

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Центр дополнительных образовательных программ по направлениям
математика, процессы управления, физика и химия

«Проектирование и создание информационно-вычислительной системы
для анализа океанографических данных
(на основании материалов международной программы АРГО)»

Дипломная работа слушателя

к.ф.-м.н. Рожиной Е.А.

выполнена под руководством

Смирнова А.А.

Санкт-Петербург

2019 год

Содержание

1. Введение	3
2. Постановка задачи	3
3. Океанографическая информация	
3.1 Основные гидрофизические характеристики Мирового океана	4
3.2 Подходы к построению Баз Океанографических Данных (БОД)	8
3.3 Проект ARGO	9
4. Информационно-вычислительная система по работе с данными ARGO	
4.1 Требования к разрабатываемой системе	12
4.2 Программные средства разработки ПО и хранения данных	13
4.3 Этапы разработки ПО	13
4.4 Методики расчетов	18
4.5 Описание программного комплекса	21
5. Заключение	24
Литература	
Ссылки	
Приложение	

1. Введение

Температура, соленость, плотность, скорость и направление течений, уровень поверхности моря являются основными физическими параметрами Мирового океана. В настоящее время накоплен большой объем временных рядов характеристик состояния океана и морской воды, который играет роль фундаментальной основы для разнообразных исследований Мирового океана – климатологических, экологических, гидрологических и других. Различными институтами накоплен значительный опыт работ по формированию баз океанографических данных (А.Х. Халиулин, 2016). Кроме того, существуют крупные проекты, обеспечивающие получение и хранение океанографических данных и предоставление их пользователям. Международный проект ARGO^[10] предоставляет данные измерений, полученные глобальной сетью дрейфующих буев-измерителей и обработанные в вычислительных центрах. Эти данные представляют собой исходную наблюдательную информацию, необходимую для различных вычислений. Формат предоставляемых данных – NetCDF^[1] (С.М. Гордеева, О.И. Шевчук, 2013). Это транспортный контейнер, разработанный для хранения и распространения данных, но неудобный для анализа больших массивов данных и произвольных подмножеств, а также вычислений. Пользователи сами декодируют NetCDF-файлы, выбирая необходимую информацию и организуя ее в соответствии с целями своих исследований. Типичной является ситуация, когда институт или другая организация разрабатывает собственный банк данных, сообразно своей специфике^[6]. Подобная задача стоит перед 280 Центром картографического производства ВМФ, который является специализированным центром сбора и хранения результатов океанографических исследований Министерства обороны РФ. Данные, хранимые в 280 Центре, а это более 20 млн. серий различных измерений (Балясников С. Б., Мاستрюков С. И., 1998), необходимо дополнить профилями, полученными буями ARGO, которые будут использоваться в научных целях.

2. Постановка задачи

Целью данной дипломной работы является проектирование и создание информационно-вычислительной системы, реализующей хранение океанографических данных и предоставление пользователям доступа к этим данным. Разрабатываемая система предназначена для использования в научных исследованиях. Особым требованием

является использование СУБД PostgreSQL и работоспособность системы без доступа к сети Internet. В рамках дипломной работы должны быть решены следующие задачи:

1. создание базы данных для хранения океанографических измерений (профилей) давления, температуры, солености, метаданных и результатов вычислений глубины, скорости звука, плотности воды и их ошибок.
2. создание программного комплекса, реализующего следующие функции:
 - загрузку данных из файлов в формате NetCDF с физического носителя, включая валидацию профилей, метаданных, контроль качества и дополнение измерений расчетными значениями;
 - поиск данных через веб-интерфейс, получение выборки (среза, подмножества) набора данных, экспорт в различные форматы по требованию пользователей, вычисление базовых элементов статистики по созданной выборке, сохранение результатов анализа;
 - визуализацию данных;
3. первичное наполнение БД реальными данными не более 10 млн. записей.

Здесь и далее используются следующие термины:

профиль – зависимость изменения гидрофизических характеристик от глубины(или давления), серия измерений, выполненных на различных глубинах одной вертикали в течение короткого периода времени; *станция* – сеанс измерений, главные характеристики: географическая точка в Мировом океане и момент времени, в который производились наблюдения; *горизонт* – глубина; *стандартные горизонты* – постоянные горизонты наблюдений, принятые для океанографических измерений и отбора проб (ГОСТ 18451-73-ГОСТ 18458-73, М., 1973).

3. Океанографическая информация

3.1 Основные гидрофизические характеристики Мирового океана

К основным видам гидрофизических характеристик Мирового океана относятся: *соленость, давление, температура*. Эти параметры являются основой для вычисления плотности морской воды и, что особо важно, скорости звука в морской воде.

Соленость – удельное количество растворенных в воде солей. Измеряется в «‰» (промилле), в ПЕС (практические единицы солёности) или PSU (англ. Practical Salinity

Units) практической шкалы солёности (англ. Practical Salinity Scale). Средняя солёность Мирового океана — 35 ‰ или 1 ПЕС. Для калибровки приборов в Бискайском заливе добывается так называемая нормальная вода с солёностью, близкой к 35 ‰ или 1 ПЕС. При максимальном испарении и наименьшем количестве осадков, наблюдается повышенное значение солёности. От поверхности до дна океана солёность уменьшается. Донные воды от экватора до арктических широт имеют солёность 34,7-34,8‰. В средних широтах избыток испарения над осадками приводит к тому, что поверхностные воды становятся более солёными, чем глубинные.

Температура океанической воды снижается с глубиной, сначала скачкообразно, а потом плавно. На глубине температура воды практически однородна, так как глубина океанов в основном заполнена водами одинакового происхождения, которые формируются в полярных областях Земли. На глубине более 3-4 тысяч метров температура колеблется от + 2 до 0 °С. Температура Мирового океана зависит от широты и распределяется на ее поверхности зонально. Наибольшая средняя температура находится около экватора и составляет 27- 28 °С. На температуру морской воды большое влияние оказывает климат окружающих территорий. Например, температура воды Красного моря, окруженного пустынями, достигает 34°С. В умеренных широтах температура морской воды зависит от времени года и даже времени суток. Кроме того, на температуру океанических вод оказывают сильное влияние и течения: теплые течения переносят воды от экватора в умеренные широты, а холодные течения — от полярных областей. Подобное перемешивание вод способствует более равномерному распределению температур в водных массах. Средняя температура поверхностного слоя Мирового океана достигает 17,5°С, с глубиной происходит ее снижение, однако, температура воды в горячих источниках на дне океана составляет 400 °С. Наибольшая средняя температура у поверхности воды в Тихом океане равна 19°С, в Индийском — 17°С, в Атлантическом океане — 16°С, в Северном Ледовитом океане — 1°С.

Давление в морях и океанах увеличивается примерно на 1 бар на каждые 10 м глубины. Его значение зависит от плотности воды. Под воздействием давления вышележащих слоев вода сжимается. Сжимаемость морской воды незначительна: при нормальной солёности (35‰) и $t^{\circ} = 15^{\circ}\text{C}$ она равна 0,0000442. В случае абсолютной несжимаемости воды объем Мирового океана увеличился бы на 11 миллионов км³, и его уровень повысился бы на 30 м.

