

Интернет-журнал «Наукovedение» ISSN 2223-5167 <http://naukovedenie.ru/>  
Выпуск 6 (25) 2014 ноябрь – декабрь <http://naukovedenie.ru/index.php?p=issue-6-14>  
URL статьи: <http://naukovedenie.ru/PDF/138TVN614.pdf>  
DOI: 10.15862/138TVN614 (<http://dx.doi.org/10.15862/138TVN614>)

УДК 004.65:004.622

**Ботыгин Игорь Александрович**

ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»  
Россия, Томск  
Доцент кафедры «Информатики и проектирования систем»  
Кандидат технических наук  
E-mail: [bia@tpu.ru](mailto:bia@tpu.ru)

**Лоскутов Виталий Валерьевич**

ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»  
Россия, Томск  
Магистрант  
E-mail: [deonold@gmail.com](mailto:deonold@gmail.com)

## **Распределенная система сбора и систематизации данных приземных гидрометеорологических наблюдений**

**Аннотация.** В статье приведены результаты разработки программного обеспечения распределенного сетевого хранилища данных приземных метеорологических наблюдений, представленных национальным вариантом международного кода для оперативной передачи данных о состоянии окружающей среды с метеорологических станций по всем регионам России, и предоставлением необходимых данных исследователям, как в визуальной форме, так и в широко распространенных экспортных форматах.

Спроектирована функционально-структурная схема распределенного сетевого хранилища метеоданных, включающая следующие модули: модуль командного центра (индикация и мониторинг вычислительных узлов), модуль диспетчера баз данных (распределение метеоданных по узлам и выполнение клиентских запросов), клиентский модуль (доступ к метеоданным исследователям), модуль терминалов (источники получения данных метеорологических наблюдений), модули хранения и обработки (узлы хранения распределенных метеоданных).

Приведены результаты практической апробации разработанного программного обеспечения сетевого хранилища данных приземных метеорологических наблюдений. Для имитации кластера информационно-вычислительных серверов в пилотном проекте использовалась многопоточность, поддерживаемая, практически, всеми операционными системами для создания процессов с параллельно выполняющимися потоками.

Выбранные для реализации сетевого хранилища инструментальные средства разработки и проведенные имитационно-программные эксперименты позволили максимально эффективно реализовать взаимодействие модулей хранилища, обеспечить его оптимальное быстродействие, устойчивость и надежность функционирования при распределении и управлении большими объемами данных.

Полученные результаты позволяют трансформировать наработки по управлению данными приземных метеорологических наблюдений и для оперативного сбора, систематизации и хранения данных гидрометеорологических наблюдений, представленных в других буквенно-цифровых кодах, а также для наблюдений других категорий.

**Ключевые слова:** информационная технология; распределенные системы; база данных; сетевое хранилище данных; приземные метеорологические наблюдения; код оперативной передачи данных приземных гидрометеорологических наблюдений; архитектура сетевого хранилища, многопоточность; распараллеливание; большие данные; сервер; сокет.

**Ссылка для цитирования этой статьи:**

Ботыгин И.А., Лоскутов В.В. Распределенная система сбора и систематизации данных приземных гидрометеорологических наблюдений // Интернет-журнал «НАУКОВЕДЕНИЕ» 2014. № 6  
<http://naukovedenie.ru/PDF/138TVN614.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ. DOI:  
10.15862/138TVN614

В настоящее время технологии работы с большими объемами данных (Big Data) находятся в сфере внимания, практически, всех специалистов ИТ-отрасли. Именно BigData-технологии, потенциально, предоставляют фантастические возможности аналитико-информационного анализа для государственных и коммерческих отраслевых приложений во всех сферах жизни общества. Сверхбольшие объемы информации характерны для космических наблюдений, при сборе и систематизации различных метеорологических данных, для крупных производственных сфер, “smart grid”, для новейших медицинских исследований, исследований в биологии и других многочисленных исследований. Набирающие всеобщую популярность сервисы и услуги глобальных телекоммуникаций (Cloud Computing) способствуют не только сверхбыстрому росту объемов данных, но и выделению направления Big Data в самостоятельную область ComputerScience-исследований.

