Haik Simon, Mohammad Al-Turany, Denis Bertini, Florian Uhlig, Carsten Preuß, Kilian Schwarz, Katharina Behnert, Yvonne Leifels, Uwe Spillmann, Walter F. J. Müller, Volker Friese, Peter Senger, Walter Schön, Christian Schmidt, Lars Schmitt, Silvia Masciocchi, Peter Malzacher, Karin Miers, Jan Reinelt, Ingo Augustin, Peter Braun-Munzinger, Karl-Heinz Langanke, Volker Lindenstruth, Andreas Grossmann, J. Wambach (ab ca. 13h)

Herr Johan Messchendorp und Herr Joachim Stroth wollen via EVO am Meeting teilnehmen, was wegen technischer Probleme nicht gelingt. Herr Messchendorp nimmt über Telefonleitung teil, Herr Stroth ist telefonisch nicht zu erreichen.

Zu Beginn einigt man sich darauf, das Meeting in deutscher Sprache abzuhalten.

Herr Lindenstruth führt in die zuvor per Mail an die Teilnehmer versandten Folien ein.

Bemerkung der Protokollantin: Zur Verkürzung des Protokolls werden nur die die Folien ergänzenden Ausführungen protokolliert.

Meetings zur Koordination und Synchronisierung der IT-Anforderungen sollen zukünftig regelmäßig stattfinden, Turnus ca. alle drei Monate.

zu Folie 2: (Organigramm):

Es gibt ein IT-Koordinationsmeeting im DVEE-Besprechungsraum jeden zweiten Dienstag um 13:00 Uhr, zu dem jeder willkommen ist.

Herr Lindenstruth betont, dass es nicht geplant ist, die Abteilungen Core-IT und Admin-IT zusammenzulegen.

Für die Abteilungen HPC+SC ist in den nächsten Jahren ein Zuwachs von 13 Stellen geplant und vom Aufsichtsrat genehmigt.

Die Kern-IT muss entsprechend der Nutzerzahlen wachsen.

zu Folie 3: (Tabelle der geschätzten Bedarfszahlen der Experimentgruppen):

Diese Zahlen sollen einmal im Jahr upgedated werden. Die Experimentgruppen sind aufgefordert ihre entsprechend korrigierten Zahlen zu benennen.

zu Folie 4: (LOEWE CSC 2011):

Auch regionale Ressourcen können genutzt werden: der Rechner in Frankfurt kann FAIR zur Verfügung gestellt werden, hat aber im Vergleich wenig Massenspeicher (nur 800 TB MSS + 1,6 PB lokale Platten).

zu Folie 5: (Cooling System architecture)

System des „Brühwürfels“; in Betrieb mit 2,5 PB Lustre Brutto.

zu Folie 6: (GSI Datacenter Roadmap)

Seit einem Jahr gibt es keinen Platz mehr im Rechenzentrum für neue Rechner.

Für die Zukunft ist geplant:

- „Mini-Cube“ in der Testinghalle als Zwischenlösung
- für den Bau eines neuen Rechenzentrums wurden 19 Mio € genehmigt durch alle Institutionen außer Senat (11 Mio fürs Gebäude und 8 Mio für neue Rechner)

Folie 7: (FAIR Green-IT Tier-0 Data Center):

Vorstellung der Pläne für das neue Rechenzentrum:

6 Stockwerke verfügbar für Rechner, für FAIR nötig wäre eines.

Der Rest kann vermietet werden. CERN hat Interesse an 2 Stockwerken, dadurch kämen Mittel zur GSI. Es gibt aber noch keine Zusage.

zu Folie 8: (Querschnitt neues Rechenzentrum):

Es gibt bereits Vorträge für das Gebäude.

zu Folie 9: (Längsschnitt neues Rechenzentrum):

Wegen der Änderungen gibt es heute ein Meeting mit den Bauleuten.

zu Folie 10: (FAIR centers outside GSI)

Appell an alle: Bitte Bedarf anmelden und Zeitpunkt, wann die Rechner benötigt werden.

zu Folie 11: (Usage of FAIR Tier-0 center)

Vorschlag: längeres Glasfaserkabel ins Rechenzentrum legen, dadurch kann man die Online-Systeme ins Rechenzentrum legen und benötigt sie nicht mehr bei den Experimenten. Dies führt zu Synergieeffekten.

Vorgespräche mit PANDA haben bereits stattgefunden; CBM ist mit diesem Plan einverstanden.

zu Folie 12: (FAIR HPC Backbone)

Eine Überlegung ist, den Unis einen direkten Zugriff zu ermöglichen. Verträge hierzu werden beauftragt. Dazu ist eine Vernetzung der lokalen Rechenzentren geplant. Es gibt ein Angebot für einen 80 Gb Link nach Frankfurt und eine kosteneffiziente Dark Fibre nach Frankfurt.

