

Аналитический комплекс ЕСИМО

Е.Д.Вязилов, Д.А.Мельников, Н.В.Чуняев

Всероссийский научно-исследовательский институт гидрометеорологической информации – Мировой центр данных
vjaz@meteo.ru, melnikov@meteo.ru

Аннотация

Представлены сведения об аналитическом комплексе Единой государственной системы информации об обстановке в Мировом океане (ЕСИМО). Дан краткий обзор аналитических систем. Показана работа с аналитическим комплексом. Сказано о перспективах развития аналитического комплекса ЕСИМО.

1. Введение

Современные проблемы информационных технологий заключаются не в недостатке информации, а в трудностях ее обработки и анализа средствами систем управления базами данных. Созданные приложения направлены главным образом на визуализацию исходных данных. Поэтому нужно использовать Business Intelligence (BI) системы или информационно-аналитические системы (ИАС). Фундаментом любой ИАС являются системы измерений, сбора и хранения данных. Основная задача этих систем, служащих прослойкой между пользователями и базами данных (БД), помочь провести обработку и анализ информации для принятия решений.

Состояние, в котором сегодня находятся аналитические данные, можно назвать «фрагментарным». Создаются разнородные приложения, которые используют данные, хранящиеся в отдельных БД. При этом возрастает число структур данных, применяемых в различных приложениях, что усложняет стандартизацию сервисов, форм визуализации. Преимуществами аналитических систем являются следующее.

Большие объемы данных становятся доступными в режиме он-лайн. В результате экономится время на поиск и обработку информации, можно в режиме он-лайн увидеть результаты обновления данных. Появляется возможность использования различных методов обработки (выявление критических значений, оценка риска, получение климатических значений, др.).

Возможность выделять ценную информацию из огромного объема накопленных данных является важнейшей задачей современных аналитических систем. Чтобы принимаемые решения были эффективными, важно наличие возможности обеспечить

лиц, принимающих решения (ЛПР), в любое время, в нужном формате, минимально необходимом объеме информации и правильно представленном виде. Наличие у ЛПР слишком большого количества показателей и большого объема данных ничуть не лучше, чем отсутствие необходимых данных. Безусловно, важных данных много, но для достижения наилучших результатов необходимо существенно ограничить объем информации, представляемой руководителям для поддержки решений.

ЛПР не может тратить больше 5% рабочего времени на анализ и подготовку решений с использованием информации о состоянии природной среды. Поэтому они не могут тратить время на анализ большого числа показателей или слишком детальных данных.

ЛПР хотят, чтобы данные уже были извлечены из самых разных источников и по ним найдены взаимосвязи, о которых раньше не знали. Стандартные отчеты, широко используемые в различных автоматизированных системах, не обеспечивают интерактивности. Способность немедленно получать ответы на основе самой разной информации улучшает принятие решений.

Данные нужно собрать, сделать их доступными для ЛПР, которые в них нуждается, и представить в виде карт, диаграмм, графиков, таблиц, текста, «приборов». Анализ должны подвергаться как исходные и диагностические данные, так и прогнозная и климатическая информация.

Для анализа данных используются различные событийные (мониторинг), интерактивные (Data-Mining, OLAP) и регламентные инструменты (отчеты, электронные таблицы, БД, средства анализа и прогноза) [4]. Можно привести несколько примеров аналитических систем.

Программа Catalyst [1] занимается слиянием и анализом разнородных данных, позволяет оперативно делиться содержимым разрозненных репозиторий, устранит неоднозначности в БД, даст возможность аналитикам изучать структуру взаимосвязей с учетом геопространственных и временных параметров.

QlikView (компания QlikTech) – BI-система нового типа, предоставляющая пользователям инструменты для самостоятельного исследования и визуального представления данных. Система обеспечивает коллективный анализ данных, позволяет сотрудникам совместно исследовать данные в реальном времени, чтобы принимать обоснованные решения. Ключевые возможности платформы — сред-

ства ассоциативного поиска дополнены инструментами сравнительного анализа, позволяющие наглядно сопоставлять выбранные пользователем показатели. Упрощен процесс разработки приложений, улучшены возможности управления системой и обеспечения информационной безопасности. Мобильный клиент системы поддерживает полноценный аналитический интерфейс, в т.ч. интерактивный анализ, средства визуализации, ассоциативный поиск, а также возможности обмениваться данными с другими приложениями.

