INDIGO - DataCloud

Better Software for Better Science

INDIGO MidnightBlue Service Catalogue

An unmatched open modular suite of software components for Data and Cloud computing is now available for resource providers and researchers from all disciplines, all around Europe

www.indigo-datacloud.eu

Table of contents

Welcome message

Dear Readers,

It is with great pleasure that I introduce to you the first edition of the INDIGO-DataCloud Service Catalogue.

When in spring 2014 we started to define the project that eventually became INDIGO-DataCloud, we were clearly aware that scientists in Europe and elsewhere had significant troubles in designing and running their applications, in archiving and processing their data, so that these could easily and efficiently exploit distributed resources. We are scientists ourselves, and we saw that e-infrastructures were there, the cloud computing paradigm had boomed already and was getting more and more popular in both public and private enterprises, and still severe technological gaps often prevented scientific communities from effectively using the resources that in many cases were available to them.

We therefore asked ourselves how we could expand and integrate the massive amount of good software solutions that were being made available at different levels by the open source community, so that we would not reinvent the wheel, and focus instead on the concrete and specific needs of scientific communities and resource centers. We gathered around the project big European e-infrastructures, an incredible amount of developers know-how coming from many years of experience in dealing with big data processing, cloud and grid technologies, we raised the interest of key industrial players. Most importantly many multidisciplinary scientific communities, including physics, bioinformatics, cultural heritage preservation, earth sciences, astronomy, life sciences, climatology and many others, enthusiastically decided to join the INDIGO-DataCloud Consortium.

What you are reading now is the first concrete result output of the INDIGO-DataCloud project, made possible by resulting from a formidable collaborative effort. by the entire collaboration: a novel, powerful set of software services that make it possible to federate hybrid resources, to easily write and run your scientific applications on the cloud, to present them through a variety of standard interfaces. They are all freely downloadable as open source components, and are currently being integrated into many scientific applications.

We believe this is an important concrete step towards the realization of a European Open Science Cloud. Feel free to browse this Catalogue, experiment with the solutions we provide, and do not forget to send

us your feedback and comments. We are looking forward to deploying the INDIGO-DataCloud software stack in many e-infrastructures and scientific applications, and are convinced that this effort will give scientists, in Europe and elsewhere, the possibility to produce more results, more easily. This is indeed the ultimate goal of the INDIGO-DataCloud, reflected in its motto: “Better Software for Better Science”.

Happy reading,

Davide Salomoni

INDIGO-DataCloud Project Coordinator

Before we start: the INDIGO challenge

The cloud world is based on services provided at different levels. Ideally, the user of an application, in our case a researcher, accesses it through a web interface, and without needing to know which of the services of this application run on the cloud. However, to develop new applications or include new functionalities, a whole chain of services is needed.

The challenge we decided to tackle in INDIGO-DataCloud was to design a new suite of services to allow research communities from all disciplines to deploy complex applications on their own and run them on the cloud, according to their needs.

The current cloud technology, based on lightweight containers and related virtualization developments, allows to design Platforms as a Service in a relatively straightforward way. We can already see many examples of this in the industry, where open source PaaS solutions (eg. OpenShift or Cloudfoundry) are being deployed to support the work of companies in different sectors. However, supporting scientific users is more complex because of the heterogeneous and distributed nature of the services shared by researchers in their collaborations, starting at the the Infrastructure as a Service level (i.e. the data centers) and because of the inherent complexity of the scientific work requirements. The key point of the INDIGO platform is to find the right agreement to unify interfaces between the PaaS and IaaS levels.

Cloud Service Layers & INDIGO Added Value

- **IaaS**, Infrastructure as a Service, allows users to instantiate new machines capable of running different services
- **PaaS**, Platform as a Service, allows users to directly launch and combine those services
- **SaaS**, Software as a Service, provides direct access to complete applications

Behind the scene of this simple layered scheme, the setup of complex infrastructure and software components for scientific research is performed within large (private or public) datacenters, typically engaged in well-established European e-infrastructures, accessed by users from different disciplines, all of them needing solutions to combine the execution of cloud services with both efficient and simple operations.

INDIGO-DataCloud provides the means to properly federate and interface with resource centers in a standardized, simple, reliable and performing way, while allowing to deploy a suite of modular micro-services for the architectural design of complex systems.

INDIGO-DataCloud MidnightBlue Service Catalogue

On August 8, 2016 the INDIGO-DataCloud project announced the general availability of its first public software release, codenamed **MidnightBlue**. The release comes after an initial phase of requirement gatherings which involved several European scientific collaborations in areas as diverse as structural biology, earth sciences, physics, bioinformatics, cultural heritage, astrophysics, life sciences, climatology. This resulted in the development of many software components addressing existing technical gaps linked to easy and optimal usage of distributed data and compute resources.