Таким образом, проблема хранения и обработки гигантских объемов данных является актуальной и требует проектирования распределенной информационной инфраструктуры ее поддержки, обеспечивающей гибкую и открытую платформу для управления тысячами серверов в кластере и развернутыми на них корпоративными приложениями. В мире современных информационных технологий существуют решения, позволяющие проектировать такие комплексные виртуальные организации – grid-системы. Идеология grid предполагает возможность доступа к ресурсам виртуальной организации независимо от географического места расположения пользователя и самих ресурсов [1-3]. Современной предпосылкой к интеграции и консолидации данных является также и концепция Cloud computing [4-6], в основе которой также лежат виртуальные и распределенные системы.

Заметим, что именно распределенные информационные системы являются базовым инструментом для обработки данных с территориально удаленных серверов, функционирующих на различных платформах и хранящих информацию в различных форматах. В современных сетевых информационных технологиях всё чаще используют распределённую обработку данных. Именно распределенные автоматизированные системы повышают эффективность удовлетворения информационно-вычислительных потребностей пользователей и обеспечивают гибкость и оперативность принимаемых ими решений.

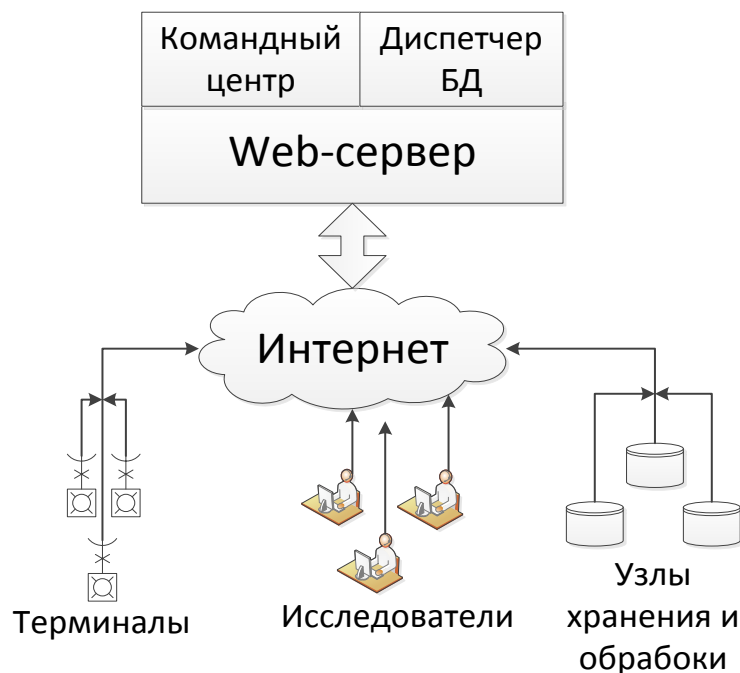
Одним из наиболее простых и эффективных решений указанной проблемы, по мнению авторов, является создание сетевого хранилища данных. По определению, сетевое хранилище подразумевает территориально-распределенное хранение информации. Таким образом, сетевое хранилище данных можно представить распределенной вычислительной системой, предназначенной для хранения наборов данных, которые не только логически связаны между собой, но и физически (территориально) распределены по разным узлам компьютерной системы. В пользовательском смысле, основной задачей такой распределенной системы, как составного элемента централизованного Интернет-сервиса, является хранение данных и предоставление доступа к ним по мере необходимости со стороны неограниченного количества сетевых пользовательских сервисов.

Апробация разработки указанной структуры сетевого хранилища данных осуществлена при проектировании программного обеспечения распределенного сетевого хранилища приземных метеорологических наблюдений, представленных кодом КН-01 [7] о состоянии окружающей среды с метеорологических станций по всем регионам России, и предоставлением необходимых данных исследователям как в визуальной форме, так и в широко распространенных экспортных форматах. Заметим, что данные метеорологических наблюдений с сети станций Гидрометцентра России, расположенных на суше, и закодированные национальным вариантом международного кода FM 12 SYNOP [8, 9] – КН-01, идеально поддерживают концепцию «трех V» (Volume – объем данных, Variety – скорость прироста данных, Velocity – скорость обработки данных), введенную для определения

«Большие данные» [10]. Поддержка концепция «четырех V» (добавляется Veracity – достоверность данных), безусловно, уже зависит от реализации информационной системы обработки данных [11]. А для концепции «пяти V» (добавляется Value – значение данных), важны полезность и выгода от использования BigData-систем [12].

