

# Платформа интеграции веб-ориентированных грид-приложений<sup>1</sup>

И.М. Алешин, К.И. Холодков, В.Г. Гетманов, В.Н. Корягин

**Аннотация.** В работе описана платформа интеграции, предназначенная для упрощения разработки веб-приложений, использующих ресурсы распределенных вычислительных систем (грид). Основу платформы составляют несколько модулей, обеспечивающих взаимодействие с компонентами грид, включая авторизацию, обмен данными, визуализацию результатов. Связь приложения с распределенными ресурсами системы осуществляется с помощью программного интерфейса. Платформа ориентирована на создание приложений, реализующих численное решение задач, сводящихся к выполнению большого количества слабо связанных относительно несложных заданий. Приведен пример интеграции одного из прикладных приложений.

**Ключевые слова:** грид-приложение, COA, связующее программное обеспечение, платформа интеграции, спектрально-временной анализ, полигармоническая модель.

## Введение

Говоря о грид-системах (или просто – грид), будем иметь в виду вычислительные комплексы, работающие под управлением связующего (или промежуточного — от англ. *middleware*) программного обеспечения. Грид является типичным примером реализации парадигмы *High-Throughput computing* (HTC), например, книга [1]. HTC-системы могут быть гетерогенными и территориально распределенными, следствием чего является отсутствие каких-либо требований, предъявляемых к оборудованию при необходимости использования открытых стандартов для программных средств. Ниже мы ограничимся обсуждением грид-систем, построенных на специально выделенных для них ресурсах (серверы, кластеры, системы хранения и т. п.), исключая из рассмотрения большой класс волонтерских грид

подобных BOINC [2], ориентированных на использование простаивающих ресурсов персональных компьютеров.

Особенности функционирования конкретных грид определяются видом связующего программного обеспечения, используемого для запуска вычислений, управления очередями, подключения к другим системам запуска и пр. Изначально грид-системы строились на основе универсальных пакетов, типичным и наиболее распространенным примером которых являются *Globus Toolkit* [3]. Для взаимодействия с этими системами пользователь должен владеть довольно сложными инструментами, в частности, уметь описывать свою задачу на языке описания заданий, который, зачастую, довольно плохо документирован и имеет тенденцию меняться от версии к версии. При этом универсальность доступного инструментария дает значительные полномочия на использование

<sup>1</sup> Работа выполнена при поддержке Программы фундаментальных исследований Президиума РАН № 43П по стратегическим направлениям развития науки на 2014 г. «Фундаментальные проблемы математического моделирования».

ресурсов. Поэтому задача авторизации пользователей, контроль запускаемых приложений и другие аспекты системной безопасности играют значительную роль. Часть связующего ПО, отвечающая за аутентификацию пользователя, проверку полномочий и журналирование, стала комплексной и сложной во взаимодействии.

Отсутствие простого способа взаимодействия с распределенными вычислительными системами существенно препятствовало их широкому распространению. Поэтому началось развитие программных компонент, которые позволяют упростить взаимодействие пользователя с грид. Это привело к появлению связующего программного обеспечения, способного гибко агрегировать несколько грид-узлов, скрывая от пользователей значительную часть технических деталей организации системы. В качестве примера можно привести метапланировщики Gridway и EGEE WMS [4]. В настоящее время при создании грид все шире применяется концепция сервис-ориентированной архитектуры (COA) [5]. Применительно к научным приложениям используется также термин «сервис-ориентированная наука» — Service-Oriented Science [6]. По сути, идея COA является парадигмой, обобщающей практику создания информационной инфраструктуры (<http://www.soa-manifesto.org>). Применительно к грид использование этой концепции приводит к требованиям модульного подхода в разработке, использованию слабо-связанных компонентов, стандартизованных интерфейсов и протоколов. В качестве примера использования SOA в связующем программном обеспечении для грид можно привести следующие программные продукты: облачная платформа организации web-сервисов Everest [7], инструментальный комплекс DISCENT [8], платформа интеграции GEMLCA [9] и ряд других приложений.

Отдельно следует отметить несколько крупных глубоко-интегрированных проектов, таких как Apache Ignite (GridGain In-memory Data Fabric)[10] или Univa/Oracle GridEngine[11]. Данные решения реализуют не только вычислительные механизмы, но и компоненты доступа к данным, механизмы потоковой параллельной обработки данных и многие другие функции. Несмотря на уделяемое разработчи-

ками большое внимание стандартизации, документированию и абстракционной чистоте компонентов этих систем, не свойственное ранним крупным проектам, данные решения всё же требуют заметных усилий в их освоении. В целом, такая многоцелевая и глубоко интегрированная система является неоправданно сложной при создании систем, ориентированных на грид-вычисления.

Здесь мы приводим описание разработанного нами инструментария, предназначенного для создания сервис-ориентированных научных приложений — платформы интеграции грид-приложений со связующим программным обеспечением. Платформа интеграции представляет собой многоуровневую систему управления ресурсами, запуском вычислений и доставкой результатов. При этом от разработчика прикладной части не требуется знания деталей устройства грид, так как взаимодействие осуществляется через программные интерфейсы соответствующих модулей в составе платформы. Естественно, что такое упрощение достигается за счет значительного ограничения круга проблем, которые могут быть решены в рамках созданного программного продукта: платформа ориентирована на решение задач, допускающих высокую степень параллельности вычислений. К моменту написания статьи на базе платформы интеграции нами были построены два веб-приложения: первое из них [12] реализует точное численное решение обратной задачи определения параметров упругой анизотропии коры и верхней мантии по сейсмическим волновым формам. Описание второго из реализованных нами приложений — спектрально-временной анализ (СВАН) на основе полигармонической модели — приведено ниже в качестве примера использования платформы.