Herr Braun-Munzinger wendet ein, dass die Daten zweimal kopiert werden müssen, eine Kopie muss physisch getrennt werden; das muss mit ins Konzept.

Zu bedenken sind auch die Betriebskosten von geschätzt 6 Mio€/Jahr, plus Personal.

Herr Lindenstruth weist auf die Möglichkeit der Vermietung, um Geld hereinzubekommen, hin, außerdem wird es Einsparung durch Zentralisierung geben. Verschiedene Ansätze müssen diskutiert werden.

Herr Braun-Munzinger sieht auch finanziellen Beitrag von FAIR, ebenso wie Mitgliedsstaaten Posten von CERN bezahlen.

Herr Lindenstruth führt aus, dass der Beschleuniger 60MW benötigt, das heißt, die Computer machen ca. 10-20% der Beschleunigerenergiekosten aus, da der

Beschleuniger im Gegensatz zu den Computern ein hohes dynamisches Lastverhalten hat.

Herr Lindenstruth gibt das Wort weiter an die Abteilungsleiter der IT. Als erste spricht Karin Miers.

zu Folie 13: (Core-IT):

Übersicht über die neuen Projekte der Core-IT:

Exchange 2010; Vorteile:

- bessere Performance
- Abwesenheitsassistent kann von außerhalb und individuell eingestellt werden

Zu Sharepoint: Frage an alle: Welche Anforderungen gibt es?

Zum Netzwerk: Es ist ein neues Konzept erforderlich, dieses soll noch in diesem Jahr erarbeitet werden; also bitte Anforderungen anmelden!

zu Folie 14: (Admin-IT)

Entscheidung für Typo 3, weil in der Helmholtzgemeinschaft vielfach vertreten (7mal).

Umzug der Webseiten wird im Spätsommer stattfinden, das Layout ist schon verabschiedet.

Herr Lindenstruth bemerkt dazu, dass OnTeam solange weiterlaufen wird, wie es benötigt wird.

Zum Dokumentenmanagementsystem bittet Frau Miers alle Benutzer, ihre Anforderungen anzumelden.

zu Folie 15: (Questions/Input)

Zusammenfassung aller anstehenden Fragen

W. Müller bemerkt, dass über das EDMS diskutiert werden muss. Herr Lindenstruth stellt fest, dass dazu eine Arbeitsgruppe u. a. mit Hr. Eickhoff gegründet werden muss.

Es folgt Peter Malzacher zu Scientific Computing:

zu Folie 16 und 17: (Focus of Scientific Computing; short/medium term challenges)

GSI muss die versprochenen Gridanforderungen von ALICE erfüllen:

Ca. 900 Jobs parallel, 440TB Storage Element, das ist eigentlich seit 1. April gefordert.

Derzeit hat das Storage Element 300TB. Entweder muss das ausgebaut werden, besser wäre es das Storage Element auf Lustre zu verlegen.

Während der Downtime von CERN ist ein Software-Umbruch bei den LHC-Experimenten vorzusehen. In dieser Zeit startet auch FAIR mit intensivem Computing. Die anstehende Software-Entwicklung wird gemeinsam mit CERN vor sich gehen. Besonders interessant ist das Geant5 Projekt, das von vornherein auf parallele Architekturen konzipiert wird. Da wollen wir mitarbeiten.

Bis dahin wird es für LHC noch einen Anstieg an Hardware-Ressourcen geben. Danach nur noch Ersatzinvestitionen.

Frau Masciocchi wendet ein, dass es zwar während des Shutdowns von ALICE keine neuen Daten geben wird, aber neue Produktion und Datenbewegungen: Die Batchfarm wird weiter verwendet werden.

Laut Herrn Lindenstruth wird es keinen großen Aufwuchs in der Shutdownzeit geben (Abschaltung CERN Ende 2012 für mehr als 18 Monate bis ca. Ende 2013). Ausbesserungen werden in dieser Zeit in erster Linie aus dem IT-Etat finanziert, also gibt es wenig Spielraum.

Herr Langanke schlägt vor, eine Finanzierung über HIM-Mittel zu prüfen.

Walter Schön stellt die High Performance Computing Gruppe vor:

zu Folie 18 und 19: (High Performance Computing):

Das Hauptproblem liegt in der Skalierbarkeit, daher werden Strategien entwickelt für Skalierbarkeit von Systemen (z. B. Lustre).

Lustre ist erfahrungsgemäß immer voll!

