

2. Eroshenkova D. A. et al. Automated determination of forest-vegetation characteristics with the use of a neural network of deep learning // *International Conference on Neuroinformatics*. – Springer, Cham, 2019. – С. 295-302.
3. Тебенькова Д. Н. и др. Мультифункциональность и биоразнообразие лесных экосистем // *Журнал "Лесоведение"*. – 2019. – №. 5.
4. Лукина Н. В. и др. Биоразнообразие и климаторегулирующие функции лесов: актуальные вопросы и перспективы исследований // *Вопросы лесной науки*. – 2020. – Т. 3. – №. 4. – С. 1-90.
5. Terekhov V. et al. Classification of Tree Species by Trunk Image Using Conventional Neural Network and Augmentation of the Training Sample Using a Telegram-Bot // *International Conference on Neuroinformatics*. – Springer, Cham, 2021. – С. 210-216.

Automatic meteorological station dataset storage based on PostgreSQL DBMS and approach to providing online access to them

Titov A.G.^{1,2}, Bogomolov V.Yu.¹

¹Institute of Monitoring of Climatic and Ecological Systems SB RAS, Tomsk, Russia

²Federal Research Center for Information and Computational Technologies, Tomsk, Russia

E-mail: titov@scert.ru

The tasks of monitoring of environmental parameters, and, in particular, meteorological parameters, imply usage of the specialized scientific equipment, such as automatic meteorological stations that transfer real time observation data (air temperature, humidity, wind speed, etc.) to the data storage. As a rule such data represent a set of text or binary files corresponding to the observation period defined. It should be noted that data obtained from the modern sensors form huge archives especially when taking into consideration time periods greater than one year. Since it's well known that calculation of climatic parameters requires time periods greater than 30 years to be used, the problem of meteorological data streams handling as well as providing unified application interface to them along with corresponding graphical user interface, is still of great importance. To realize the task described, the PostgreSQL DBMS was chosen since it has built-in capabilities that allow working with georeferenced data. Thus, meteorological database consists of 6 main tables.

The PostgreSQL database structure developed for meteorological data obtained from automatic meteorological stations provides technological basis for regional system designed for collection, storage and online dissemination of the datasets required by the end user. The transition made from the source data in text and binary formats to the relational DBMS allows using unified SQL queries to perform effective search and retrieval of the data the user needs, as well as to develop graphical user interface based on web mapping technologies.

Acknowledgments. This study was supported under project # FWRG-2021-0004 of a RAS Basic Research Program.

Организация хранения данных автоматизированных метеорологических комплексов в СУБД PostgreSQL и обеспечение доступа к ним

Титов А.Г.^{1,2}, Богомолов В.Ю.¹

¹Институт мониторинга климатических и экологических систем СО РАН, Томск, Россия

²Федеральный исследовательский центр информационных и вычислительных технологий, Томск, Россия

E-mail: titov@scert.ru

ВВЕДЕНИЕ

Задачи мониторинга параметров окружающей среды, и, в частности, метеорологических параметров, подразумевают использование оборудования, такого как автоматизированные метеорологические комплексы (АМК), передающих данные наблюдений (температура атмосферы, влажность, скорость ветра на различных вертикальных уровнях, и т. д.) в режиме квазиреального времени, в систему хранения

ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНАЯ ИНФРАСТРУКТУРА РЕГИОНАЛЬНЫХ ИССЛЕДОВАНИЙ ОКРУЖАЮЩЕЙ СРЕДЫ

данных. Как правило, эти данные представляют собой набор текстовых или бинарных файлов, соответствующих конкретному периоду наблюдений. Следует отметить, что данные, получаемые от современных метеорологических датчиков, характеризуются высокой частотой измерений (десятки тысяч записей в час), что обуславливает большой объем потоковых данных (десятки и сотни гигабайт), получаемых за интервалы времени длительностью больше года. Принимая во внимание, что для расчета оценок климатических характеристик используются 30-летние и более интервалы, задача организации хранения данных метеорологических наблюдений, обеспечения унифицированного интерфейса доступа к ним, а также соответствующего графического интерфейса пользователя, является весьма актуальной.

