EOSC-Hub Technology Area description

Workflow management and user interfaces and Data analytics

Marcin Płóciennik (PSNC)

eosc-hub.eu

@EOSC_eu

Dissemination level: Public/Confidential *If confidential, please define:*
Disclosing Party: (those disclosing confidential information)
Recipient Party: (to whom this information is disclosed, default: project consortium)

EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 777536.

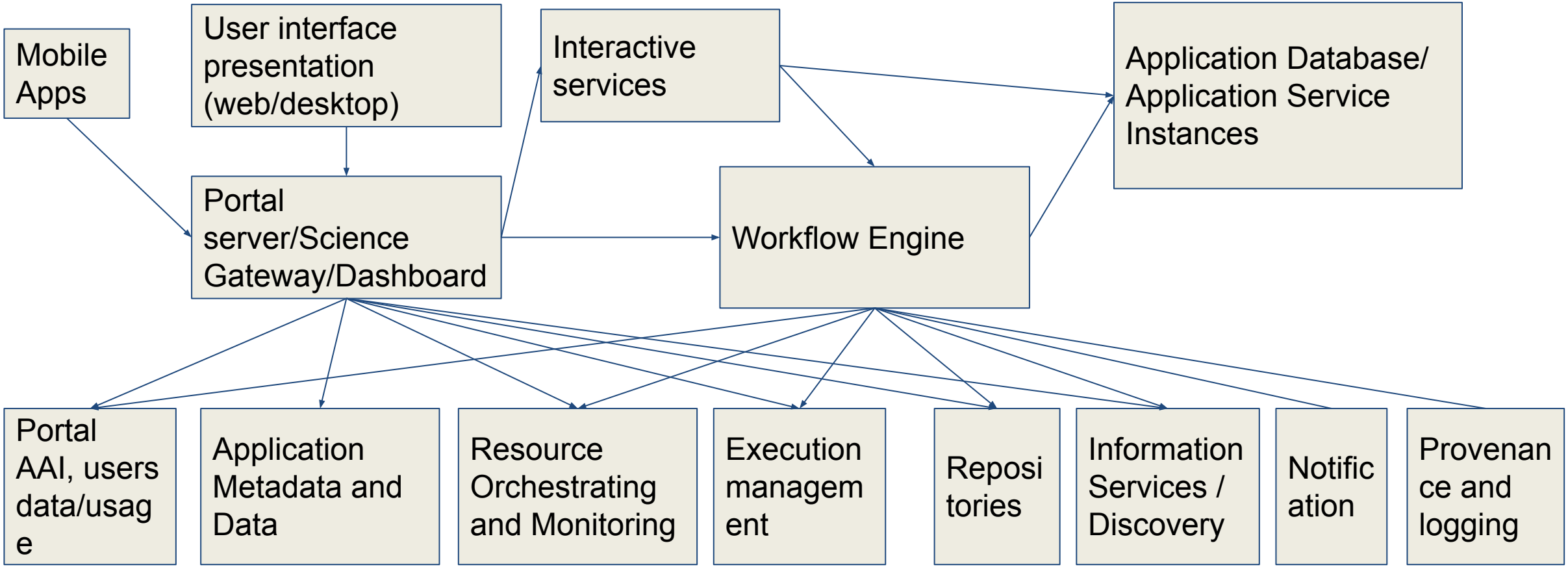
The macro-features already identified

- Graphical user interfaces: Portals/Dashboards/Science Gateways
- Marketplace - user facing part of service ordering procedure
- Data analytics service
- Big data analytics
- …. more in to be described as next steps

Portals

- description:
  - end user facing services, offering a wide range of functionalities.
  - mostly these are dedicated community services (thematic services) that compose the whole end user environments or general purpose portals supporting wide range of the use cases and underlying services and infrastructure support.
  - some of the portals are in the form of the dashboards or Science Gateways (SG) made by a set of interconnected services, pluggable components and libraries. SG are mainly high level user interfaces allowing seamless access to computing, storage, AAI and number of the functionalities from most of the EOSC Tech areas that needs to be exposed to users - can be defined as core part of Virt. Res. Env.
  - Usually integrated with workflow/pipeline engines
  - usually composed by a set of frontends, APIs, libraries and toolkits in order to cover at most as possible a wider range of different Mobile/Web front-ends and underlying OSes and programming languages.

