Mitglied der Helmholtz-Gemeinschaft

JÜLICH FORSCHUNGSZENTRUM

Fenix: Realising a new paradigm for collaborative supercomputing research infrastructures

D. Pleiter | MaX International Conference 2018 | Trieste | 29 January 2018

Fenix Goals

Establish HPC and data infrastructure services for multiple research communities

- Encourage communities to build community specific platforms
- Delegate resource allocation to communities

Develop and deploy services that facilitate federation

- Based on European and national resources

Science community driven approach

- Infrastructure realisation and enhancements based on co-design approach
- Science communities providing resources to realise infrastructure
  → HBP SGA Interactive Computing E-Infrastructure
- Resource allocation managed by community

Distinctive architectural features

- Interactive Computing Services
- Elastic Scalable Computing Services
- Federated data infrastructure tightly integrated with supercomputing resources

Disclaimer

The Fenix infrastructure is still in a design and development phase. Several aspects presented in this talk are to be considered tentative

Consortium of Fenix Resource Providers

Currently involved centres

- BSC (ES)
- CEA (FR)
- CINECA (IT)
- CSCS (CH)
- JSC (DE)

Consortium features

- European HPC centres that provide resources within PRACE-2.0
- Strong links to key science drivers

Foreseen extensibility

- Open for more partners and stakeholders

Research Communities

Brain research

- Scalable brain simulations and challenging data analytics requirements
- Building-up knowledge base as part of Neuroinformatics Platform

Materials science

- Data sets from simulations but also experiments
- European community already engaged in enabling data sharing

Genomics

- Explosion of data volumes
- Some groups start to exploit HPC infrastructures

Physical science experiments

- Data from large-scale experiments, e.g. ERIC
- Need for scalable simulations for interpreting experimental results or to process data

Common Features and Requirements

Variety of data sources

- Distributed data sources
- Heterogeneous characteristics

HPC systems as source and sink of data

- Scalable model simulations creating data
- Data processing using advanced data analytics methods

Aim for data curation, comparative data analysis and for building-up knowledge bases

→ Need for infrastructure to facilitate data sharing and high-performance data processing

Architectural Concept (1/2)