Министерство регионального развития Российской Федерации открыло доступ к информационной системе «Оценка эффективности деятельности органов исполнительной власти субъектов Российской Федерации» («Рабочий стол руководителя», <http://ef.minregion.gas-u.ru/minreg2/>). Система предназначена для осуществления сбора, хранения и аналитической обработки показателей. В открытом доступе находится информация по уровню и динамике эффективности деятельности органов исполнительной власти субъектов Российской Федерации в целом и с разбивкой по сферам; для каждого региона, а также рангу региона по этому показателю относительно других субъектов; объему и доле неэффективных расходов и объемам выделяемых грантов по субъектам. Эта ИАС предоставляет возможности ознакомления с информацией как по показателям, где предусмотрены различные режимы просмотра данных: матрица, таблица, график, карта, информация, так и по регионам, где доступны режимы таблицы и справки. Содержащаяся в системе информация об эффективности деятельности регионов представлена по достигнутому уровню и по динамике эффективности, массив информации охватывает временной период с 2007 г. Любая интересующая информация доступна к выгрузке в форматах Microsoft Word и Microsoft Excel. Информация предоставляет возможности для проведения сравнительного межрегионального анализа по используемым для оценки эффективности деятельности субъектов Российской Федерации показателям и по общему уровню и динамике эффективности, а также кросстемпорального анализа трансформации положения субъекта по указанным параметрам.

Компания «НЕОЛАНТ» по заказу Федеральной службы охраны России создала систему «Кризис-СЭФ» – «Мониторинг и анализ кризисных ситуаций в социально-экономической, политической, экологической и техногенной сферах жизнедеятельности страны». Система разработана для федеральных и региональных органов государственной власти и правоохранительных органов. «Кризис-СЭФ» обеспечивает информационно-аналитическую поддержку принятия управленческих решений по организации и координации деятельности федеральных и региональных органов исполнительной власти РФ. Основное предназначение этой системы в предотвращении кризисов в политической, социально-экономической, экологической и техногенной сферах. Достигается это за счет непрерывного монито-

ринга и анализа огромного массива информации и представления ее в доступном и привычном для человеческого восприятия визуальном виде. Перед руководством в режиме реального времени представят ясная и полная картина происходящего с рекомендациями на основе анализа всех влияющих факторов.

Крупные фирмы – разработчики программных продуктов включают в свой портфель аналитические инструменты. Такие инструменты есть у компаний Oracle, Microsoft [6], IBM. Эти инструменты жестко привязаны к конкретным системам управления базами данных и стоят очень дорого.

Недостатками создания и внедрения аналитических инструментов являются [2,3]:

- аналитическая система выдает только то, что она обрабатывает, а гипотезу о принятии решения она не выдвигает и как доказать эту гипотезу с помощью анализируемых данных она тоже не знает, т.е. ЛПР все-таки нужен посредник - аналитик для работы с такими системами, особенно на первом этапе использования аналитических инструментов; аналитики будут заранее готовить аналитические представления для конкретных задач и очень быстро настраивать инструмент на новые задачи;

- результаты, полученные из системы, не всегда согласуются с результатами, полученными другими способами, это связано, как правило, с тем, что используются различные источники данных и методы их обработки;

- большинство аналитических систем слишком сложны для ЛПР, которые должны на практике их применять;

- для решения некоторых задач отсутствуют данные, которые нужно загрузить в БД системы, лучше это делать универсальным способом, например, использовать средства интеграции распределенных и неоднородных данных.

Пользователи Единой государственной системы информации об обстановке в Мировом океане (ЕСИМО) широко используют информацию из этой системы и при этом чаще всего вручную анализируют огромные объемы данных. Для автоматизации этого процесса в ЕСИМО разработан аналитический комплекс (АК, <http://ak.esimo.ru>).

