

УДК 622.1:004.65

DOI: 10.25635/2313-1586.2024.04.127

**Ногин Святослав Александрович**  
младший научный сотрудник,  
Институт горного дела УрО РАН,  
620075 г. Екатеринбург,  
ул. Мамина-Сибиряка, 58  
e-mail: [nsa@igduran.ru](mailto:nsa@igduran.ru)

**Nogin Svyatoslav A.**  
Junior Researcher,  
Institute of Mining UB RAS,  
620075 Ekaterinburg,  
58 Mamina-Sibiryaka Str.  
e-mail: [nsa@igduran.ru](mailto:nsa@igduran.ru)

## ОРГАНИЗАЦИЯ ДАТАСЕТА НА ОСНОВЕ ДАННЫХ МАРКШЕЙДЕРСКИХ НАБЛЮДЕНИЙ\*

## ON THE ISSUES OF ORGANIZING THE SURVEY DATA

### Аннотация:

В связи с развитием средств автоматизации в сфере обработки и анализа данных геодезических наблюдений возрастает значение того, в каких форматах представлены используемые массивы данных и какие средства предоставляют базы данных для операций над ними. В статье рассматривается случай наземной наблюдательной станции и приводится разработка для него формата хранения данных на основе выбранной нереляционной (NoSQL) модели баз данных – расширяемых хранилищ записей. Описано объединение данных исходно различных форматов в единый датасет. Показано, как в полученном наборе данных реализуется логическая модель данных, принятая в расширяемых хранилищах записей. Приведен процесс реализации этой модели средствами табличного процессора на примере данных по Сарановскому и Северопесчанскому месторождениям.

*Ключевые слова:* датасет, маркшейдерские наблюдения, данные маркшейдерских наблюдений, расширяемые хранилища записей, хранение данных, автоматизированная обработка данных, базы данных.

### Abstract:

In connection with the development of automation in the field of processing and analysis of geodetic observation data, the importance of the formats in which the used data arrays are represented and also the means of providing databases for operations on them. The paper considers the case of a ground observation station and presents the development of a data storage format based on a selected non-relational (NoSQL) database model, which is Wide Column Stores. The merging of data in initially different formats into a single dataset is described. It is shown how the logical data model adopted in wide column stores is realized in the obtained dataset. The process of realization of this model by means of a table processor on the examples of data on Saranovskoye and Severopeschanskoye deposits are given.

*Key words:* dataset, surveying observations, surveying data, wide column stores, data storage, automated data processing, databases.

### Состояние проблемы

В ходе работы с геодезическими данными, являющимися результатами традиционных методов измерений – нивелирования и дальномерных измерений, на всех ее этапах (накопления, обработки, анализа данных и использования полученных результатов) происходит активное внедрение средств автоматизации, обусловленное как развитием применяемых при обработке и анализе ЭВМ и программного обеспечения (ПО), такого как геоинформационные системы (ГИС), так и распространением средств автоматических измерений, а зачастую и возрастающими требованиями к объемам осуществляемого мониторинга.

В целях мониторинга, постобработки и анализа результатов геодезических измерений разработаны различные программные комплексы. Они предоставляют множество возможностей: запись в режиме реального времени данных, поступающих с геодезических и геотехнических датчиков, выявление выбросов в них, определение предельно допустимых значений измеренных параметров (например, смещений при сборе данных с тахеометров или наклонов для инклинометров), 2D- и 3D-визуализация векторов смещений и величин деформаций, определение трендов деформаций и прогнозирование на их

\* Статья подготовлена в рамках Госзадания №075-00412-22 ПР, тема 3 (2022-2024) (FUWE-2022-0003), рег. № 1021062010536-3-1.5.1.

основе. Наличие доступа к исходным по отношению к дальнейшей обработке данным обеспечивает корректное выполнение запросов конечного пользователя.

Примерами автоматизации могут служить обработка баз данных тахеометрической съемки [1] и организация базы данных и работа с ней средствами программы Geomarks [2] (разработка ОАО «Атомэнергопроект»), в которой формируется база данных геодезического мониторинга с данными по вертикальным и горизонтальным смещениям на полигонах АЭС, а также аналогичными данными по геодезическим знакам, заложенным в охраняемых зданиях, сооружениях и технологическом оборудовании АЭС. В Geomarks проводится анализ целостности данных в базе данных, возможна ее нормализация (обработка выбросов, интерполяция пропущенных значений, слияние временных рядов). Сама база данных ориентирована на работу с данными площадных измерений.