Service-oriented provisioning of resources

- Focus on infrastructure services suitable for different science communities

Support for community specific platforms

- Encourage and facilitate community efforts

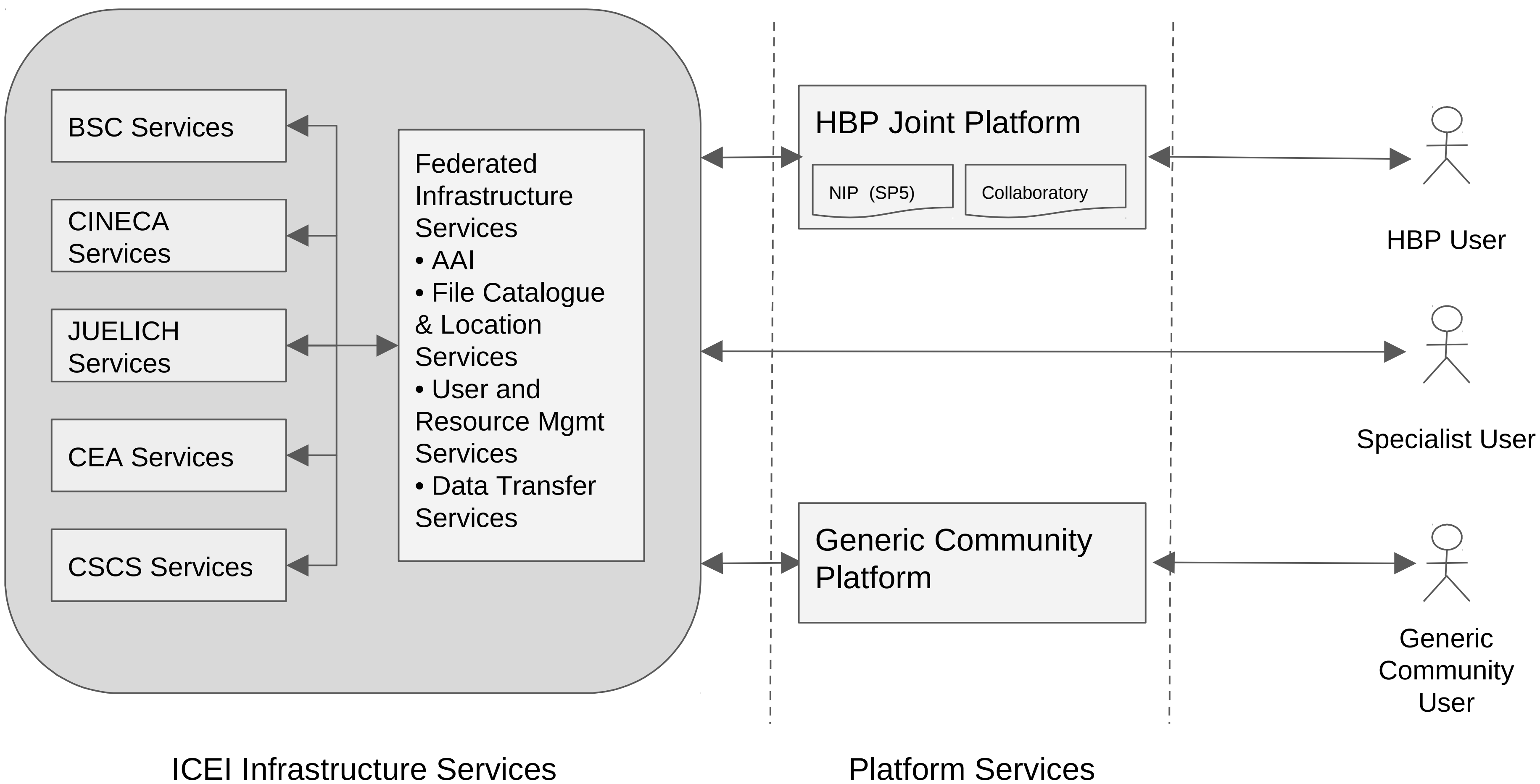
Federation of infrastructure services

- Enhance availability of infrastructure services
- Broaden variety of available services
- Optimise for data locality

Differentiation from Cloud service providers

- Limited level of virtualisation
- Business model: Account for provisioning of capabilities instead of (elastic) consumption of resources

Architectural Concept (2/2)



Overview over Planned Fenix Services

Computing services

- Interactive Computing Services
- (Elastic) Scalable Computing Services
- VM Services

Data services

- Federated Archival Data Repositories
- Active Data Repositories
- Data Mover Services
- Data Location and Transport Services

Other

- Authentication and Authorisation Services
- User and Project Management Services
- Monitoring Services

Interactive Computing Services

Interactivity

- Capability of a system to support distributed computing workloads while permitting
  - Monitoring of applications
  - On-the-fly interruption by the user
- Interactive processing of data

Architectural requirements

- Interactive access
- Tight integration with scalable compute resources
- Fast access to storage resources

Support for interactive user frameworks

- Jupyter notebook, R, Matlab/Octave

(Elastic) Scalable Computing Services

Different options for service provisioning

- Access to highly scalable compute resources with possible longer wait times
- Elastic access to a limited amount of compute resources

Possible realisation of elastic provisioning

- Free resources by means of checkpoint/resume mechanisms
- Reserve (small) amount of nodes

Considered use case

- Coupling of neuro-robotics experiments to brain simulations

Open co-design questions

- Upper limit for acceptable response times
- Scaling range

Virtual Machine Services

Use case

- Deployment of community services running 24/7
- Examples: HBP Collaboratory, AiiDA daemon

Requirements

- Allow users to flexibly create and manage VM services similar to a cloud environment
- Provide stable infrastructure services
- Integration in AAI

Architectural Concepts: Data Store Types

Archival Data Repository

- Data store optimized for capacity, reliability and availability
- Used for storing large data products permanently that cannot be easily regenerated

Active Data Repository

- Data repository localized close to computational or visualization resources
- Used for storing temporary slave replica of large data objects

Possibly: Upload buffers

- Used for keeping temporary copy of large, not easy to reproduce data products, before these are moved to an Archival Data Repository

Architectural Concepts: HPC vs. Cloud

State-of-the-art: HPC

- Highly-scalable parallel file systems
  - Scale to $O(10^5)$ clients
  - Optimised for parallel read/write streams
- Interface(s): POSIX
  - Well established interface
  - Wealth of middleware relying on this interface

State-of-the-art: Cloud

- Solutions for widely distributed storage resources
  - Optimised for flexibility
- Various interfaces: Amazon S3, OpenStack Swift
  - Typically web-based stateless interfaces
- Advantages compared to POSIX
  - Suitable for distributed environments (e.g. support for federated IDs)
  - Simple clients
  - Rich mechanisms for access control