Плотность морской воды: отношение массы воды к занимаемому ей объему. В морских глубинах плотность нелинейно растет с увеличением давления. Другие факторы, влияющие на плотность воды — температура и солёность. При уменьшении температуры

морской воды и повышении ее солености, увеличивается плотность воды (рис. 2.5). Например, при значениях солености 35 ‰ и температуре 0°C плотность морской воды составляет 1,02813 г/см³ (масса каждого кубического метра такой морской воды на 28,13 кг больше, чем такого же объема дистиллированной воды).

Скорость звука. В отличие от воздуха, вода слабо поглощает энергию звуковых волн, при некоторых условиях звук в морской воде может распространяться на очень большие расстояния (порядка тысяч километров). Скорость звука не зависит от его частоты и обычно находится в пределах от 1400 до 1550 м/с. Средняя скорость звука приблизительно равна 1500 м/с, что в 4-5 раз выше скорости звука в атмосфере. Знание зависимости скорости звука от глубины в заданном регионе необходимо для подводной связи, навигации, обнаружения различных объектов, ультразвукового сканирования рельефа дна и многих других практических и научных целей. С увеличением температуры, солености и давления (глубины) происходит возрастание скорости звука, причем наибольшее влияние оказывает температура. Так, при изменении температуры на 1°C скорость звука в морской воде увеличивается на 4,4 м/с, при увеличении солености на 1 ‰ - на 1,2 м/с и при возрастании глубины на 100 м - на 1,75 м/с.

К явлениям, связанным с распространением звука в морской воде, относятся реверберация, рефракция, подводный звуковой канал. Рефракция в морской воде вызывает искривление направления распространения звуковой волн в акустически неоднородной среде. Характер преломления определяется знаком и величиной градиента скорости звука. Наибольшие градиенты скорости звука в море наблюдаются в вертикальной плоскости, поэтому в этой плоскости присутствует и наибольшая рефракция. Звуковой канал-слой морской среды, в котором звуковые волны благодаря рефракции распространяются, не выходя за пределы слоя, с малыми потерями при отражении от поверхности моря и дна, благодаря чему звук может преодолевать необычно расстояния-сотни и даже тысячи км. Реверберация представляет собой убывание энергии звукового сигнала из-за его отражения на неоднородностях окружающей среды. Неоднородностями в толще воды являются воздушные пузырьки, взвешенные частицы, неровности поверхности моря и дна, живые организмы.

Все эти факторы имеют особую важность при проектировании гидроакустической аппаратуры и должны приниматься во внимание при ее использовании в океане. Одной из задач океанологии является выявление для разных районов и сезонов типичных зависимостей скорости звука по глубине. В одной из крупнейших работ, атласе значений скорости звука в северной части Атлантического океана (US Naval Oceanographic Office, 1967) обработка результатов проводилась в квадратах 110x110 км по четырем сезонам с

осреднением за 90 суток. Результатом этой работы являются наборы графиков вертикального распределения скорости звука, полученные по точкам на стандартных горизонтах, и огибающие осредненных кривых. Подобные графики необходимы для представления масштабов изменчивости скорости звука (Рис.1): пространственной, временной и пространственно-временной.

Наблюдаемые кривые распределения скорости звука могут быть разделены на типы. *Типизация кривых* сводится к отнесению реальных распределений скорости к той или иной модели среды. При этом модели среды описываются определенным набором опорных точек. Районирование кривых это процесс определения пространственно-временных границ, определяющих множество типизированных кривых (Матвиенко, В.Н., 1981).

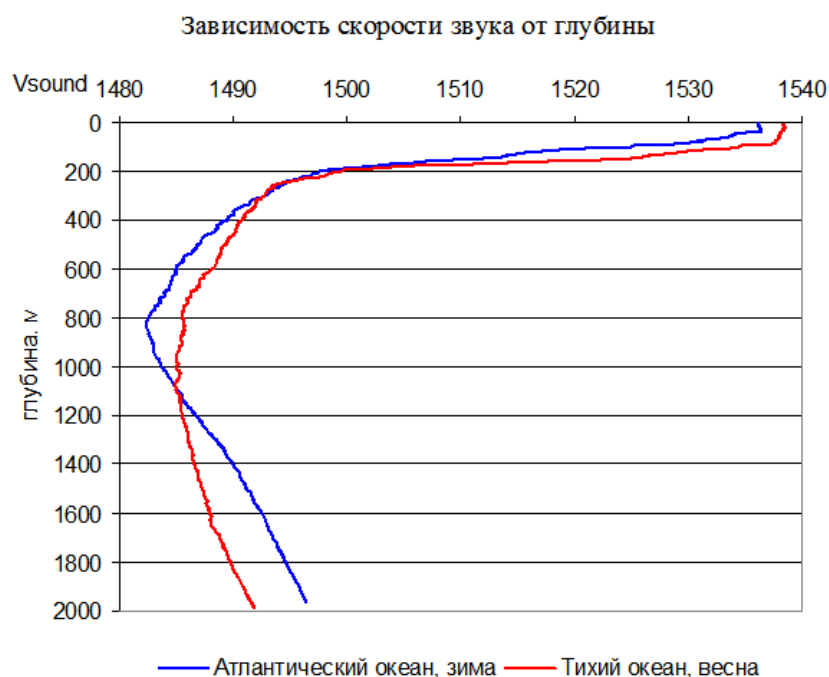


Рис 1. Показаны кривые вертикального распределения значений скорости звука в различные сезоны в Тихом и Атлантическом океане. Профили получены буями ARGO №5904716 в марте 2016г. (красный график) и №3901880 в январе 2018г. (синий)

Мониторинг гидрофизических характеристик, накопление измеренных данных и возможность анализа больших массивов измерений для осреднения результатов определения скорости звука по районам и сезонам имеют первоочередное значение для задач гидроакустики.

3.2 Подходы к построению баз океанографических данных

Различными исследователями к настоящему времени выработан ряд основных требований к созданию баз океанографических данных (БОД). Это, во-первых, полнота использования исходных наблюдательных данных. Метаданные должны быть максимально полными и соответствовать стандартам, принятым в мировых океанографических центрах. В работе А.А.Кораблева и др. (2005) приводятся требования к построению БОД, согласно международным стандартам хранения океанографической информации:

1. Обязательны проверка качества информации, контроль станций и сравнительный контроль данных.
2. Основная единица хранения - океанографическая *станция*. Это сеанс измерений, главными характеристиками которого являются место, дата и время.
3. Метаданные хранятся в двух связанных таблицах «Станции» и «Информация о станции»
4. Профили океанографических параметров хранятся в отдельных таблицах. Все таблицы однотипны и содержат по пять полей: абсолютный номер станции, последовательный номер горизонта, значение горизонта, значение параметра, флаг качества.

Процедура разработки информационной системы включает в себя:

1. определение состава метаданных и связей между таблицами БД;
2. создание сопровождающих баз данных;
3. разработку конвертеров данных из начальных источников в форматы хранения и загрузку данных отдельных БД;
4. обновление метаданных; контроль качества;
5. интерполяцию на стандартные горизонты;
6. типизацию одноградусных квадратов на основании месячных распределений градиентов океанографических параметров средствами объективного анализа;
7. определение зон повышенной изменчивости;

и другие этапы.