2. Общие сведения об аналитическом комплексе

Для аналитиков создано интерактивное средство с простым интерфейсом и инструменты для работы с интегрированными в ЕСИМО данными. Аналитики работают, не используя сложные SQL-запросы, добавляя при необходимости новые информационные ресурсы и изменяя состав аналитических представлений (АП) без помощи ИТ-специалистов.

В составе АК включен механизм использования языка T-SQL, который позволяет делать выборки из БД. Для включения карты в аналитическое представление в АК имеются средства применения картографических сервисов (Web Map Service) путем

указания URL–адреса этого сервиса. При этом сервис может находиться как в каталоге ЕСИМО, так и в каталоге открытых слоев. Если необходимого сервиса нет, то его надо создать, используя имеющиеся в ЕСИМО средства. Для этого результаты агрегации в АК оформляются как информационный ресурс ЕСИМО. С помощью ГИС-сервера строится новый слой, который включается в каталог WMS-сервисов.

В АК используется единый контролируемый доступ к информации всех пользователей, построенный на основе системы JOSSO на портале ЕСИМО. С АК работают администратор, аналитик и руководитель – пользователь. Актуальность данных обеспечивается организацией ежедневного аудита состояния информационных ресурсов (ИР) и отражения этой информации в виде жизненного цикла каждого ИР. Жизненный цикл ИР включает создание или изменение ИР на Поставщике данных, передача данных на Сервер интеграции для последующего распределения по нескольким базам интегрированных данных (БИД) или ftp-серверам, включение ИР в БИД, использование данных в ГИС, АК и других приложениях.

Использование интегрированных ресурсов ЕСИМО в АК позволяет улучшить согласованность данных и уменьшить расходы на их поддержку. Функция оповещения конечных пользователей об обновлении данных в АК повышает производительность труда ЛППР.

3. Работа с аналитическим комплексом

Для развития аналитики в ЕСИМО выделено более 20 прикладных задач (ПЗ), каждая ПЗ может иметь несколько аналитических представлений. Каждое представление включает от трех (таблица, график или карта, текст) до десяти элементов представлений.

Организация работы по созданию ПЗ определяется заявкой и технической спецификацией на ПЗ [5]. Основными задачами разработки ПЗ являются:

- определение системы показателей, отслеживающих те или иные процессы морской деятельности и организация их мониторинга;
- выделение ИР, их загрузка в АК;
- разработка АП;
- регулярное использование АП;
- внедрение показателей в процессы принятия решений.

Основными функциями для получения аналитической информации являются поиск данных, агрегация данных, очистка данных, создание демонстрационного варианта АП, встраивание полученных представлений в портал, обновление данных в АП.

Агрегация данных. Для начала следует более внимательно присмотреться к данным и определить, какие параметры среды нужно использовать. Допустим, необходимо выявить тенденции изменения на временном масштабе один год. А исходные данные представлены в виде срочных или среднесуточных значений. Здесь нужно предварительно провести

агрегацию данных до среднемесячных значений. Можно поэкспериментировать с различными способами агрегации данных. Например, агрегирование по квадратам поможет создать временные ряды из случайных наблюдений в океане.

Очистка данных. После того как данные собраны и агрегированы, необходимо провести очистку данных для подготовки их к визуализации. Еще на этапе выявления данных можно обнаружить, что в исходных данных бывают пропуски в наблюдениях, поэтому необходимо обеспечить полноту значений полей и при необходимости провести интерполяцию данных во времени. Поскольку для обработки временных рядов необходима 100% полнота данных, разрабатывается план их вычисления. Нужен инструмент, позволяющий восстановить отсутствующие значения показателей путем интерполяции. Если некоторые временные ряды являются короткими (менее пяти значений), то с помощью средств очистки данных АК эти временные ряды отсеиваются.

Создание демонстрационного варианта АП. После уточнения алгоритмов и экспериментов с различными временными рядами появляется демонстрационный вариант АП.

Встраивание полученных представлений в портал. Теперь, когда имеется работоспособные АП, их можно интегрировать в портал ЕСИМО или АРМ Руководителя.

