Data Life Cycle Lab Earth and Environment

LSDMA All-Hands Meeting Mar 25, 2015

Marek Szuba, Jörg Meyer

LSDMA

The Team

- DKRZ
  - Carsten Ehbrecht
  - Stephan Kindermann
  - Michael Lautenschlager

- KIT
  - Parinaz Ameri
  - Uğur Çayoğlu: will start his PhD in August 2015
  - Ahmad Maatouki: defended his Master thesis end of January 2015
  - Jörg Meyer
  - Marek Szuba

Services for Climate Research

- GLORIA
  - MongoDB infrastructure on LSDF (7TB)
  - campaign will start in spring 2015
  - replication/redundancy required

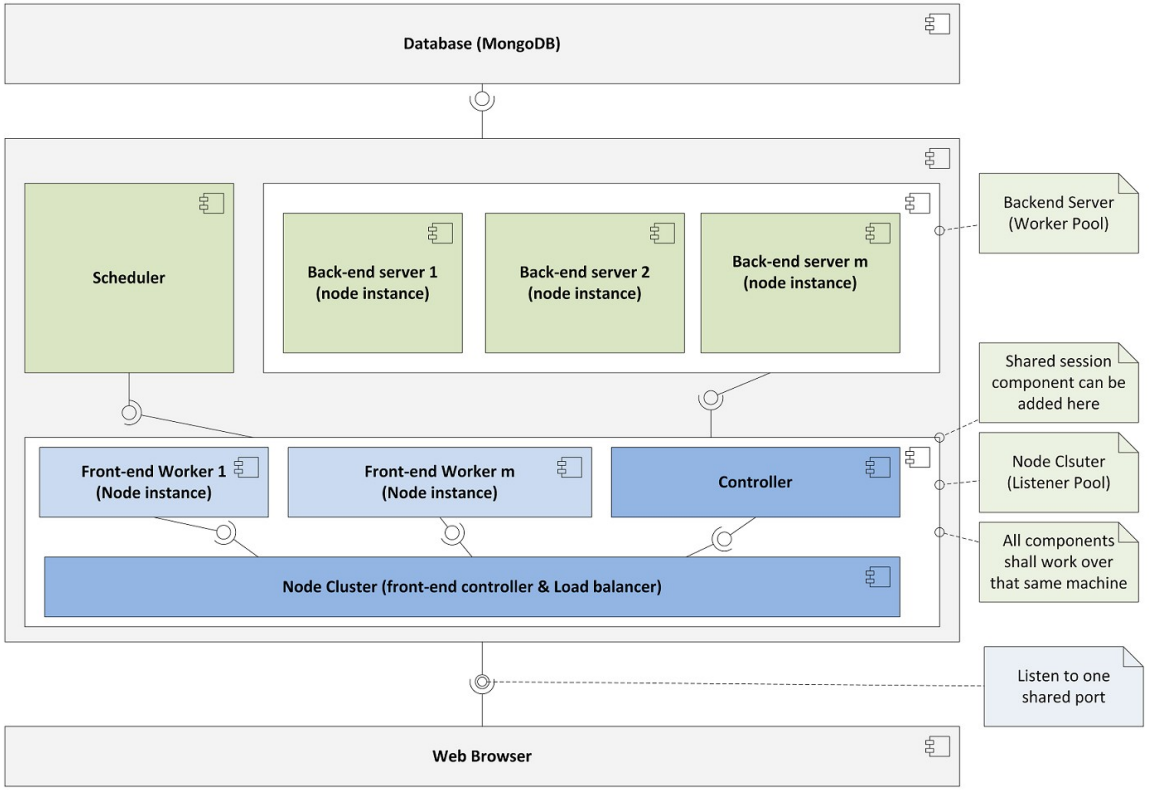
- Satellite data
  - MongoDB with metadata (geolocations) of 22 instruments
  - improved geo-matcher

- EUDAT B2SAFE
  - Safe replication of ENES data
  - iRODs + PIDs (EPIC-handles)