Разработчики БОД уделяют должное внимание качеству данных и метаданных. Так, в работе (А.Х. Халиулин, 2016) приводится следующий пошаговый алгоритм, реализованный при разработке БОД МГИ:

1. импорт данных и приведение их в заданный формат;
2. проверка метаданных:
 - проверка попадания станции в бассейн;
 - проверка скорости судна между станциями;
 - сравнение последнего горизонта и глубины места и пр.;
3. автоматическое проставление флагов качества.

Контроль качества данных может включать в себя:

1. оценку попадания измеренного значения в пределы, определяемые на основе климатических оценок;
2. проверку на «спайки» - резкие выбросы в вертикальных профилях проверяемых параметров морской среды;
3. проверку на градиенты на предмет превышения пороговых величины (В.Н. Еремеев, 2014).

3.3 Международный проект ARGO

В настоящее время наблюдения за состоянием Мирового океана приобрели массовый характер, когда число производимых в море измерений возросло многократно. Широко применяются зондирующие устройства, непрерывно или дискретно регистрирующие характеристики морской воды. Затем измеренные данные передаются в центры обработки и хранения информации при помощи искусственных спутников Земли. В конце 90-х годов был инициирован международный проект ARGO по мониторингу основных характеристик Мирового океана с помощью автоматических дрейфующих буйев. Каждый буй производит измерения примерно раз в 10 суток (см. Рис. 2), наибольшая глубина погружения (горизонт) составляет 2000 м. Каждый буй дрейфует в течение 10 суток на заданной глубине, затем опускается на горизонт 2000 м. С горизонта 2000 м буй всплывает на поверхность, измеряя давление, температуру, соленость. Измерение солёности основано на электропроводности (кондуктометрия). Также происходит сбор информации о течениях на двух горизонтах (заданной глубине дрейфа и поверхности) в течении работы цикла. Функционирование буя продолжается до тех пор, пока не закончится запас батареи (рабочий период составляет около 4 лет).

Все данные наблюдений с буйев поступают через приемные спутниковые станции (АРГОС) в два центра данных АРГО (США и Франция) и в национальные центры данных АРГО.

С 2007 года численность буйев достигла планируемого количества - более 3000 (см. Рис. 3) буйев (Магазов С.С., 2015), которые производят около 100 000 профилей измерений в год. Каждый профиль может содержать до 2000 горизонтов измерений. Данные предоставляются пользователям, как в реальном времени, так и после обработки, калибровки и проверки качества. Формат предоставления данных – NetCDF (Network Common Data Form) – машиннонезависимый (кроссплатформенный) двоичный формат самоописываемых расширяемых файлов, являющийся стандартом для обмена научными данными.

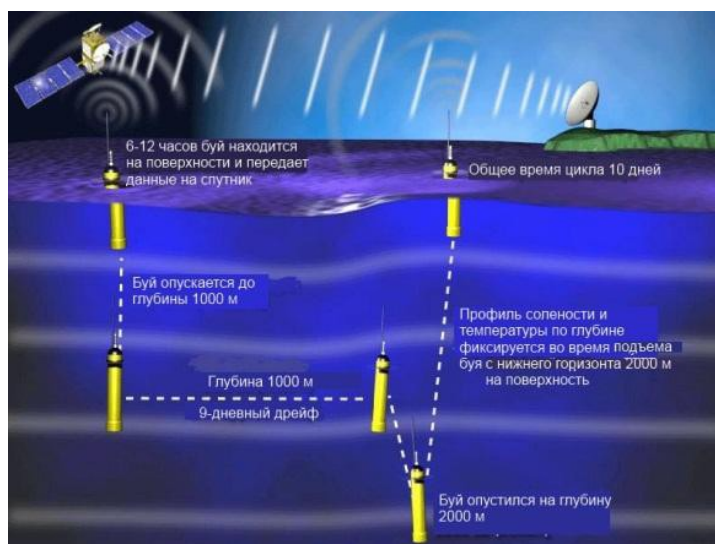


Рис. 2. Технологический цикл работы буйа.

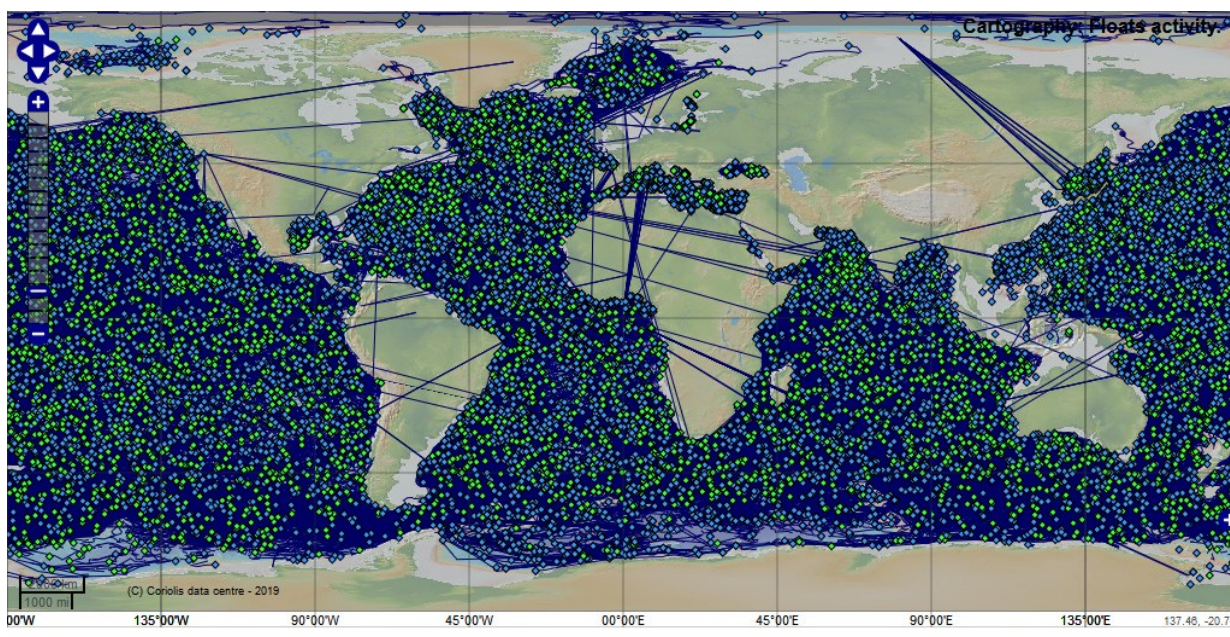


Рис. 3. Плотность покрытия буйами ARGO акватории Мирового океана по данным wwz.ifremer.fr на 4 февраля 2019 г.