Обновление данных. Данные обновляются в ИР ЕСИМО в соответствии с регламентом. В АК важно обеспечить мониторинг обновления данных.

Схема организации аналитической обработки включает:

- поставщиков данных (БД, структурированные и неструктурированные файлы), которые поддерживают ИР ЕСИМО;
- единое хранилище данных – БИД, для пополнения БИД используется Сервер интеграции, который в режиме реального времени по событию (изменению или замещению данных на Поставщике данных) реплицирует данные от Поставщиков данных в БИД;
- ГИС-сервер для подготовки картографических слоев на основе результатов обработки данных;
- средства АК.

Подготовка данных для АК включает поиск необходимых ИР в СРБД ЕСИМО или, если необходимых ресурсов нет, то подготовку и загрузку ИР в СРБД.

В процессе анализа аналитик может продлить временной ряд вперед и назад, подключить другие показатели, сравнить показатели с индикаторными значениями. С помощью АК можно пройти все этапы от обнаружения данных до создания новых представлений.

Если в традиционных отчетах таблицы, графики, карты выводятся отдельно и их можно только смотреть, то АК позволяет все эти визуальные компоненты объединить, сделать их интерактивными. Пользователь может сам фильтровать данные, сворачивать представление данных и углубляться в

них, менять масштаб карты и фильтровать в ней данные, включать и выключать слои.

4. Перспективы развития АК

Основными направлениями развития АК являются разработка конструктора схем и создание динамических АП, таких как, например, представленных по адресам <http://news.mail.ru/society/7970463/?frommail=1>, http://www.footprintnetwork.org/en/index.php/GFN/page/2010_living_planet_report/, <http://eco.ria.ru/ecoinfogr/20091125/195466388.html>.

За основу конструктора схем можно принять, используемую в гидрометеорологии схему отражения синоптического состояния погоды. Схема может иметь до 18 показателей, рис.1. Каждый показатель должен иметь свой прибор (термометр, барометр-анероид, флюгер, гигрометр, датчик скорости, водомерная рейка, рейка для измерения высоты волн). Аналитик должен иметь возможность настроить соответствующие «приборы» и критические значения параметров, а после этого АП автоматически должно отражать состояние погодных условий в виде светофора на приборе (зеленый, желтый, красный, малиновый).

Должно существовать три схемы (с кнопкой переключения) для текущих, аналитических и прогнозистических данных. Климатические значения (среднее, минимальное и максимальное) должны отражаться на всех трех схемах в виде горизонтальных линий, пересекающих прибор и представленных другим цветом.

Создание динамических представлений базируется на следующем. Многие АП могут состоять из десятков элементов (однородных графиков, карт), которые логически и физически связаны и могут показываться на экране поочередно. Просмотр множества однородных вариантов графиков, диаграмм должен быть организован по нескольким объектам, показателям, статистикам и т.п.

5. Заключение

Использование ИР ЕСИМО позволяют быстрее приступить к обработке данных в АК. Аналитик может создать АП за несколько минут, преобразовать форму представления данных одним щелчком мыши, добавить эффективные синхронизированные последовательности анимации, позволяющие быстро выявить тенденции или аномалии и более убедительно аргументировать свои предложения благодаря расширенным инструментам представления результатов анализа.

АК обеспечивает качественно новые возможности анализа, позволяя пользователям получать доступ и загружать данные практически из любого источника. Аналитики получили возможность самостоятельно создавать АП, улучшать их и передавать друг другу, им больше не требуется помощь ИТ-отдела. Теперь речь идет об исследовании данных, а не только об их предоставлении.

