Subject Areas A & B Introduction

Kilian Schwarz

KickOff Meeting
Pilotmaßnahme ErUM-Data
February 21, 2019
LMU/München

subject area A

The idea of subject area A and the funding we applied for is the target oriented collective and multidisciplinary development of tools and structures in order to enable the integration, the efficient usage as well as the control of heterogeneous resources in the fields of application of Nuclear, Particle, and Astroparticle Physics.

The objective is here the comprehensive functionality of these tools and structures and therefore the achievement of largest possible synergies among the experiments.

Subject Area A

➢Developments for the provision of technologies for the use of heterogeneous computing resources

A1) Werkzeuge zur Einbindung	A2) Effiziente Nutzung
• Scheduling von Cloud-Jobs • Container-Technologien • Checkpointing • Zugang zu Experiment-Datenbanken	• Transiente Datencaches • Transparenter Zugriff auf verteilte Daten
A3) Workflow-Steuerung • Identifikation und Steuerung • In-Pilot Job-Monitoring • Accounting • Optimierung durch Data-Mining	

Area A:

Developments for the provision of technologies for the
use of heterogeneous computing resources

A1:

tools for including heterogeneous resources into scientific computing workflows

participants:
- Frankfurt/GSI (ALICE):
  - including HPC resources into Grid computing (first use case: ALICE T2@GSI)
    by using Singularity Containers
- Freiburg (ATLAS):
  - scheduling of Cloud jobs (virtual machines) by using ROCED
  - extension to more backend batch systems, Cloud APIs, VM configurations
  - extension to using Singularity Containers
- KIT (CMS, Belle II, Auger):
  - worklow management und scheduling in (heterogenous) Cloud environments
  - multi VO and multi site scheduling using ROCED extension COBalD
- Wuppertal (ATLAS):
  - Container usage for IT services.
  - Focus are leightweight solutions in terms of setup and operation

job hiring in progress (Frankfurt, Freiburg)
manpower hired (Wuppertal, KIT)

Area A:
Developments for the provision of technologies for the
use of heterogeneous computing resources

A2:
efficient usage of heterogeneous resources

participants:
- Frankfurt/GSI (ALICE):
    - developing XRootD based „disk caching system on the fly"
    - ingredients: XRootD forward/caching proxy + XRootD PlugIns
- Freiburg (ATLAS):
    - contribution to „disk caching on the fly"
    - scalable storage solutions with Dynafed
- KIT (CMS, Belle II, Auger):
    - Distributed disk caching using XrootD
    - lightweight integration of data caches into heterogeneous resources
- München (ATLAS, Belle II):
    - XRootD based disk caching „XCache"
- DESY (ILC):
    - smart caching in WLCG data federation using dCache technology

job hiring in progress (Frankfurt, Freiburg)
manpower hired (KIT, München)

Area A:
Developments for the provision of technologies for the
use of heterogeneous computing resources

A2:
efficient usage of heterogeneous resources
coordinated work plan among participating institutes has been created

common work packages:
- A2.1: development of XrootD based „disk caching on the fly" on various platforms
- A2.2: benchmarking against other caching solutions, e.g. XCache
- A2.3: integration and test of XRootD caches in various centres for various experiments

common mile stones:
- G1: comparison of different caching solutions in order to find the optimal solution
- G2: porting the optimal solution to other sites
- G3: providing the optimal caching solution as open source product
       including documentation

Area A:

Developments for the provision of technologies for the use of heterogeneous computing resources

A3:

identification and control of workflows on heterogeneous resources

participants:
- Freiburg (ATLAS):
    - on demand scheduling of VMs with resource monitoring, accounting and fair-share
    - „unified queue" for Tier centres and opportunistic resources
- München (ATLAS, Belle II):
    - analysis of job log files by using methods of machine learning in order to find anomalies in batch or storage systems
- Wuppertal (ATLAS):
    - Container usage for IT services

job hiring in progress (Freiburg)
manpower hired (Wuppertal, München)

subject area B

The idea of subject area B is to make use of the technologies developed in subject area A and to bundle and adjust them for various target systems.

Extensive tests under realistic conditions, e.g. reliability and scalability, are necessary.

For this especially the current Tier1/2 sites are important partners since here the necessary know how and the corresponding hardware for tests are available.

Subject Area B

➢Application and test of virtualized software components
in the environment of heterogeneous computing resources

B1) Tests der Technologiekomponenten	**B2) Job- und Ressourcenmanagement**
Implementierung und Tests auf verschiedenen Plattformen von • Speicher- und Cachinglösungen und • virtualisierter Dienste (Datenbanken, Monitoring, Accounting).	Jobverteilung und Überwachung in der Umgebung heterogener Computingressourcen unter Einbeziehung von Containervirtualisierung.
B3) Virtualisierung von Nutzerjobs	**B4) Kombinierte Tests**
• Erfassung der Anforderungen, • Bestimmung und Erzeugung der Laufzeitumgebung, • Erstellung des Containers und von Metadaten und • Checkpointing von Containervirtualisierung.	Testen von Gesamtsystemen (Speicher, Dienste, Ressourcemanagement) auf verschiedenen Plattformen in Bezug auf • Installations- und Wartungsaufwand, • Performance, • Skalierbarkeit und • Robustheit.