	Research Community	Case Study/Application
Biological & Medical science	ELIXIR	Galaxy as a Cloud service
	EuroBioImaging	Medical Imaging Biobanks
	INSTRUCT	Molecular dynamics simulations
	WeNMR	HADDOCK portal
Social science & Humanities	Galleries, Libraries, Archives, Museums	eCulture science Gateway
	DARIAH	Big Data in Arts and Humanities
Environmental and Earth science	LifeWatch	Monitoring and Modelling Algae Bloom in a Water Reservoir
	ENES	Climate models inter comparison data analysis
	EMSO	MOIST-multidisciplinary oceanic information system
		TRUFA (Transcriptomes User-Friendly Analysis)
Physics & Astrophysics	LBT	Astronomical Data Archives
	CTA	Archive System for the Cherenkov Telescope Array
EGI Virtual Teams Competence Centres	Chipster	BILS
	READemption	Human Brain Project
	JAMS	BBMRI-ERIC CC
	HAPPI	DARIAH CC
	INERTIA	EPOS CC
	DRIHM	Disaster Mitigation
	CANFAR	LoFAR

Figure 2 Use cases analysed by INDIGO"

Category A: Computational Requirements, marked as CO#x
Category B: Requirements related to Storage, marked as SO#x
Category C: Requirements on Infrastructure global services, marked as PL#x

#REQ	Description
CO#1	Deployment of Interface SaaS
CO#2	Deployment of Customized computing back-ends as batch queues
CO#3	Deployment of user-specific software
CO#4	Automatic elasticity of computing batch queues
CO#5	Terminal access to the resources.
CO#6	Privileged access
CO#7	Execution of workflows
CO#8	Provenance information
CO#9	Cloud bursting
CO#10	Data-aware scheduling
CO#11	Provisioning of efficient Big Data Analysis solutions exploiting server-side and declarative approaches
CO#12	Execution across multiple centres.
CO#13	On-line processing of data
CO#14	Special hw configuration - MPI, multicore, GPGPU
SO#1	Shared storage accessible like a POSIX filesystem
SO#2	Persistent data storage
SO#3	Long-term availability of results
SO#4	Local user storage
SO#5	Availability of reference data
SO#6	Interoperability with application domain specific software and services (e.g. IS-ENES/ESGF)
SO#7	Metadata management / Database as a Service
SO#8	Share data capabilities
SO#9	Data replication
SO#10	Distributed storage
SO#11	Dropbox-like storage
PL#1	Global-level AAI
PL#2	On-line access to data
PL#3	Network configuration
PL#4	Monitoring and operation

Figure 3 List of requirements provided by the different user communities

The community requirements were translated into a set of INDIGO components, which are now released into a consistent and modular suite, offered as a contribution toward the definition and implementation of an efficient European Open Science Cloud.

INDIGO-DataCloud MidnightBlue release provides **open source components** and high-level features that address specific data and compute solutions, as described below.

Data Center Solutions. INDIGO is providing many new features allowing data and compute resource centers to increase efficiency and services for customers, such as:

- **Improved scheduling** for allocation of resources by the popular open source cloud platforms, OpenStack and OpenNebula. This provides both better scheduling algorithms and support for spot-instances.
- Support for improved IaaS resource orchestration capabilities using standards orchestration engines through the use of the TOSCA standard, for both OpenStack and OpenNebula.
- Improved QoS capabilities of storage resources for better support of high-level storage requirements, such as flexible allocation of disk or tape storage space and support for data life cycle.
- Improved and transparent support for Docker containers. This includes for example the introduction of native container support in OpenNebula.

Data Solutions. INDIGO provides a complete set of data-related features offering advanced access to distributed data. They include:

- Distributed Data Federation through several protocols, in order to support both legacy application and advanced standard interfaces such as CDMI or just simple web interfaces.
- The possibility to federate diverse storage technologies (such as Posix, Object Storage, CEPH, etc) in a seamless way, letting users exploit data and storage resources wherever they are available.

Automated Solutions. INDIGO provides a rich set of high-level automated functionalities allowing users to easily specify and deploy complex data and compute resource requirements. Some of the most innovative are:

- Improved capabilities in the geographical exploitation of cloud resources. End users do not need to know where resources are located, because the INDIGO PaaS layer hides the complexity of both scheduling and brokering.
- Standard interface to access PaaS services. INDIGO uses the TOSCA standard to hide the difference on the different way of implementing services at the PaaS level.
- Support for data requirements in cloud resource allocations: computational resources can be requested and allocated where data is stored.
- Integrated use of resources coming from both public and private cloud infrastructures.
- Deployment, monitoring and automatic scalability of existing applications.
- Integrated support for high-performance Big Data analytics.
- Support for dynamic and elastic clusters of resources. HTCondor, Torque and Mesos cluster are supported.