## 1. Программная архитектура

При создании платформы мы ориентировались на геофизические проблемы, требующие проведения расчетов высокой степени параллельности. Говоря точнее, имеются в виду задачи, решения которых сводится к выполнению большого количества запусков относительно несложных заданий, абсолютно независимых друг от друга. В англоязычной литературе для

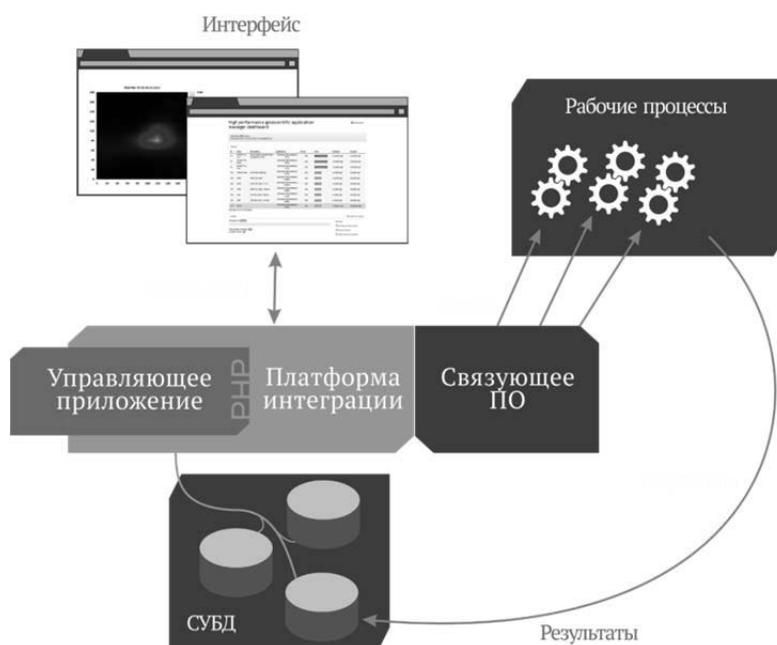


Рис. 1. Общая схема организации вычислений с использованием платформы интеграции

Ключевую роль играет управляющее приложение, которое средствами платформы интеграции взаимодействует с пользователем, связующим ПО грид и СУБД результатов

обозначения такого рода расчетов часто используется термин "Bag-of-tasks" (например, [13]). Вообще говоря, задачи такого рода лучше всего адаптируются в грид независимо от предметной области, а ориентация на геофизические приложения выражается в специфике компонент, осуществляющих методы анализа результатов расчетов, их визуализацию и прочее.

Платформа интеграции ориентирована на создание web-приложений и реализована на языке программирования PHP и платформе CodeIgniter (<http://codeigniter.com>), что определяет выбор языка разработки прикладных приложений. Выбор языка достаточно произволен и PHP был выбран потому, что он имеет C-подобный синтаксис и является популярным, особенно, среди начинающих разработчиков. Для реализации подобной задачи можно использовать и другие средства: Python [14], Scala [7], Java [11] и ряд других.

Назначение платформы состоит в управлении ресурсами грид на основе взаимодействия со связующим программным обеспечением системы. Основные задачи, реализованные в платформе таковы: получение сведений о состоянии ресурсов, организация очереди задач, мониторинг их выполнения, хранение резуль-

татов и обеспечение доступа к ним. Кроме того, платформа служит для обеспечения взаимодействия с распределенной системой как разработчика приложения, так и пользователей, что достигается предоставлением соответствующих программных интерфейсов. Общая схема организации вычислений с использованием платформы интеграции изображена на Рис. 1.

Имея целью создать максимально гибкое, легко расширяемое и масштабируемое решение, мы следовали шаблону проектирования "Модель-вид-контроллер" (МВК), в котором модель данных, пользовательский интерфейс и поведение приложения разделены таким образом, чтобы модификация одного из компонентов оказывала минимальное воздействие на остальные [15]. Применительно к нашей задаче удобно выделить две сущности – «приложение» и «проект». «Приложение» состоит из проблемно-специфичных видов, расширений контроллеров, схемы базы данных, исполняемого вычислительного модуля. Модель этой сущности описана в Табл. 1. Сущность «Проект» – это заданный пользователем набор исходных параметров вычислительной задачи и присвоенным ему названием, описанием и уникальным идентификатором (Табл. 2). Каждому

Табл. 1. Описание ключевых свойств модели «Приложение». Необходимо заметить, что ключевым свойством является «код» приложения, по которому все компоненты платформы обращаются к компонентам приложения

Свойство	Описание
Код	Цифробуквенное обозначение для удобства работы с компонентами приложения
Название	Краткое название приложения
Описание	Описание методов или конкретной реализации приложения, любая другая информация (необязательно)

Табл. 2. Описание ключевых свойств модели «Проект»

Свойство	Описание
Название	Краткое название проекта (необязательно)
Описание	Описание служит для указания каких-либо особенностей в исходных параметрах или информации о происхождении исходных данных или любых других данных (необязательно)
Состояние	Отражает текущее состояние проекта. Предусмотрен ряд состояний: «Активен» – проект создан, но не отправлен на вычисления, соответствующий экземпляр СУБД результатов готов к использованию, «Неактивный» - то же, но СУБД результатов остановлена, «Вычисляется» - выполняются расчёты, «Выполнено» - расчёты выполнены, результаты переданы в СУБД и «Выполнено (неактивно)» - то же, но СУБД с результатами остановлена.