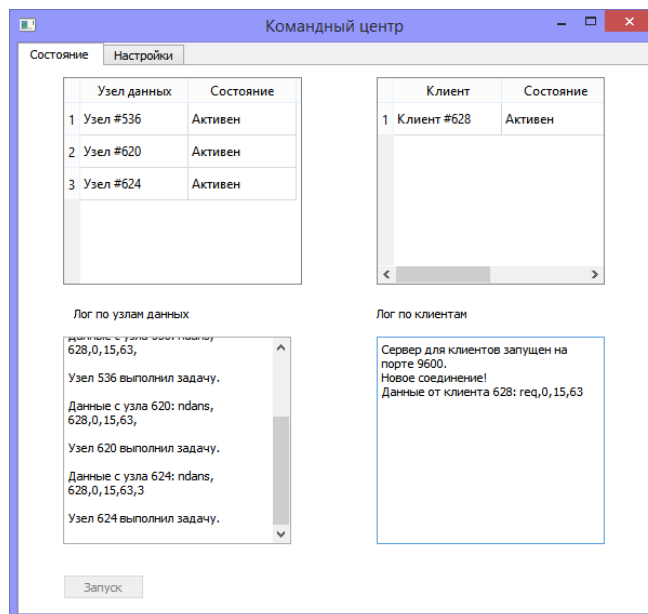
При проектировании распределенного сетевого хранилища метеоданных были выделены следующие модули (рис.1):

- модуль командного центра (индикация и мониторинг вычислительных узлов);
- модуль диспетчера баз данных (распределение метеоданных по узлам и выполнение клиентских запросов);
- клиентский модуль (доступ к метеоданным исследователям);
- модуль терминалов (источники получения данных метеорологических наблюдений);
- модули хранения и обработки (узлы хранения распределенных метеоданных).



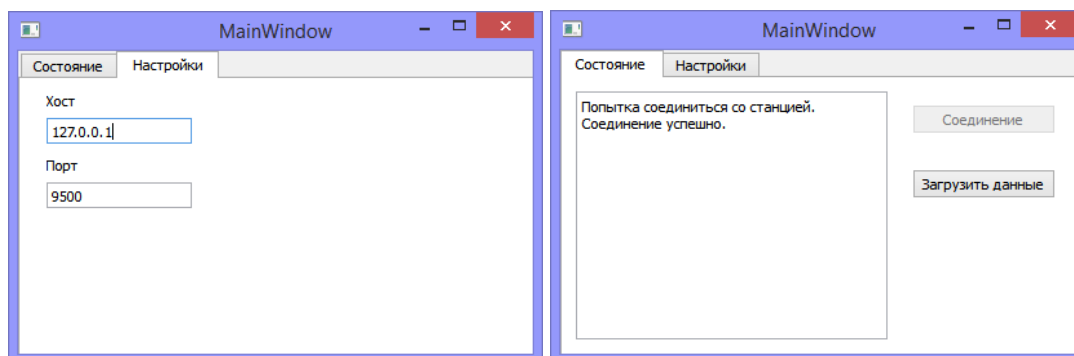
*Рис.1. Структурная схема сетевого хранилища  
(разработано авторами)*

Центральным звеном сетевого хранилища метеорологических данных является модуль командного центра (рис. 2.). Именно на модуль командного центра возложен весь функционал по управлению и функционированию хранилища метеоданных. Командный центр, являющийся местом для подключения узлов хранения данных и терминалов ввода, отслеживает подключение и отключение каждого из них (мониторинг работы узлов хранения), осуществляет отправку информации с терминалов диспетчеру баз данных.