В файлах, представляемых пользователям центрами данных ARGO, объединяются измерения, полученные разными буями и в разное время, объем файлов от 1.5 до 200 Мб. За декабрь 2018 г. суммарный объем предоставленных данных, прошедших обработку и проверку, составил 450 Мб. На ftp-сервере Institut français de recherche pour l'exploitation de la mer (ifremer.fr) для удобства пользователей файлы уже рассортированы по океанам^[2], а наиболее свежие данные за последние три месяца находятся в директории /argo/latest_data/. Файлы содержат информацию:

1. тип данных;
2. формат версии;
3. версия справочника;
4. нуль-пункт времени;
5. время создания, изменения файла;
6. данные о типе буя, его номере и серийном номере;
7. список параметров станции;
8. номер цикла буя;
9. направление (всплытие или погружение);
10. центр обработки данных;
11. идентификаторы данных в Центре обработки;
12. способ предоставления данных (после 24 часов после получения или после обработки);
13. данные о приборах, способе измерений;
14. момент измерений (UTC), флаг качества, ошибка момента;
15. широта и долгота места измерения, флаг качества координат;
16. система позиционирования;
17. измеренные давление, температура и соленость, флаги качества и ошибки;
18. калиброванные (adjusted) давление, температура и соленость, флаги качества и их ошибки;
19. данные о калибровке, истории обработки и др.

Содержится также информация о размерностях массивов данных, единицах измерения, длинные и короткие названия данных, формат данных, отсылки к справочным таблицам ARGO, значения, размещаемые в случае отсутствия реальных значений параметров и форматы. Типичный список переменных (собственно, массивов данных) и их атрибутов в файлах ARGO в формате NetCDF, приводится в Приложении 1. Таким образом, предоставляется исчерпывающая информация, необходимая для интерпретации и анализа наблюдений, а также контроля качества.

4. Информационно-вычислительная система по работе с данными ARGO

4.1 Требования к разрабатываемой системе

Общие требования к построению информационно-вычислительной системы океанографических данных были сформулированы согласно специфике самих данных и рабочих процессов организации-заказчика, оставляя место и возможности для расширения видов хранимых данных и связей между ними:

1. Метаданные должны быть максимально полными и соответствовать стандартам, принятым в мировых океанографических центрах;
2. Все данные должны храниться локально, в местной БД с возможностью быстрого доступа к любому срезу;
3. Массовая загрузка больших объемов измерений в формате netCDF за короткий срок для данных ARGO;
4. Система должна работать без доступа к сети Internet;
5. Пополнение БД должно выполняться администратором БД с физического носителя;
6. СУБД PostgreSQL;
7. Возможность расширения системы для данных других форматов из других источников;

Функциональные требования к информационно-вычислительной системе океанографических данных:

1. Проверка качества метаданных и данных;
2. Вычисление глубин (горизонтов) в метрах для представленных измерений;
3. Вычисление плотности морской воды и скорости звука на данной глубине;
4. Интерполяция на стандартные горизонты;
5. Вычисление недостающих параметров.

Функциональные требования к интерфейсу информационно-вычислительной системы океанографических данных:

1. Реализация доступа пользователей через веб-интерфейс.

2. Предоставление сводной информации о наличии накопленных данных (карта точек, статистика по годам и т.д.)
3. Выборка данных по параметрам: тип данных (в рамках данной работы только один – гидрофизические); источник данных (в рамках данной работы только один – ARGO); участок акватории, задаваемый координатами центра и полушириной; диапазон дат; месяцы, сезоны года.
4. Статистические расчеты по выбранному подмножеству данных.
5. Выгрузка параметров выборки, выбранных данных и результатов анализа в одном из пользовательских форматов.

4.2 Программные средства разработки ПО и хранения данных

Выбор средств разработки в основном был определен форматом распространения данных NetCDF. Данные в файле netCDF хранятся в форме массивов. Для доступа к данным, хранимым в формате NetCDF, разработаны программные библиотеки, например netcdf4-python^[3] – интерфейс к C-библиотеке NetCDF. Соответственно, выбран язык программирования Python версии 3 и Django^[4] в качестве фреймворка для веб-разработки.

Django - свободный фреймворк для веб-приложений на языке Python, использующий шаблон проектирования MVC. Разработчики отмечают, что Django заметно упростил ряд сложностей в разработке веб-приложений. Преимуществами Django является быстрота, безопасность, масштабируемость и пригодность для разработки научных вычислительных платформ^[5]. Django ORM это эффективный инструмент для взаимодействия с СУБД, избавляет от необходимости работы на низком или специфическом для конкретной СУБД уровне.

СУБД PostgreSQL была выбрана по требованию заказчика, ее технические возможности вполне достаточны для операций с БД, содержащих порядка млрд. записей.

Поскольку на данный момент сервер еще не закуплен организацией-заказчиком, проект реализован на личном ноутбуке, работающем под Windows 7 Home.

4.3 Этапы разработки ПО

Создание БД

Первым этапом является изучение первичных данных, отбор необходимых для размещения в БД и определение представлений и связей. Переменные, хранимые в файлах

ARGO и представленные в Приложении 1, можно разделить на относящиеся к станциям (общие для всех горизонтов одного профиля):

время: JULD, JULD_QC, JULD_LOCATION;

координаты: LATITUDE, LONGITUDE, POSITION_QC;

способ измерения: VERTICAL_SAMPLING_SCHEME, DIRECTION ;

способ представления данных: DATA_MODE и номер цикла: CYCLE_NUMBER;

к информации о станции (общие для всех станций выполненных конкретным буем):

номер буя: PLATFORM_NUMBER

тип буя: PLATFORM_TYPE

серийный номер: FLOAT_SERIAL_NO,

и к калиброванным измерениям, их флагам качества и ошибкам:

PRES_ADJUSTED, PRES_ADJUSTED_QC, PRES_ADJUSTED_ERROR

TEMP_ADJUSTED, TEMP_ADJUSTED_QC, TEMP_ADJUSTED_ERROR

PSAL_ADJUSTED, PSAL_ADJUSTED_QC, PSAL_ADJUSTED_ERROR.

Все остальные, содержащие некалиброванные измерения и данные об обработке, ее историю, информацию о калибровке, центре данных, даты создания и изменения файла и проч., не являются необходимыми для настоящего проекта и не включаются в БД, но могут быть получены из исходного файла, хранимого в архиве.

Основные представления были определены с учетом существующих требований к построению БД (Кораблев А., и др., 2005). На Рис. 4 представлена UML-диаграмма, показывающая содержание основных таблиц БД и связи.

