Computing core of the software package for “cloud” analysis of climate change and the environment

Okladnikov I.G. (igor.okladnikov@gmail.com)
Institute of Monitoring of Climatic and Ecological Systems, Siberian Branch of Russian Academy of Sciences, Tomsk, Russia (http://www.imces.ru)
Institute of Computational Technologies, Siberian Branch of Russian Academy of Sciences, Tomsk branch, Tomsk, Russia (http://ict.nsc.ru)

The computing core of the information and computing software package based on a dedicated software framework for carrying out scientific research related to a statistical processing and analysis of spatial geophysical data archives, obtained both from observations and modeling, is presented. During the development, the accumulated experience in a development of information-computational web GISs for processing of large amounts of spatial data was used. Basic components of the computing core are represented by modules for searching, selecting, and processing spatial data arrays, and by a core manager governing processing and data flows. Its modular structure provides a possibility of operative expansion of the functionality of the software complex using various procedures for mathematical and statistical analysis, processing and graphical representation of results in a form of graphs, diagrams and fields on a map of a respective territory. The work was performed in the framework of the state budget theme No. AAAAA-A17-117013050037-0. Development of the core’s manager was performed in the framework of ICT SB RAS state target program No. 0316-2018-0002.

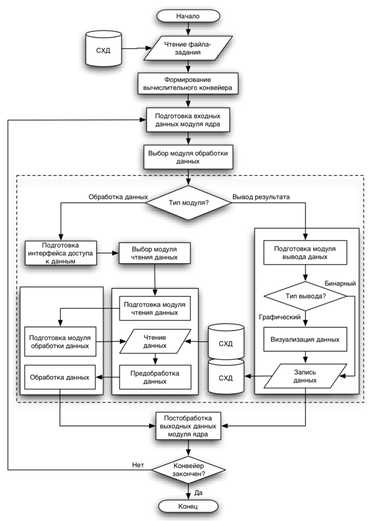
Introduction

Climate change has stimulated the development of observation systems and environmental modeling, which has already led to the emergence of archives of spatial observational and modeling data reaching petabyte volumes. The analysis of such data becomes impossible without appropriate computational and informational support. The creation of effective tools for computational and informational support for the analysis of large amounts of data and the organization of their effective use to obtain new knowledge, and their use for practical applications is one of the areas of modern climate science. In fact, to solve this problem, it is necessary to create a software environment based on modern information and telecommunication technologies [1, 2] that meets the requirements of spatial data infrastructure (SDI, [3, 4]), which implies the use of modern technologies for processing geophysical data that allow integrating various software solutions for performing “cloud” processing using remote high-performance computing resources. Development of thematic information and computing systems that form a necessary infrastructure should be based on the use of Web-GIS technologies [5-8]. Their usage is a promising way to improve the efficiency of multidisciplinary regional and global research in a field of Earth sciences, including the analysis of climate change and its impact on the spatio-temporal behavior of local ecosystems. One of the modern information and computing software packages for statistical processing and “cloud” analysis of spatial geophysical data archives is being developed at IMCES SB RAS. This paper presents the computing core of this software, its key components and capabilities.

Core

The modular computing core is a key component of the software package for "cloud" distributed processing of large sets of spatial climatic data. It consists of a set of software components (modules) connected through a unified program interface (API) and performs search, selection, processing, and visualization of spatial data. The software package can be deployed on several interconnected computational nodes. A separate computing core is installed on each node and connected with its own spatial data archive and metadata database [9]. These archives of spatial data can be represented by modeling data (reanalysis, global and regional climatic and meteorological models) and observational data (meteorological stations, satellite observation data, etc.) of various spatial resolution, given at different spatio-temporal domains and stored in NetCDF or HDF format [10].

The computing core was developed using the GNU Data Language (GDL, http://gnudatalanguage.sourceforge.net/) and Python (http://python.org) programming languages, which provide proven procedures for mathematical processing and visualization of spatial data, as well as programming interfaces for reading / writing files in NetCDF and ESRI Shapefile formats, and access to PostgreSQL DBMSs. To manage core’s input-output, form a computational pipeline and control the execution of modules, the computing core manager is used. It provides management of core’s modules and transfer of intermediate results between them. Data processing is performed in the framework of a computational pipeline, which is represented by a sequence of calls to computing core modules with transfer of data from outputs of one module to inputs of another. The computational pipeline is formed using a special task file in XML format, prepared by the web portal based on the results of user actions in the graphical interface. The task file contains key characteristics of spatial data sets to be processed, a description of a computational pipeline in the form of a sequence of calls to computational modules, and descriptions of intermediate data arrays transferred between them, as well as parameters for writing processing results to a graphic file.



