

УДК 681.518:378

В.Я. Воропаева¹ (канд. техн. наук, доц.), В.Ф. Шапо² (канд. техн. наук, доц.)1) ГВУЗ «Донецкий национальный технический университет», г. Донецк
кафедра автоматизации и телекоммуникаций2) Одесская национальная морская академия, г. Одесса
кафедра теории автоматического управления и вычислительной техникиE-mail: voropaeva@meta.ua, stani@te.net.ua**МЕТОД АНАЛИЗА ВЛИЯНИЯ РАБОТЫ ПРОГРАММНЫХ СРЕДСТВ БИЗНЕС-АНАЛИТИКИ НА ЗАГРУЗКУ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПРЕДПРИЯТИЯ**

Предложен метод расчета объема данных, передающихся на сервер бизнес-аналитики по корпоративной компьютерной сети и из территориально-распределенных подразделений предприятия. Проанализированы источники и порядок формирования данных для бизнес-анализа. Получены математические зависимости, позволяющие рассчитать время подготовки данных для бизнес-анализа, время передачи данных для бизнес-анализа по локальной корпоративной компьютерной сети и из территориально-распределенных подразделений предприятия по сети Интернет. Даны рекомендации по выбору пропускной способности сегментов локальной корпоративной компьютерной сети и внешних Интернет-каналов для решения задач бизнес-анализа и определения характеристик сервера бизнес-аналитики.

Ключевые слова: бизнес-аналитика, Интернет-канал, сервер, корпоративная компьютерная сеть, пропускная способность.

Общая постановка проблемы

В последние два десятилетия мировая экономика развивается в целом быстрыми темпами, несмотря на периодически повторяющиеся кризисные периоды. Возникают новые направления ведения бизнеса, а привычные – динамично меняются, приспосабливаясь к новым требованиям: более разнообразной и гибкой работе с заказчиками и поставщиками, транспортными компаниями и налоговыми органами и т.д. Все больше предприятий становятся транснациональными, имея множество офисов и производственных подразделений в различных странах мира.

Отвечая требованиям времени, появляется и ряд новых методик и принципов ведения бизнеса: комплексное планирование ресурсов предприятия, планирование производственных ресурсов, управление кадровыми ресурсами, внедрение систем документооборота, бизнес-аналитика и т.д. Их практическая реализация невозможна без применения современных информационных технологий. Десятки зарубежных и отечественных производителей создают программные комплексы разной направленности, сложности и стоимости, реализующие ведущие мировые разработки и подходы. Современные информационные системы (ИС) предприятий и организаций все шире используют программное обеспечение классов ERP (Enterprise Resource Planning, планирование ресурсов предприятия), MRP (Manufacturing Resources Planning, планирование ресурсов производства), CRM (Customer Relationships Management, управление взаимоотношениями с клиентами), BI (Business Intelligence, бизнес-аналитика), ECM (Enterprise/Electronic Content Management, электронное управление документооборотом предприятия) и расширяющийся спектр офисных программных продуктов.

© Воропаева В.Я., Шапо В.Ф., 2013

Важной составляющей анализа, планирования, оптимизации бизнеса является и моделирование бизнес-процессов, также реализованное в ряде программных разработок. В связи с этим весьма актуальными являются задачи анализа и расчета трафика в корпоративных компьютерных сетях (ККС), связанного с появлением множества новых программных комплексов, выбор оптимальной пропускной способности ККС в целом и ее отдельных сегментов, анализ загрузки и выбор производительности серверного и сетевого оборудования [1, 2].

Решение задач и результаты исследований

В данной работе предлагается метод анализа влияния программных средств бизнес-аналитики на загрузку ИС предприятия и расчета ряда характеристик при передаче данных.