В работе [1] представлено программное обеспечение для хранения и обработки суточных метеорологических данных, для их последующего использования в проектах SWAT (Soil and Water Assessment Tool, [2]). База метеорологических данных реализована на основе реляционной СУБД MS Access. Структура представленной базы данных достаточно очевидна, одна таблица используется для хранения описаний метеорологических станций, оставшиеся 5 таблиц используются для хранения значений метеорологических параметров, таких как температура, количество осадков и т. д. При этом каждому параметру соответствует своя таблица, а переменной величиной является идентификатор станции. Интерфейс пользователя позволяет просматривать и редактировать как данные описаний станций, так и данные наблюдений, а также импортировать и экспортировать их в форматы *.xls, .csv, txt. Также интерфейс позволяет запускать вычислительные процедуры, такие как «Среднесуточная температура точки росы за каждый месяц» и пр.

В техническом докладе австралийских авторов [3] представлена разработанная ими на основе ПО MS Access структура базы метеорологических данных, которые были получены начиная с 1931 года. В данной работе каждой автоматизированной метеорологической станции соответствует своя таблица, первичным ключом которой является поле временной метки (timestamp). Статистическая обработка данных производится ежечасно, с автоматической рассылкой отчетов в государственные организации.

Интересной является работа [4], в которой описана попытка применить стандарты систем управления потоками данных (Data Stream Management System) к хранению и предоставлению доступа к метеорологическим данным. В данном проекте для решения этой задачи применялся продукт SQLStream (<https://sqlstream.com/>), при этом использовались две отдельные базы данных: одна база данных PostgreSQL, для хранения статической информации, такой как описание метеорологических станций (идентификатор, географические координаты, высота над уровнем моря, и т. п.), и вторая база данных SQLStream, пополняемая в режиме квазиреального времени. Фактически, база данных SQLStream состоит из одной таблицы Data, с полями (<идентификатор станции>, <наименование переменной>, <значение переменной>). Результаты на момент выполнения данного исследования оказались несколько неоднозначными. В то время как система управления потоками данных оказалась эффективнее в плане