2000 Cores kommen in den folgenden 1-2 Wochen und werden schnellst möglich installiert.

zu Folie 20: („test cube“):

„Brühwürfel-System“ kommt in die Testinghalle.

zu Folie 21: (Cluster for mini cube):

Herr Schön stellt zwei Alternativen vor:

- Möglichkeit A: mit Abstand kosteneffizienteste Lösung, aber dafür weniger Ausfallsicherheit.
- Möglichkeit B: teurer, aber sicherer. RAID-Controller kann mehr Platten verwalten und in Betrieb ausgetauscht werden.

Der Preisunterschied liegt allerdings bei Faktor 1,5 bis 2.

Er weist ausdrücklich darauf hin, dass es bei Wahl der billigeren Alternative zu Ausfällen kommen kann!

Herr Lindenstruth ergänzt, dass es bei riesigen Speichermengen immer zu Ausfällen kommen kann. Also bitte Verständnis haben. Er ist eher gegen eine teure Hybridlösung.

Haik Simon fragt, ob es Konzepte gibt für bestimmte Bereiche. Soll es verschiedene Storage-Klassen geben?

Herr Lindenstruth antwortet mit ja, und die Überlegung sei; auch Tapes zu berücksichtigen.

Herr Langanke regt an, die Finanzierung auch für andere Institute zu öffnen.

Frau Masciocchi lobt das HPC-Team und gibt evtl. Überlastung zu bedenken.

Laut Herrn Lindenstruth wird HPC die erste Abteilung sein, die personellen Aufwuchs bekommt. Geld aus den Baukosten kann man evtl. für Personal abzweigen.

Haik Simon reklamiert die fehlende Benutzerverwaltung. Diese soll auf die Agenda fürs nächste Treffen.

zu Folien 22 und 23: (Compute nodes for „mini cube“, Migration Lustre I - Lustre II):

Der „mini cube“ soll im September stehen, also muss die Ausschreibung für die Rechner SOFORT erfolgen.

Wunsch ist: 2UTwin von SuperMicro

Im Moment. 2Gb/Core; Herr Lindenstruth bemerkt, dass dieser Wert heruntergehen wird.

Das ist schlecht für Theorie und Konstruktionsrechnung, hier wäre HiMem Queue möglich.

Vorschlag zur Diskussion in der Community:

- 2GB/ Kern
- 32 GB/Knoten
- Infiniband
- 1TB zusätzlich kaufen
- GPU's - ja oder nein und wenn ja dann welche?
- vielleicht GPU's aus Frankfurt nutzen

C. Schmidt meldet Bedarf an Grafikkarten mit hoher Grafikkartendichte und die möglichst sofort.

Herr Lindenstruth antwortet, dass die Rechner von Walter Schön ausbaufähig sind mit 1GPU/Server, höhere Dichte wird deutlich teurer. Es gibt Rechner in Frankfurt.

In der Folge wird über die Hersteller der Grafikkarten diskutiert, hier steht ATI (günstig) gegen Nvidia (teuer). Christian Schmidt gibt zu bedenken, dass für Lattice Nvidia-Grafikkarten benötigt werden, da die fertige Software aus USA, Cuda/Nvidia, auf ATI-Grafikkarten nicht läuft.

Hier besteht offensichtlich höherer Diskussionsbedarf, deshalb wird beschlossen, eine Arbeitsgruppe zu bilden, die sich am kommenden Freitag, 20.Mai, 12:00 Uhr zu einem weiteren Meeting trifft.

Zu dieser Arbeitsgruppe sollen gehören: Thomas Neff, Christian Schmidt, Silvia Masciocchi, Haik Simon, Johan Messchendorp, Volker Friese, Jochen Markert, Joachim Stroth, Walter Schön, Volker Lindenstruth, Peter Malzacher.

Thema soll sein: Die Festlegung der dringend benötigten Ausschreibungskriterien für den „Mini-Cube“.

Es wird beschlossen, vor der Festlegung des Termins für das nächste Meeting erst einmal das Ergebnis der Arbeitsgruppe abzuwarten, dann soll es einen „Jour-Fixe“ geben, Abstand 3 Monate, mit der Bildung von Arbeitsgruppen dazwischen.

Frau Masciocchi führt aus, dass beim parallelen Laufen der Systeme Lenny, Etch und Squeeze Probleme beim Arbeiten auftreten. Dies sei ein zentrales Thema und müsse unbedingt hier diskutiert werden.

Herr Malzacher erklärt, es solle in Zukunft ein Hauptsystem geben, die anderen Systeme werden auf virtuellen Maschinen laufen.

Walter Schön weist darauf hin, dass auch diese Systeme gewartet werden müssen und daher deren Anzahl zu kontrollieren ist.

Zum Schluss bittet Herr Lindenstruth darum, alle Zahlen auf Folie 2 genau zu überprüfen und beim nächsten Mal substantiell neue Zahlen zu nennen.

Ende: 14:08 Uhr

Protokoll: Katharina Behnert und Kilian Schwarz