

# Diskussion – Strategie für Ressourcen

# Rahmenbedingungen

- Realisierung für „ExaByte Computing“ im Umfeld der Experimente der Teilchen-, Astroteilchen, Hadron- und Kernphysik in den nächsten 5-6 Jahren
- Finanzpolitisches Umfeld – Geldgeber
  - BMBF – Verbundforschung (ErUM-Data)
  - Helmholtz Gemeinschaft
  - Max-Planck-Gesellschaft
- BMBF: ErUM-Data als neue Förderlinie
  - Existierendes ErUM-Data Pilotprojekt
  - (Weiter-)Entwicklung von Werkzeugen und wissenschaftlicher Software
  - Hardware im Rahmen der Säule „Föderierte Infrastruktur“

# Struktur

- Aufgrund des sehr großen Ressourcen-Bedarfs, werden große Installationen (zumindest für Massenspeicher – „data lakes“) entstehen müssen.
- Gerade CPU-Leistung kann auch „föderiert“ angeboten werden, d.h. verteilt auf mehrere Standorte sein.
- „Pledged“ Ressourcen
  - Große Installation am GridKa (KIT), DESY, GSI/FAIR, (MPG?)
  - Föderierte Zentren für die Analyse an verschiedenen Standorten (Helmholtz Zentren, Universitäten und auch andere Rechenzentren wie z.B. Leibniz, GWDG, etc)
- Science Clouds und ggf. kommerzielle Anbieter
- Breite opportunistische/parasitäre Nutzung verschiedenster Ressourcen
- Nutzung von zeitweise verfügbaren Ressourcen

# Welche Rolle spielen die Helmholtz-Zentren und die Universitäten?

- Helmholtz-Zentren liefern den Hauptanteil der Ressourcen
  - GridKa(KIT), DESY, GSI/FAIR, (Jülich?) (auch MPG)
  - Annahme: Ressourcen werden für alle beteiligten Experimente zusammen betrieben
    - Gibt es spezifische Gründe dies nicht zu tun?
- Universitäten werden bei vielen Experimenten für alle Experimentspezifischen Aufgaben benötigt
  - Datenmanagement, Software, Koordination mit dem Experiment, Nutzerunterstützung
  - Ausbildung
  - Ressourcen an Unis -> nächste Seite

# Ressourcen an Universitäten

- LHC Run-3
  - Computingmodelle wie bisher
  - Tier-2 Zentren an Universitäten liefern wie bisher ihren Service und Ressourcen
  - Finanzierung der Ressourcen ist unklar, angestrebt wird ErUM-Data
  - Standorte, die in Run-4 keine „pledged“ Ressourcen mehr liefern wollen, sollten keine neuen Ressourcen beschaffen
- LHC Run-4
  - Neue Computingmodelle mit „Data Lakes“ an großen Zentren
  - Universitäten könnten sich föderiert mit anderen Zentren an den Aufgaben beteiligen und eine nennenswerte Rolle spielen
  - Anbindung an große Rechenzentren (Helmholtz, Leibniz, GWDG ...) notwendig
  - Opportunistischer Beitrag ist für jeden Cluster möglich

# Funktion der Zentren

- Der Großteil der Ressourcen wird für zentrale Aufgaben der Experimente benötigt. Diese müssen von den Zentren abgedeckt werden
  - Datenspeicherung
  - Rekonstruktion der Daten
  - Simulationsrechnungen (auf allen Ressourcen)
  - Vorprozessierung der Daten für die Analyse
  - Datenarchivierung
- Spezialisierte Ressourcen für die Analyse
  - National Analysis Facilities/Föderierte Analysis Facilities/Science Cloud?
    - Wichtiges Ziel von ErUM-Data!
  - Dedizierte Cluster für lokale Nutzung wird es an vielen Standorten geben (nicht „pledged“)

# Übergang zum HL-LHC innerhalb der nächsten 5 Jahre

- Viele Entwicklungsarbeiten notwendig, um den Ressourcenbedarf zu reduzieren
  - Fortschreiben der Entwicklungen des ErUM-Data Pilotprojekts
  - Entwicklungsarbeit für existierende Tools
  - Bedarf für einen ErUM-Data Verbundantrag von Teilchen-, Astroteilchen, Hadron- und Kernphysik
  - Hierzu gehören auch realitätsnahe Implementierungen und Skalierungstests
    - Könnte im Kontext von LHC Run-3 laufen
  - Bedarf (FTE) und Expertise
- Baldige Festlegung der beteiligten Zentren/Standorte für die Ressourcen
  - Helmholtz Zentren - KIT (GridKA), DESY, GSI/FAIR
  - Max-Planck-Gesellschaft
  - Föderierte Zentren (inkl. Universitäten)
- Lösung der Finanzierungsprobleme für die Ressourcen
  - ErUM-Data?
  - Gespräche zwischen Geldgebern (BMBF, Helmholtz, MPG) notwendig

# Wollen wir kommerzielle Ressourcen nutzen?

- Sinnvoll, wenn dies wirtschaftlich günstig ist
- Sinnvoll, wenn eine Ressourcenknappheit ausgeglichen werden kann



# Strategische Punkte

- Große Rechenzentren werden benötigt
  - Starke Unterstützung für die Helmholtz-Zentren KIT, DESY, GSI diese Rolle zu übernehmen
  - Finanzierung: Helmholtz Gemeinschaft und ErUM-Data bieten hierfür den richtigen Rahmen
- Föderierte Zentren/Science Cloud werden/wird benötigt
  - Speziell für die Analyse
  - Starke Unterstützung für das Engagement auch kleinerer Zentren hierzu beizutragen. Gerade spezialisierte Ressourcen könnten so verfügbar gemacht werden.
  - Finanzierung: zumindest teilweise aus ErUM-Data
- Nutzung opportunistischer Ressourcen ist essenziell
- Essenziell ist die Beteiligung der Gruppen in DE an der (Weiter-)Entwicklung der benötigten Werkzeuge – „Enabling ExaByte Computing“
  - ErUM-Data Verbund ab 2021
- Essenziell ist die (Weiter-)Entwicklung der wissenschaftlichen Software für die Optimierung der Ressourcennutzung
  - ErUM-Data Verbund ab 2021
- Entscheidung über Ressourcenbedarf, Standorte und Finanzierung für Run-4 ist spätestens 2024 notwendig
  - Ergebnisse aus den Entwicklungsarbeiten notwendig für die Entscheidung