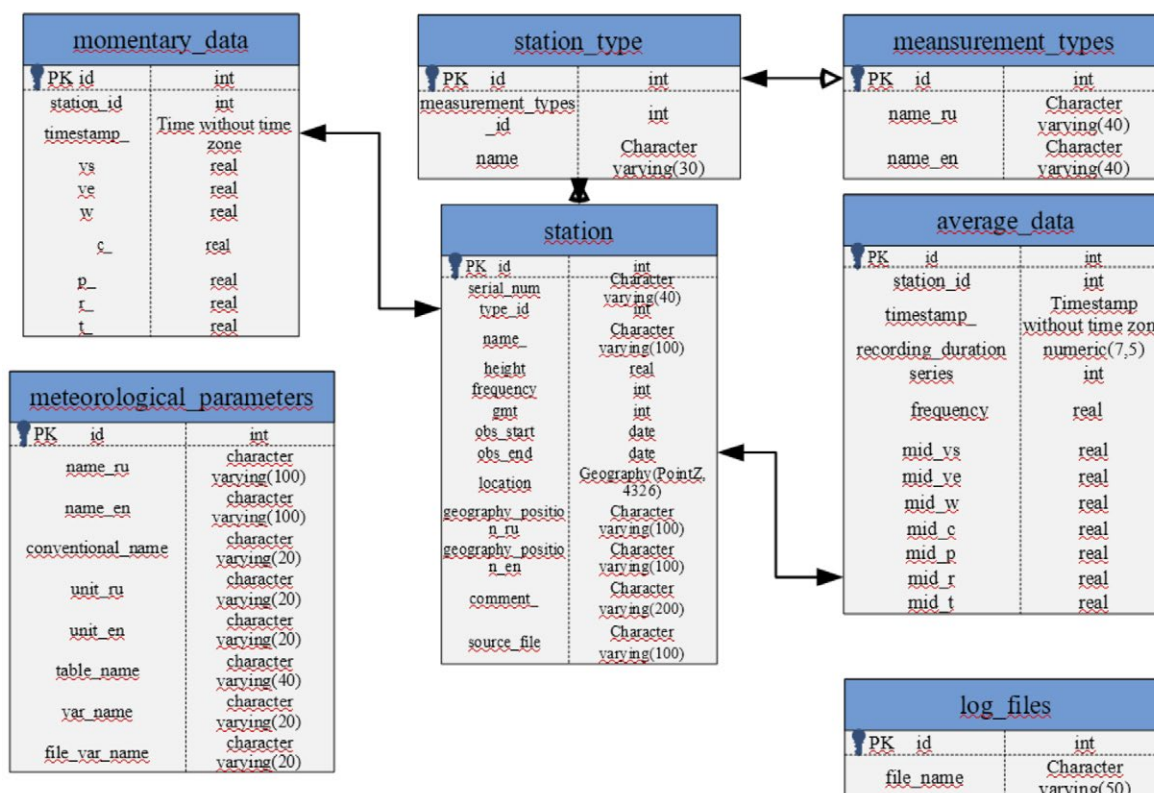


Рис. 1.

Структура базы метеорологических данных АМК.

id	name_en	name_ru
2	Thermodynamic temperature	Термодинамическая температура
3	Average atmospheric pressure	Среднее значение атмосферного давления
4	Average value of relative air humidity	Среднее значение относительной влажности воздуха
5	The average value of the speed of the horizontal wind component	Среднее значение скорости горизонтальной компоненты ветра
6	Maximum value of the horizontal wind component	Максимальное значение горизонтальной компоненты ветра
7	Average value of the vertical component of the wind	Среднее значение вертикальной компоненты ветра
8	Mean value of the meridional (southern) component of the horizontal wind	Среднее значение меридианальной (южной) компоненты горизонтального ветра
9	Mean zonal (eastern) component of the horizontal wind	Среднее значение зональной (восточной) компоненты горизонтального ветра
10	Temperature	Температура
11	The value of the meridional (southern) component of the horizontal wind	Значение меридианальной (южной) компоненты горизонтального ветра
12	The value of the zonal (eastern) component of the horizontal wind	Значение зональной (восточной) компоненты горизонтального ветра
13	Vertical wind speed	Вертикальная компонента ветра
14	Atmosphere pressure	Атмосферное давление
15	Relative humidity	Значение относительной влажности воздуха

var_name	unit	unit_en	table_name	file_var_name	conventional_name
mid_t	градусы Цельсия	degree Celsius	average_data	T	T
mid_p	мм.рт.ст.	mm Hg	average_data	P	P
mid_r	%	%	average_data	r	NULL
v	м/с	m/s	average_data	V	avg_Uuv
vmax	м/с	m/s	average_data	Vmax	NULL
mid_w	м/с	m/s	average_data	w	NULL
mid_vs	м/с	m/s	average_data	Vs	NULL
mid_ve	м/с	m/s	average_data	Ve	NULL
t_	градусы Цельсия	degree Celsius	momentary_data	T	T
vs	м/с	m/s	momentary_data	Vs	avg_v
ve	м/с	m/s	momentary_data	Ve	avg_u
w	м/с	m/s	momentary_data	w	avg_w
p_	мм.рт.ст.	mm Hg	momentary_data	P	P
r_	%	%	momentary_data	r	NULL

Рис. 2.
Таблица meteorological_parameters.

получения постоянных обновлений от метеорологических станций, ее оказалось гораздо сложнее администрировать, а также потребовалось выполнить большой объем дополнительных работ для обеспечения рутинных операций SQL.