экземпляру сущности «проект» соответствует экземпляр сущности «приложение». Важной частью платформы является управляющее приложение, которое обеспечивает управление грид-системой, взаимодействие с пользователем и прикладным приложением, а также реализует контроллеры для сущностей «проект» и «приложение».

Модель в нашей реализации играет пассивную роль и сводится к набору параметров, характеризующих состояние ресурсов распределенной системы и процесс выполнения заданий. Соответствующая структура данных спроецирована на реляционную схему и реализована в СУБД MySQL [16]. Доступ к элементам модели со стороны управляющего приложения осуществляется с использованием технологии объектно-реляционного отображения [17]. Обращение к элементам модели и со стороны разработчика, и со стороны пользовательского интерфейса обеспечивается специально созданными программными интерфейсами PHP и HTTP+JSON, соответственно.

Практически, разработка прикладного приложения на базе платформы интеграции состоит из нескольких шагов. Во-первых, необходимо разместить в определённых директориях на машине запуска и контроля исполняемый модуль прикладного приложения и схему базы данных для хранения результатов расчетов. В качестве «машины запуска и контроля» может

выступать любой сервер в составе грид. Для этого на нем должны функционировать мета-планировщик и управляющее приложение. Во-вторых, необходимо разработать пользовательский интерфейс для вида внесения исходных данных и настройки параметров вычислений. При этом важно обеспечить корректную обработку входных данных на стороне контроллера, реализовав дополнения к базовым функциям платформы (Табл. 3). В-третьих, необходимо разработать дополнения к части контроллера «приложение», которое обеспечивает визуализацию результатов. Все дополнения и виды должны быть размещены в определённых директориях платформы. Наконец, необходимо отразить новое приложение в списке приложений платформы, создав, тем самым, экземпляр сущности «приложение».

Одной из важных технических особенностей платформы является использование реляционной СУБД для хранения результатов вычислений. Вместе с новой сущностью «проект» создается экземпляр системы управления базой данных, структура которой определяется спецификой задачи. Экземпляры СУБД располагаются на машине диспетчеризации СУБД – любом подключенном к Интернет сервере, обладающем достаточным объёмом дискового пространства и пропускной способности канала доступа и дисковой системы. Внесение результатов расчетов в базу осуществляется сред-

Табл. 3. Ключевые базовые функции платформы интеграции  
 Звездочкой (\*) отмечены функции, которые дополняются прикладным приложением

Функция	Краткое описание
Создание проекта*	Обработка исходных данных от пользователя, запуск экземпляра СУБД, создание таблиц для хранения результатов, конфигурирование копии исполняемого модуля
Оценка времени выполнения проекта*	Пробный запуск сконфигурированного исполняемого модуля для выполнения малой части всех вычислений с целью оценки производительности
Запуск проекта	Отправка необходимых файлов на вычислительные узлы, запуск непосредственно вычислений
Визуализация*	Запрос к БД результатов, подготовка визуального представления
Остановка проекта	Остановка вычислений и очистка вычислительных узлов
Удаление проекта	Остановка проекта, затем удаление экземпляра БД, удаление соответствующей копии исполняемого модуля и его конфигурации.

ствами платформы по окончании вычислений. Такой подход удобен для передачи результатов расчета на локальный компьютер пользователя, обеспечивает простоту масштабирования системы хранения, расширяет возможности индивидуальной настройки параметров СУБД.

Основной метод визуализации результатов, реализованный в платформе, заключается в построении одномерных графиков и двумерных диаграмм. Эти средства были выбраны как наиболее распространенные, простые в реализации и достаточно универсальные. Так, в пилотном приложении [10], реализованном на основе нашей платформы, этими средствами строятся двумерные сечения многомерной функции, а в приложении, описанном в следующем разделе – спектрально-временные диаграммы. Помимо графического отображения результатов, платформа предоставляет интерфейс непосредственного обращения к полям базы данных, который представляет собой простой интерпретатор SQL-команд, обеспечивающих поиск и отбор результатов расчета без возможности их модификации.

## 2. Грид-приложение

### «Полигармоническая модель»

В качестве примера применения платформы интеграции опишем созданный на ее основе web-интерфейс спектрально-временного анализа (СВАН) геомагнитных данных. Спектрально-временная диаграмма (СВАН-диаграмма) представляет собой частотный спектр сигнала, вычисляемого в скользящем

временном окне. Оценка временных изменений спектрального состава вариаций геомагнитного поля является мощным инструментом исследования геомагнитных возмущений и широко применяется в геофизике [18-20]. Вместо традиционного расчета СВАН-диаграмм с применением дискретного преобразования Фурье (ДПФ) мы использовали полигармоническую модель, в которой фрагмент записи компоненты геомагнитного поля аппроксимируется комбинацией небольшого числа гармоник. Такой подход получил применение, например, при обработке речевых сообщений [21, 22], описании излучаемых радарными радиоимпульсов [23], геоинформационном моделировании [24]. Использование полигармонической модели имеет ряд преимуществ по сравнению с традиционным расчетом спектра [25]. В частности, она позволяет ослабить ограничение разрешающей способности, которое имеет место в ДПФ (Приложение).