Area B:
 Application and test of virtualized software components
 in the environment of heterogeneous computing resources

B1:

tests of technology components

participants:
- Frankfurt (ALICE):
    - tests and documentation of Singularity containers for Grid jobs (A1)
    - tests and documentation of „disk caching on the fly" (A2)
    - both with respect to simple installation, performance, scalability
- München (ATLAS, Belle II):
    - tests with XCache for ATLAS (A2)
    - testing job log file analysis procedure via ML for Belle II (A3)
- Wuppertal (ATLAS):
    - Container usage for IT services

job hiring in progress (Frankfurt)
manpower hired (München, Wuppertal)

Area B:
 Application and test of virtualized software components
 in the environment of heterogeneous computing resources

B2:

job and resource management

participants:

- Freiburg (ATLAS):
  - optimising the operation parameters of the „unified queue" developed in A3
    - job distribution based on monitoring of complex systems
    - provisioning of VMs and Containers
    - VM and Container management
    - handling of job meta data
    - dynamic integration of temporary resources

job hiring in progress (Freiburg)

Area B:
 Application and test of virtualized software components
 in the environment of heterogeneous computing resources

B3:
virtualisation of user jobs

participants:
- München (ATLAS, Belle II):
    - testing job log analysis via ML
- Wuppertal (ATLAS):
    - Container usage for IT services

job hiring in progress ()
job hiring accomplished (München, Wuppertal)

Area B:
 Application and test of virtualized software components
 in the environment of heterogeneous computing resources

B4:
combined tests

participants:
- Frankfurt/GSI (ALICE):
    - combined tests of Container jobs@Grid getting data from „disk caching on the fly"
- Freiburg (ATLAS):
    - based on tools for monitoring, accounting, bench marking from area A complete
      workflows for complex systems shall be tested
- München (ATLAS, Belle II):
    - testing job log analysis via ML and XCache
- Wuppertal (ATLAS):
    - Container usage for IT services
- CERN (ATLAS):
    - tests with Rucio in context with Cloud/HPC and other experiments (CMS, Belle II)
- DESY (ILC):
    - testing caching with dCache
- GridKa:
    - testing disk caching solutions
job hiring in progress (Frankfurt, Freiburg)
job hiring accomplished (München, Wuppertal)

Overview Work Packages

Institute	Experiments	A1	A2	A3	B1	B2	B3	B4	C1	C2	C3	C4	D1	D2	D3
Aachen	CMS, Auger, ν-Exp.								X			X	X		X
Erlangen	CTA, IceCube									X		X			
Frankfurt	ALICE, CBM	X	X		X			X					X		
Freiburg	ATLAS	X	X	X		X		X							
Hamburg	CMS									X		X			
Karlsruhe	CMS, Belle II, Auger	X	X								X			X	
Mainz	ATLAS								X	X					
München	ATLAS, Belle II		X	X	X		X	X			X				
Wuppertal	ATLAS	X		X	X		X	X							
Associated															
CERN	ATLAS							X		X			X	X	
DESY	(ILC in WP D2)		X					X						X	
GridKa								X							
GSI		X	X					X							
Jülich	Panda													X	
Münster	ALICE, CBM									X	X				

Overview of FTEs area A

5.5 FTEs in total for
Area A: development of technologies for usage of heterogeneous resources

Standort	PI	FTE	Experiment	AP A1	AP A2	AP A3
KIT Süd	Quast	1,5	CMS	X	X	X
KIT Nord	Weber/Haungs	0,5	Auger	X		
Freiburg	Schumacher	1,5	ATLAS	X	X	X
München	Duckeck	0,5	ATLAS		X	X
Frankfurt FIAS	Lindenstruth	1	ALICE	X	X	
Wuppertal	Zeitnitz	0,5	ATLAS	X		X
Assoziiert						
Bonn	Bechtle	-	ATLAS	X		
DESY	Fuhrmann	-	CMS		X	
GSI	Schwarz	-	ALICE	X	X	

Overview of FTEs area B

4.7 FTEs in total for
Area B: application and test of virtualised software components in the environment of heterogenous computing resources

Standort	PI	FTE	Experiment	AP B1	AP B2	AP B3	AP B4
Aachen	Stahl	-	CMS				X
Bonn	Bechtle	1	ATLAS		X	X	X
Freiburg	Schumacher	0,5	ATLAS		X		X
Göttingen	Quadt	1	ATLAS				X
Frankfurt	Lindenstruth	0,5	ALICE	X			X
KIT	Quast	-	CMS/Auger	X	X		X
München	Kuhr	0,7	ATLAS/Belle II	X		X	X
Wuppertal	Zeitnitz	1	ATLAS	X		X	X
Assoziiert							
CERN	Elsing	-	ATLAS				X
DESY	Gülzow/Fuhrmann	-	-				X
GSI	Schwarz	-	ALICE				X
KIT/Gridka	Petzold	-	-				X