В течение более 20 лет для построения ИС предприятий и хранения и обработки данных использовалось множество семейств СУБД различных производителей, версий и областей применения (от простейших однопользовательских до сложных сетевых): Dbase/FoxBase/FoxPro, Clarion, Clipper, Paradox, Microsoft Access, Open/Libre Office Base, Oracle, Sybase, Informix, DB2, 1С и множество других. По данным [3] в настоящее время имеется 153 типа форматов файлов баз данных и 553 типа форматов файлов данных, среди которых наиболее популярны форматы pdf, djvu, электронных таблиц xls/xlsx/ods, текстовых процессоров и редакторов doc/docx/odt/txt/rtf, интернет-приложений xml/html/html, баз данных dbf, mdb и ряд других. В итоге в ИС любого крупного предприятия накоплены огромные объемы разнородных данных, несовместимых по форматам файлов и применяемым СУБД. Их уничтожение может привести к остановке работы предприятия, в то же время имеется острая необходимость в установке новых программных продуктов, удовлетворяющих современным требованиям, замене устаревшего серверного парка. Кроме этого, анализ разнородных данных, находящихся в разных источниках и на разных носителях в ККС или ИС, весьма затруднен. Это может привести к отставанию в конкурентной борьбе и потере рыночных позиций предприятия.

Для решения проблемы анализа данных существует множество программных систем и вспомогательных средств бизнес-аналитики, некоторые из которых перечислены ниже: Crystal Reports (Seagate Software), Reports Developer (Oracle), InfoMaker (Sybase), Express Server, Express Objects и Express Analyser (Oracle), Seagate Info (Seagate Software), Hyperion Essbase (Hyperion Solutions), WebFOCUS Business Intelligence Suite (Information Builders), Platform for Enterprise Business Intelligence (Cognos), BusinessObjects 2000 (Business Objects), SQL Server Analysis Services (Microsoft), SAS System (SAS Institute), QlikView.

Использование программных средств бизнес-аналитики позволяет существенно ускорить и автоматизировать процесс анализа данных, но приводит к существенному увеличению нагрузки на ККС и серверную подсистему ИС предприятия, поэтому задача анализа объемов и типов данных, передаваемых в системах бизнес-аналитики, расчета и выбора характеристик аппаратной части (сетевого оборудования и серверов) весьма актуальна.

Кроме объёмов данных, постоянно возрастают и скоростные и емкостные характеристики отдельных дисковых накопителей и систем хранения данных. Появились и развиваются интерфейсы SAN (сеть хранения данных, Storage Area Network), NAS (сетевые хранилища данных, Network Attached Storage), SATA 3, Infiniband, SAS (Serial Attached SCSI, последовательная реализация интерфейса SCSI), Fibre Channel, iSCSI, AoE, твердотельные диски (Solid State Drive, SSD) и др. Происходит взаимное проникновение технологий и интерфейсов передачи данных в различные области с целью повышения гибкости и удобства их применения, при этом повышаются надежность, совместимость и скорости передачи данных. Так, в настоящее время в зависимости от года выпуска, производителя, области применения и модели устройства в SSD-накопителях среднее время доступа к данным составляет 0,1-0,3 мс, скорость чтения/записи/передачи данных – 260-550 Мбайт/с. В

жестких дисках среднее время доступа к данным составляет 6-20 мс, а скорость чтения/записи/передачи данных – 80-160 Мбайт/с. В среднем SSD имеют в 2-3 раза более высокую скорость чтения/записи/передачи данных и в 80-100 раз меньшее время доступа к данным [4, 5], но и в несколько раз большую по сравнению с жесткими дисками удельную стоимость хранения данных.