[Computing core manager]

Task file (XML DTD)

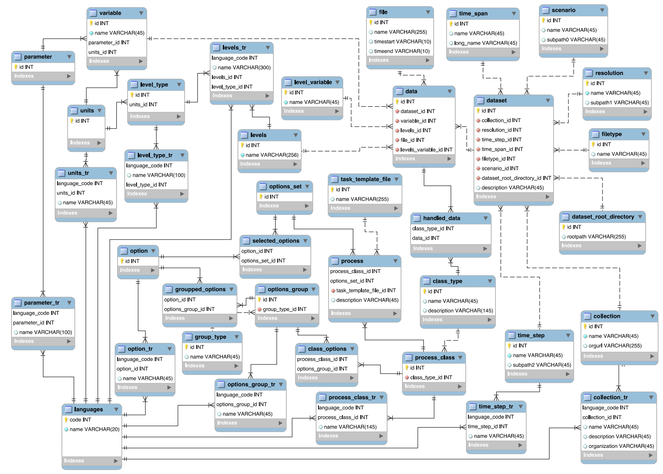
```
<!ELEMENT task (service, metadb, computing, (data|destination)+, processing+)>
<!ATTLIST task
    uid ID #REQUIRED
    owner CDATA #IMPLIED
    description CDATA #IMPLIED
<!ELEMENT service (timePeriod)>
<!ATTLIST service
    processingClass CDATA #REQUIRED
    calculationParameter CDATA #REQUIRED
    trend CDATA #REQUIRED
    isBaseLayer CDATA #REQUIRED>
<!ELEMENT timePeriod (dateStart, dateEnd)>
<!ATTLIST timePeriod
    type CDATA #REQUIRED>
<!ELEMENT dateStart EMPTY>
<!ATTLIST dateStart
    year CDATA #REQUIRED
    month CDATA #REQUIRED
    day CDATA #REQUIRED>
<!ELEMENT dateEnd EMPTY>
<!ATTLIST dateEnd
    year CDATA #REQUIRED
    month CDATA #REQUIRED
    day CDATA #REQUIRED>
<!ELEMENT metadb EMPTY>
<!ATTLIST metadb
    host CDATA #REQUIRED
    name CDATA #REQUIRED
    user CDATA #REQUIRED
    password CDATA #REQUIRED>
<!ELEMENT computing EMPTY>
<!ATTLIST computing
    system CDATA #REQUIRED
    type (LOCAL|REMOTE) #REQUIRED
    ip CDATA #IMPLIED>
<!ELEMENT  data  (description?,  dataset?,  variable?,  region?,  levels?,  time?,
param*)>
<!ATTLIST data
    uid ID #REQUIRED
    type (dataset|array|datafile|parameter) #REQUIRED
    kind (map|stations|lines) #IMPLIED>
<!ELEMENT description EMPTY>
<!ATTLIST description
    title CDATA #REQUIRED
    name CDATA #REQUIRED
    units CDATA #REQUIRED
    units_sfx CDATA #IMPLIED
    source CDATA #IMPLIED>
<!ELEMENT dataset EMPTY>
<!ATTLIST dataset
    name CDATA #REQUIRED
    resolution CDATA #REQUIRED
    time_step CDATA #REQUIRED
    scenario CDATA #REQUIRED
<!ELEMENT variable EMPTY>
<!ATTLIST variable
    name CDATA #REQUIRED
    tempk2c CDATA "no">
<!ELEMENT region ((point, point, point+)|(tile+))>
<!ATTLIST region
    units (degrees|pathrow) #REQUIRED
<!ELEMENT point EMPTY>
<!ATTLIST point
    lon CDATA #REQUIRED
    lat CDATA #REQUIRED>
<!ELEMENT tile EMPTY>
<!ATTLIST tile
    path CDATA #REQUIRED
    row CDATA #REQUIRED>
<!ELEMENT levels EMPTY>
<!ATTLIST levels
    values CDATA #REQUIRED>
<!ELEMENT time (segment+)>
<!ATTLIST time
    uid CDATA #IMPLIED
    template CDATA #FIXED "YYYYMMDDHH"
```

```
<!ELEMENT segment EMPTY>
<!ATTLIST segment
    name CDATA #REQUIRED
    beginning CDATA #REQUIRED
    ending CDATA #REQUIRED>
<!ELEMENT param (#PCDATA)>
<!ATTLIST param
    uid ID #REQUIRED
    type (integer|float|string|intarr|fltarr|strarr|binary) #REQUIRED>
<!ELEMENT destination (file, projection?, graphics?)>
<!ATTLIST destination
    uid ID #REQUIRED
    type (image|raw) #REQUIRED>
<!ELEMENT file EMPTY>
<!ATTLIST file
    name CDATA #REQUIRED
    type (eps|geotiff|shape|geotiff24|netcdf|bin|xml|png) #REQUIRED>
<!ELEMENT projection (p0lon?, p0lat?, rot?, limits)>
<!ATTLIST projection
    name CDATA #REQUIRED>
<!ELEMENT p0lon (#PCDATA)>
<!ATTLIST p0lon
    units CDATA #FIXED "degrees">
<!ELEMENT p0lat (#PCDATA)>
<!ATTLIST p0lat
    units CDATA #FIXED "degrees">
<!ELEMENT rot (#PCDATA)>
<!ATTLIST rot
    units CDATA #FIXED "degrees">
<!ELEMENT limits (limit, limit, limit, limit)>
<!ATTLIST limits
    units CDATA #FIXED "degrees">
<!ELEMENT limit (#PCDATA)>
<!ATTLIST limit
    role (left|right|top|bottom) #REQUIRED>
<!ELEMENT graphics (width, height, legend?, colortable?, colorscale?,
smoothing?, steps?)>
<!ATTLIST graphics
    title CDATA #REQUIRED
    kind      (shaded|contour|station|tile|multiline|vector|regress2d)
#REQUIRED>
<!ELEMENT width (#PCDATA)>
<!ATTLIST width
    units CDATA #FIXED "pixels">
<!ELEMENT height (#PCDATA)>
<!ATTLIST height
    units CDATA #FIXED "pixels">
<!ELEMENT  legend  (units?,  nlabels?,  ncolors?,  limited?,  minimum?,
maximum?, file?)>
<!ATTLIST legend
    title CDATA #REQUIRED
    kind (embedded|file|none) #REQUIRED
    type (continuous|discrete) #REQUIRED
    position (right|left|top|bottom|none) #REQUIRED>
<!ELEMENT units (#PCDATA)>
<!ELEMENT ncolors (#PCDATA)>
<!ELEMENT nlabels (#PCDATA)>
<!ELEMENT limited (#PCDATA)>
<!ELEMENT minimum (#PCDATA)>
<!ELEMENT maximum (#PCDATA)>
<!ELEMENT colortable (#PCDATA)>
<!ELEMENT colorscale (#PCDATA)>
<!ELEMENT smoothing (#PCDATA)>
<!ELEMENT steps (#PCDATA)>

<!ELEMENT processing (input+, output*)>
<!ATTLIST processing
    uid ID #REQUIRED
    class CDATA #REQUIRED
    module_path CDATA #IMPLIED>
<!ELEMENT input EMPTY>
<!ATTLIST input
    uid ID #REQUIRED
    data CDATA #REQUIRED>
<!ELEMENT output EMPTY>
<!ATTLIST output
    uid ID #REQUIRED
    data CDATA #REQUIRED>
```