Portals/Gateways/Dashboards - architecture



EOSC-hub Portals

- standard, protocols and api

Standard	Short description
TOSCA	TOSCA to describe resources to allocate
JSAGA	using SAGA standard to access DCI
OCCI	OCCI standard to create Virtual Resources
JSON	Any REST API call uses JSON streams for input and output

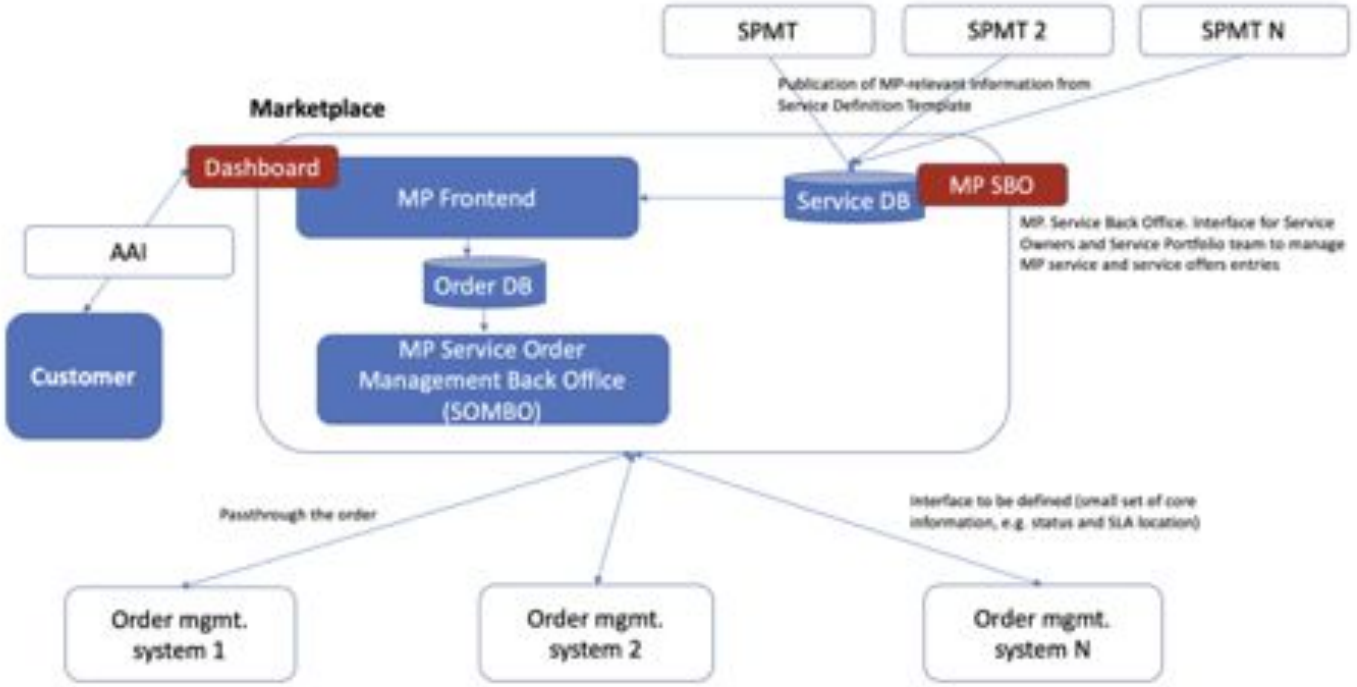
Protocol/API	Short description
OAuth2	
SAML/OIDC	
REST APIs	Interfaces to the underlying services