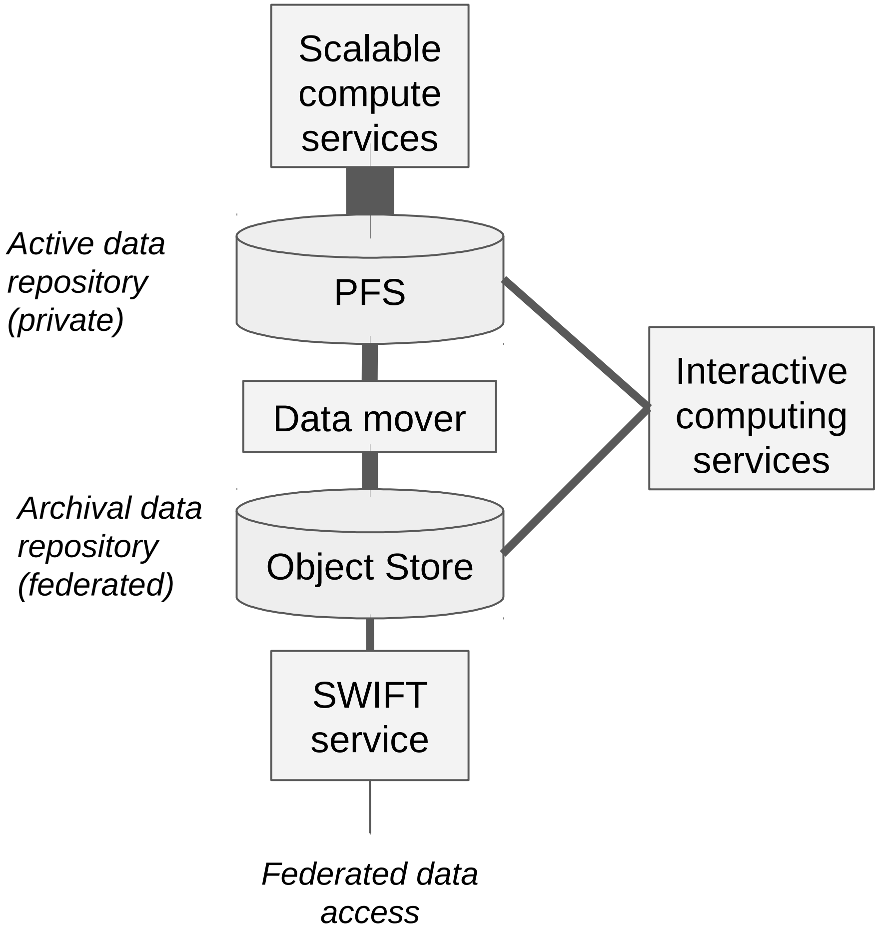
Storage Architecture

Concept

- Federate archival data repositories with Cloud interfaces
- Non-federated active data repositories with POSIX interface accessible from HPC nodes

Envisaged implementation: Mandate same technology at all sites

- Current candidate: OpenStack SWIFT



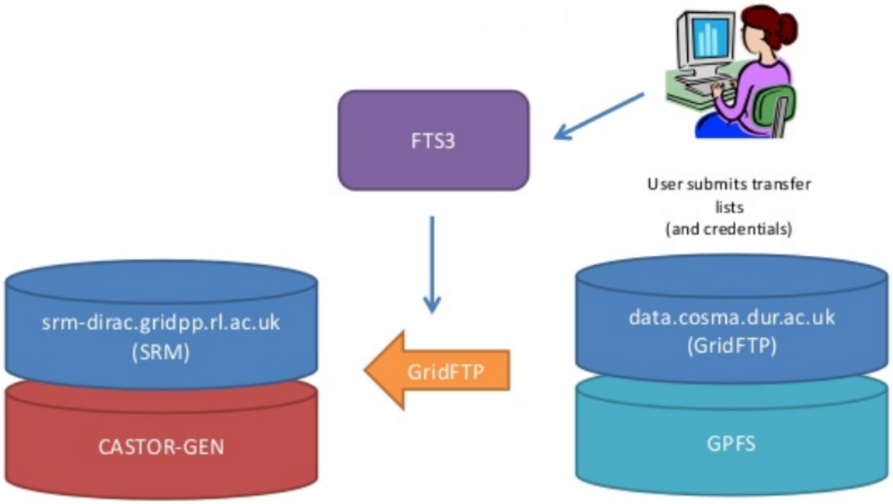
Data Location and Transfer Services

Objectives

- Enable identification of physical replicum of data object based on a Peristent Identifier by querying a central service
- Facilitate easy replication of data objects within the federated data infrastructure

Challenges

- Established technology candidates (e.g., FTS3), but incompatibilities wrt protocol and AAI



Authentication and Authorisation Infrastructure

Requirements

- All Fenix services must be in the same AAI domain
- Users should be able to authenticate with Fenix infrastructure services and community platform services in a seamless way
- The AAI must be extendable to other Fenix Communities
- Coherent authorisation