B2SAFE
Replicate Research Data Safely

Satellite data – Scalable oriented cluster library for Node.js

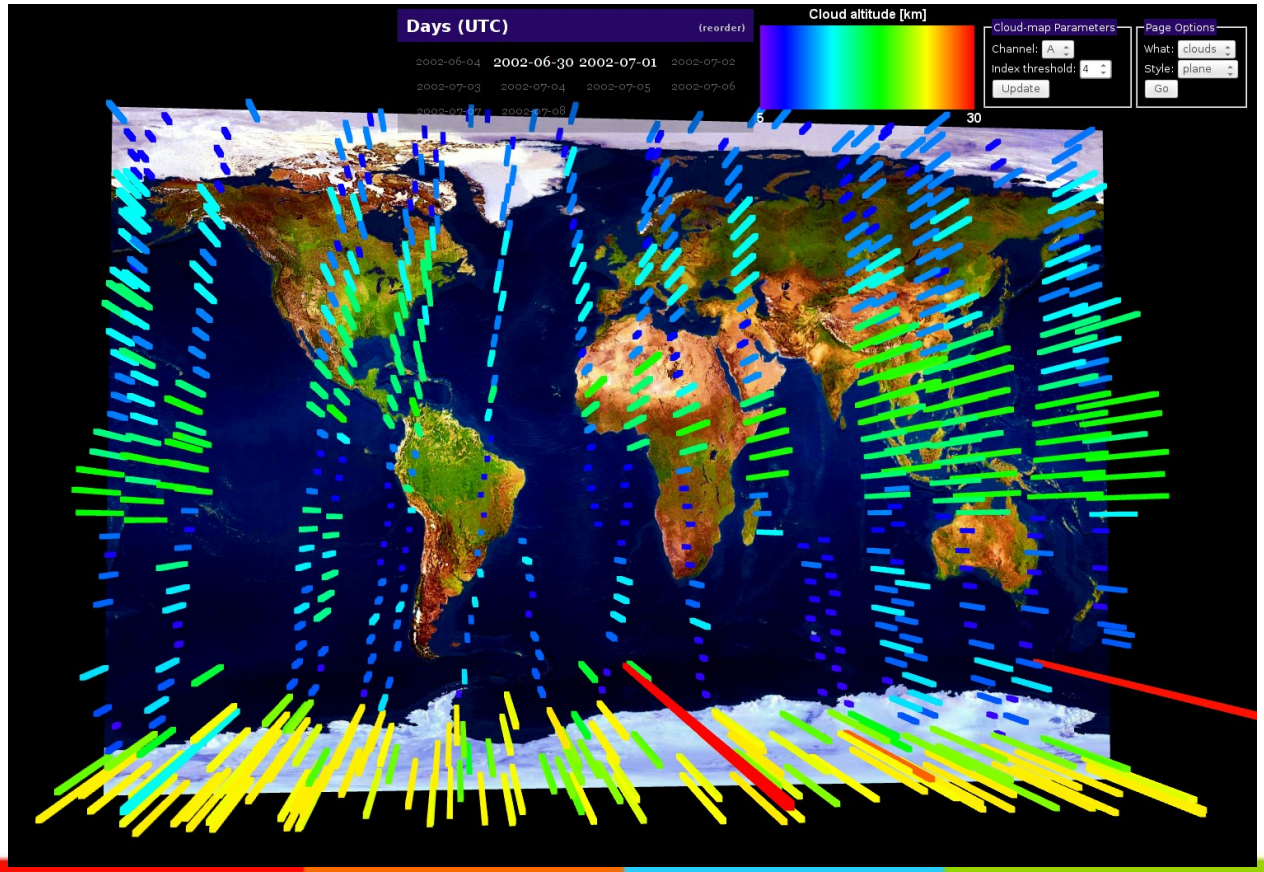
- Web service API for MongoDB and predefined use cases
- parallel preprocessing of MongoDB data
- implementation complete
- *to be published in 2015*



Real-time 3D Visualization of Earth-observing-satellite Data

- Visualization of climate data with web browser application
- Access to input data from MongoDB (via scalable Node.js cluster)
- uses WebGL