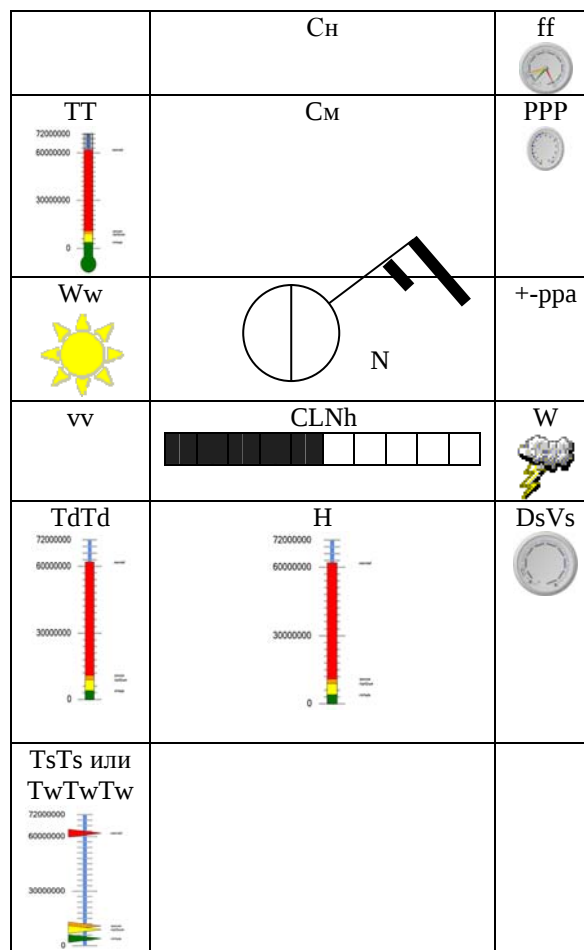


Рис. 1 Схема отражения состояния погодных условий

Обозначения: Cn - облака верхнего яруса, даются условными значками согласно символике; ff - скорость ветра в м/с; TT - температура воздуха в градусах; Cm - облака среднего яруса, даются условными значками согласно символике; PPP - давление воздуха (десятки, единицы и десятые доли мб); + - pp - величина барической тенденции; a - характеристика барической тенденции; vv - горизонтальная видимость в м; CLNh - количество облаков верхнего или нижнего яруса (отражается путем закрашивания квадратов, соответствующих числу баллов); T - общее количество облаков, отмечается путем закрашивания черным цветом круга в объеме соответствующем количеству баллов (от 0 до 10); WW - погода в срок (цифры кода переводятся в символику на основе иконок); W - погода между сроками (цифры кода переводятся в символику на основе иконок); TdTd - точка росы в градусах C; h - высота основания облаков над поверхностью в м; Ds - генеральное направление перемещения судна, в градусах; Vs - средняя скорость перемещения судна в км/час; TsTs - разница между температурой воздуха и воды с десятками градуса; TwTwTw - температура воды в градусах.

Литература

- [1] Бобровский С. Копаемся в шпионских данных // PC Week. 2012. [Электронный ресурс]. – Режим доступа:
<http://www.pcweek.ru/idea/blog/idea/1143.php>
<http://getinfo.ru/article464.html>, свободный. – Загл. с экрана.
- [2] Вейн Раш. BI в режиме самообслуживания: всё ближе к цели // PC Week/RE №31-32 (733-734). 31 августа — 6 сентября 2010.
- [3] Гореткина Е. Gartner: рынок бизнес-аналитики поляризуется // PC WEEK. PC Week/RE №10 (760) 19 апреля 2011.
<http://www.pcweek.ru/business/article/detail.php?ID=130134>.
- [4] Колесов А. Время управления эффективностью бизнеса уже настало? // PC Week/RE №42 (744). 9—15 ноября 2010.
- [5] Методические рекомендации по подготовке спецификаций комплексного информационного обеспечения морской деятельности в Российской Федерации ресурсами ЕСИМО. Обнинск: ФГБУ «ВНИИГМИ-МЦД». 209. 52 с.
- [6] MS SQL Server 2012: как выстраивается BI // C-News. 29.05.12. [Электронный ресурс]. – Режим доступа:
http://www.cnews.ru/reviews/index.shtml?2012/05/29/491201_2
<http://getinfo.ru/article464.html>, свободный. – Загл. с экрана.

Analytical complex ESIMO

E.D.Vyazilov, D.A.Melnikov, N.V.Chunyaev

Information on analytical complex of Unified state system for information on World Ocean (ESIMO) is presenting. The brief review of analytical system is giving. Показана The operations with analytical complex is showing. It is saying about perspectives of developments analytical complex ESIMO.