Основные модели данных:

DRIFTERS – данные о буях, типы буев содержатся в ARGO Reference table 16;

SESSIONS – метаданными станций, связана с DRIFTERS через FK поле drifter_id, значения флагов качества JULD_QC и POSITION_QC содержатся в таблице ARGO Reference table 2, к данным ARGO добавлено поле для имени исходного NetCDF-файла;

MEASUREMENTS – профили, связаны с SESSIONS через FK поле session_id, флаги качества *_QC так же содержатся в таблице ARGO Reference table 2, к данным ARGO добавлены вычисляемые значения глубины, плотности, скорости звука и их ошибок;

WATER_DATA_INTERPOLATED – таблица для данных, интерполированных на стандартные горизонты, связана с SESSIONS через FK поле session_id, эта таблица будет необходима при развитии проекта для интеграции данных ARGO с другими источниками, например, судовыми измерениями;

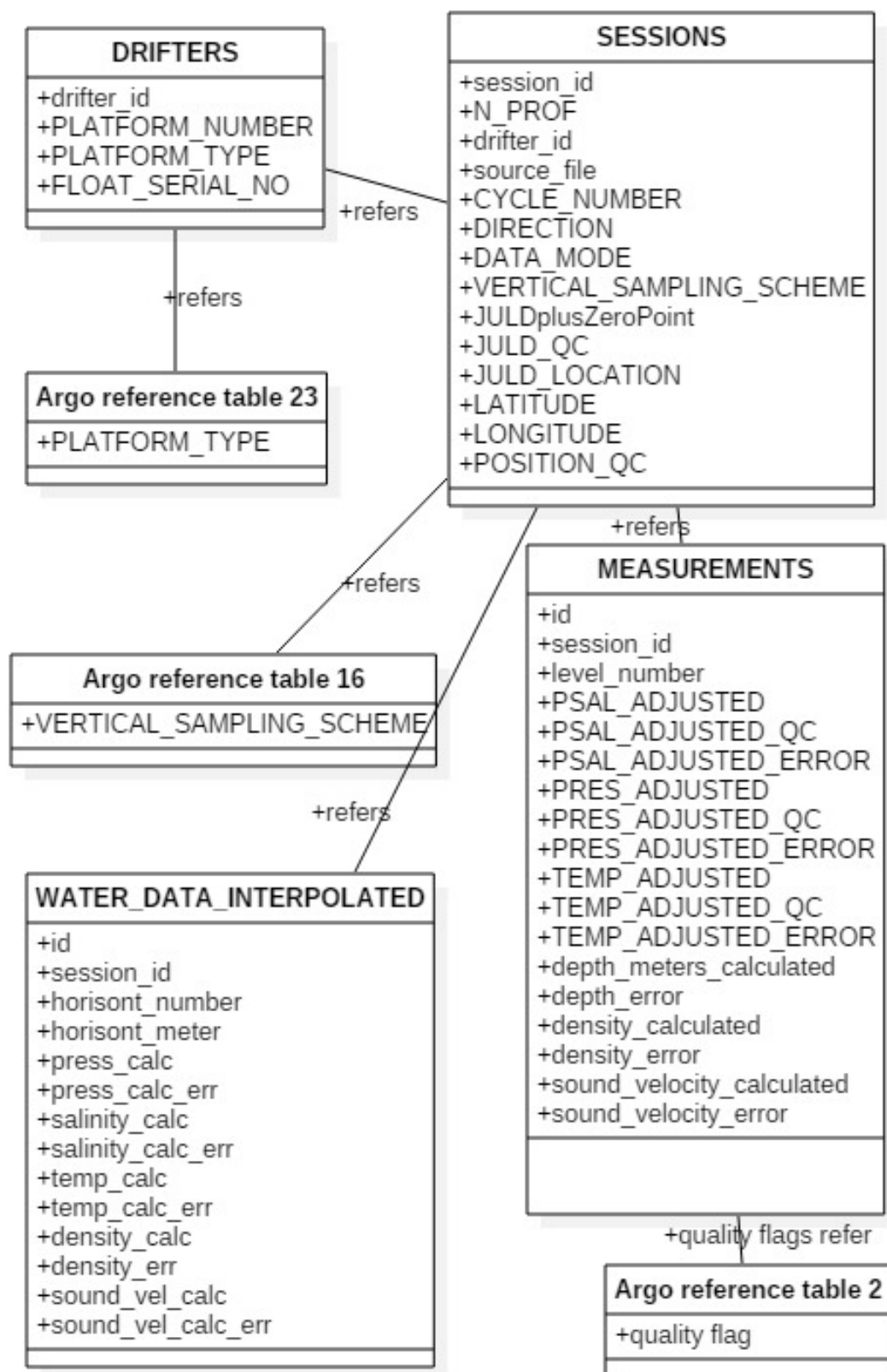


Рис. 4. Основные представления БД разрабатываемого проекта и связи.

Создание средства чтения NetCDF, проверки данных и загрузки в БД

Чтение данных производится вызовом конструктора Dataset, который инициализируется именем NetCDF файла. Созданный таким образом объект класса Dataset содержит в себе всю структуру данных. Полями экземпляра класса Dataset являются словари массивов данных (variables), их размерностей (dimensions). Обращение производится по ключу:

к массиву данных - `dataset.variables[key]`;

к размерностям массива данных - `dataset.variables[key].dimensions`;

к атрибуту - `dataset.variables[key].имя_атрибута`,

Метод `ncattrs()` дает доступ к множеству атрибутов конкретной переменной:

`dataset.variables[key].ncattrs()`. Основными размерностями являются:

N_PROF - количество станций (и, соответственно, измеренных профилей) и

N_LEVELS - количество уровней в профиле (горизонтов).

После создания объекта класса Dataset начинается заполнение БД: в цикле от 0 до N_PROF-1 происходит запись в таблицу DRIFTERS (или получение уже существующего объекта БД). Затем запись метаданных станции и связь с соответствующей записью в таблице DRIFTERS. Во внутреннем цикле от 0 до N_LEVELS-1 запись об измерениях на соответствующем уровне данной станции при условии, что значения параметров не пусты. Отсутствующие значения в файлах ARGO маркируются значением атрибута `_FillValue` и равны 99999 для измеряемых параметров (999999 для моментов времени). Кроме того, не игнорируются горизонты с флагом качества равным 4 (очень плохое качество) или 9 (отсутствующее значение) или со значениями, находящимися за пределами допустимых диапазонов (например, отрицательная соленость). Вычисляются согласно методикам, приведенным в п. 2.4, и записываются значения глубины, плотности и скорости звука. Вычисления производятся при помощи библиотеки NumPy^[7], предоставляющей общие математические и числовые операции в виде пре-скомпилированных, быстрых функций для манипуляции с массивами и матрицами.

Создание пользовательского веб-интерфейса

В Django реализуется схема «Модель (доступ к данным) - Шаблон (отображение) - Представление (уровень логики)».

Модели данных определяются в файле `models.py` путем создания классов-наследников от `models.Model`. В процессе миграции Django создает соответствующие моделям таблицы в БД. Django автоматически генерирует поле `id` для каждой таблицы. Так

же, как и при работе с любой базой данных, можно создать связи между таблицами, установив Foreign Key.

Шаблоны служат для отображения страниц, располагаются с специальной директории templates. В данном случае определен базовый шаблон, от которого остальные шаблоны наследуют элементы навигации и оформление. View.py служит для создания обработчиков, вызываемых соответствующими http-запросами, установленными в файле urls.py. Веб-формы описываются в forms.py классами-наследниками от forms.Form. Пользовательский интерфейс состоит из следующих частей:

1. Пункт меню «Выбрать» - поиск данных. Параметры поиска: источник данных (на данном этапе разработки – только данные проекта ARGO), участок акватории, период времени, сезоны и горизонты. Предусмотрена визуализация полученных выборок и экспорт в CSV (Рис. 4).
2. Пункт меню «Описание данных» - страница с общей статистикой по БД, описанием структуры таблиц, справочными таблицами ARGO.
3. Пункт меню «Методы» открывает страницу с ссылками на источники формул и алгоритмов расчетов, контрольными значениями и калькуляторами для проверки.
4. Пункт меню «Буи» - все представленные в БД буи, выборка по номеру буя.
5. Пункт меню «Станции» - список станций, получение метаинформации и профиля.