Построение СВАН-диаграммы на основе полигармонической модели осуществляется следующим образом. задается размер временного окна  $\tau$ , для которого определяется набор  $L$  частот. Значения частот и соответствующих им амплитуд подбираются из условия минимума функционала (2). После этого окно смещается на время  $dt$  и подбор оптимальных параметров осуществляется для нового интервала. Выбор оптимальных частот производится перебором возможных значений, с учетом условия (1). Для этого интервал частот  $[\omega_1, \omega_2]$  разбивается на  $K + 1$  частей. Каждому из возможных значений частоты соответствует целое число  $k$ , такое что

$$\omega_k = \omega_1 + k\Delta\omega, \quad 0 \leq k \leq K.$$

Вместо набора частот удобно использовать индекс — порядковый номер сочетания  $\{k_1, k_2, \dots, k_L\}$  в лексикографически упорядоченном списке всех возможных таких сочетаний. Эффективное вычисление значений частот по индексу осуществляется с помощью алгоритма, описанного в статье [27].

Основная проблема при построении СВАН-диаграмм связана с нелинейной зависимостью функционала (2) от частот. Использование локальных методов минимизации, основанных на линеаризации исходных выражений вблизи некоторой точки, может привести к неверному результату из-за наличия несколько локальных минимумов, как это показано в статье [26]. Очевидный метод поиска решения — непосредственный перебор всех возможных сочетаний частот. Пусть заданы  $L$  — количество частот, используемых в модели, и граница частотного интервала  $\omega_1, \omega_2$ . Предположим также, частоты могут иметь одно из  $K + 1$  значений, расположенных в этом интервале. Решение ищется прямым перебором всех возможных комбинаций на равномерной сетке с узлами и шагом по частоте

$$\Delta\omega = (\omega_2 - \omega_1)/K.$$

Для каждой комбинации частот по формулам (3) вычисляются амплитуды гармоник и значение функционала.

При возрастании числа используемых частот  $L$  и узлов сетки  $K$  объем вычислений, необходимых для поиска оптимальных значений частот, существенно возрастает. С учетом условия (1), число возможных комбинаций  $N$  равно, числу сочетаний из  $K + 1$  по  $L$ :

$$N = (K + 1)!/L!(K + 1 - L)!.$$

Это число, хотя и меньше исходного  $(K + 1)^L$ , при увеличении  $L$  растет экспоненциально. Действительно, когда  $L = (K + 1)/2$  и число сочетаний максимально, из формулы Стирлинга имеем

$$N \sim 2^{2L}/\sqrt{\pi L}, \quad L \gg 1.$$

К счастью, решение системы (3) для фиксированного набора частот никак не связано с расчетами при других значениях этих параметров. Это позволяет эффективно использовать

распределенные вычислительные системы для построения СВАН-диаграммы.

Как отмечалось выше, если числа  $L, K$  достаточно велики, то для проведения расчетов необходимо использование распределенных вычислений. Некоторые общие аспекты организации параллельных вычислений при использовании полигармонической модели для построения СВАН-диаграмм рассмотрены в работе [28]. Ниже мы приводим описание соответствующего сервиса, реализованного в виде веб-приложения на основе описанной выше платформы интеграции.

Для реализации расчета СВАН-диаграммы в грид было разработано простое приложение, основанное на разработанной платформе. Это означает создание ряда компонент: веб-интерфейс ввода исходных данных, SQL-схема хранения результатов расчета в базе данных и, наконец, исполняемый модуль, осуществляющий расчет целевого функционала для заданного набора частот.

Веб-интерфейс прикладного приложения состоит из единственного вида (Рис. 2). Он предназначен для загрузки файла с магнитограммой и задания входных параметров: величины шага по времени  $dt$ , временного окна для вычисления спектра  $N$ , диапазона циклических частот  $\omega_1, \omega_2$ , количества частот, используемых в модели  $L$  и число частотных узлов  $K$ .

Так как время расчета существенно зависит от входных параметров — в нашем случае это параметры  $L$  и  $K$  — в интерфейсе предусмотрена возможность оценки времени, необходимого для полного выполнения. Реализация основана на функции, входящей в число встроенных средств разработанной нами платформы.

На этапе создания прикладного приложения пользователь должен предоставить файлы исполняемый модулем и настроечными параметрами к нему. Исполняемый модуль должен быть скомпилирован с учетом оборудования и программного обеспечения вычислительных узлов грид. В нашем случае исполняемый модуль представляет собой С-код, собранный компилятором GCC (<http://gcc.gnu.org>). Предоставленные пользователем файлы размещаются на машине запуска и контроля. Туда же с помощью веб-интерфейса, помещается файл с

# High performance geoscientific application manager | New POLYFREQ1 form

Dashboard

## Define your new Polyfrequency project

### General parameters

Project distinctive name:

Project descriptive phrase (e. g. Test pattern #1 w/noise):

Project temporary id:

### Project parameters

datafile:  No...n

dt:  H:  dH:  K:  L:   $\omega_1$ :   $\omega_2$ :

Estimate Estimation provides approximate time requirements  I confirm that the above is ok

Рис. 2. Веб-интерфейс этапа внесения исходных данных в приложении расчёта СВАН-диаграмм

Предусмотрена возможность загрузки файла с магнитограммой и задания параметров запуска. Уникальный временный идентификатор проекта (Project temporary id) используется при разработке в диагностических целях.

анализируемой магнитограммой. Сразу после этого на машине диспетчеризации СУБД автоматически создается экземпляр базы данных, предназначенный для хранения результатов вычислений, а загруженные данные средствами платформы размещаются на каждом из вычислителей в составе грид. Распределение заданий по вычислительным узлам происходит автоматически, встроенными механизмами Gridway. Внесение результатов производится средствами платформы и осуществляется с помощью удаленного SQL-соединения к экземпляру СУБД.

