

## ТЕХНОЛОГИИ ОБРАБОТКИ, ХРАНЕНИЯ И ВИЗУАЛИЗАЦИИ BIG DATA В ЗАДАЧАХ МОДЕЛИРОВАНИЯ КЛИМАТА АРКТИКИ

Никитин Н.О.<sup>1</sup>, Калюжная А.В.<sup>1</sup>

<sup>1</sup> – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики. Санкт-Петербург. Россия, [nikolay.o.nikitin@gmail.com](mailto:nikolay.o.nikitin@gmail.com)

**Аннотация.** Предложены подходы к обработке массивов данных, которые создаются в ходе климатических расчетов в Арктическом регионе, выполняемых комплексом моделей (гидродинамические, ледовые, волновые, атмосферные), а также их хранению и визуализации. Описана реализация распределенного хранилища, реализованного на основе семантического подхода, а также ГИС для визуализации больших массивов данных из распределённой системы хранения.

Ключевые слова: большие данные, семантическое хранение, динамическая визуализация.

В настоящее время проблемы обработки больших массивов данных, которые производят современные численные модели природных явлений, всё ещё далеки от решения. Использование в моделях измельченных горизонтальных расчетных сеток с обширным пространственным покрытием, большого количество вертикальных уровней, высокого временное разрешение в совокупности к стремительно росту объёмов полученных в ходе результатов моделирования файлов. Особенно остро эти аспекты проявляются в задачах моделирования Арктического региона из-за необходимости хранения массивов ледовых данных и высоких требованиях к пространственному и временному разрешению моделей. Использование классических подходов к хранению и обработке данных не позволяет производить оперативное извлечение временных и пространственных срезов, а также визуализировать поля в динамике. Это делает целесообразным интерпретацию данных гидрометеорологического моделирования как BigData.

В качестве возможного подхода предлагается технология семантического хранения, основанная на унифицированном хранении данных в виде файлов формата NetCDF. Этот формат обеспечивает возможность гибкой настройки структуры хранения данных, поэтому размерность элементарных блоков массивов с данными была модифицирована особым образом для обеспечения быстрого извлечения как временных рядов, так и пространственных полей [1].

Представление семантики хранимых данных осуществляется на основе позиции в древовидной системе организации хранения. Она задает географический регион, источник (модель или реанализ) и конкретную переменную, для которой хранится каталог данных. Для каждого каталога строится кэшированное описание, хранящее как метаданные, так и временные метки файлов. Такой подход позволяет существенно повысить производительность извлечения данных.

Хранилище реализовано с использованием архитектуры с распределённым хранилищем и централизованным сервером в соответствии с идеологией платформы Exarch [2]. Экспериментальные исследования позволяют говорить о линейной зависимости времени извлечения данных от числа полей в файлах, а также хорошей параллелизуемой этого процесса.

Для решения задачи динамической визуализации данных хранилища предложена геоинформационная система TerraXT, обеспечивающая гибкую визуализацию полей данных и временных рядов измерений [3]. Система позволяет выполнять раздельную совместную визуализацию векторных и скалярных полей на трехмерном «глобусе», выполнять наложение произвольного числа слоев данных, визуализировать попутные

наблюдения, совмещать графики в нескольких точках, строить срезы по глубине. При этом данные могут поступать как из локального файлового каталога, так и из удаленного хранилища. Запросы к хранилищу обеспечивают минимизацию объема передаваемой по сети информации (все операции извлечения и разреживания рядов и полей выполняются на сервере). Поддерживаются сенсорный режим управления, в т. ч. на нестандартных устройствах (видеостена, сенсорный стол).

Использование представленных технологий позволяет унифицировать подход к работе с разнородными результатами моделирования большого объема, сложности и неоднородности как данными единого формата, а также обеспечить анализ результатов работы комплекса моделей и расчет производных массивов данных в сжатые сроки.

#### **Литература**

1. Mbogo G. K., Rakitin S. V., Visheratin A. High-performance meteorological data processing framework for real-time analysis and visualization //Procedia computer science. – 2017. – Т. 119. – С. 334-340.
2. Visheratin A. A. et al. Exarch: semantics-based modular distributed data storage //Proceedings of the 2017 International Conference on Cloud and Big Data Computing. – ACM, 2017. – С. 22-26.
3. Golubev K. et al. Floodvision: A Tool for Fast and Comfortable Scenario-Based Visual Analysis of a Large Climate Datasets //Procedia computer science. – 2017. – Т. 119. – С. 298-306.

## **PROCESSING, STORAGE AND VISUALISATION TECHNOLOGIES FOR BIG DATA AS A PART OF CLIMATE MODELLING IN THE ARCTIC**

**Nikitin N.<sup>1</sup>, Kalyuzhnaya A.<sup>1</sup>**

<sup>1</sup> – *Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint-Petersburg, Russian Federation, nikolay.o.nikitin@gmail.com*

**Abstract.** Approaches to the processing of data arrays that are created in the course of climate calculations in the Arctic region, performed by a set of models (hydrodynamic, ice, wave, atmospheric), as well as their storage and visualization, are proposed. The implementation of a distributed repository implemented based on a semantic approach and system for visualization of repository data is described.

Key words: big data, semantic storage, dynamic visualization.