Portals

- interoperability guidelines
  - standards from all underlying layers (PaaS- Orchestration, Federated AAI, Data Management)
- samples of services offering the feature
  - FutureGateway EOSC-hub Service http://fgsg.egi.eu is a FG demonstrator which currently provides two applications: R-Studio and Chipster., tryFG instantiate a complete FG environment to practice with the FutureGateway.
  - WebNMR portals - Online portals for structural biology analytics (providing access to a variety of tools for structural biology and computational modelling covering NMR, cryo-electron microscopy, and integrative modelling): DISVIS, POWERFIT, HADDOCK, GROMACS, AMPS-NMR, CS-ROSETTA, UNIO, FANTEN, AMBER, Extended descriptions of the suite portals available at https://confluence.egi.eu/display/EOSC/WeNMR+VA

Marketplace, user facing part of service ordering procedure

- description + architecture:
  - dedicated platform for presentation, ordering and access to services.
  - it is meant to support Service Management, and along with Service Order Management Back Office it provides Service Order Management in EOSC-hub.



Marketplace, user facing part of service ordering procedure

- standard, protocols and api

Standard	Short description
eInfraCentral based EOSC-hub STD	Service model. Unified approach to present services in the Marketplace, based on the MP-relevant subset of EOSC-hub STD

Protocol/API	Short description
ARGO Messaging System - HTTP API	(Implements the Google PubSub protocol). MP uses this message oriented service to retrieve list of EOSC-hub services along with their metadata (part of STD) relevant in the MP scope.
eInfraCentral API - REST API	Used to retrieve information about services registered in eInfraCentral Catalogue
JIRA webhooks	User-defined callback over HTTP. Used to retrieve information from JIRA in the scope of MP Projects (represented by JIRA Epic) and MP Orders (JIRA tasks)
JIRA REST API	Used to create and synchronise representation of MP Projects and Orders in JIRA