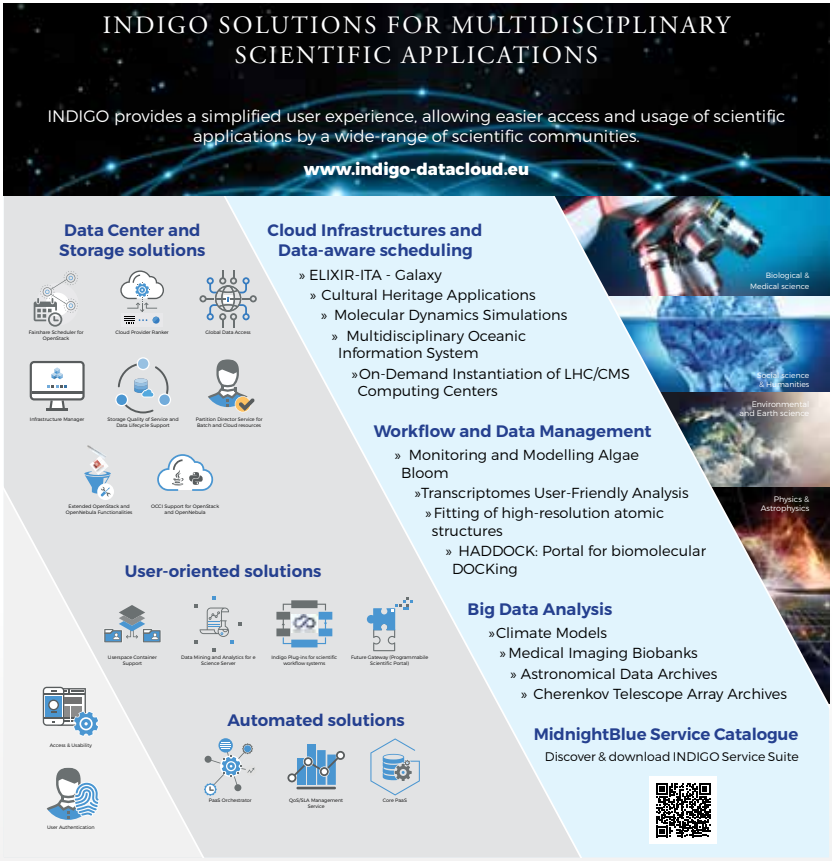
High-level User Oriented Solutions, integrating scientific applications in programmable front-ends and in mobile applications. Researchers and data managers are able to access resources through:

- Toolkits (libraries) allowing usage of the INDIGO platform from Scientific Gateways and desktop applications.
- An open source Mobile Application Toolkit for iOS and Android, serving as the base for the development of Mobile Apps.
- User-friendly front ends for building programmable, general-purpose multi-domain Science Gateways.

All the INDIGO components are integrated into a comprehensive **Authentication and Authorization Architecture**, with support for user authentication through multiple methods (SAML, OpenID Connect and X.509), support for distributed authorization policies and a Token Translation Service, creating credentials for services that do not natively support OpenID Connect.

The **INDIGO-DataCloud software** is released under the Apache 2.0 software license and can be deployed on both public and private cloud infrastructures.

INDIGO - DataCloud Software Releases Documentation is available from the official INDIGO - DataCloud Repositories http://repo.indigo-datacloud.eu., for operating systems packages and Docker containers.



Why a Catalogue?

The purpose of this catalogue is to give resource providers and researchers from all disciplines all around Europe a practical guide to identify the best INDIGO-DataCloud services to download and use to improve the quality of the research they are performing.

Index of Services

In the following pages we have listed INDIGO-DataCloud high-level services available with this first public release, with links for direct download of each components and additional information regarding related documents and cases of adoption. An online, interactive catalogue is also available at www.indigo-datacloud.eu .

1. Identity and Access Management

Short Service Name

IAM

Solution Type

Common Solution

Installation Area

The IAM service provides a layer where identities, enrolment, group membership, attributes and policies to access distributed resources and services can be managed in a homogeneous and interoperable way. It supports the federated authentication mechanisms behind the INDIGO AAI.

The IAM service provides user identity and policy information to services so that consistent authorization decisions can be enforced across distributed services.

Identity and Access Management is provided through multiple methods (SAML, OpenID Connect and X.509) by leveraging on the credentials provided by the existing Identity Federations (i.e. IDEM, EDUGAIN, etc). Distributed authorization policies and Token Translation Service will guarantee selected access to the resources as well as data protection and privacy.