Анализу больших данных на горных предприятиях также посвящены работы [3, 4], в которых предлагается схема формирования базы данных и взаимодействия искусственного интеллекта (ИИ) с данными для оценки состояния горного предприятия, однако отсутствуют примеры реализации всей предложенной схемы и ее внедрения в современную практику.

Таким образом, несмотря на наличие различных представлений данных маркшейдерских наблюдений для функций хранения, редактирования и автоматизированной обработки в ПО, для случая поверхностной наблюдательной станции сохраняется необходимость разработать формат набора данных, учитывающий то, что такая станция представляет собой сеть профилей. Помимо обеспечения перечисленных функций, также ставится задача поддержки совместимости данных текущих наблюдений с архивными данными отчетов по наблюдательным станциям указанного типа, что позволило бы проводить анализ за весь срок наблюдений, работая с единственным набором данных для выбранного объекта мониторинга.

#### *Выбор формата хранения*

В последние годы, помимо получивших широкое распространение реляционных БД, в практику вошли и другие типы БД, которые стали известны как NoSQL (сокращение от «not only SQL») [5]. Среди них, в частности, выделяют расширяемые хранилища записей (ширококолоночные хранилища записей, Wide Column Stores, WCS). Данные в них можно представить как таблицы, включающие строки с уникальными идентификаторами и столбцы (рис.1, рис.2). Основная особенность этой модели заключается в том, что одной строке может соответствовать произвольное множество столбцов, которые могут быть объединены в семейства столбцов. Данные представляются как пары «ключ-значение», где значение соответствует составному ключу, состоящему из ключа строки, имен семейства столбцов и столбца, а также необязательной временной метки. Внутри поля таблицы допустимо записывать несколько различных значений, отличающихся временными метками [5, 6, 7]. Можно показать, что WCS обладают всеми необходимыми признаками БД.

Здесь необходимо оговорить то, что существуют различные подходы к определению «базы данных» (БД) [8]. Приведем некоторые из них.

Так, согласно первому подходу, БД является массивом информации о какой-либо сфере деятельности, предназначенной для коллективного использования и допускающей компьютерную обработку [9]. В соответствии с другим, она должна представлять собой совокупность взаимосвязанных данных при минимальной избыточности, допускающей их использование оптимальным образом для одного или нескольких приложений в определенной предметной области [10]. В рамках данного определения БД является множеством связанных файлов. Есть точка зрения, что БД – это совокупность данных, организованных по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования данными, независимая от прикладных программ [11]. БД

может пониматься как контейнер файлов, предоставляющий операции добавления, изменения и удаления данных и файлов [12]. Наконец, может быть дано следующее определение [13]: БД является совокупностью сведений (о реальных объектах, процессах и т.п.), относящихся к определенной теме или задаче, организованную таким образом, чтобы обеспечить удобное представление как совокупности в целом, так и любой ее части.