Заметим, что время расчета амплитуд для одной комбинации частот в нашей задаче мало, поэтому на один вычислительный узел передавался набор заданий, определяемый диапазоном индексов наборов частот. Величина уско-

рения, достигаемого за счет использования распределенных вычислений, существенно зависит от размера этого пакета. Это обусловлено, главным образом, значительными накладными расходами на процедуру постановки в очередь и запуска заданий.

На Рис. 3 представлен график ускорения расчетов СВАН-диаграмм при использовании грид в зависимости от числа сочетаний, передаваемых на один узел одновременно. Была использована модель с пятью модами, частотный интервал делился на 32 отрезка. Вычисления производились для 450 временных интервалов. Отметим, что такой выбор параметров значительно перекрывает предполагаемый объем вычислений в геомагнитных исследованиях. Из рисунка видно, что оптимальный размер

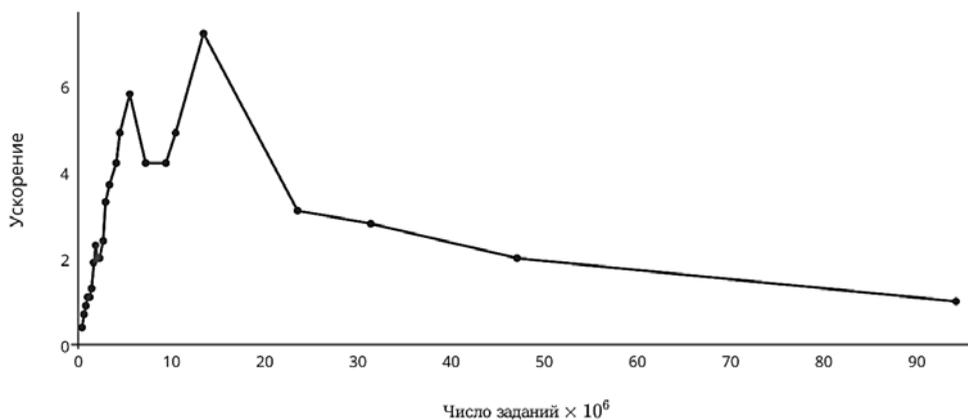


Рис.3. Зависимость среднего (по нескольким запускам) ускорения расчета СВАН-диаграммы с применением полигармонической модели от числа заданий, передаваемых одновременно на один грид-узел (прочие параметры расчета были фиксированы)

## High performance geoscientific application manager | Data query

Dashboard

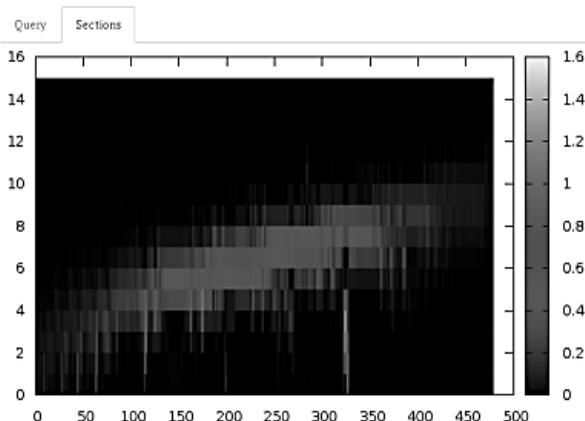


Рис. 4. Вид графического представления результатов расчета.

В описываемом приложении используется для отображения СВАН-диаграмм

## High performance geoscientific application manager | Data query

Dashboard

idx	cost	k
201376	0.0286959	0
201028	0.0310214	1
201033	0.0305358	2
201027	0.0295046	3
201001	0.0276082	4
10	0.0251104	5
31473	0.0214866	6
120647	0.0235438	7
120647	0.0297467	8
135597	0.0366564	9

Рис.5. Вид, предназначенный для анализа данных при помощи простых запросов

пакета в нашей системе равен, приблизительно, 12 млн. частотных комбинаций. Полный объем вычислений составляет около 100 млн. комбинаций и занимает около часа, при расчете на одном ядре и около 10 минут — в грид.

Как было описано ранее, в состав платформы входят средства, предназначенные для анализа и визуализации результатов расчетов. Одно из них используется для построения СВАН-диаграммы (Рис. 4), а другое предоставляет возможность непосредственного отображения рассчитанных данных (Рис. 5). Имеется также возможность выгрузки результатов на локальный компьютер пользователя в графическом виде или в *ASCII*-файл. При необходимости бо-

лее детального анализа пользователь может получить копию базы данных с расчетами.

Для проведения тестовых вычислений использовалась созданная ранее грид-инфраструктура, состоящая из двух территориально распределённых кластеров и управляющего сервера [29]. Компоненты машины запуска и контроля и машины диспетчеризации СУБД в нашем случае располагались на одном компьютере. Гибкость использования ресурсов каждого кластера обеспечивалась применением технологии виртуализации. Подробное описание инфраструктуры и деталей организации вычислений приведено в статье [12].

### Заключение

Таким образом, разработана программная платформа для грид-приложений. Платформа обеспечивает возможность решения задач определённого класса без необходимости погружения в технологические особенности распределённых вычислений. Создание приложений на основе платформы сводится к реализации веб-интерфейса и средств стыковки с промежуточным программным обеспечением. Помимо удобств, использование веб-интерфейса обеспечивает дополнительную безопасность за счёт ограничения кода, исполняемого пользователем. Это позволило, в частности, отказаться от аутентификации пользователей стандартными для связующего программного обеспечения средствами.