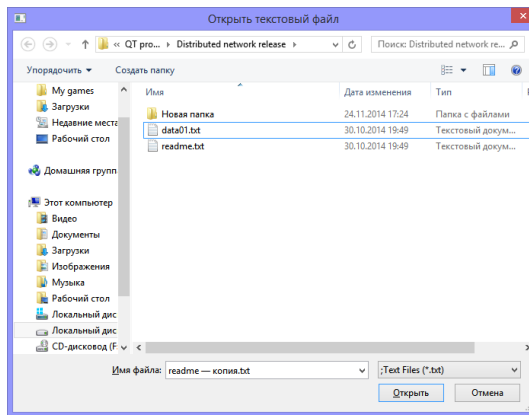
*Рис. 2. Окно командного центра  
(разработано авторами)*

В модуле командного центра создаются серверные сокеты для узлов хранения и терминалов, параметры которых указываются в настройках (рис. 3.). На узле хранения необходимо указать адрес и порт командного центра для корректной работы. Центр принимает запрос от терминала ввода под определённым номером. Далее формируется соответствующий установочный запрос, который отправляется на каждый узел хранения данных. Центр принимает ответы каждого узла и перенаправляет их диспетчеру базы данных для работы.



*Рис.3. Формы узлов хранения  
(разработано авторами)*

Данные для узла хранения должны быть отформатированы в коде КН-01, каждая телеграмма на отдельной строке в текстовом файле. К каждому узлу прилагается собственная база данных, которая пополняется с помощью функции «Загрузить данные» (рис. 4.).



**Рис. 4.** Загрузка данных в узле хранения  
(разработано авторами)

После прочтения текста телеграммы из текстового файла, приложение узла хранения разбирает его по метаописанию КН-01 и записывает в специальном файле базы данных. В каждой строке файла находится содержимое одной телеграммы, в столбцах записывается их порядковый номер и числовое содержание (рис. 5.).

id	date_day	date_time	date_windme...	location_regi...	location_stati...	cloudlow_inc...	cloudlow
57	26	615	1	26	64	2	1
58	27	715	1	26	64	2	1
59	28	815	1	26	64	2	1
60	29	915	1	26	64	2	1
61	30	15	1	26	64	2	1
62	31	115	1	26	64	2	1
63	1	115	1	26	65	2	1
64	2	215	1	26	65	2	1
65	3	315	1	26	65	2	1
66	4	415	1	26	65	2	1
67	5	515	1	26	65	2	1
68	6	615	1	26	65	2	1
69	7	715	1	26	65	2	1

**Рис. 5.** Структура записей файла базы данных  
(разработано авторами)

Названия столбцов выбираются в соответствии с назначением соответствующего поля телеграммы КН-01, что обеспечивает не только удобство работы с метеоданными, но и позволяет быстрее получить доступ к ним.

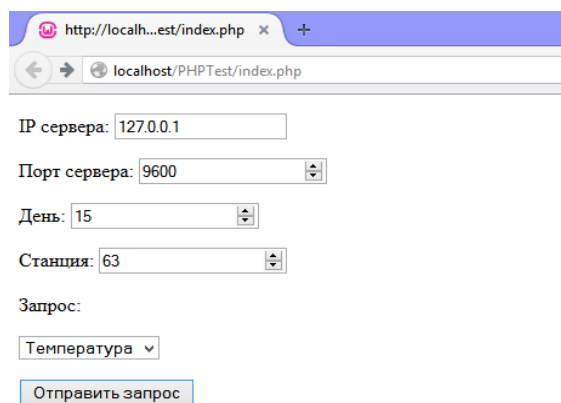
Непосредственной связи терминалов ввода с сетевым хранилищем метеоданных в системе не предусмотрено. Предоставлена возможность ввода метеонаблюдений в коде КН-01 с использованием специализированного приложения, доступного через модуль терминала.