Marketplace, user facing part of service ordering procedure

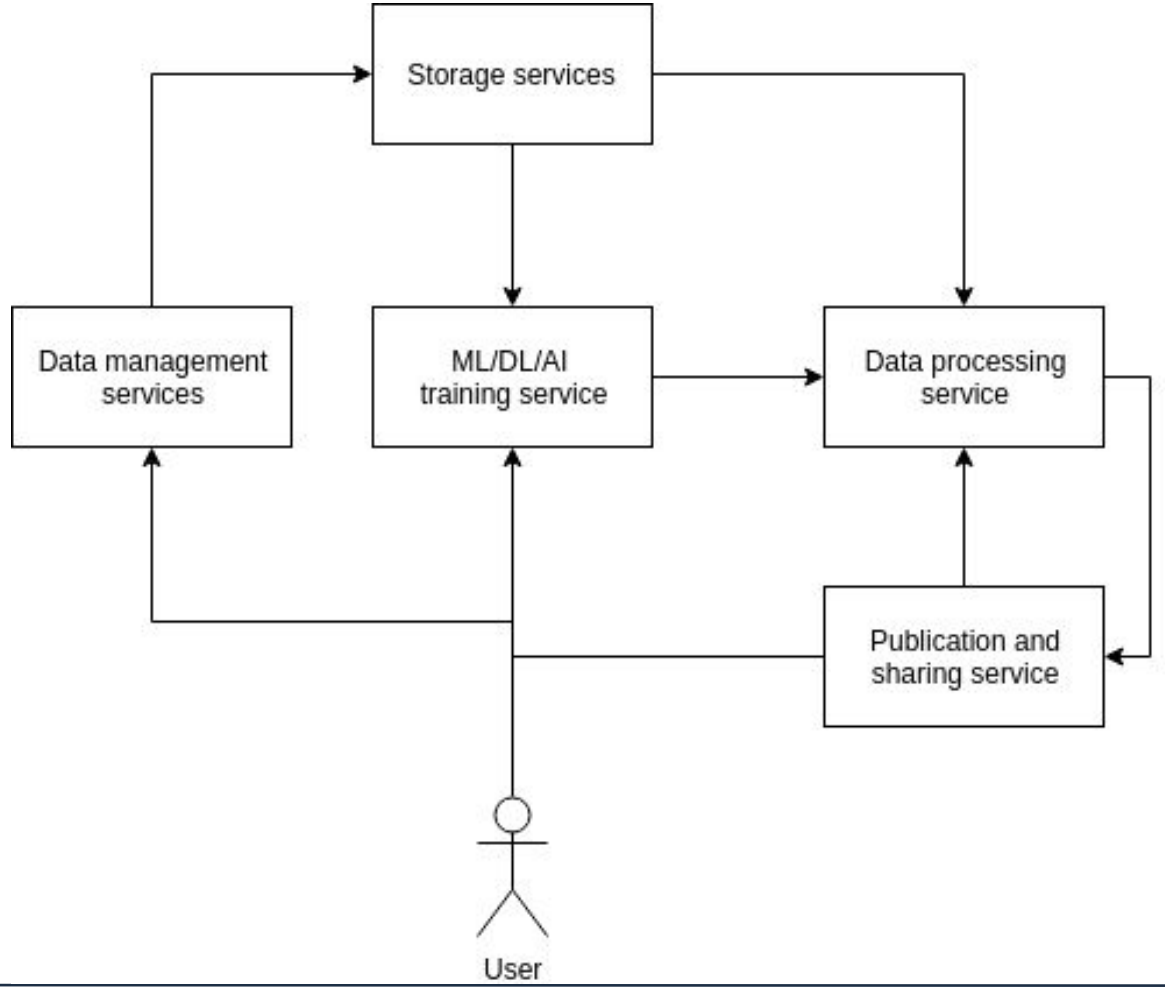
- interoperability guidelines
  - MP API to integrate with SPMT solutions and JIRA-like solutions
  - integrating with JIRA based Order Management solutions requires a dedicated JIRA instance, allowing webhook integration and information about the configuration of JIRAs objects (IDs of JIRA ticket fields, ticket statuses, worklows info etc.) in order to configure the mapping between the MP and JIRA.
- samples of services offering the feature
  - EOSC-hub Marketplace
  - SPMT solutions:
    - EOSC-hub SPMT: AGORA SPMT (https://grnet.github.io/agora-sp/)
    - eInfracentral Catalogue (https://github.com/eInfraCentral/docs)
  - Order Management solutions:
    - JIRA based EOSC-hub Order Management System

Data analytics services using ML/DL/AI

- description + architecture:
  - The data analytics service provide with the required tools and mechanisms to encapsulate and execute ML/DL/AI models across different platforms, covering the whole machine learning development life-cycle.
  - includes both CLI libraries as well as a SDK for developers to ease building and composing application architectures to be deployed on the Cloud.
  - The service can be divided into the following specialized components:
    - ML/DL/AI training service providing the basis for building, training, testing and validating a model
    - Data processing as a service based on a developed model, providing advanced functionalities like continuous delivery of models, automatic data processing (event-driven), etc.
    - Model publication, sharing and reuse service,
    - Storage services (not part of the analytics service), needed to store and retrieve the datasets to analyze, the results and any other assets produced or required by the framework.
    - Data management services (not part of the analytics service) to orchestrate the required data transfer and data movement tasks.

Data analytics service using ML/DL/AI

- architecture:
  -



Data analytics service using ML/DL/AI

Standard	Short description
TOSCA	OASIS Topology and Orchestration Specification for Cloud Applications is a standard to describe application topologies of cloud services and applications, comprising its components, dependencies, relationships and other details. TOSCA can be used to provide a standard way to deploy a given service over different resource providers and e-Infrastructures
OPEN API SPECIFICATIONS	OpenAPI is a specification that allows programmers to define standard REST interfaces, machine discoverable and interoperable.

Protocol/API	Short description
OpenID Connect	OpenID Connect is an authentication layer on top of OAuth 2.0, an authorization framework. The standard is controlled by the OpenID Foundation.
OAuth2	OAuth2 is the standard in which OpenID Connect is based, used for authorization.
REST APIs	RESTful services are a must. APIs for this service should be machine-discoverable.