В качестве примера приведено веб-приложение, обеспечивающее доступ к ресурсам грид при расчете СВАН-диаграмм с использованием полигармонической модели. Описанный веб-интерфейс предоставляет возможность редактировать параметры вычислений, оценить время выполнения задания, отслеживать степень завершенности задания, а также предоставляет возможность проанализировать полученные данные. Пробные расчеты показали возможность использования приложения в геомагнитных исследованиях.

Авторы благодарны С.Д. Иванову за обсуждение работы и полезные замечания.

### Приложение

#### Построение спектрально-временной диаграммы сигнала на основе полигармонической модели

Пусть на временном промежутке  $\tau$  задан сигнал  $z(t_i)$ , определенный для моментов времени

$$t_i = i\Delta t, \quad \Delta t = \tau / (N - 1), \quad 0 < i < N,$$

$i, N$  — натуральные числа. Наша цель состоит в аппроксимации последовательности  $z(t_i)$  небольшим числом тригонометрических функций вида

$$\{a_l \cos \omega_l t, b_l \sin \omega_l t\}, \quad l = 1, \dots, L.$$

Здесь  $L$  — число используемых гармоник  $\omega_l = f_l / 2\pi$  из интервала циклических частот  $[\omega_A, \omega_B]$ , величины которых, не ограничивая общности, можно считать упорядоченными:

$$\forall i \in [1, L]: \quad \omega_i < \omega_{i+1}, \quad \omega_A \leq \omega_i \leq \omega_B. \tag{1}$$

При заданном числе гармоник задача сводится к поиску частот и амплитуд, минимизирующих функционал

$$F(\vec{\omega}; \vec{a}, \vec{b}) = \frac{1}{N} \sum_{i=0}^N \left[ z(t_i) - \sum_{l=1}^L (a_l \cos \omega_l t_i + b_l \sin \omega_l t_i) \right]^2. \tag{2}$$

Здесь символы со стрелками являются  $n$ -мерными векторами. Если предположить, что количество и величины частот заданы, то в силу квадратичной зависимости целевого функционала (2) от амплитуд гармоник  $\vec{a}, \vec{b}$ , определение последних очевидным образом сводится к решению системы линейных уравнений

$$\sum_{l=1}^L (A_{ml} a_l + C_{ml} b_l) = Z_m^{(a)}, \quad \sum_{l=1}^L (C_{lm} a_l + B_{ml} b_l) = Z_m^{(b)}. \tag{3}$$

Здесь введены обозначения:

$$\begin{aligned}
 A_{ml} &= \frac{1}{N} \sum_{i=0}^N \cos \omega_m t_i \cos \omega_l t_i = A_{lm}, \\
 B_{ml} &= \frac{1}{N} \sum_{i=0}^N \sin \omega_m t_i \sin \omega_l t_i = B_{lm}, \\
 C_{ml} &= \frac{1}{N} \sum_{i=0}^N \cos \omega_m t_i \sin \omega_l t_i, \\
 Z_m^{(a)} &= \frac{1}{N} \sum_{i=0}^N z(t_i) \begin{pmatrix} \cos \omega_k t_i(4) \\ \sin \omega_l t_i \end{pmatrix}.
 \end{aligned}$$

Покажем на примере модели с двумя частотами, что применение полигармонической модели позволяет улучшить частотное разрешение. Используя известные тригонометрические соотношения и учитывая [30], что

$$\begin{aligned}
 \frac{1}{2n} \sum_{k=0}^n \cos kx &= \frac{1}{2n} \sin \left( \frac{n+1}{2} x \right) \cos \left( \frac{nx}{2} \right) \cos \left( \frac{x}{2} \right) \equiv \Phi_1(n, x), \\
 \frac{1}{2n} \sum_{k=1}^n \sin kx &= \frac{1}{2n} \sin \left( \frac{nx}{2} \right) \sin \left( \frac{n+1}{2} x \right) \cos \left( \frac{x}{2} \right) \equiv \Phi_2(n, x),
 \end{aligned} \tag{6}$$

элементы матриц  $A, B, C$  можно представить в виде

$$\begin{aligned}
 A_{ml} &= \Phi_1(N, \Omega_{ml}^-) + \Phi_1(N, \Omega_{ml}^+), \\
 B_{ml} &= \Phi_1(N, \Omega_{ml}^-) - \Phi_1(N, \Omega_{ml}^+), \\
 C_{ml} &= \Phi_2(N, \Omega_{ml}^-) + \Phi_2(N, \Omega_{ml}^+), \\
 \Omega_{ml}^\pm &= (\omega_m \pm \omega_l) \Delta t.
 \end{aligned}$$

Из последних формул следует, что для дискретных сигналов при выполнении условий

$$N \gg 1, \quad |\Omega^- \Delta t| \gg 1 \tag{7}$$

матрица коэффициентов в уравнениях (3) близка к диагональной:

$$A_{ll}, B_{ll} \sim \frac{1}{2} + O\left(\frac{1}{N}\right), \quad A_{lm}, B_{lm}, C_{lm} \sim O\left(\frac{1}{N}\right) \quad (l \neq m). \tag{8}$$