Metadata database

To describe georeferenced data sets and their processing routines, and provide effective functioning of the software complex, a dedicated metadata storage in the form of a MySQL database was developed:



References

[1] Stefano Nativi, Mohan Ramamurthy, Bernd Ritschel. EGU-ESSI Position Paper. https://sites.google.com/a/imaa.cnr.it/egu-essi/Home/documents
[2] Gordov E.P., Lykossov V.N. Development of information-computational infrastructure for integrated study of Siberia environment // Computational technologies. 2007. V. 12. Special issue 2. P. 19-30.
[3] Steiniger S., Hunter A.J.S. Free and open source GIS software for building a spatial data infrastructure. / In: Bocher E., Neteler M., (eds.), Geospatial Free and Open Source Software in the 21st Century. LNGC, Heidelberg, Springer. 2012. P. 247-261.
[4] Koshkarev A.V., Ryakhovskii A.V., and Serebryakov V.A. Infrastructure of distributed environment of storage, search and transformation of geospatial data // Open Education. 2010. V. 5. P. 61-73.
[5] Frans J. M. van der Wel. Spatial data infrastructure for meteorological and climatic data. // Meteorol. Appl. V. 12. 2005. P. 7–8.
[6] Gordov E.P., Fazliev A.Z., Lykosov V.N., Okladnikov I.G, Titov A.G. Development of web based information-computational infrastructure for Siberia Integrated Regional Study / Environmental Change in Siberia // Earth Observation, Field Studies and Modelling, Balzter, Heiko (Ed.) 1st Edition, 2010, 300 p.
[7] Ataeva O.M., Kuznetsov K.A., Serebryakov V.A., Filippov V.I. Portal for integration of spatial data “GeoMeta”. CC RAS, Moscow. 2010. 106 p.
[8] Gordov E.P., Okladnikov I.G., Titov A.G. Application of web mapping technologies for development of information-computational systems for georeferenced data analysis // Vestnik NSU, Series: Information technologies. 2011. Volume 9, Issue 4. P. 94-102.
[9] Okladnikov I.G., Gordov E.P., Ryazanova A.A., and Titov A.G. Elements of software package for “cloud” analysis of the environment and climate change // IOP Conf. Series: Earth and Environmental Science. 2019. V.386. 012037. doi:10.1088/1755-1315/386/1/012037.
[10] Okladnikov I.G., Gordov E.P., and Titov A.G. Development of climate data storage and processing model // IOP Conf. Series: Earth and Environmental Science. 2016. V. 48. 012030. doi:10.1088/1755-1315/48/1/012030