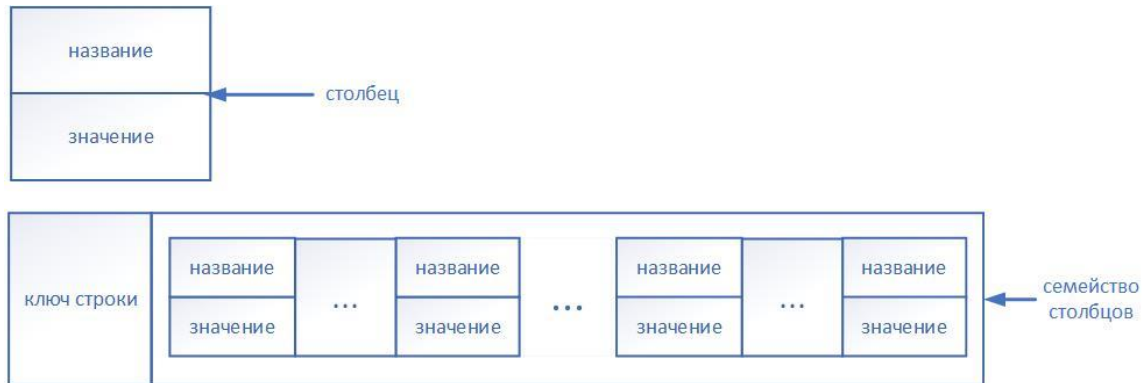


Рис.1. Схема модели WCS

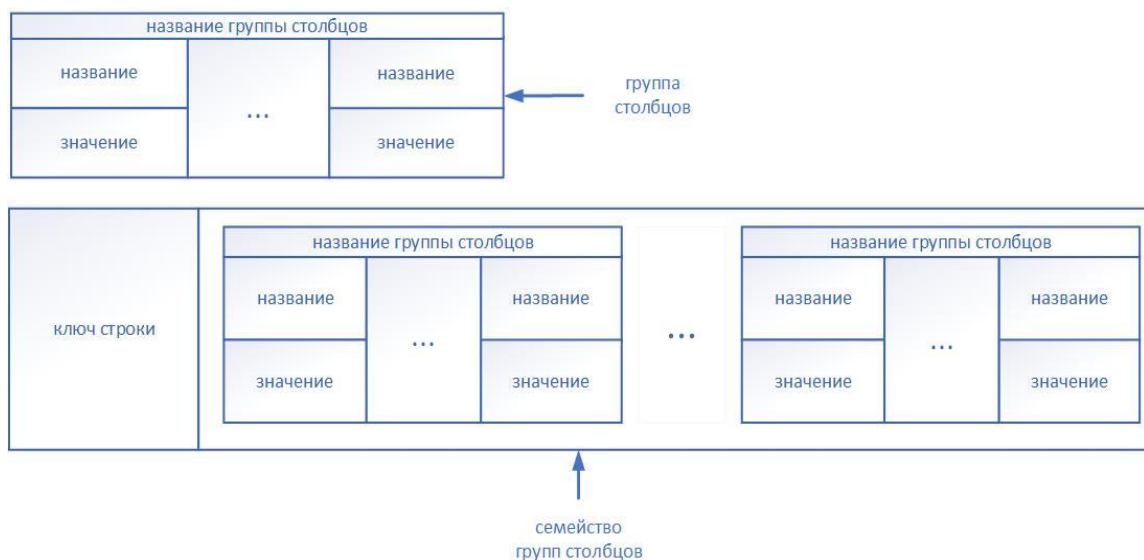


Рис.2. Схема модели WCS с хранением семейств столбцов

Из указанного выделяются следующие общие признаки баз данных: наличие упорядоченной структуры совокупности данных, построение БД согласно заданной системе правил, хранение и обработка данных в ней средствами ПО на ЭВМ, что подразумевает выбор определенного формата файла. Таким образом, легко видеть, что WCS могут рассматриваться как базы данных и, как показано далее, именно на их основе возможно предложить формат массива данных для рассматриваемого случая поверхностной наблюдательной станции. Благодаря наличию ключа у каждой строки возможно учесть порядок расположения реперов в каждом профиле. Кроме того, допустимы как пустые поля, так и значения разных типов, что позволяет упростить учет пропусков наблюдений.

Поскольку для анализа данных о сдвигении первостепенна интеграция архивных неоцифрованных данных с данными текущих наблюдений, возможно отказаться от специализированных NoSQL систем управления базой данных (СУБД), также вследствие того что обеспечиваемые ими возможности, такие как многопользовательский и удален-

ный доступ, поддержка требований ACID в рамках транзакций, обработка больших объемов данных и прочие, в рамках решаемых задач не являются необходимыми. Использование структуры данных из WCS при создании датасета средствами табличного процессора позволяет добиться такой интеграции, осуществлять хранение и добавление данных и обеспечить возможности дальнейшей обработки и анализа, которые решаются с помощью инструментов работы с электронными таблицами, такими как программные библиотеки `openpyxl` и `Pandas`.

### Практическая реализация датасета

С учетом упомянутых требований и современной практики обработки результатов маркшейдерских съемок можно предложить следующий процесс построения датасета (на примере применения табличного процессора Microsoft Excel).

Данными, получаемыми после камеральной обработки данных геодезического мониторинга, являются высотные отметки реперов и длины интервалов между реперами в профиле. Важно заметить, что ведомости наблюдений архивных отчетов также содержат этот набор данных, в силу чего можно рассматривать его как минимально необходимый. При этом данные архивных документов должны быть обработаны, чтобы их можно было рассматривать как часть датасета. Остальные параметры, служащие для оценки процесса сдвижения – оседания, горизонтальные и вертикальные деформации – вычисляются на основе этого набора.