2. Fairshare Scheduler for OpenStack

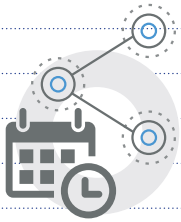
Short Service Name

Synergy

Solution Type

Data Center Solution

Installation Area



Synergy, the INDIGO Fairshare Scheduler for Openstack, is an extensible all-purpose management service for integration with OpenStack Infrastructures. It is implemented by a collection of independent pluggable tasks and executed periodically (e.g. cron jobs) or interactively (e.g. RESTful API). Synergy can be used to allocate a set of dynamic OpenStack resources to be shared among different projects. Moreover, Synergy offers a queuing mechanisms for requests until relevant resources are available. It can oversee the instantation of both virtual machines and containers managed via the nova-docker service.

3. Partition Director Service for Batch and Cloud resources

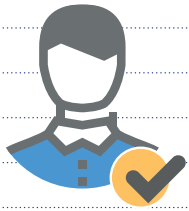
Short Service Name

Dynpart

Solution Type

Data Center Solution

Installation area



Dynpart, the Partition Directore Service for Batch and cloud resources, facilitates the management of a hybrid data center that provides both batch-system based services and cloud-based services. Physical computing resources, in fact, can act as member of batch system cluster or as compute node in a cloud environment.

Dynpart can easily manage such mutual exclusive approach of physical resources making the data center dynamic and flexible.

4. Cloud Provider Ranker

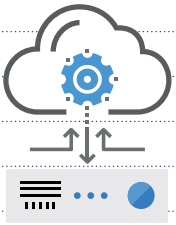
Short Service Name

CloudProviderRanker

Solution Type

Data Center Solution

Installation Area



CloudProviderRanker is a standalone REST WEB Service which ranks cloud providers basing on rules implemented with the Drools framework[1].

CloudProviderRanker uses a Business Rules Management System (BRMS) engine implementing an experimental basic ranking algorithm based on user priorities for resource selection.

- The CloudProviderRanker checks if preferences have been specified; if this is the case, then they have absolute priority over any other provider's information (like monitoring data).
- If preferences are not specified, for each provider the rank is calculated as a sum of SLA's rank and a combination of monitoring data, each of them conveniently normalized with weight specified in a Ranker's configuration file.

1 http://drools.org/

5. Global Data Access

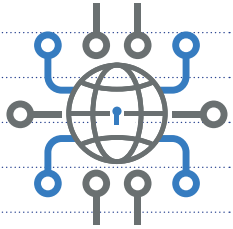
Short Service Name

OneData

Solution Type

Data Solution

Installation Area



Global Data Access is the global data management system providing easy access to distributed storage resources and supports a wide range of use case, from data management to data-intensive scientific computations.

INDIGO communities

Physics & Astrophysics

«**We tried Onedata** for the Large Binocular Telescope (LBT) use case and configured this service for a simulated distributed archive. We tried to install and configure all Onedata components (Onezone, Oneprovider and Oneclient) on dedicated virtual machines, deploying docker images. The main goal was to use Onedata to store and distribute data to different sites according to defined data policies».

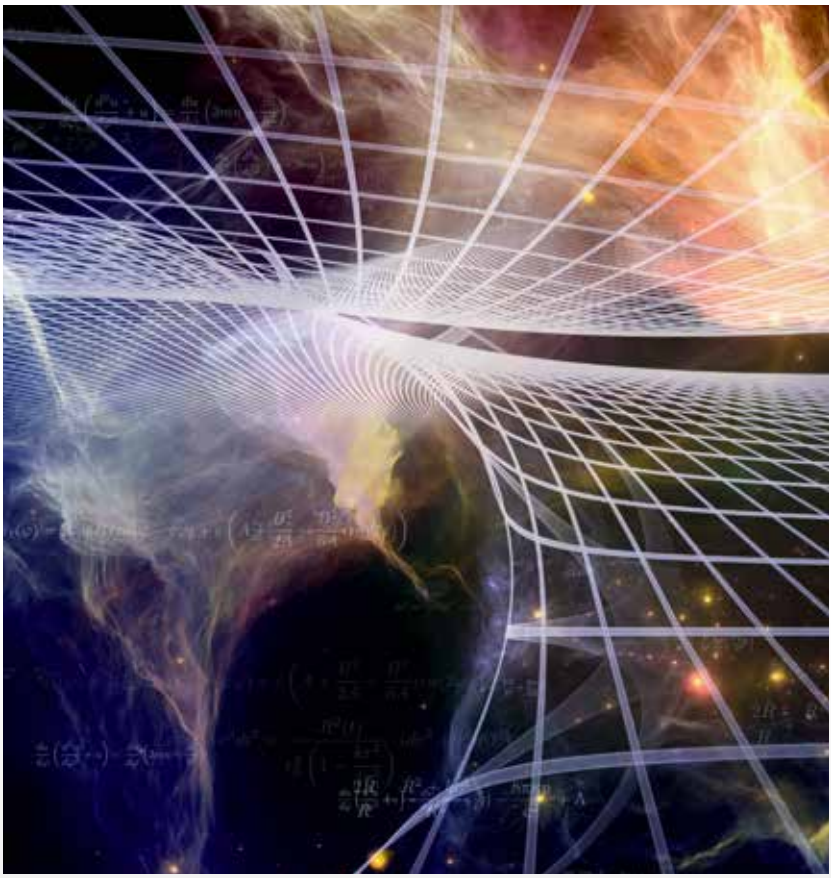
«Onedata provides an intuitive configuration interface and a very flexible framework to store data using global distributed storage providers».