Недиагональными членами уравнений (3) в этом случае можно пренебречь. Смысл первого из условий (7) очевиден, а второе условие, фактически, определяет разрешающую способность метода по частоте. Действительно, приведенные выше оценки (8) перестают быть справедливыми при значениях второго аргумента функций  $\Phi_{1,2}(N, \Omega)$ , близких к

$$\Omega_1 \approx i \frac{\pi}{2N} + k, \quad \Omega_2 \approx i \frac{\pi}{N} + k, \quad i, k \in Z,$$

соответственно. Здесь  $Z$  — множество целых чисел. В указанных точках функции  $\Phi_{1,2}(N, \Omega)$  имеют локальные экстремумы, величины которых зависят только от комбинации  $N\Omega$ . Это хорошо видно из формул (6), выписанных для малых значений  $\Omega$  и достаточно больших  $N$ , чтобы можно было пренебречь единицей в аргументах синуса:

$$\Phi_1(N, \Omega) \approx \frac{1}{2} \frac{\sin(N\Omega)}{N\Omega}, \quad \Phi_2(N, \Omega) \approx \frac{1}{2} \frac{(1 - \cos(N\Omega))}{N\Omega}.$$

Более того, в первом приближении значения функций в указанных выше экстремумах не зависят от числа отсчетов  $N$ , то есть коэффициенты в уравнениях (3) становятся одного порядка. В этом случае оценки (8) перестают быть справедливыми, а условие разрешимости системы приводит к ограничению разрешающей способности по частоте: разность соседних частот не должна быть намного меньше величины

$$|\Delta f| = |f_l - f_m| \approx \frac{1}{N\Delta t}.$$

Однако, в нашем случае, в отличие от дискретного преобразования Фурье, указанное условие не является строгим, и допустимая разность частот может быть в несколько раз меньше величины  $1/N\Delta t$ . Для демонстрации этого факта рассмотрим модель с двумя частотами. В интересующем нас случае определитель системы (3)  $\Delta$  равен приближенно

$$\Delta(x) \approx \frac{1}{16} \left(1 - \frac{\sin^2 x}{x^2}\right) \left(1 - \frac{\sin^2 x}{x^2} - \frac{(1 - \cos x)^2}{x^2}\right) \quad (x = N\Omega^-).$$

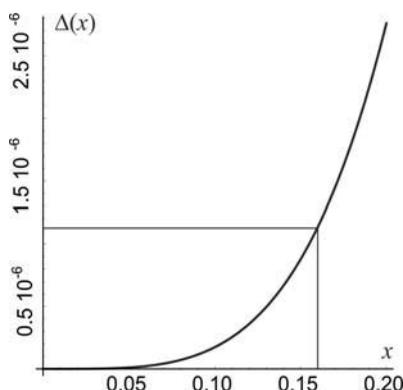


Рис. 6. Зависимость дискриминанта системы уравнений (3)  $\Delta$  от величины  $x = 2\pi N|f_2 - f_1|$  для модели с двумя частотами. Линиями помечены значение  $x = 1/(2\pi)$  и соответствующее ему значение дискриминанта.

Из рисунка видно, что при  $x = 1/(2\pi)$  ( $N\Delta f = 1$ ) определитель  $\Delta(x)$  заметно отличен от нуля и остается таким вплоть до значений аргумента близких к  $x \sim 0.05$ , что означает возможность двух-трехкратного увеличения частотного разрешения по сравнению с методом дискретного преобразования Фурье.

### Литература

1. A.S. Tanenbaum and M. van Steen. Distributed Systems: Principles and Paradigms. Pearson Prentice Hall, 2007.
2. David P Anderson. Public computing: Reconnecting people to science. In Conference on Shared Knowledge and the Web, pages 17–19, 2003.
3. Ian Foster and Carl Kesselman. Globus: A metacomputing infrastructure toolkit. International Journal of High Performance Computing Applications, 11(2):115–128, 1997.
4. José Luis Vázquez-Poletti, Eduardo Huedo, Rubén S Montero, and Ignacio Martín Llorente. A comparison between two grid scheduling philosophies: EGEE WMS and GridWay. Multiagent and Grid Systems, 3(4):429–439, 2007.
5. Erwin Laure, A Edlund, F Pacini, P Buncic, S Beco, F Prelz, A Di Meglio, O Mulmo, M Barroso, Peter Z Kunszt, et al. Middleware for the next generation grid infrastructure. Technical report, CERN, 2004.
6. Ian Foster. Service-oriented science. Science, 308(5723):814–817, 2005.
7. Sukhoroslov O., Afanasiev A. Everest: A Cloud Platform for Computational Web Services. In Proceedings of the 4th International Conference on Cloud Computing and Services Science (CLOSER 2014). SCITEPRESS - Science and Technology Publications, 2014, pp. 411–416.
8. Ю.А. Бычков, Г.А. Опарин, А.Г. Феоктистов, В.Г. Богданова, и А.С. Корсуков. Сервис-ориентированный подход к организации распределенных вычислений с помощью инструментального комплекса DISCENT. Информационные технологии и вычислительные системы, (2):7–15, 2014.
9. Grid execution management for legacy code architecture.