В данной работе представлен прототип универсального решения на основе СУБД PostgreSQL, разработанного в ИМКЭС СО РАН, для географически распределенной системы автоматизированных метеорологических комплексов.

СТРУКТУРА БАЗЫ ДАННЫХ

Для реализации поставленной задачи была выбрана СУБД PostgreSQL, которая имеет встроенную функциональность для работы с геопривязанными данными. База метеорологических данных состоит из 6 основных таблиц (Рис. 1).

Таблица **measurement_types** содержит базовые типы измерений: атмосфера, почва, океан. Таблица **station_type** содержит имеющиеся типы автоматизированных метеорологических станций/комплексов (например, «АМК-03»), также она содержит столбец **measurement_types_id** – идентификатор типа измерений. Таблица **station** содержит следующие столбцы: **id** – уникальный идентификатор станции; **serial_num** – серийный номер станции; **type_id** – идентификатор типа станции; **name_** – текстовое название; **height_ta** – высота размещения станции; **frequency** – частота измерения; **gmt** – время по Гринвичу; **obs_start** и **obs_end** – время начала и завершения всех записей для станции; **location** – географические координаты; **geography_position** – географическое месторасположение станции (адрес и координаты); **comment** – столбец для заметок. Следует отметить, что используемый термин «станция» однозначно определяется 5 параметрами: **serial_num**, **type_id**, **height_ta**, **frequency**, **location**. При изменении хотя бы одного из этих параметров (смена месторасположения, замена самого прибора, корректировка высоты), будет заведена новая станция с уникальным идентификатором (**id**). В таблице **momentary_data** хранятся непосредственно данные мгновенных измерений со всех метеорологических станций, она таким образом является самой большой по объему. Таблица **average_data** содержит усредненные данные наблюдений, пересылаемые станцией однократно за установленный интервал времени. На Рис. 2 представлены поля таблицы **meteorological_parameters**, где **var_name** – наименования столбцов из таблиц **momentary_data** и **average_data**, **unit** – физические единицы измерения. Эта таблица содержит метаданные, необходимые для функционирования графического интерфейса пользователя.

ЗАКЛЮЧЕНИЕ

Разработанная структура базы данных для метеорологических данных, получаемых с датчиков автоматизированных метеорологических комплексов, является технологической основой для региональной системы сбора, хранения и предоставления доступа к этим данным. Переход от исходных текстовых и бинарных файлов, генерируемых АМК, к реляционной СУБД, позволяет использовать унифицированные SQL-запросы для поиска и выборки данных, необходимых пользователю, и разработать на их основе интуитивно-понятный графический интерфейс.

Благодарности. Данная работа была поддержана проектом № FWRG-2021-0004 базовой программы РАН.

ЛИТЕРАТУРА:

1. *Hrast Essenfelder, Arthur. SWAT Weather Database: A Quick Guide. 2018. doi: 10.13140/RG.2.1.4329.1927. URL: https://www.researchgate.net/profile/Arthur-Hrast-Essenfelder-2/publication/294535100_SWAT_Weather_Database/data/5756f68e08aef6cbe35f4e5b/WeatherDatabase-QuickGuide.pdf*
2. *Oñate-Valdivieso, F.; Bosque Sendra, J. Semidistributed hydrological model with scarce information: Application to a large south american binational basin. J. Hydrol. Eng. 2014, 19, 1006–1014.*
3. *F. Zhou, D.J. Smith and S. Khan. An Automated Weather Database System / CSIRO Land and Water, CSIRO Land and Water Technical Report 19/02, May 2002. URL: https://www.researchgate.net/publication/271770165_An_Automated_Weather_Database_System*
4. *Axel Lewenhaupt, Richard Nysäter, Magnus Olsson, and Tobias Reinhammar. Weather data application using SQLstream. 2015. URL: <https://www.kth.se/social/files/558023ecf27654759601639d/10.pdf>*