Andrea Bignamini,
Researcher at the Astronomical Observatory of Trieste, Italy

Physics & Astrophysics

«**OneData allows us** to implement a first prototype of distributed archive for the Cherenkov Telescope Array (CTA) project. The distributed architecture of the CTA Archive will allow to lower costs with respect to a single huge data centre including easy manageability and maintenance».

Eva Sciacca,
Researcher at INAF – Astrophysics Observatory of Catania, Italy



6. PaaS Orchestrator

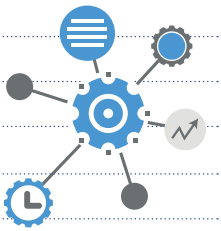
Short Service Name

Orchestrator

Solution Type

Automated Solution

Installation Area

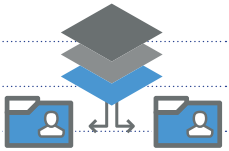


PaaS Orchestrator is the core component of the PaaS layer. It collects high-level deployment requests from the software layer, and coordinates the resource or service deployment over dynamic Mesos clusters or directly over IaaS platforms.

7. Userspace Container Support

Short Service Name

udocker



Solution Type

User-oriented Solution

Installation Area

Udocker is the tool to execute simple Docker containers in user space without requiring root privileges.

INDIGO udocker does not require any type of privileges nor the deployment of services by system administrators, miming a subset of the docker capabilities with minimal functionalities:

- Basic download and execution of docker containers by non-privileged users in Linux systems were docker is not available.
- Access and execution of docker containers in Linux batch systems and interactive clusters that are managed by other entities such as grid infrastructures or externaly managed batch or interactive systems.

It can be downloaded and executed entirely by the end user.

8. Data Mining and Analytics for eScience Server

Short Service Name

Ophidia

Solution Type

User-oriented Solution

Download

Ophidia is the framework that supports data-intensive analysis to exploit parallel computing techniques and smart data distribution methods. It exploits an array-based storage model and a hierarchical storage organisation to divide and distribute multidimensional scientific datasets over multiple nodes.

INDIGO communities

Physics & Astrophysics

«**I gained practical experience** with the Ophidia framework when working on the INAF-CMCC use case on integration into Ophidia of the Flexible Image Transport System (FITS) format. This use case aims at extending Ophidia's data-intensive analysis to the astronomical field where also multidimensional scientific datasets are commonly used. We experienced the powerful functionalities of the INDIGO platform, which provides native support in terms of array data types, in the development of a workflow for the reduction of the astronomical images.

I also tested Ophidia as end-user by downloading the virtual machine that implements all its features. I found the terminal user friendly and the documentation well explained.»

Elisa Londero,
Researcher at the Astronomical Observatory of Trieste, Italy

9. Infrastructure Manager

Short Service Name

IM

Solution Type

Data Center Solution

Installation Area

Infrastructure Manager deploys complex and customized virtual resources on different IaaS Cloud platforms (such as AWS, OpenStack) providing an abstraction layer to the definition and the provision of such resources.

IM enables computing resource orchestration using OASIS standard languages (i.e. TOSCA protocol).

Moreover, it eases the access and the usability of IaaS clouds by automating the VMI (Virtual Machine Image) selection, deployment, configuration, software installation, monitoring and update of the virtual infrastructure.

INDIGO communities

Biological & Medical science

«The Infrastructure Manager gives you full flexibility to write your own recipes to deploy and configure your cluster. The user can also use already available RADLs to deploy a cluster. The interface is user friendly, and tutorials are helpful. The user can also store his/her credentials for not only one but multiple resource providers (Amazon, Google, EGI etc), which is a big plus"

Zeynep Kurkcuoglu,

Bijvoet Center for Biomolecular Research, the Netherlands

INDIGO Champion for the "Virtualization of the HADDOCK portal" use case

10. OCCI support for OpenStack and OpenNebula

Short Service Names

OOI, rOCCI, jOCCI, pOCCI

Solution Type

Data Center Solution

Installation Area

An implementation of the Open Grid Forum's Open Cloud Computing Interface[2] (OCCI) for OpenStack[3]. Extended AWS support for rOCCI. Python and Java libraries for OCCI support.

OCCI is a framework that allows management of resources in arbitrary cloud management frameworks through the OCCI standard.