Anticipated solution

- Federation of Identify Providers (IdP)
- Central Fenix IdP Service based on OpenStack technology (and/or UNICORE)
  - Acts as proxy to forward attributes

Resource Allocation Model

Actors

- Fenix Resource Providers
- Fenix Communities
- Fenix Users

Role of Fenix Resource Providers

- Provide fixed amount of resources for given period to Fenix Communities
- Define rules for resource allocation (e.g., peer-review process)

Fenix Users

- Submit proposal for resources to relevant Fenix Community

Fenix Community

- Review proposal and award available resources to Fenix Users

Fenix Credits

Fenix Credit =
Currency for authorising resource consumption

Different types of resources

- Scalable compute resources (N_{node} × time)
- Interactive computing services (N_{node} × time)
- Active data repositories (capacity × time)
- Archival data repositories (capacity)
- Virtual Machines

Credit attributes

- Value and type of resource
- Fenix Resource Provider
- Validity period

User Management

Model

- Scientist identifies itself through virtual identity issued by accepted Identity Provider
- Scientist registers with Fenix Community to become a Fenix User

Workflow

- Scientist obtains virtual identity
- Scientist applies for membership in a Fenix Community and accepts Fenix Community Usage Agreement
- Fenix Community decides on application

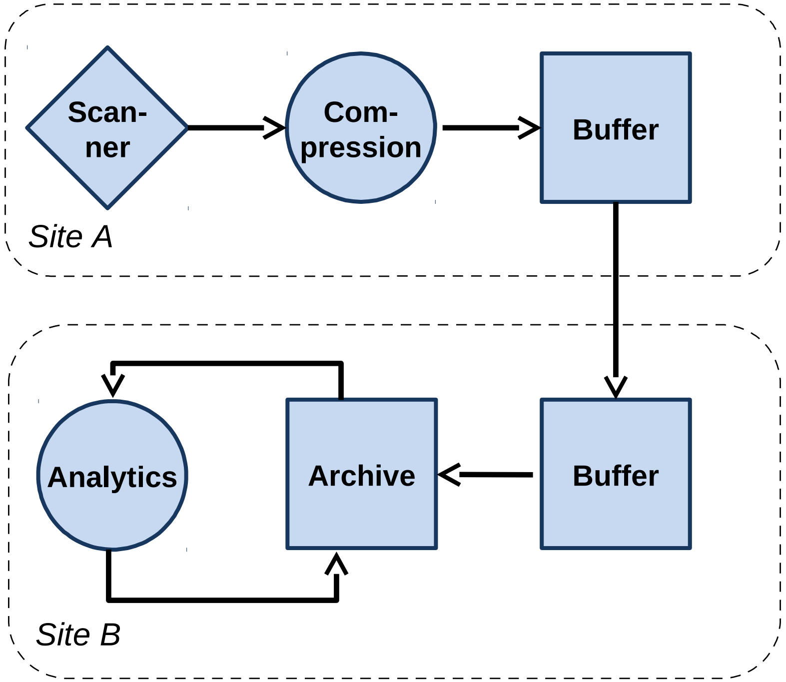
Use Case Analysis

Analysis of workflow based on abstract infrastructure model

- Data ingest
- Data repository
- Processing station
- Data transport

Use case/workload specific annotation of components

- Data transport
  - Maximum/average required bandwidth
  - Interface requirements
- Data repository
  - Maximum capacity requirements
  - Access control requirements
- Processing station
  - Data processing hardware architecture requirements
  - Required software stacks



Summary and Outlook

Strong science drivers towards data-oriented, federated HPC infrastructures

- Examples: Brain research, materials science

Many opportunities and challenges

- Federation of services including AAI
- POSIX vs. Cloud storage technologies
- Integration of interactive computing services
- New models for allocating HPC and data resources to research communities

Fenix

- Group of (currently) 5 European supercomputing centres committing to federate relevant services
- First step towards realisation of Fenix planned in context of HBP SGA ICEI (Interactive Computing E-Infrastructure)

Credits

BSC

- Javier Bartolome, Sergi Girona and others

CEA

- Gilles Wiber, Hervé Lozach, Jacques-Charles Lafoucriere, Jean-Philippe Nomine and others

CINECA

- Carlo Cavazzoni, Debora Testi, Giuseppe Fiameni, Michele Carpen, Roberto Mucci and others

CSCS

- Colin McMurtrie, Roberto Aielli, Sadaf Alam, Stefano Gorini, Thomas Schulthess and others

Jülich Supercomputing Centre

- Alex Peyser, Anna Lührs, Björn Hagemeier, Boris Orth, Dorian Krause, Thomas Eickermann, Thomas Lippert and others