Таким образом, данные метеонаблюдений с терминалов попадают на обработку в командный центр, который и обеспечивает их передачу диспетчеру баз данных. Диспетчер баз данных, в свою очередь, классифицирует метеоданные и отправляет узлам хранения (серверы для хранения данных). На узле хранения, получаемые от диспетчера метеоданные размещаются в специализированной базе данных в виде текстовых файлов. В этих файлах данные приземных

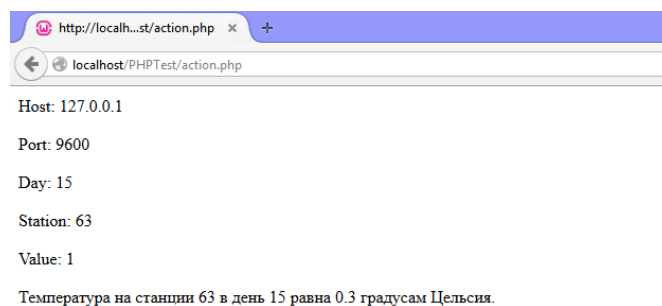
гидрометеорологических наблюдений представлены в строгом соответствии со схемой кода КН-01 и его разделами. Файлы метеонаблюдений, в свою очередь, расположены в каталогах по станциям региона, годам и месяцам наблюдения.

Диспетчер поддерживает таблицу со всеми активными узлами в данный момент, с описанием – какой узел что хранит. Для эффективного распределения полученных метеоданных от командного центра, диспетчер выполняет поиск по таблице активных узлов хранения, выбирает наиболее подходящий таким образом, чтобы узел не содержал разнотипной информации. В настоящий момент структуризация данных метеонаблюдений ориентирована на административно-территориальные регионы Российской Федерации.

Взаимодействие исследователей с системой происходит через веб-браузер (рис. 6., рис. 7.), либо с помощью специального клиентского модуля (рис. 8).

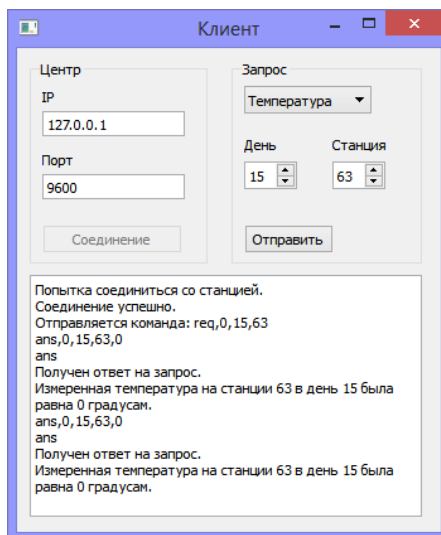


*Рис. 6. Формирование запроса веб-приложением  
(разработано авторами)*



*Рис. 7. Ответ веб-приложения  
(разработано авторами)*

При использовании клиентского модуля соединение компьютера исследователя с командным центром осуществляется с помощью стека протоколов TCP/IP (рис. 8.). Профессиональный клиентский модуль имеет интерфейс для соединения с командным центром и формирования интересных запросов.



*Рис. 8. Окно клиентского модуля  
(разработано авторами)*

После процедуры авторизации исследователя, включающей как идентификацию пользователя, так и определение права на получение запрашиваемой информации, запрос исследователя передается диспетчеру баз данных. Сформированному запросу присваивается определённый номер. Полученные сообщения под данным номером рассматриваются клиентским модулем как ответ на запрос. Диспетчер баз данных определяет узлы, к которым необходимо обратиться для получения требуемой информации. Запросы диспетчера к узлу хранения содержат номер запроса, тип и границы запрашиваемых данных, а также контрольную сумму.

Узлы хранения данных осуществляют поиск требуемой информации в своих базах данных. Происходит просмотр базы данных на совпадения. Если таковые имеются, в центр последовательно отправляются сообщения с ответными данными с указанным номером запроса. Эти данные впоследствии отправляются и на компьютер исследователя. Кроме того, имеется возможность передачи полученных данных в распространенных экспортных форматах.

Несмотря на то, что формат данных КН-01 используется на всей территории Российской Федерации, универсального средства распределенного сетевого хранилища приземных метеорологических наблюдений, представленных кодом КН-01 о состоянии окружающей среды с метеорологических станций по всем регионам России и предоставлением необходимых данных по запросу, не существует. Прежде всего, при разработке данного программного обеспечения, ставилась задача обеспечения возможности распределения и управления большими объемами данных. Выбранные для реализации сетевого хранилища средства позволили максимально эффективно реализовать взаимодействие модулей системы и обеспечить его оптимальное быстродействие.