Для хранения набора данных по выбранному объекту создается книга Excel, каждой профильной линии отводится лист книги. В каждом листе содержатся (рис. 3)

- идентификатор репера – натуральное число, уникальным образом в рамках создаваемой базы определяющее репер. Это позволяет рассматривать идентификатор как ключ строки;
- наименование репера, которое в связи с утратами и восстановлениями реперов может повторяться;
- группа столбцов координат  $x$ ,  $y$ ;
- две группы колонок  $d_n$ ,  $h_n$  дистанций и высотных отметок, где  $d$  – колонка дистанций;  $h$  – высота над уровнем моря;  $n$  – номер серии; в случае отсутствия замеров или данных по ним одна из колонок может оставаться пустой. В каждой группе количество столбцов зависит от числа проведенных серий;
- строка дат наблюдений, записанных в соответствующих столбцах  $d_n$ ,  $h_n$ ;
- строки данных, заполняемых в порядке увеличения расстояния от начала профиля при заложении репера: в каждой строке с идентификатором репера записывается его высота для указанной серии наблюдений, между строками высот – строки с дистанциями.

	реперы	координаты		колонок высот		колонок дистанций				
1	id	name	x	y	d0	h0	d1	h1	d2	h
2						07.1979		1980		
3	1	OP1				442,317		442,317		
4					52,963		52,958		52,958	
5	2	OP2				446,039		446,039		
6					56,064		56,054		56,041	
7	3	OP3	603 101,0	6 492 927,6		450,817		450,817		
8					49,599		49,604		49,589	
9	4	1	603 068,2	6 492 890,4		454,587		454,587		
10					20,893		20,890			
11	5	2				456,349		456,351		
12					25,019		25,014		45,892	
13	6	3	603 037,0	6 492 856,7		458,279		458,281		

Рис. 3. Таблица с данными наблюдений (фрагмент)

Необходимо учитывать наличествующую информацию о повреждениях реперов, чтобы отделить движения массива от подвижек реперов, им не вызванных. Решение этой проблемы также можно осуществить средствами БД: создается вторая книга Excel, содержащая идентификаторы для однозначного отождествления реперов, их наименования, годы существования по наличным данным, колонку сведений о повреждениях и утратах реперов с указанием дат этих событий, что позволяет внести соответствующие поправки в ходе дальнейшего анализа (рис.4).

Отсутствие требований атомизации атрибутов в рамках WCS позволяет при этом перечислять разнотипные данные.

идентификаторы	реперы		годы существования		пометки
	id	name	год закладки	последняя фиксация	
1	72	ОР-1а	2006		
2	44	ОР-1	1979		не найден (2009–2012)
3	73	19	2006		не найден (2020)
27	66	20	1979	1980	уничтожен строительством емкостей резервуара
28	67	21	1979	1980	уничтожен строительством емкостей резервуара
29	74	ОР4	2006	2007	
30	68	22	1979	1993	не найден (1984, 1987–1990), погнут (1988)
31	77	ОР4	2013	2018	
32	75	ОР5	2006	2018	
33	76	ОР6	2006	2019	
34	69	ОР-4	1979	1993	
35	70	ОР-5	1979	1993	не найден (1987), без дистанции (1990)
36	71	ОР-6	1979	1993	не найден (1987–1990)
37					

Рис. 4. Второй файл БД (фрагмент)

Отличием модели данных в построенном датасете от модели WCS является то, что для отдельной колонки допустимы объединения разных строк – в случае записи дистанции, измеренной для реперов, между которыми ранее или позднее выбранной даты были заложены другие реперы. Это позволяет наглядно проследить изменения в составе профильной линии с течением времени, сохраняя неизменной структуру датасета.

Для заполнения датасетов по предложенной схеме были выбраны два объекта: Северопесчанское и Сарановское месторождения. Схема рабочего процесса над данными по этим объектам с использованием построенных датасетов представлена на рис. 5.