to be presented at European Geoscience Union 2015



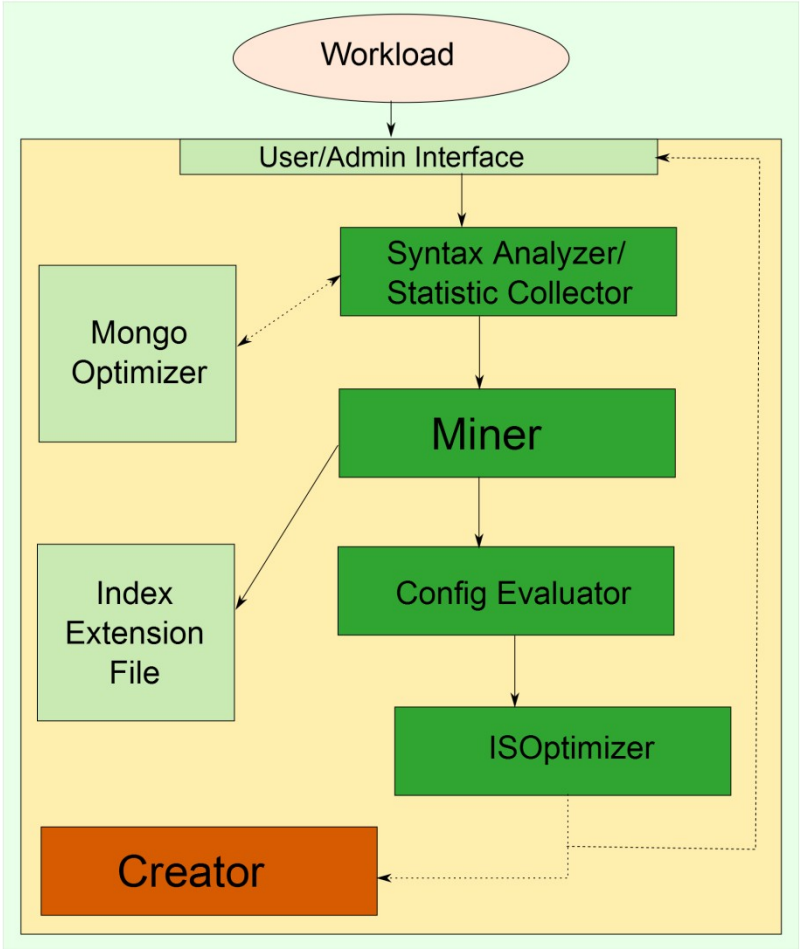
Index Selection Tool

Motivation:

- Index selection problem
  - Database with n=10 fields:
    4,268,523,260 possible indexes
    ($\sum (2^k)\ n! / (n-k)!$)
  - Investigated for RDMS

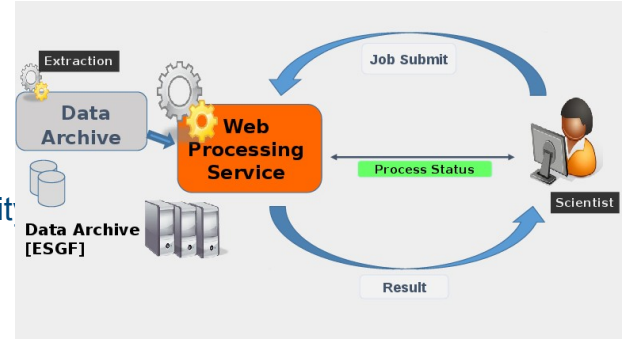
Goal:

- automatic creation of indexes for MongoDB (schema free NoSQL DB)
- dynamically adopt to changes in usage pattern
- climate data db as real-world use-case



Geospatial data life cycle framework Birdhouse

- Birdhouse: Web Processing Services for climate data
  - code: https://github.com/bird-house doc: http://bird-house.github.io/
  - based on:
    - Malleefowl: base processes and mandatory in a bird-house
    - Emu: a few test cases to try out
    - Hummingbird: provides CDOs and CFChecker as a service
    - Flyingpigeon: a collection of processes useful for the impact communit
    - Phoenix: the simple web browser application for WPS



- Recent improvements:
  - New processes:
    - extract coordinates: generate time series out of gridded data for given geographic coordinates.
    - analog days: detection of days with analog surface pressure patterns. based on NCEP data.
    - CFChecker: checks NetCDF metadata compliance according to Climate and Forecast (CF) conventions.
  - Climate indices:
    - Several realizations of the simple (one variable) indices from ocgis/icclim.
  - Data Access:
    - to avoid time consuming transports of data, first realizations have been tested to connect a Birdhouse WPS directly to the ESGF data archive.
  - Deployment: WPS services available as Docker images, using Anaconda Scientific Python Distribution for dependencies.