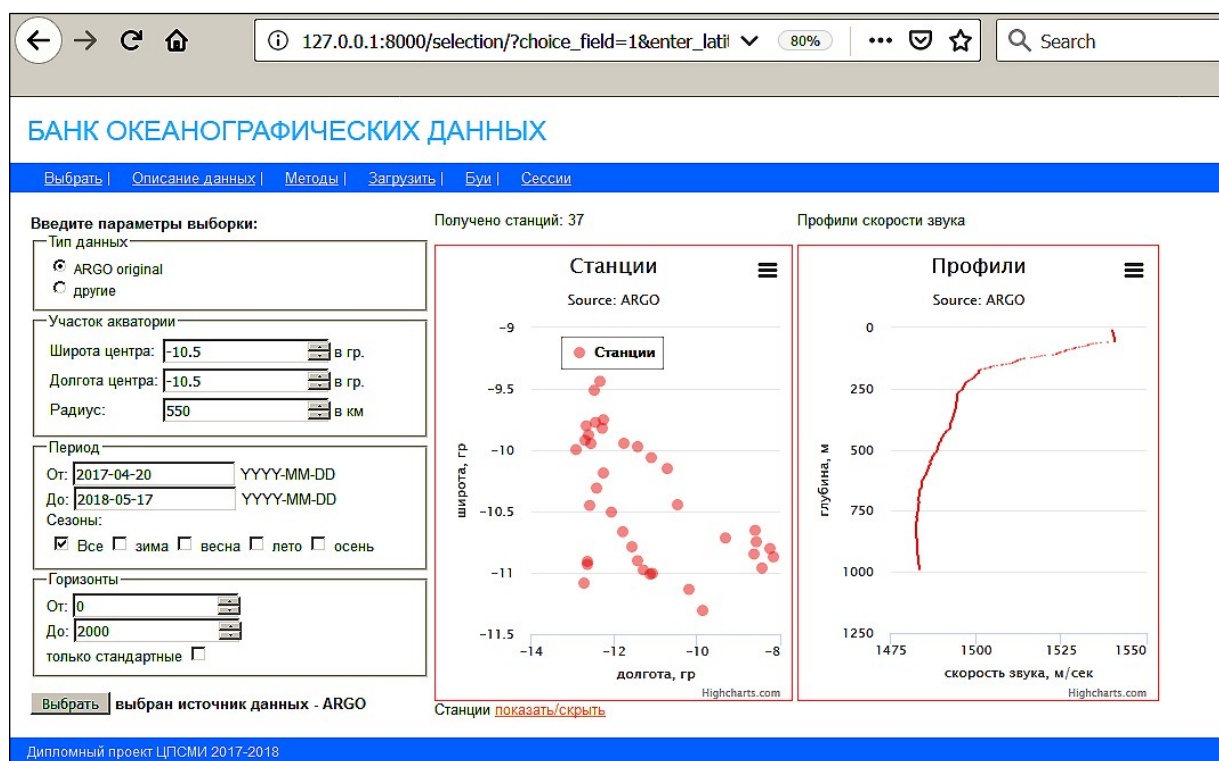


Рис. 4. Внешний вид формы выборки данных и визуализация результатов.

4.4 Методики расчетов

В данных, поставляемых ARGO, отсутствуют значения горизонтов в метрах и для дальнейших преобразований их необходимо вычислить. Глубина погружения зависит от гидростатического давления и широты местоположения системы измерения (П.А.Калашников, 1985):

$$H = (99.404 + 4.983 \cdot 10^{-4} \varphi - 2.06 \cdot 10^{-4} \varphi^2 + 1.492 \cdot 10^{-6} \varphi^3) \cdot p - 2.204 \cdot 10^{-2} \cdot p^2 \quad (1)$$

где H - глубина в метрах; p - гидростатическое давление, МПа (100 децибар); φ - географическая широта в угловых градусах (модуль широты).

Плотность морской воды может быть вычислена через измеренные значения температуры, давления и солёности по международному уравнению состояния морской воды УС-80, принятому за стандарт вычислений (ГСССД 76-84, 1984). Это уравнение разработано Объединенной группой экспертов по океанографическим таблицам и стандартам ЮНЕСКО, Международным советом по исследованию моря (МСИМ), Национальным комитетом по океанографическим исследованиям (СКОР) и Международной ассоциацией физических наук об океане (МАФНО). Плотность ρ морской воды как функция практической солёности S , температуры t и гидростатического давления p определяется уравнением

$$\rho(S, t, p) = \frac{\rho(S, t, 0)}{1 - \frac{p}{K(S, t, p)}}$$

где

$\rho(S, t, p=0)$ – плотность морской воды при стандартной атмосфере (101325 Па);

$K(S, t, p)$ - средний модуль упругости;

Плотность $\rho(S, t, p=0)$ определяется выражением:

$$\rho(S, t, 0) = \rho_w + \sum_{i=0}^4 (S \cdot b_i + S^{3/2} c_i) t^i + d_0 S^2$$

где

$$b = \begin{bmatrix} 8.24493 \cdot 10^{-1} \\ -4.0899 \cdot 10^{-3} \\ 7.6438 \cdot 10^{-5} \\ -8.2467 \cdot 10^{-7} \\ 5.3875 \cdot 10^{-9} \end{bmatrix}; \quad c = \begin{bmatrix} -5.72466 \cdot 10^{-3} \\ 1.0227 \cdot 10^{-4} \\ -1.6546 \cdot 10^{-6} \\ 0.0 \\ 0.0 \end{bmatrix}; \quad d_0 = 4.8314 \cdot 10^{-4};$$

ρ_w - плотность стандартной среднеокеанической чистой воды (относящиеся к ней величины обозначены индексом W):

$$\rho_w = \sum_{i=0}^5 a_i t^i, \text{ где } a = \begin{bmatrix} 999.842594 \\ 6.793952 \cdot 10^{-2} \\ -9.095290 \cdot 10^{-3} \\ 1.001685 \cdot 10^{-4} \\ -1.120083 \cdot 10^{-6} \\ 6.536332 \cdot 10^{-9} \end{bmatrix}$$

средний модуль упругости вычисляется по формуле:

$$K(S, t, p) = \sum_0^2 n_i p^i$$

$$n = \begin{bmatrix} K(S, t, 0) \\ A \\ B \end{bmatrix}$$

где

$$K(S, t, 0) = \sum_{i=0}^4 (e_i + S f_i + g_i S^{3/2}) t^i, A = \sum_0^3 (h_i + S \cdot I_i) t^i + j_0 S^{3/2}, B = \sum_0^2 (k_i + S \cdot m_i) t^i$$