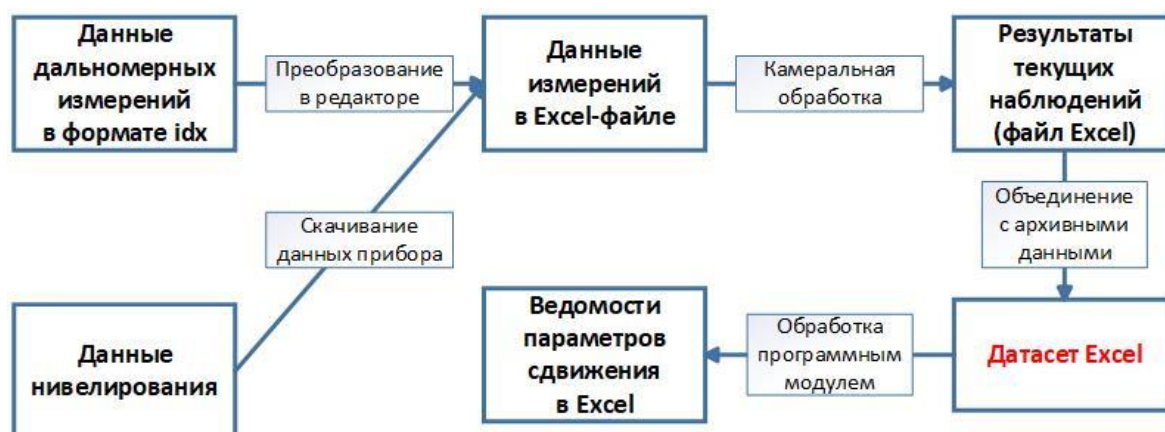


Рис. 5. Интеграция процессов над данными в рамках табличного процессора

Данные по Сарановскому месторождению включают период 1979 – 2023 гг. (с учетом дополнения новыми данными) по семи наземным профильным линиям, профилю на территории промплощадки, специальным наблюдательным станциям, включающим стеновые реперы для охраняемых зданий. Для Северопесчанского месторождения

взяты наблюдения последних лет по сохранившимся профильным линиям. По предложенной схеме для выбранных объектов созданы наборы данных, удобные как для заполнения и последующего редактирования оператором, так и для дальнейших обработки и анализа программными средствами, что повышает эффективность их использования [1], а кроме того, при необходимости, упрощает импорт данных датасета в специализированную СУБД, поддерживающую расширяемые хранилища записей.

#### Заключение

Согласно разработанной на основе логической модели данных, принятой в базах данных WCS (wide column stores, расширяемые хранилища записей), методике построены датасеты, содержащие обработанные данные маркшейдерских наблюдений по Сарановскому месторождению хромитов и Северопесчанскому железорудному месторождению [14, 15]. Непосредственно на их основе построены ведомости, отвечающие требованиям нормативных документов, и матричные диаграммы параметров сдвигения, что позволяет провести интеграцию рабочих процессов с данными наблюдений, как представлено на рис. 5. Это позволяет улучшить представление о развитии процесса сдвигения и получить данные для дальнейшего анализа.

#### Список литературы

1. Ребрин Е. Ю., 1993. Представление и обработка данных маркшейдерских съемок для построения планов горных работ на персональных компьютерах. *Известия Уральского государственного горного университета*, № 3, С. 110 – 116.
2. Веселов В.В., Сячинов А.Н., Широков В.В., Онищенко Д.Н., 2020. *Программный комплекс GeoMarks : программа для ЭВМ.* № рег. 2020660055: дата рег.:26.08.2020: дата публ. 26.08.2020. Бюл. № 9. АО «АТОМ-ЭНЕРГОПРОЕКТ». URL:[https://www.elibrary.ru/download/elibrary\\_43965872\\_47766549.PDF](https://www.elibrary.ru/download/elibrary_43965872_47766549.PDF) (дата обращения: 20.05.2024).
3. Ческидов В.В., 2017. Современные задачи комплексного мониторинга состояния горнотехнических объектов. *Сергеевские чтения. Выпуск 19. Геоэкологическая безопасность разработки месторождений полезных ископаемых*. Москва: РУДН, 590 с.
4. Ческидов В.В., Маневич А.И., Липина А.В., 2019. Получение и анализ больших данных в практике мониторинга состояния горнотехнических сооружений. *Горная промышленность*, № 2 (144), С. 86 – 88.
5. Кузнецов С.Д., Посконин А.В., 2013. Распределенные горизонтально масштабируемые решения для управления данными. *Труды Института системного программирования РАН*, Т. 24, С. 327– 358.
6. Wiese L., Waage T., Brenner M., 2019. CloudDBGuard: A Framework for encrypted data storage in NoSQL Wide Column Stores. *Data and Knowledge Engineering*, Vol. 126, P. 101732.
7. Chen J.-K., Lee W.-Z., 2020. The Transformation of Relational Database to Wide Column Store Database. *International Symposium on Computer, Consumer and Control (IS3C)*, Taichung City, Taiwan, 2020, pp. 384-386, doi: 10.1109/IS3C50286.2020.00105.
8. Воройский Ф.С., 2006. *Информатика. Энциклопедический систематизированный словарь-справочник: введение в современные информационные и телекоммуникационные технологии в терминах и фактах*. Москва: Физматлит, 945 с.
9. Зафиевский А.В., Короткин А.А., Лататуев А.Н., 2012. *Базы данных: учебное пособие*. Ярославль: Ярославский государственный университет имени П. Г. Демидова, 164 с.
10. Бураков П.В., Петров В.Ю., 2010. *Введение в системы баз данных*. Санкт-Петербург: СПбГУ ИТМО, 129 с.