Meta Data Catalogues and Quality

Catalogue for climate data (observations and simulations)

- Collaboration: DLCL + IMK + Simlab Climate and Environment
- Status: catalogue for simulation data (SimLab); needs to be extended

Automatic checks on meta data quality

- correctness of syntax
- completeness
- consistency
- compliance with conventions
- correctness of **semantic** (community specific)

- Collaboration: DLCL Earth and Environment (DKRZ + KIT) + DLCL Key Technologies + DSIT + (RDA, EUDAT2020)
- Status: first meetings and brain stormings

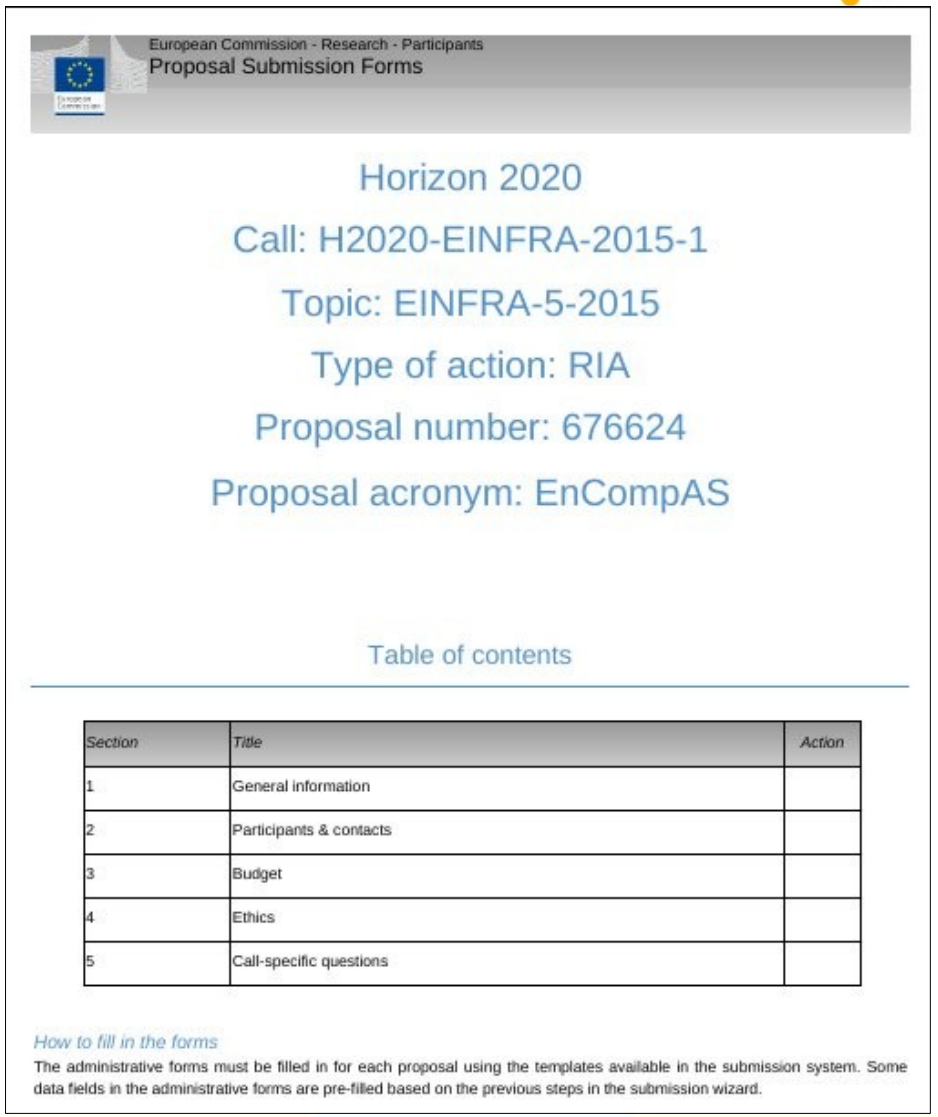
EUDAT2020

- EUDAT2020 Kickoff today
- KIT
  - Scientific communities environments and requirements
    - survey on data and computing landscapes, environments, and service requirements
- DKRZ
  - B2FIND: meta data catalogue for research data

EU-Proposal

Centre of Excellence
for Environmental Application Software
(EnCompAS)

- submitted on Jan, 14th
- together with
  SimLab Climate and Environment



European Commission - Research - Participants
Proposal Submission Forms

Horizon 2020
Call: H2020-EINFRA-2015-1
Topic: EINFRA-5-2015
Type of action: RIA
Proposal number: 676624
Proposal acronym: EnCompAS

Table of contents

Section	Title	Action
1	General information	
2	Participants & contacts	
3	Budget	
4	Ethics	
5	Call-specific questions	

How to fill in the forms

The administrative forms must be filled in for each proposal using the templates available in the submission system. Some data fields in the administrative forms are pre-filled based on the previous steps in the submission wizard.