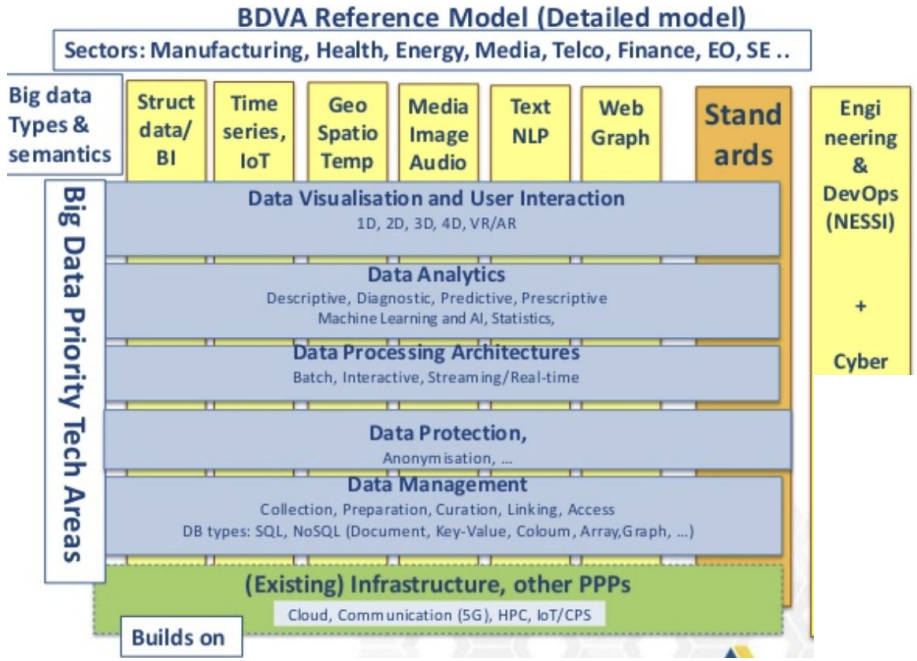
Data analytics service using ML/DL/AI

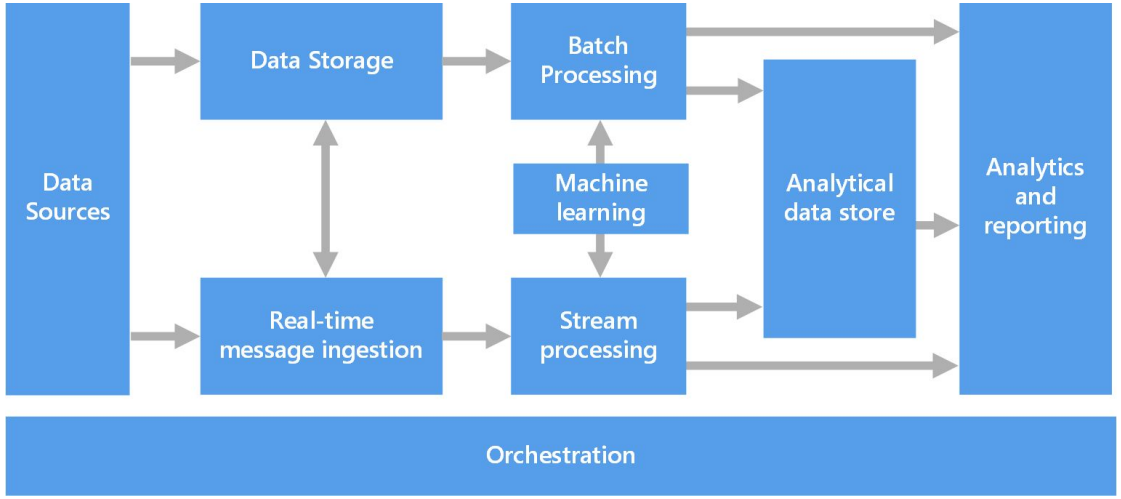
- interoperability guidelines
  - From the user facing perspective it should be desirable that all related services use and follow the same API semantics, as a way to ensure interoperability. Moreover, authentication and authorization should follow the schemas already in place.
  - The use of the same underlying technologies should also be considered. The use of Docker containers to encapsulate the user applications is widespread and should be evaluated, as other relevant initiatives may arise.
- samples of services offering the feature
  - ModelHub.ai, not covering all the ML lifecycle phases http://modelhub.ai/
  - DEEP-Hybrid-DataCloud services, covering all the ML lifecycle phases https://deep-hybrid-datacloud.eu
  - Kipoi, focused only on genomics http://kipoi.org/