10. Apache Ignite. The Apache Foundation. URL: <http://ignite.incubator.apache.org>
11. Univa Grid Engine. Univa Corporation. URL: <http://univa.com>
12. И.М. Алешин и К.И. Холодков. Применение распределённых вычислительных систем к расчёту апостериорных распределений. Геофизические исследования, 15(4):73–80, 2014.
13. Walfredo Cirne, Daniel Paranhos, Lauro Costa, Elizeu Santos-neto, Francisco Brasileiro, Jacques Sauv e, Campina Grande, Fabr cio Alves, Barbosa Silva, Cat lica Santos, Carla Osthoff Barros, Laborat rio Nacional, Computa o Cient fica, Cirano Silveira, and Hewlett Packard. Running bag-of-tasks applications on computational grids: The mygrid approach. In In ICPP, page 407, 2003.
14. Disco. Nokia Research Center URL: <http://discoproject.org>
15. John Deacon. Model-view-controller (mvc) architecture. Technical report, JDL, 2009.
16. AB MySQL. MySQL: the world's most popular open source database. MySQL AB, 1995.
17. Josh Juneau. Object-relational mapping. In Java EE 7 Recipes, pages 369–408. Springer, 2013.
18. А.Н. Зайцев, В. И. Одинцов, В. В. Иванов. Спектральные особенности восточной и западной электроструй в период магнитной бури 24 марта 1991 г. Геомагнетизм и аэрномия, 39(1):35–41, 1991.
19. В.И. Одинцов, Н. М. Ротанова, Ю. П. Цветков, Ан Ченчанг. Спектральный анализ аномального магнитного поля земли для разновысотных съёмок. Геомагнетизм и аэрномия, 40(2):59 – 66, 2000.
20. A Dmitriev, A. Belov, R Gorgutsa, V Ishkov, V. V. Kozlov, R Nymmik, V Odintsov, A Petrukovich, G Popov, E Romashets, et al. The development of the Russian space weather initiatives. Advances in Space Research, 31(4):855–860, 2003.
21. А.Б. Барабанов. Идентификация параметров полигармонической модели речевого сигнала. Материалы XII Всероссийского совещания по проблемам управления (ВСПУ-2014), С. 3038–3049, 2014.
22. А. А. Мельников. Быстрый алгоритм идентификации параметров модели голосового сигнала. Материалы XII Всероссийского совещания по проблемам управления (ВСПУ-2014), с. 3090–3101, 2014.
23. David C Rife and Robert R Boorstyn. Multiple tone parameter estimation from discrete-time observations. Bell System Technical Journal, 55(9):1389–1410, 1976.
24. Д.И. Якушев. Геоинформационное моделирование пространственно-временных геофизических процессов с полигармонической структурой. Автореферат на соискание учёной степени доктора технических наук, 2008.
25. В.Г. Гетманов. Алгоритм разделения близких по частоте источников вибрации. Колебания и вибрационная активность машин и конструкций, С. 157–160. Наука, 1988.
26. В.Г. Гетманов. Об алгоритме поиска по частоте в задаче оценивания параметров моделей полигармонических сигналов. Автометрия, 45(3):83–89, 2009.
27. Н.Е. Тимошевская. О нумерации перестановок и сочетаний для организации параллельных вычислений в задачах проектирования управляющих систем. Известия Томского политехнического университета, 307(6):18–19, 2004.
28. Борзунов Г.И. Гетманов, В. Г. Алгоритм параллельных вычислений для задачи спектрально - временного анализа на базисных полигармонических функциях. Информационные технологии, 21(9), 2015.
29. К.И. Холодков, И.М. Алешин, В.Н. Корягин, О. В. Сухорослов, А.Н. Шогин. Опыт развертывания грид-инфраструктуры для поддержки вычислительных веб-сервисов. Научно-техническая информация. Серия 1, (4):15–19, 2012.
30. А.П. Прудников, Ю.А. Бычков, and О.И. Маричев. Интегралы и ряды: Элементарные функции, т. 1. "Наука," М.: 1981.

**Алешин Игорь Михайлович.** Заведующий лабораторией Федерального государственного бюджетного учреждения науки Институт физики Земли им. О.Ю. Шмидта Российской академии наук; ведущий научный сотрудник Федерального государственного бюджетного учреждения науки Геофизический центр Российской академии наук. Окончил Московский государственный университет им. М.В. Ломоносова в 1990 году. Кандидат физико-математических наук. Автор 48 печатных работ. Область научных интересов: геоинформатика. E-mail: [ima@ifz.ru](mailto:ima@ifz.ru)

**Холодков Кирилл Игоревич.** Научный сотрудник Федерального государственного бюджетного учреждения науки Институт физики Земли им. О.Ю. Шмидта Российской академии наук. Окончил Московский технологический университет (МИРЭА) в 2011 году. Кандидат технических наук. Автор 17 печатных работ. Область научных интересов: распределенные вычисления. E-mail: [keir@ifz.ru](mailto:keir@ifz.ru)

**Гетманов Виктор Григорьевич.** Главный научный сотрудник Федерального государственного бюджетного учреждения науки Геофизический центр Российской академии наук и Федерального государственного бюджетного учреждения науки Институт физики Земли им. О.Ю. Шмидта Российской академии наук. Окончил Московский инженерно-физический институт 1968 году. Доктор технических наук. Автор 157 печатных работ. Область научных интересов: методы цифровой обработки и анализа рядов геофизических данных. E-mail: [v.getmanov@gcras.ru](mailto:v.getmanov@gcras.ru)

**Корягин Владимир Николаевич.** Ведущий научный сотрудник Федерального государственного бюджетного учреждения науки Институт физики Земли им. О.Ю. Шмидта Российской академии наук. Окончил Московский государственный университет им. М.В. Ломоносова в 2002 году. Кандидат экономических наук. Автор 32 печатных работ. Область научных интересов: геоинформатика. E-mail: [vlad@ifz.ru](mailto:vlad@ifz.ru)