11. Карпова И.П., 2020. *Базы данных*: Учебное пособие. Санкт-Петербург: Питер, 240 с.
12. Дейт К. Дж., 2005. *Введение в системы баз данных*. 8-е издание, пер. с англ. Москва: Издательский дом "Вильямс", 1328 с.
13. Михеева В.Д., Харитоновна И.А., 2004. *Microsoft® Access 2003*. Санкт-Петербург: БХВ-Петербург, 1072 с.
14. Свидетельство о государственной регистрации базы данных № 2023622239 Российская Федерация. *Данные наблюдений за процессом сдвижения на Северопесчанском месторождении хромитов за 2021–2022 гг.*: № 2023621896: заявл. 21.06.2023: опубл. 05.07.2023 / С.В. Усанов, С.А. Ногин, Ю.П. Коновалова [и др.]; заявитель Федеральное государственное бюджетное учреждение науки Институт горного дела Уральского отделения Российской академии наук. EDN KIBLDT.
15. *Данные наблюдений за процессом сдвижения на Главном Сарановском месторождении хромитов за период 1979–2021 гг.*: база данных: свидетельство о государственной регистрации № 2022622821: дата 10.11.2022 / Усанов С.В., Ногин С.А., Коновалова Ю.П., Панжин А.А., Харисова О.Д., Усанова А.В., Панжина Н.А., Прошин В.А., Владимиров Л.С.; ИГД УрО РАН. [https://www.fips.ru/registers-doc-view/fips\\_servlet?DB=DB&DocNumber=2022622821&TypeFile=html](https://www.fips.ru/registers-doc-view/fips_servlet?DB=DB&DocNumber=2022622821&TypeFile=html) (дата обращения: 02.08.2024 г.).