Приведены результаты практической апробации разработанного программного обеспечения сетевого хранилища данных приземных метеорологических наблюдений. Для имитации кластера информационно-вычислительных серверов в пилотном проекте использовалась многопоточность, поддерживаемая, практически, всеми операционными системами для создания процессов с параллельно выполняющимися потоками.

Полученные результаты позволяют трансформировать наработки по управлению данными в формате SYNOP-FM12 и для оперативного сбора, систематизации, хранения и



данных гидрометеорологических наблюдений, представленных в других буквенно-цифровых кодах (FM18 – BUOY, FM-32 – PILOT, ... и т.д., вплоть до FM-75 – CLIMAT-TEMP). Кроме того, возможна адаптация системы и для наблюдений других категорий – AGRO, BUFR, RADR и др.

Безусловно, исследователям, помимо непосредственного визуального представления данных, необходимы и средства, хотя бы, их первичной математической обработки, желательно с графическим представлением результатов обработки запрошенных данных. Но – это уже разработка специализированных компьютерных инфраструктур, удаленно предоставляющих требуемые не только информационные, но и вычислительные сервисы и услуги.

## ЛИТЕРАТУРА

1. Распределённые вычисления [Электронный ресурс]. – Режим доступа: <http://ru.wikipedia.org/?oldid=65233585> (дата обращения 05.12.2014).
2. Грид [Электронный ресурс]. – Режим доступа: <http://ru.wikipedia.org/?oldid=66132964> (дата обращения 05.12.2014).
3. Grid computing [Электронный ресурс]. – Режим доступа: [https://en.wikipedia.org/wiki/Grid\\_computing](https://en.wikipedia.org/wiki/Grid_computing) (дата обращения 05.12.2014).
4. The NIST Definition of Cloud Computing. Hub [Электронный ресурс] – Режим доступа: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf> (дата обращения 05.12.2014).
5. What is cloud? [Электронный ресурс] – Режим доступа: <http://www.ibm.com/cloud-computing/us/en/what-is-cloud-computing.html> (дата обращения 05.12.2014).
6. What is cloud computing? [Электронный ресурс] – Режим доступа: <http://www.salesforce.com/cloudcomputing/> (дата обращения 05.12.2014).
7. Код для оперативной передачи данных приземных метеорологических наблюдений с сети станций Росгидромета. [Электронный ресурс] – Режим доступа: <http://method.meteorf.ru/norma/rec/code.pdf> (дата обращения 05.12.2014).
8. SYNOP Data Format (FM-12). [Электронный ресурс] – Режим доступа: <http://weather.unisys.com/wxp/Appendices/Formats/SYNOP.html> (дата обращения 05.12.2014).
9. Land station surface synoptic code FM 12-IX SYNOP. [Электронный ресурс] – Режим доступа: <http://atmo.tamu.edu/class/atmo251/LandSynopticCode.pdf> (дата обращения 05.12.2014).
10. Крис Канаракус. Машина Больших Данных. [Электронный ресурс] – Режим доступа: <http://www.osp.ru/nets/2011/04/13010802/> (дата обращения 05.12.2014).
11. The Big Data & Analytics Hub [Электронный ресурс] – Режим доступа: <http://www.ibmbigdatahub.com/infographic/four-vs-big-data> (дата обращения 05.12.2014).
12. Bernard Marr. Big Data: The 5 Vs Everyone Must Know. [Электронный ресурс] – Режим доступа: <https://www.linkedin.com/today/post/article/20140306073407-64875646-big-data-the-5-vs-everyone-must-know> (дата обращения 05.12.2014).

**Рецензент:** Мицель Артур Александрович, профессор, доктор технических наук, профессор кафедры Автоматизированные системы управления Томского государственного университета систем управления и радиоэлектроники

**Botygin Igor Aleksandrovich**

National Research Tomsk Polytechnic University  
Russia, Tomsk  
E-mail: [bia@tpu.ru](mailto:bia@tpu.ru)

**Loskutov Vitaliy Valer'evich**

National Research Tomsk Polytechnic University  
Russia, Tomsk  
E-mail: [deonold@gmail.com](mailto:deonold@gmail.com)

## **A distributed system for surface hydro-meteorological data collection and systematization**

**Abstract.** The following article contains the results of development of a distributed network storage for surface hydro-meteorological data observations, encrypted in a national variant of international code for transferring environmental data to the regions of Russian Federation and supplying all the necessary data for researchers in both visual form and other widely-spread export formats.