This allow the users to exploit same CLI and APIs to manage diverse Cloud Management Framework in a transparent way.

2 http://occi-wg.org/

3 https://www.openstack.org/

11. Extended OpenStack and OpenNebula Functionalities

Short Service Names

OPIE, nova-docker, Keystone AAI Support

TOSCA in HEAT, ONEDock, DockerHub Sync

OpenStack Client & Nova Client

OpenStack Identity Authentication Library

Solution Type

Data Center Solution

Installation Area

This service provides a set of common solutions ranging from authentication and authorization (implementing the INDIGO AAI service) to Docker support for OpenStack and OpenNebula open source cloud models.

Improved scheduling for allocation of resources by popular open source cloud platforms, i.e. OpenStack and OpenNebula.

- Enhancements addressing both better scheduling algorithms and support for spot-instances.
- Dynamic partitioning of resources among "traditional batch systems" and cloud infrastructures (for some LRMS) is also supported.

Support for standards in IaaS resource orchestration engines through the use of the TOSCA standard.

- This functionality overcomes the problems of portability and usability that ways of orchestrating resources in cloud computing frameworks widely differ among each other.

Improved IaaS orchestration capabilities for popular open source cloud platforms, i.e. OpenStack and OpenNebula.

- Enhancements includes the development of custom TOSCA templates to facilitate resource orchestration for end users, increased scalability of deployed resources and support of orchestration capabilities for OpenNebula.

12. Core PaaS

Short Service Names

Kubernetes, Monitoring, Accounting

Solution Type

Automated Solution

Installation Area

Core PaaS provides the basic functionalities and tools to steer the performance of all the PaaS services available in the infrastructures. In particular:

- Availability and scalability of the core services
- Monitoring of the Computational and Storage resources and of the PaaS μServices
- Accounting of the resource usage in terms of computing and storage

13. Storage Quality of Service and Data Lifecycle support

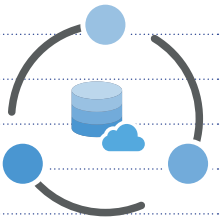
Short Service Name

CDMI

Solution Type

Data Solution

Installation Area



This solution implements the INDIGO-DataCloud CDMI Server, a set of functionalities aimed at Improving QoS capabilities of storage resources for better support of high-level storage requirements, such as flexible allocation of disk or tape storage space and support for data life cycle.

- This is an enhancement also with respect to what is currently available in public clouds, such as Amazon Glacier and Google Cloud Storage.

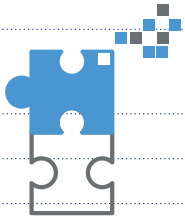
CDMI provides the official reference implementation of the SNIA Cloud Data Management Interface (CDMI), an ISO standard, and also a Spring Boot application port of the SNIA CDMI-Server. The CDMI server has been extended to support Quality-of-Service (QoS) and Data Life-cycle (DLC) operations for multiple storage back-ends like dCache, Ceph, GPFS, Gemss+TSM, StoRM and HPSS.

14. Future Gateway (Programmable Scientific Portal)

Short Service Names

fgAPIServer, fgAPIServerDaemon

PortalSetup, fgTools



Solution Type

User-oriented Solution

Installation Area

Future Gateway is a programmable interface of a RESTful API Server, compliant with CSGF APIs specifications, able to provide an easy acces to the PaaS layer by leveraging on recent Web technologies.

The FutureGateway consists of a set of software components able to build, or assist existing web portals or other community oriented interfaces to become Science Gateways. FutureGateway allows the access to distributed computing resources such as grid, cloud and HPC. FutureGateway comes from a four years experiences gained with a similar component. In particular the following key features have been identified:

- Provide a more flexible way accessing the distributed computing services.
- Leave to the FutureGateway adopters the choice of the backward portal technology.
- Provide the most simple way to develop ScienceGateway applications.

The FutureGateway is composed by a set of tools:

- FutureGateway API Server
- FutureGateway API Server Daemon
- FutureGateway jSAGA Adaptors
- FutureGateway Portal Setup
- fgTools

15. INDIGO Plug-ins for scientific workflow systems

Short Service Names

indigoclient, indigoKepler

Solution Type

User-oriented Solution

Installation Area

Workflow management systems provide an infrastructure for the set-up, performance and monitoring of a defined sequence of tasks, arranged as a workflow by using resources and services made available by providers and other communities.

The INDIGO Plug-ins for scientific workflow systems are aimed at interacting with the already available INDIGO solutions, facilitating the deployment of complex scientific workflow.

This plugin could be used in order to allow to exploit the INDIGO APIs and submit requests for execution of applications directly from the workflow manager to the INDIGO APIs.