### References

1. Rebrin E. Yu., 1993. Predstavlenie i obrabotka dannykh marksheiderskikh s'emok dlya postroeniya planov gornykh работ na personal'nykh komp'yuterakh [Representation and processing of surveying data for building mining plans on personal computers]. *Izvestiya Ural'skogo gosudarstvennogo gornogo universiteta*, № 3, P. 110 – 116.
2. Veselov V.V., Syachinov A.N., Shirovok V.V., Onishchenko D.N., 2020. Programmnyi kompleks GeoMarks: programma dlya EVM . [Program complex GeoMarks: computer program]. № reg. 2020660055: data reg.:26.08.2020: data publ. 26.08.2020. *Byul. № 9. АО "АТОМ-ENERGOPРОЕКТ"*. URL:[https://www.elibrary.ru/download/elibrary\\_43965872\\_47766549.PDF](https://www.elibrary.ru/download/elibrary_43965872_47766549.PDF) (data obrashcheniya: 20.05.2024).
3. Cheskidov V.V., 2017. Sovremennye zadachi kompleksnogo monitoringa sostoyaniya gornotekhnicheskikh ob"ektov [Current tasks of complex monitoring of the state of mining objects]. *Sergeevskie chteniya. Vypusk 19. Geoekologicheskaya bezopasnost' razrabotki mestorozhdenii poleznykh iskopaemykh*. Moscow: RUDN, 590 p.
4. Cheskidov V.V., Manevich A.I., Lipina A.V., 2019. Poluchenie i analiz bol'shikh dannykh v praktike monitoringa sostoyaniya gornotekhnicheskikh sooruzhenii [Obtaining and analyzing big data in the practice of monitoring the state of mining structures]. *Gornaya promyshlennost'*, № 2 (144), P. 86 – 88.
5. Kuznetsov S.D., Poskonin A.V., 2013. Raspredelemnnye gorizonta'l'no masshtabiruemye resheniya dlya upravleniya dannymi [Distributed horizontally scalable solutions for data management]. *Trudy Instituta sistemnogo programmirovaniya RAN*, Vol. 24, P. 327– 358.
6. Wiese L., Waage T., Brenner M., 2019. CloudDBGuard: A Framework for encrypted data storage in NoSQL Wide Column Stores. *Data and Knowledge Engineering*, Vol. 126, P. 101732.
7. Chen J.-K., Lee W.-Z., 2020. The Transformation of Relational Database to Wide Column Store Database. *International Symposium on Computer, Consumer and Control (IS3C)*, Taichung City, Taiwan, 2020, pp. 384-386, doi: 10.1109/IS3C50286.2020.00105.
8. Voroiiskii F.S., 2006. *Informatika. Entsiklopedicheskii sistematizirovanniy slovar'-spravochnik: vvedenie v sovremennye informatsionnye i telekommunikatsionnye tekhnologii*

v terminakh i faktakh [Informatics. Encyclopedic systematized dictionary-reference: introduction to modern information and telecommunication technologies in terms and facts]. Moscow: Fizmatlit, 945 p.

9. Zafievskii A.V., Korotkin A.A., Latatuev A.N., 2012. Bazy dannykh: uchebnoe posobie [Databases: Study guide]. Yaroslavl': Yaroslavskii gosudarstvennyi universitet imeni P. G. Demidova, 164 p.

10. Burakov P.V., Petrov V.Yu., 2010. Vvedenie v sistemy baz dannykh [Introduction to database systems]. Sankt-Peterburg: SPbGU ITMO, 129 p.

11. Karpova I.P., 2020. Bazy dannykh: Uchebnoe posobie [Databases: Study guide]. Sankt-Peterburg: Piter, 240 p.

12. Deit K. Dzh., 2005. Vvedenie v sistemy baz dannykh [Introduction to database systems]. 8-e izdanie, per. s angl. Moscow: Izdatel'skii dom "Vil'yame", 1328 p.

13. Mikheeva V.D., Kharitonova I.A., 2004. Microsoft® Access 2003. Sankt-Peterburg: BKhV-Peterburg, 1072 p.

14. Svidetel'stvo o gosudarstvennoi registratsii bazy dannykh № 2023622239 Rossiiskaya Federatsiya. Dannye nablyudenii za protsessom sdvizheniya na Severopeschanskom mestorozhdenii khromitov za 2021–2022 gg. [Data of observations of the shear process at Severopeschansky chromite deposit for 2021-2022]: № 2023621896: zayavl. 21.06.2023: opubl. 05.07.2023 / S.V. Usanov, S.A. Nugin, Yu.P. Konovalova [i dr.]; zayavitel' Federal'noe gosudarstvennoe byudzhethoe uchrezhdenie nauki Institut gornogo dela Ural'skogo otdeleniya Rossiiskoi akademii nauk. 1. EDN KIBLDT.

15. Dannye nablyudenii za protsessom sdvizheniya na Glavnom Saranovskom mestorozhdenii khromitov za period 1979–2021 gg. [Data of observations of the shear process at the Main Saranovskoye chromite deposit for the period 1979 – 2021]: baza dannykh: svidetel'stvo o gosudarstvennoi registratsii № 2022622821: data 10.11.2022 / Usanov S.V., Nugin S.A., Konovalova Yu.P., Panzhin A.A., Kharisova O.D., Usanova A.V., Panzhina N.A., Proshin V.A., Vladimirov L.S.; IGD UrO RAN. [https://www.fips.ru/registers-doc-view/fips\\_servlet?DB= DB&DocNumber=2022622821&TypeFile=html](https://www.fips.ru/registers-doc-view/fips_servlet?DB= DB&DocNumber=2022622821&TypeFile=html) (data obrashcheniya: 02.08.2024 g.).