The designed functional scheme of distributed network meteorological data storage includes the following modules: the command center module (calculating nodes indication and monitoring), the database dispatcher module (distribution of the meteorological data to nodes and executing client requests), client module (the access to data), terminal module (sources of receiving meteorological observations data), modules of processing and storage (nodes for storing distributed data).

The results of practical approbation of the developed software for network data storage are presented in the following work. The multi-threading was used for imitation of the information and calculation server cluster. The multi-threading is supported by almost every modern operation system to create processes with parallel thread execution.

The instrumental choices for designing the network storage allowed to effectively implement the interaction of storage modules, ensure its speed and stability during the distribution and managing big data volumes.

The results allow to transform the pilot project for surface meteorological data management for the operational data collection, systematization and storage of other data types encoded into other numerical/symbolic codes or any other types of observation.

**Keywords:** information technology; distributed systems; database; network data storage; hydro-meteorological surface observations; code for transmitting surface hydro-meteorological observation data; network storage architecture; multithreaded; parallelization; big data; server; socket.

## REFERENCES

1. Raspredelemnnye vychisleniya [Elektronnyy resurs]. – Rezhim dostupa: <http://ru.wikipedia.org/?oldid=65233585> (data obrashcheniya 05.12.2014).
2. Grid [Elektronnyy resurs]. – Rezhim dostupa: <http://ru.wikipedia.org/?oldid=66132964> (data obrashcheniya 05.12.2014).
3. Grid computing [Elektronnyy resurs]. – Rezhim dostupa: [https://en.wikipedia.org/wiki/Grid\\_computing](https://en.wikipedia.org/wiki/Grid_computing) (data obrashcheniya 05.12.2014).
4. The NIST Definition of Cloud Computing. Hub [Elektronnyy resurs] – Rezhim dostupa: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf> (data obrashcheniya 05.12.2014).
5. What is cloud? [Elektronnyy resurs] – Rezhim dostupa: <http://www.ibm.com/cloud-computing/us/en/what-is-cloud-computing.html> (data obrashcheniya 05.12.2014).
6. What is cloud computing? [Elektronnyy resurs] – Rezhim dostupa: <http://www.salesforce.com/cloudcomputing/> (data obrashcheniya 05.12.2014).
7. Kod dlya operativnoy peredachi dannykh prizemnykh meteorologicheskikh nablyudeny s seti stantsiy Rosgidrometa. [Elektronnyy resurs] – Rezhim dostupa: <http://method.meteorf.ru/norma/rec/code.pdf> (data obrashcheniya 05.12.2014).
8. SYNOP Data Format (FM-12). [Elektronnyy resurs] – Rezhim dostupa: <http://weather.unisys.com/wxp/Appendices/Formats/SYNOP.html> (data obrashcheniya 05.12.2014).
9. Land station surface synoptic code FM 12-IX SYNOP. [Elektronnyy resurs] – Rezhim dostupa: <http://atmo.tamu.edu/class/atmo251/LandSynopticCode.pdf> (data obrashcheniya 05.12.2014).
10. Kris Kanarakus. Mashina Bol'shikh Dannykh. [Elektronnyy resurs] – Rezhim dostupa: <http://www.osp.ru/nets/2011/04/13010802/> (data obrashcheniya 05.12.2014).
11. The Big Data & Analytics Hub [Elektronnyy resurs] – Rezhim dostupa: <http://www.ibmbigdatahub.com/infographic/four-vs-big-data> (data obrashcheniya 05.12.2014).
12. Bernard Marr. Big Data: The 5 Vs Everyone Must Know. [Elektronnyy resurs] – Rezhim dostupa: <https://www.linkedin.com/today/post/article/20140306073407-64875646-big-data-the-5-vs-everyone-must-know> (data obrashcheniya 05.12.2014).