Однако даже высокие скорости работы различных интерфейсов и накопителей далеко не всегда позволяют получить комфортное для работы пользователя время реакции сервера на запросы к системе бизнес-аналитики при больших объемах обрабатываемых данных (до нескольких десятков Гбайт). Поэтому многие ведущие производители программных систем бизнес-аналитики [6, 7] используют технологии семейства in-memory, позволяющие размещать анализируемые данные непосредственно в оперативной памяти сервера. Это дает существенный прирост производительности и сокращение времени реакции сервера и системы бизнес-аналитики на запросы пользователей, но выдвигает гораздо более строгие требования к производительности подсистемы оперативной памяти и ее объему, который может составлять многие десятки или даже сотни Гбайт. Некоторые проблемы, связанные с выбором аппаратной конфигурации различных типов серверов, были рассмотрены в работе [8]. Неоптимальный выбор конфигурации серверов по соотношению цена/производительность может привести либо к серьезным неоправданным финансовым затратам, либо к получению недостаточной производительности для решения нужной задачи, поэтому решение данной проблемы также является весьма актуальным.

В связи с существованием в современных предприятиях территориально-распределенных подразделений возникает необходимость предоставления их пользователям доступа к ККС, в том числе и к системам бизнес-аналитики. Однако для оптимизации использования дорогостоящих лицензий на программное обеспечение вне зависимости от того, зависит их стоимость от числа пользователей, числа процессоров в сервере бизнес-аналитики или числа ядер в каждом из процессоров, оптимально использование единственного сервера с привязкой к нему всех лицензий и предоставление доступа к нему как из локальной ККС, так и из территориально-распределенных подразделений.

Для оценки требуемого объема оперативной памяти при использовании технологий семейства in-memory нужно определить число пользователей, использующих систему бизнес-аналитики; количество типов данных, которые использует каждый пользователь в бизнес-анализе; объем данных каждого типа. Тогда полный объем данных, переданных по локальной ККС и из удаленных подразделений, который необходимо разместить в оперативной памяти сервера для работы системы бизнес-аналитики, составит

$$V_{ud_l} = \sum_{j_l=1}^{u_l} \sum_{i_l=1}^{d_l} V_{i_l j_l}, \quad (1)$$

$$V_{ud_g} = \sum_{j_g=1}^{u_g} \sum_{i_g=1}^{d_g} V_{i_g j_g}, \quad (2)$$

$$V_{ud} = V_{ud_l} + V_{ud_g} = \sum_{j_l=1}^{u_l} \sum_{i_l=1}^{d_l} V_{i_l j_l} + \sum_{j_g=1}^{u_g} \sum_{i_g=1}^{d_g} V_{i_g j_g}, \quad (3)$$

где V_{ud} – полный объем данных, передаваемый всеми пользователями программной системы бизнес-аналитики; V_{ud_l} – объем данных, передаваемый пользователями системы программной бизнес-аналитики по локальной ККС; V_{ud_g} – объем данных, передаваемый пользователями программной системы бизнес-аналитики из удаленных подразделений по сети Интернет; u_l – число пользователей программной системы бизнес-аналитики, получающих к ней доступ

по локальной ККС; d_l – число типов данных, анализируемых каждым пользователем, получающих к ней доступ по локальной ККС; u_g – число пользователей программной системы бизнес-аналитики, находящихся в удаленных подразделениях и использующих Интернет для доступа к ККС; d_g – число типов данных, анализируемых каждым пользователем, находящимся в удаленных подразделениях и использующих Интернет для доступа к ККС.

Кроме данных V_{ud} , обрабатываемых на сервере программной системой бизнес-аналитики, некоторый объем оперативной памяти сервера V_{os} занимает операционная система (занимаемый объем зависит от ее версии), некоторый объем V_{bi} – самой программной системой бизнес-аналитики, некоторый объем V_c используется для операций кеширования данных, некоторый объем V_r должен резервироваться, в т.ч. и в связи с возможной утечкой памяти, также зависящей от версии операционной системы и исполняемого на сервере программного обеспечения. Тогда минимально необходимый объем оперативной памяти сервера V_n можно определить по формуле:

$$V_n = V_{ud} + V_{os} + V_{bi} + V_c + V_r. \tag{4}$$

Выбор объема оперативной памяти сервера бизнес-аналитики V_{sm} должен выполняться в соответствии с неравенствами:

$$\begin{cases} V_{sm} > V_n; \text{объем оперативной памяти сервера для решения данной задачи достаточен,} \\ V_{sm} \leq V_n; \text{объем оперативной памяти сервера для решения данной задачи недостаточен.} \end{cases} \tag{5}$$

Структурная схема взаимодействия отдельных компонентов ИС предприятия при использовании программной системы бизнес-аналитики представлена на рис. 1.