Коэффициенты:

$$e = \begin{bmatrix} 19652.21 \\ 148.4206 \\ -2.327105, \\ 1.360477 \cdot 10^{-2} \\ -5.155288 \cdot 10^{-5} \end{bmatrix}, f = \begin{bmatrix} 54.6746 \\ -0.603459 \\ 1.09987 \cdot 10^{-2} \\ -6.1670 \cdot 10^{-5} \\ 0 \end{bmatrix}, g = \begin{bmatrix} 7.944 \cdot 10^{-2} \\ 1.6483 \cdot 10^{-2} \\ -5.3009 \cdot 10^{-4} \\ 0 \\ 0 \end{bmatrix}$$

$$h = \begin{bmatrix} 3.239908 \\ 1.43713 \cdot 10^{-3} \\ 1.16092 \cdot 10^{-4} \\ -5.77905 \cdot 10^{-7} \end{bmatrix} I = \begin{bmatrix} 2.2838 \cdot 10^{-3}, \\ -1.0981 \cdot 10^{-5} \\ -1.6078 \cdot 10^{-6} \\ 0 \end{bmatrix} j_0 = 1.91075 \cdot 10^{-4}$$

$$k = \begin{bmatrix} 8.50935 \cdot 10^{-5} \\ -6.12293 \cdot 10^{-6} \\ 5.2787 \cdot 10^{-8} \end{bmatrix} m = \begin{bmatrix} -9.9348 \cdot 10^{-7} \\ 2.0816 \cdot 10^{-8} \\ 9.1697 \cdot 10^{-10} \end{bmatrix}$$

Международное уравнение состояния УС-80 действительно в диапазонах солёности от 0 до 42; температурах от -2 до 40°C; значениях давления от 0 до 1000 бар. Средне-квадратическое отклонение плотности от экспериментальных данных в этих диапазонах равно 0,009 кг/м³. Для диапазона солёности 30-40 среднеквадратическое отклонение плотности равно 0,005 кг/м. Контрольные значения приведены в Таблице 1.

Таблица 1. Опорные значения контрольных величин (ГСССД 76-84, 1984)

S, psu	t, °C	p, бар	ρ, кг/м ³
0	5	0	999,967
		1000	1044,128
	25	0	997,048
		1000	1037,902
35	5	0	1027,676
		1000	1069,489
	25	0	1023,343
		1000	1062,538

Для вычисления скорости звука в воде $V(S, t, p)$ существует несколько алгоритмов, один из них – “UNESCO equation” (Fofonoff N.P., 1983). Для вычисления звука используется формула (Chen and Millero, 1977):

$$V(S, t, p) = C_w(t, p) + A(t, p) \cdot S + B(t, p) \cdot S^{3/2} + D(t, p) \cdot S^2,$$

где

$$C_w(t, p) = \sum_{i=0}^3 \sum_{j=0}^5 c_{i,j} t^j p^i, \quad A(t, p) = \sum_{i=0}^3 \sum_{j=0}^5 a_{i,j} t^j p^i,$$

$$B(t, p) = \sum_{i=0}^1 \sum_{j=0}^1 b_{i,j} t^j p^i, \quad D(p) = \sum_{i=0}^1 d_i p^i$$

Коэффициенты в полиномах равны следующим константам:

$$c = \begin{matrix} 1402.388 & 5.03711 & -5.80852 \cdot 10^{-2} & 3.3420 \cdot 10^{-4} & -1.47800 \cdot 10^{-6} & 3.1464 \cdot 10^{-9} \\ 0.153563 & 6.8982 \cdot 10^{-4} & -8.1788 \cdot 10^{-6} & 1.3621 \cdot 10^{-7} & -6.1185 \cdot 10^{-10} & 0 \\ 3.1260 \cdot 10^{-5} & -1.7107 \cdot 10^{-6} & 2.5974 \cdot 10^{-8} & -2.5335 \cdot 10^{-10} & 1.0405 \cdot 10^{-12} & 0 \\ -9.7729 \cdot 10^{-9} & 3.8504 \cdot 10^{-10} & -2.3643 \cdot 10^{-12} & 0 & 0 & 0 \end{matrix}$$

$$a = \begin{matrix} 1.389 & -1.262 \cdot 10^{-2} & 7.164 \cdot 10^{-5} & 2.006 \cdot 10^{-6} & -3.21 \cdot 10^{-8} \\ 9.4742 \cdot 10^{-5} & -1.2580 \cdot 10^{-5} & -6.4885 \cdot 10^{-8} & 1.0507 \cdot 10^{-8} & -2.0122 \cdot 10^{-10} \\ -3.9064 \cdot 10^{-7} & 9.1041 \cdot 10^{-9} & -1.6002 \cdot 10^{-10} & 7.988 \cdot 10^{-12} & 0 \\ 1.100 \cdot 10^{-10} & 6.649 \cdot 10^{-12} & -3.389 \cdot 10^{-13} & 0 & 0 \end{matrix}$$

$$b = \begin{matrix} -1.922 \cdot 10^{-2} & 7.3637 \cdot 10^{-5} \\ -4.42 \cdot 10^{-5} & 1.7945 \cdot 10^{-7} \end{matrix}, \quad d = \begin{matrix} 1.727 \cdot 10^{-3} \\ -7.9836 \cdot 10^{-6} \end{matrix}$$

Формула справедлива для следующих диапазонов параметров:

$$0^{\circ}\text{C} < T < 40^{\circ}\text{C}, \quad 0\text{‰} < S < 40\text{‰}, \quad 0 < P < 10000 \text{ децибар}:$$

Среднеквадратическое отклонение – 0.19 м/сек.

Контрольное значение - 1731.995 м/сек для S=40, t =40°C, P = 10000 децибар.

4.5 Описание программного комплекса

Назначение ПО

Данная информационно-вычислительная система предназначена для ведения специализированного банка данных в области океанологии и смежных областях

Возможности ПО

Информационно-вычислительная система предоставляет возможности:

1. загрузить в БД ряды данных;
2. выполнить поиск данных по нескольким условиям и выборку подмножества (среза);
3. просмотреть метаданные выборки, например характеристики измерительной аппаратуры или типы буёв;
4. получить результаты статистических расчетов по срезу данных;
5. выгрузить данные среза в любом из нескольких predetermined форматов для их дальнейшей обработки в других программных системах.

Ограничения ПО

Операторам для работы с ИС требуется веб-браузер с поддержкой JS и протокола HTTP версии 4 и старше, то есть практически любой, какой может найтись на обычном офисном рабочем месте.

Требования к аппаратным средствам:

1. Серверный компьютер начального уровня: CPU с количеством вычислительных ядер 4 или больше; выделенное для нужд ИС дисковое пространство от 2Тб и больше; оперативной памяти от 30Gb и больше; подключение к ЛВС на скорости 100Mbps и больше.
2. Система бесперебойного электропитания
3. ЛВС для связи между компьютерами операторов (пользователей) и сервером

Требования к программным средствам:

1. СУБД PostgreSQL
2. Веб-сервер (Apache, nginx)
3. Среда выполнения программ Python версии 3
4. Framework Django для Python
5. Операционная система для запуска всего вышеназванного

Требования организационные:

1. Доступ к системе — серверу и рабочим местам операторов — должен быть ограничен административными средствами.
2. Операторы для допуска к работе обязаны пройти обучение под надзором разработчика.
3. Сервер и рабочие места операторов должны быть отключены от сети Интернет.
4. Технический персонал обязан своевременно выполнять регламентные работы на оборудовании и ПО, делать резервное копирование базы данных.

