

INDIGO - DataCloud RIA-653549

The INDIGO DataCloud project





INDIGO-DataCloud is co-founded by the Horizon 2020Framework Programme

INDIGO-DataCloud

- An H2020 project approved in January 2015 in the EINFRA-1-2014 call
 - 11.1M€, 30 months (from April 2015 to September 2017)
 - Who: 26 European partners in 11 European countries
 - Coordination by INFN (D. Salomoni project director)
 - Including developers of distributed software, industrial partners, research institutes, universities, e-infrastructures
 - What: develop an open source Cloud platform for computing and data ("DataCloud")
 - For: multi-disciplinary scientific communities
 - Where: deployable on hybrid (public or private) Cloud infrastructures





The INDIGO-DC approach



- Access data and use resources as a "a "big pool" of computing and storage, without the need to know their type or location
- Indigo-DC developments targets 3 software levels
 - Infrastructure level (WP4)
 - Computing centers offering resources in an Infrastructure as a Service (IaaS) mode
 - Developments targeted to OpenStack and OpenNebula
 - Platform layer (WP5)
 - Set of services whose objective is to leverage disparate hardware resources coming from the Infrastructure level (Grid of distributed clusters, public and private clouds, HPC systems) to enhance the user experience
 - Advanced tools for computing and for processing large amounts of data
 - User Interfaces layer (WP6)
 - User friendly frontends such as Scientific Gateways, desktop and mobile interfaces

INDIGO - DataCloud

The INDIGO-DC architecture





Some Infrastructure level (WP4) items ...

- ... that could be useful for our needs
- Improved scheduling for OpenStack and OpenNebula (Cloud Middleware Frameworks)
 - Synergy (Padova): see next slide
 - Partition director (CNAF): move resources from Grid to Cloud and viceversa, maintaining the overall shares
 - Spot instances (CSIC): special instances that use free resources; they are killed if resources are needed for 'normal' instances
- Support for containers (Padova)
 - Another more efficient approach (wrt virtualization) to implement isolation between applications sharing the same hw resources
- All these services can be used also 'stand-alone', i.e. not only when invoked by the higher level Indigo-DC servicesS



Synergy (scheduling for OpenStack)

- The problem in the Cloud environments
 - Static partitioning: the resource allocation to the user projects can be done only by granting fixed quotas
 - One project cannot exceed its own quota even if there are unused resources allocated to other projects
 - The user request fails (and is lost) if no resource can satisfy it
- Synergy allows the IaaS administrators to allocate a subset of resources (dynamic resources) to be shared among different projects, besides the ones statically partitioned
 - Dynamic resources are handled by Synergy which allows the definition of fair-share policies (then enforced by Synergy implementing the algorithm used in the SLURM batch system)
 - Synergy also provides a persistent priority queuing mechanism for handling user requests that can not be immediately fulfilled





INDIGO - DataCloud

Indigo Releases

- On Aug 8th, Indigo-1 (codenamed MidnightBlue) was released
 - Data center solutions, data solutions, automated solutions, high level user oriented services, common services
 - Support for Ubuntu 14.04 and CentOS 7
 - Support for at least OpenStack Liberty and OpenNebula v. 4.14
- Monthly updates for bug fixes
- Indigo-2 will be released at the end of March 2017

Padova in INDIGO



- Mostly focusing on developments useful for our Cloud
- Implementation of the Synergy service (WP4)
 - Also with some contributions by IN2P3
- Implementation of CloudProvider Ranker (aka RuleEngine): WP5
 - Rule-based engine that manages the ranking among the resources that are available to fulfil the requested services
- Support of container for OpenStack: WP4
 - Assessment and troubleshooting of nova-docker, in collaboration with LIP and CSIC
- Testbed related activities: WP3
- Financial management: WP1
- Involved people
 - A. Dorigo, F. Fanzago, E. Konomi, M. Segatta, M. Sgaravatto, M. Verlato, L. Zangrando
 - L. lacono, A. Lombardo, G. Salvato



Backup slides

Indigo-1 content (1/3)



- Data center solutions
 - Improved scheduling for allocation of resources
 - Synergy
 - Opie (support for spot instances, i.e. VMs that are killed/suspended if resources are needed to run 'normal' instances)
 - Partition Director: move Grid resources to Cloud and viceversa when needed
 - Support for improved resource orchestration capabilities using standards orchestration engines
 - Improved QoS capabilities of storage resources
 - Improved and transparent support for Docker containers
 - Nova-docker, ONEDock, ...
- Data Solutions
 - Distributed Data Federation through several protocols
 - Possibility to federate diverse storage technologies (such as Posix, Object Storage, CEPH, etc) in a seamless way
 - Products: ONEDATA, CDMI server, dCache

Indigo-1 content (2/3)



- Automated Solutions
 - Improved capabilities in the geographical exploitation of Cloud resources
 - Standard interface (TOSCA) to access PaaS services
 - Support for data requirements in Cloud resource allocations
 - Deployment, monitoring and automatic scalability of applications
 - Support for dynamic and elastic clusters of resources
 - Products: Orchestrator, ONEData, Kubernets, IM, ...

Indigo-1 content (3/3)



- High-level user oriented services
 - FutureGateway, INDIGO OpenMobile Toolkit, etc.
- Common Solutions
 - Authentication and Authorization Architecture, with support for user authentication through multiple methods (SAML, OpenID Connect and X.509),
 - Support for distributed authorization policies
 - Token Translation Service, creating credentials for services that do not natively support OpenID Connect.