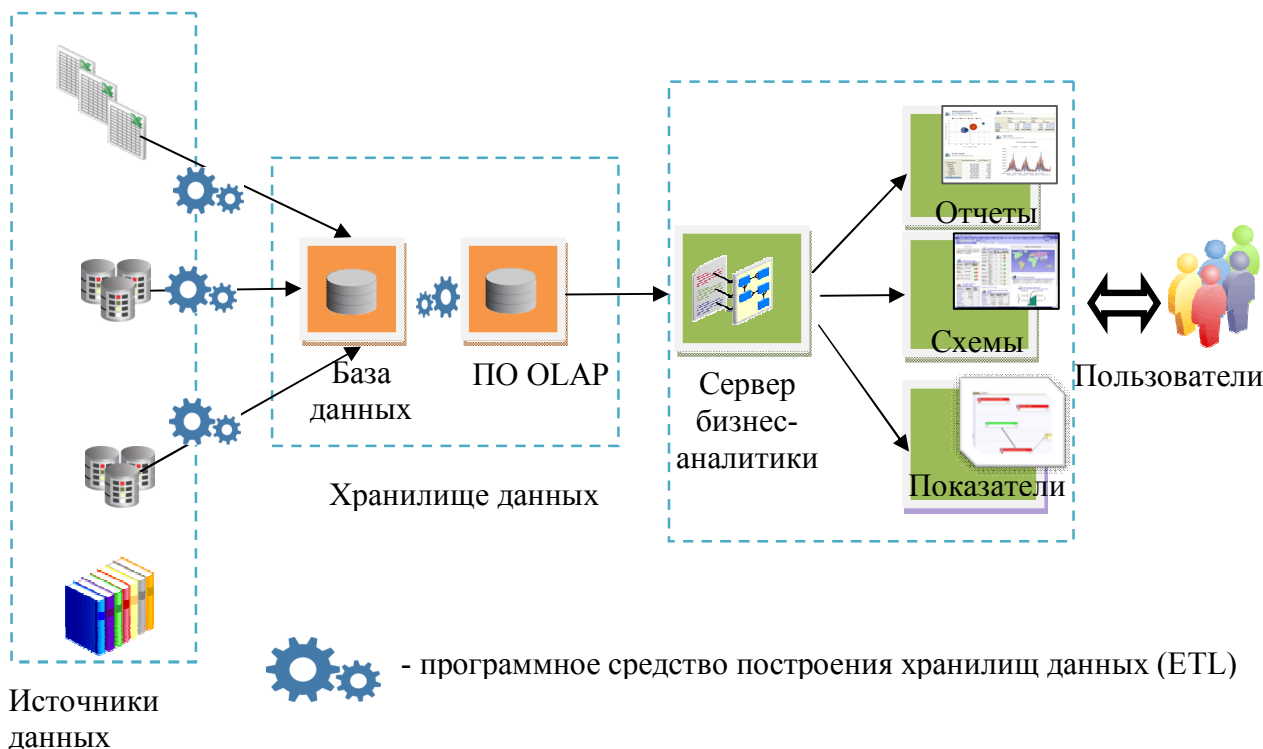


Рисунок 1 - Взаимодействие компонентов ИС предприятия при работе системы бизнес-аналитики

Для повышения эффективности обработки данных, поступающих из разных источников, уменьшения нагрузки на ККС предприятия и сервер бизнес-аналитики, для создания единой базы данных на предварительном этапе целесообразно выполнить стандартную процедуру ETL (извлечение, преобразование и загрузка данных, – Extracting, Transforming, Loading) — важный процесс в управлении хранилищами данных. В соответствии с названием выделим три его основных этапа:

1. Извлечение данных из различных внешних источников.
2. Преобразование данных к нужному формату и их очистка.
3. Загрузка данных в хранилище данных.