Требования к ИС

Информационно-вычислительная система должна обеспечивать:

1. Загрузку данных:
 - приём от оператора файлов в формате NetCDF4, распространяемых ЦОД проекта ARGO, содержащих данные замеров температуры, солёности и давления воды на различных глубинах, плюс метаданные, описывающие место и время выполнения замеров, характеристики измерительных приборов, буёв и т.п.;

- распаковку и декодирование принятых от оператора файлов NetCDF4, проверка наличия в них целевых данных;
- анализ качества данных, проверка флагов качества (устанавливаются ЦОД ARGO), проверка выхода за диапазоны допустимых значений и градиентов значений;
- заполнение лакун, интерполяция пропущенных или некачественных значений замеров;
- вычисление вторичных (производных) характеристик: плотность, глубина, скорость звука;
- интерполяцию всех значений на стандартные горизонты (глубины)
- занесение в БД сведений проверенных, исправленных и вычисленных значений и метаданных;
- обновление статистической информации о БД в целом;

2. Поиск данных

- предоставление оператору формы запроса на поиск данных с указанием на область акватории, диапазон дат, сезонность, диапазон глубин
- выполнение по указаниям заполненной формы выборку данных в БД
- вывод краткого статистического отчета по выборке, списка вошедших в выборку буйков и их характеристики
- предоставление оператору рядов данных по любому из этих буйков

3. Выгрузку данных

- выбор формата выгрузки
- формирование файла с данными в указанном формате
- Предлагать оператору получить (принять) файл с данными

Входные данные:

Файлы, предоставляемые Центрами обработки данных ARGO, в формате NetCDF.

Выходные данные:

Ряды вычисленных гидрофизических характеристик.

5. Заключение

В процессе проектирования и создания информационно-вычислительной системы по работе с океанографическими измерениями были изучены возможности языка программирования Python и фреймворка для разработки веб-интерфейса Django, получен опыт работы с СУБД PostgreSQL, Django ORM и SQL-запросами, разработана информационно-вычислительная система, оперирующая океанографическими данными международного проекта ARGO. Данный программный комплекс позволяет оценивать гидроакустическую обстановку районов Мирового океана. Эта система — первый этап формирования банка данных, в перспективе – расширение типов и источников данных, совершенствование методов математической обработки.

Автор выражает признательность международному проекту ARGO, благодаря которому были получены океанографические данные:

“These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (<http://www.argo.ucsd.edu>, <http://argo.jcom-mops.org>). The Argo Program is part of the Global Ocean Observing System. Argo (2000). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. <http://doi.org/10.17882/42182>”

Автор благодарит компанию JetBrains за любезное разрешение на использование среды разработки PyCharm.

В работе были использованы средства визуализации от Highcharts^[10].

Литература

Балясников С. Б., Матрюков С. И. Состояние и проблемы развития банка океанографических данных Министерства обороны РФ // Навигация и гидрография, ГНИНГИ, 1998б стр. 121

С.М. Гордеева, О.И. Шевчук Руководство по статистической обработке глобальных архивов информации (на примере альтиметрических данных) // СПб.: РГГМУ, 2013. – 44 с., ISBN 978-5-86813-341-1

ГОСТ 18451-73-ГОСТ 18458-73, Океанология термины и определения, М., 1973

ГСССД 76-84 Таблицы стандартных справочных данных. Морская вода. Плотность в диапазонах температур -2... 40 С, давлений 0... 1000 БАР и соленостей 0... 42, составитель - П.А.Калашников. утв. Гос. Ком. СССР по стандартам 12.12.1984 г.

В.Н. Еремеев, А.Х. Халиулин, А.В. Ингеро́в, Е.В. Жук, Е.А. Годин, Т.В. Пластун Современное состояние банка океанографических данных МГИ НАН Украины: программно-математическое обеспечение // Мор. гидрофиз. журн., 2014, № 2

П.А.Калашников. Первичная обработка гидрологической информации. // Л. Гидрометиздат, 1985, с.47, 80, 113.

Картография. Выпуск 3. Картографирование океанов // М.: Издательство “Прогресс”, 1988г, с.264

А.А.Кораблев, А.В.Пнюшков А.В.Смирнов Технология создания баз океанографических данных на примере Северо-Европейского бассейна Арктики // Научные записки РГГМУ. -2005.- т. 1.- с.89-108.

Магазов С.С. Информационные ресурсы глобальных систем мониторинга океана для задач гидроакустики // 2015, Евразийское научное объединение, Том 1, № 10 (10), с. 29-42

Матвиенко, В.Н., Тарасюк Ю. Ф. Дальность действия гидроакустических средств // Ленинград, “Судостроение”, 1981

Халиулин А.Х., Годин Е.А., Ингеро́в А.В., Жук Е.В., Галковская Л.К., Исаева Е.А. Банк океанографических данных морского гидрофизического института: информационные ресурсы для поддержки исследований прибрежной зоны Черного моря // Экологическая безопасность прибрежной и шельфовой зон моря, 2016, № 1 (12), С. 89-95

Chen, C.-T.; Millero, F. J. 1977. Speed of sound in seawater at high pressures. // J. Acoust. Soc. of Amer., 61 (5), 1129-1135.

Fofonoff N.P. Algorithms for computation of fundamental properties of seawater / N.P. Fofonoff, R.C. Millard Jr. – UNESCO , 1983. – 44. – 54 p.

Oceanographic Atlas of the North Atlantic Ocean. Section VI. Sound Velocity. // US Naval Oceanographic Office, Washington, 1967.

Internet-ресурсы

1. Network Common Data Form (NetCDF), <https://www.unidata.ucar.edu/software/netcdf/>
2. ftp://ftp.ifremer.fr/ifremer/argo/geo/atlantic_ocean/
3. NetCDF4 module, <http://unidata.github.io/netcdf4-python/>
4. The Web framework for perfectionists with deadlines | Django, <https://www.djangoproject.com/>
5. Monty Python «Плюсы и минусы Django», 2017, <https://python-scripts.com/django-obzor>
6. Оперативная база данных программы АРГО, ФГБУ "ДВНИГМИ", <http://ferhri.org/2017-07-26-02-10-13/2017-07-26-03-34-33/2017-07-26-03-47-04.html>
7. NumPy, <http://www.numpy.org/>
8. Технология работы с данными для построения карт гидрологических элементов, Арктический и антарктический НИИ, http://www.aari.ru/resources/a0013_17/barents/atlas_barents_sea/_Atlas_Barents_Sea_months/Text/Data_work.htm
9. Highcharts, <https://www.highcharts.com>
10. Международный проект ARGO, <http://www.argo.ucsd.edu/>