- Indigo-DC client API is Java based library that can perform calls to Future Gateway API. This library provide basic means for accessing resources provided by Indigo-DC project. It can be used by Java based applications as Workflow manager like Kepler.

16. QoS/SLA Management Service

Short Service Name

QoS/SLA

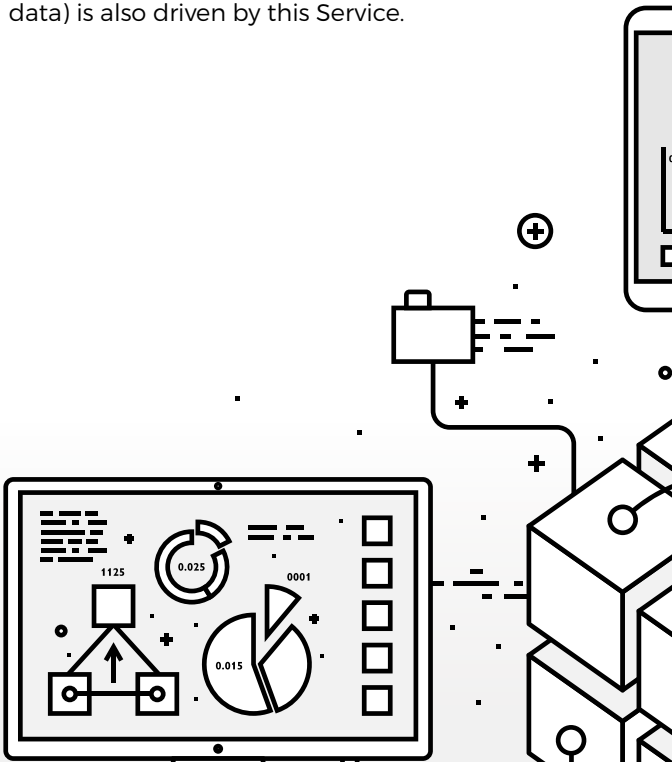
Solution Type

Automated Solution

Installation Area

QoS/SLA Management Service allows to define a proper Service Level Agreements among customares and providers (resources or serivices). Moreover the service can drive the decisions of the PaaS Orchestator and the Cloud Provider Ranker (already presented) on the basis of the signed SLA.

Support for data requirements in resource allocations (to request specific QoS categories when requesting computing resources and/or storing or accessing data) is also driven by this Service.



17. INDIGO Mobile Toolkit

Short Service Name

Mobile Toolkit

Solution Type

User-oriented Solution

Installation Area

INDIGO Mobile Toolkit provides the Software Development Kit (SDK) suitable to build mobile applications able to exploiting INDIGO PaaS services, making managers and researchers able to interact with all the layers of a computing farm (services and resources) using mobile devices. The toolkit is provided for Android.

EGI and INDIGO-DataCloud Integration

EGI is a federation of computing and storage resource providers united by a mission to support research and development It is governed by the participants represented in the EGI Council and coordinated by the EGI Foundation.

The INDIGO team has recently completed a deliverable outlining the possible roadmaps for technical integration with different infrastructures, including EGI. The full "Technological Methods of Integration with other e-Infrastructures" deliverable is available online. One of the chapters is dedicated to potential routes of technical integration between EGI & INDIGO.

Authentication and authorisation

Authentication and authorisation in EGI is commonly based on personal X.509 certificates and on the user's membership in a Virtual Organisation: the community attributes are managed by the Virtual Organisation Management System (VOMS) that issues an attribute certificate (AC) describing the user membership. EGI AAI is evolving and currently alternate AAI technologies are available for users: username/password authentication based on SAML and OIDC protocols, support for third parties attribute management, and credential translation services.

The INDIGO Identity and Access Management (IAM) provides a set of advanced interoperable solutions:

- A login service (based on OpenID-Connect) able to federate different identity providers;
- A group membership service that will allow to group users in organisations;
- An authorisation service that provides the tools to define and enforce authorisation policies over the protected resources;
- Support for controlled delegation of privileges across services;
- A Token Translation Service, bridging the gap between Authentication mechanisms that do not support OpenID-Connect.

The INDIGO IAM supports the legacy authentication technologies as well as the protocols recently added to the EGI AAI offer. The flexibility of the INDIGO IAM allows an easier integration and support for potentially all use cases by providing end users with a unified view of identities and privileges, together with the tools needed to enable a secure composition of services from multiple providers in support of scientific applications.

Computing

Grid-job submission via the INDIGO platform

Integrating INDIGO capabilities with the EGI grid infrastructure will allow users to submit jobs and manage virtual machines via already available portals using the advanced features provided by the INDIGO platform. Users communities could have an easy and straightforward way of instantiating a full community portal in a cloud environment, letting the INDIGO PaaS layer to deal with high-availability and scaling e.g. to accommodate unpredictable workloads. Moreover, already available portals could easily exploit the new INDIGO IaaS and PaaS features by implementing simple APIs provided by the INDIGO SaaS layer.