Выполнение процедуры ETL должно быть автоматизировано с помощью соответствующих программных средств, наиболее известные из которых перечислены далее: SAS Data Integration Server, SAP BusinessObjects Data Integrator, службы интеграции Microsoft SQL Server, IBM WebSphere DataStage, Informatica PowerCenter, Oracle Data Integrator, Pentaho, Talend Open Studio, Scriptella, JasperETL, Clover ETL и др.

На следующем этапе предварительно подготовленные данные должны обрабатываться в соответствии с одной из технологий семейства OLAP (OnLine Analytical Processing, аналитическая обработка в реальном времени), являющимися компонентами программных решений бизнес-аналитики и позволяющими подготовить итоговую информацию на основе больших массивов данных, структурированных по многомерному принципу. Эта процедура также автоматизируется путем использования соответствующих программных продуктов OLAP, наиболее популярные из которых перечислены ниже: Microsoft SQL Server и Analysis Services, Hyperion Essbase, Cognos PowerPlay, BusinessObjects, MicroStrategy, SAP BW, Cartesis Magnitude, Oracle Express, OLAP Option, Applix TM1, Mondrian, Palo и др.

Общее время T_{el1} , затрачиваемое в локальной ККС на подготовку и обработку данных при их последовательном сборе из разных источников, можно оценить следующей формулой:

$$T_{el1} = \sum_{i=1}^n T_i, \quad (6)$$

где T_i – время сбора (подготовки) данных от i -го источника данных.

Если подготовка и обработка данных из разных источников выполняется в локальной ККС параллельно (при этом возникает необходимость в приобретении дополнительного серверного оборудования и программных лицензий), то общее время T_f , затрачиваемое на подготовку и обработку данных, можно оценить следующей формулой:

$$T_{el1} = \max(T_i). \quad (7)$$

Общее время T_l , затрачиваемое на подготовку и обработку данных в системе бизнес-аналитики с использованием ККС предприятия, можно оценить следующей формулой:

$$T_l = T_{el1} + T_{db} + T_{el2} + T_{olap} + T_{bis} + T_{cn} + T_{osi}, \quad (8)$$

где T_{el1} – время, затраченное на первичную подготовку данных; T_{db} – время, затраченное на формирование базы данных; T_{el2} – время, затраченное на подготовку данных для формирования OLAP-хранилища; T_{olap} – время, затраченное непосредственно на формирование OLAP-хранилища; T_{bis} – время, затраченное при обработке данных на сервере бизнес-аналитики; T_{cn} – общее время, затраченное при передаче данных от OLAP-хранилища (хранилища данных) на сервер бизнес-аналитики по одному или нескольким сегментам ККС; T_{osi} – время, затраченное на формирование сервером бизнес-аналитики отчетов, схем и показателей бизнеса.

При обращении к серверу бизнес-аналитики из территориально-распределенных подразделений предприятия полное время обработки T_{g1} данных составит:

$$T_{g1} = T_f + T_l + T_b, \quad (9)$$

где T_f – время передачи данных из удаленного подразделения в центральное для обработки сервером бизнес-аналитики; T_b – время передачи результатов из центрального подразделения в удаленное после обработки сервером бизнес-аналитики.

Время подготовки данных в каждом удаленном подразделении может быть определено по формулам (6) и (7). При параллельной обработке данных в удалённых подразделениях также может возникнуть необходимость в приобретении дополнительного серверного оборудования и программных лицензий.