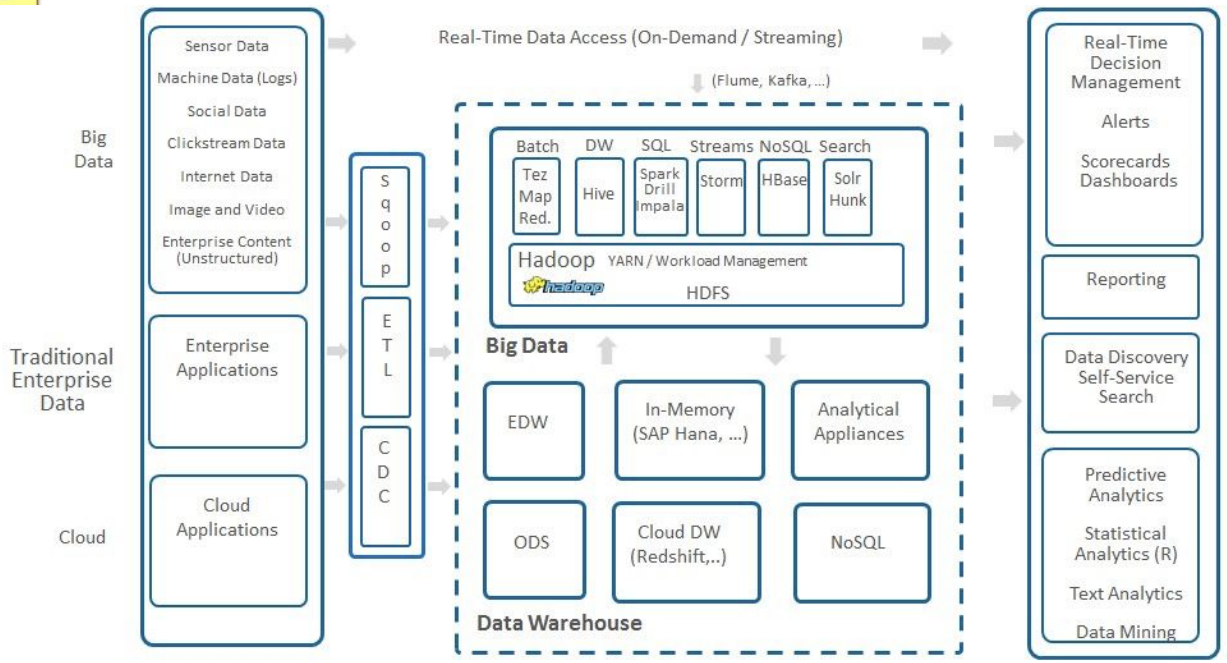
Big Data analytics service

- description
  - Big data analytics services in general are described as “complex process of examining large and varied data sets or big data, to uncover information including hidden patterns, unknown correlations.”
  - The deluge of data, however, poses several challenges that must be tackled accordingly to cope with bigger data volumes, heterogeneous formats and different frequency in data generation; the initial 3Vs (volume, variety and velocity) used to characterize Big Data, have later also been complemented with new ones able to capture complementary aspects, such as veracity, value and variability.
  - The scientific dimension adds another level of complexity with respect to metadata management, provenance, vocabularies/conventions as well as scientific (parallel) libraries, formats and tools.