Application/service deployment on cloud via the INDIGO platform

Scientific communities that need to run a long running service or an application on the EGI Federated Cloud will be able to deploy them in a simple and transparent way, using both APIs and web user interfaces, thanks to the functions provided by the INDIGO platform. The integration in the EGI Federated Cloud will occur at both IaaS and PaaS levels, allowing end users to dynamically specify the composition of the cluster of services requested for a given use case or application such, as for example, databases, load balancers, automatic scalability features or proximity to certain datasets.

The INDIGO PaaS will then compose the requests and automatically instantiate the needed services, taking care of monitoring the status of each service, automatically scaling the computational resources needed, and restarting the services in case of failure. This could exploit for example the IaaS resources available via the EGI Fed Cloud.

Storage

EGI grid storage infrastructure

The storage management of the EGI grid infrastructure is based on a specific class of services called “storage elements”. The INDIGO project extends this concept and makes it possible to access files on the EGI grid storage elements in terms of both transfers and read access, exploiting the protocols supported by EGI and providing a unified view of storage resources. End users utilizing the INDIGO platform will be able to import/export data transparently from/to the EGI infrastructure and process it using applications and/or services deployed through the INDIGO PaaS layer. The INDIGO PaaS layer will take care of performing the needed data transfers on behalf of the users, before/after executing the requested application/service.

EGI cloud storage infrastructure

The storage layer of the EGI Federated Cloud environment is mainly providing **Block** and **Object** Storage capabilities.

Block Storage: for this type of storage, the INDIGO project enables seamless use of Block storage resources available at each site. Computing resources instantiated through the INDIGO PaaS are able to transparently connect to distributed cloud block storage, presenting it to applications as if it was local. Replication and caching facilities are also provided by the INDIGO unified storage layer. Moreover, the use of local block storage resources can be facilitated through the automation provided by TOSCA Templates and the INDIGO PaaS layer. This way, users can compose a cluster of services requesting block storage as part of the cluster configuration. All the needed preparatory steps are handled by the INDIGO services.

Object Storage**:** in order to access this type of storage in the EGI Federated Cloud, a CDMI interface is suggested as the standard protocol at each site, regardless of the site-specific technological implementations. The INDIGO project is extending a standard CDMI implementation to also support advanced features such as QoS and Data Lifecycle Management so that, for example, different types of storage technologies (cold or inline storage) or time-dependent access policies to data can be required by users. Moreover, all the available CDMI endpoints can be easily federated into a single geographical storage federation, that is also able to provide seamless access to the storage in order to run legacy applications.

The INDIGO components will contribute to the technical realisation of the EGI Open Data Platform, which is a big milestone of the EGI strategy.

Network virtualisation

Regarding networking, integration of INDIGO products may add a new set of capabilities for EGI Federated Cloud users. Contributions by INDIGO will allow them to set up, manage and use private networks on-demand in participating cloud sites - a feature that is currently unavailable. Read the **full deliverable** for a detailed overview.

Technical Support

Most complex software contains bugs, and we are not an exception. One of the features of free and open source software is the ability to report bugs, helping to fix or improve the software you use. The INDIGO-DataCloud project uses the GGUS (Global Grid User Support) tool as its user support system. It provides sophisticated search functionality, report generation, interfaces to bug tracking systems used by different middleware components, and automatic ticket reminder including escalation indication. Please use the INDIGO-DataCloud Catch-All GGUS Support Unit or directly contact us through the indigo-su@lists.indigo-datacloud.eu mailing-list.

Share the INDIGO Experience

Developers, researchers and IT enthusiasts: feel free to write to info@indigo-datacloud.eu to ask for more information on how to deploy your PaaS-based solution for your work. For automatic notifications, you can register to the INDIGO-DataCloud RSS release feed or subscribe to the INDIGO-DataCloud Announce Mailing list. You can also socialize with us via Twitter, Facebook and LinkedIn. Finally, you can also subscribe to **INDIGO Newsletters** and receive communications about the project, such as new releases, community events and other events where to meet the INDIGO team, tutorials, workshops, webinars, guides, and more.

- INDIGO-Datacloud Website: https://www.indigo-datacloud.eu
- Technical Support: https://www.indigo-datacloud.eu/indigo-support-and-technical-services
- Twitter: https://twitter.com/indigodatacloud
- Facebook: https://www.facebook.com/indigodatacloud
- Linkedin: https://it.linkedin.com/in/indigodatacloud

Updates and new releases of the INDIGO services will be implemented in the forthcoming months.
The first scientific applications and use cases adopting this first INDIGO release are expected starting from September 2016.

European Commission

INDIGO - DataCloud receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement RIA 653549