Величина T_f может быть определена следующим образом:

$$T_f = T_{sip} + T_i + T_{ipc}, \quad (10)$$

где T_{sip} – время передачи данных от удаленного подразделения до его Интернет-провайдера; T_i – время передачи данных по сети Интернет между провайдерами; T_{ipc} – время передачи данных от Интернет-провайдера, к которому подключено центральное подразделение с расположенным в нем сервером бизнес-аналитики.

Требуемая пропускная способность W_g внешнего канала связи, используемого для передачи данных из удалённых подразделений по сети Интернет на сервер бизнес-аналитики, составит

$$W_g = \frac{V_{udg}}{T_f}. \quad (11)$$

Значение величины W_g позволяет определить, какую часть пропускной способности внешнего Интернет-канала занимает передача данных для решения задачи бизнес-аналитики и насколько его пропускная способность в общем соответствует спектру решаемых с его помощью задач.

Требуемая пропускная способность W_l сегмента (сегментов) ККС, используемого для передачи данных от хранилища данных на сервер бизнес-аналитики, составит

$$W_l = \frac{V_{udl}}{T_f}. \quad (12)$$

Требуемые пропускные способности, рассчитанные по формулам (11) и (12), позволяют также определить дополнительную нагрузку на коммутаторы соответствующих сегментов ККС и на маршрутизатор, через который осуществляется связь с территориально-распределенными подразделениями.

Формулы (1)-(12) позволяют определить основные характеристики (объемы передаваемых данных, время передачи данных, пропускные способности сегментов сети и каналов связи), необходимые для оптимального выбора и эксплуатации сервера бизнес-аналитики и соответствующей программной платформы.

Выводы

1. Выполнен анализ факторов, влияющих на производительность и стоимость программно-аппаратных средств бизнес-аналитики. Проанализированы источники и порядок формирования данных для бизнес-анализа.

2. Получены математические зависимости, позволяющие рассчитать время подготовки данных для бизнес-анализа, время передачи данных для бизнес-анализа по

локальної корпоративної комп'ютерної мережі та з територіально-розподілених підрозділів підприємства по мережі Інтернет.

3. Дані рекомендації по вибору пропускної спроможності сегментів локальної корпоративної комп'ютерної мережі та зовнішніх Інтернет-каналів для рішення задач бізнес-аналізу та визначення характеристик сервера бізнес-аналітики.

4. Представлено метод розрахунку обсягу даних, передаваних на сервер бізнес-аналітики по корпоративній комп'ютерній мережі та з територіально-розподілених підрозділів підприємства, формалізуючий процедуру вибору пропускних спроможностей сегментів локальної корпоративної комп'ютерної мережі та зовнішніх каналів зв'язу з урахуванням рішення задач, пов'язаних з бізнес-аналізом.

Список использованной литературы

1. Anopriyenko A. Bandwidth usage maximization for enhancement of data exchange efficiency in TCP/IP-based networks / A. Anopriyenko, Ababneh Hasan, John Samuel Ndueso // Informatics, Cybernetics and Computer Science (ICCS-2007). Scientific Papers of Donetsk National Technical University. – 2007. – Volume 8 (120). – P. 331-339.
2. Воропаева В.Я. Оптимизация построения и эксплуатации программно аппаратных средств в информационных системах учебных заведений / В.Я. Воропаева, В.Ф. Шапо, В.Ю. Воловщиков // Наукові праці Донецького національного технічного університету. Серія: «Обчислювальна техніка та автоматизація». – 2010. – Вип. 18 (169). – С. 5-11.
3. Расширения и форматы файлов [Электронный ресурс]. – Режим доступа: <http://open-file.ru>.
4. Мировой рынок бизнес-аналитики глазами Gartner [Электронный ресурс]. – Режим доступа: <http://www.pcweek.ua/themes/detail.php?ID=141801>.
5. Инструменты конкурентной борьбы [Электронный ресурс]. – Режим доступа: http://www.pcweek.ua/themes/detail.php?ID=139162&sphrase_id=40544.
6. Компания Интел [Электронный ресурс]. – Режим доступа: <http://www.intel.ru>.
7. Международная сеть изданий Best of Media "Tom's Hardware Guide" [Электронный ресурс]. – Режим доступа: <http://www.thg.ru>.
8. Воропаева В.Я. Анализ требований к созданию серверной подсистемы при построении информационных систем предприятий / В.Я. Воропаева, В.Ф. Шапо // Наукові праці Донецького національного технічного університету. Серія: «Обчислювальна техніка та автоматизація». – 2009. – Вип. 148 (17). – С. 14 – 21.