Big Data analytics services - existing architectures





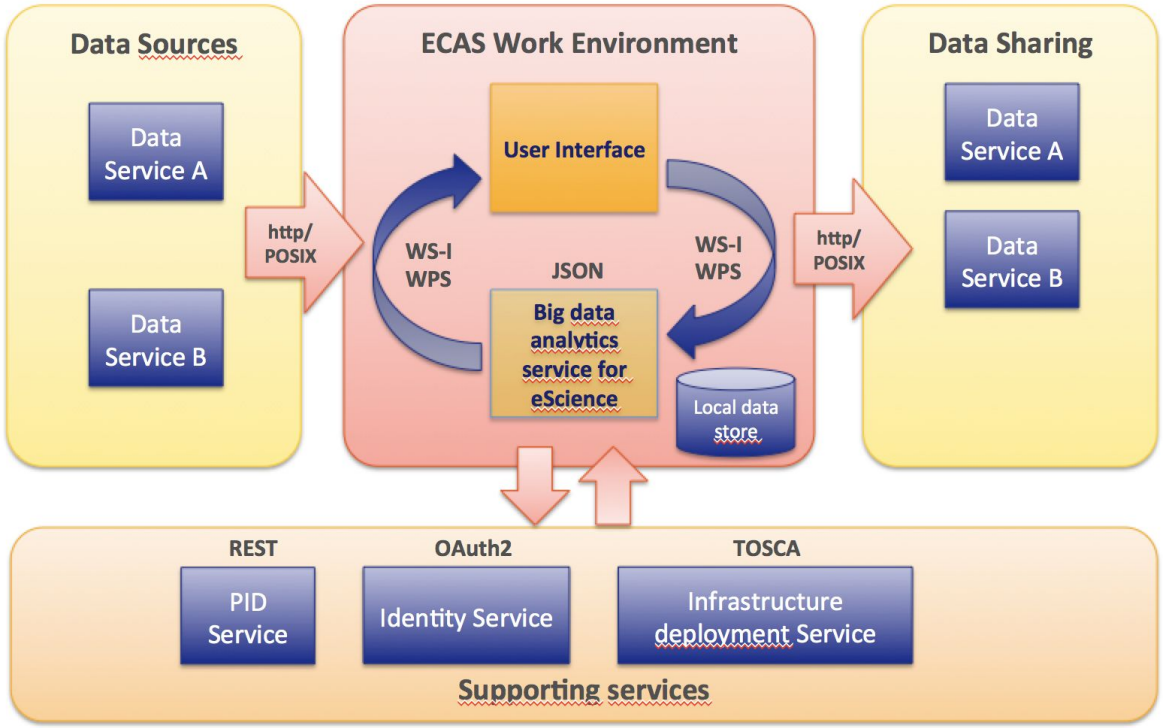


Architectures/images bu
BDVA, MIcrosoft, Gartner & Forrester

Big Data analytics service ECAS thematic service

Standard
JSON
TOSCA

Protocol/API
- SOAP API (WS-I)
- OGC WPS
- Python (PyOphidia)



Big Data analytics service/thematic services

- samples of services offering the feature
  - The ENES Climate Analytics Service (ECAS)

    Ophidia represents the core component of ECAS. It is a big data analytics framework for scientific data. It provides declarative, server-side, and parallel data analysis, jointly with an internal storage model able to efficiently deal with multidimensional data and a hierarchical data organization to manage large data volumes. It provides about 100 array-based functions and more than 50 datacube-based operators to enable OLAP tasks.

Future activities

- Missing bits in the description of macro-features already identified
  - Big Data Analitics part to be described as Thematic services
- Other macro-features that will be better described in the future
  - Should we describe as separate micro functionalities (or leave in Gateways/Portals part)
    - interactive services to create and share documents that contain live code, equations, visualizations and narrative text ????
    - Workflow engines ????

Thank you for your attention!

Questions?

Contact

eosc-hub.eu @EOSC_eu