Надійшла до редакції:
18.04.2013

Рецензент:
д-р техн. наук, проф. Скобцов Ю.О.

В. Я. Воропаева, В. Ф. Шапо

ДВНЗ «Донецький національний технічний університет»

Одеська національна морська академія

Метод аналізу впливу роботи програмних засобів бізнес-аналітики на завантаження інформаційної системи підприємства. Запропоновано метод розрахунку об'єму даних, що передаються на сервер бізнес-аналітики по корпоративній комп'ютерній мережі та з територіально-розподілених підрозділів підприємства. Проаналізовано джерела й порядок формування даних для бізнес-аналізу. Отримано математичні залежності, які дозволяють розрахувати час підготовки даних для бізнес-аналізу, час передавання даних для бізнес-аналізу по локальній корпоративній комп'ютерній мережі та з територіально-

розосереджених підрозділів підприємства по мережі Інтернет. Надано рекомендації по вибору пропускної здатності сегментів локальної корпоративної комп'ютерної мережі й зовнішніх Інтернет-каналів для рішення задач бізнес-аналізу й визначення характеристик сервера бізнес-аналітики.

Ключові слова: бізнес-аналітика, Інтернет-канал, сервер, корпоративна комп'ютерна мережа, пропускна здатність.

V. Y. Voropaeva, V. F. Shapo
Donetsk National Technical University
Odessa National Maritime Academy

A Method of Analyzing the Influence of Business-intelligence Software on the Enterprise's Information System Loading. *We analyze various data sources and file types, which have appeared as a result of information technologies development and have been formed by software applications of different generations, activity fields, architectures and developers. We also consider the drawbacks and problems connected with the use of these data and reducing the efficiency of user's work, computer networks and information systems of enterprises. These problems complicate the development of an enterprise and affect its information security, they may also lead to the growth of financial expences and time losses. We consider the possible ways of their solution and the procedure of generating and preparing data for business analysis. We analyze briefly business-intelligence software and show its role in providing an enterprise's efficiency. The introduction of this software implies some extra tasks connected with choosing the performance of server and communication equipment, its loading optimization and optimization of corporate computer network loading. Besides, we performed a brief analysis of current data transmission interfaces, data storage and technologies saving the time of data processing and transmission. The subtasks of the optimal choice of performance and prices of hardware and software licenses are classified. We provide a method of calculating the volumes of data of different kinds and their total volume, transmitted to the business-analytics server through the corporate computer network and from remote departments of the interprise via Internet. We analyzed software applications in the server's random-access memory and suggested a mathematical correlation for defining its capacity taking into account the number of users, the types and volumes of the data used by them, their work using local networks or connection with remote subdivisions. We analyzed the procedures of preparing data for business-intelligence software system and defined the sequence of fulfilling these procedures. We obtained mathematical dependences, which allow calculating the time of preparing various kinds of data for business analysis (when they are collected from different sources), the time of data transmission through local computer networks and Internet. The paper provides recommendations as for the choice of bandwidths of the segments of local computer networks and external Internet-channels depending upon the type and volume of transmitted data and that will help to solve the tasks of business-analysis.*

Keywords: *Business intelligence, Internet channel, server, corporate computer network, bandwidth.*