

DATA ARCHITECTURE AS A BASE FOR ENTERPRISE ARCHITECTURE

Vassil Milev, Stefan Kalchev

Abstract: The report presents the role, place and objectives of the data architecture in enterprise architecture. They are examined in the light of the information architecture and business process requirements. The roles of data architecture, basic principles, characteristics and development stages are described. It can serve as guidance material for IT departments which begin to implement projects in synchronization of business and IT [13, 14, 15, 18].

Keywords: Enterprise Architecture, Information System, Date Architecture, Architecture Framework, Zachman, TOGAF

АРХИТЕКТУРАТА НА ДАННИТЕ КАТО ФУНДАМЕНТ НА АРХИТЕКТУРАТА НА ПРЕДПРИЯТИЕ

Васил Милев, Стефан Калчев

Резюме: The report presents the role, place and objectives of the data architecture in enterprise architecture. They are examined in the light of the information architecture and business process requirements. The roles of data architecture, basic principles, characteristics and development stages are described. It can serve as guidance material for IT departments which begin to implement projects in synchronization of business and IT [13, 14, 15, 18].

Ключови думи: Enterprise Architecture, Information System, Date Architecture, Architecture Framework, Zachman, TOGAF

Доклада е публикуван в :



ВЕЛИКОТЪРНОВСКИ
УНИВЕРСИТЕТ "СВ. СВ. КИРИЛ
И МЕТОДИЙ"

ФАКУЛТЕТ "МАТЕМАТИКА И ИНФОРМАТИКА"



ЮБИЛЕЙНА МЕЖДУНАРОДНА НАУЧНА КОНФЕРЕНЦИЯ
50 ГОДИНИ ВТУ "СВ. СВ. КИРИЛ И МЕТОДИЙ"

ВЪВЕДЕНИЕ

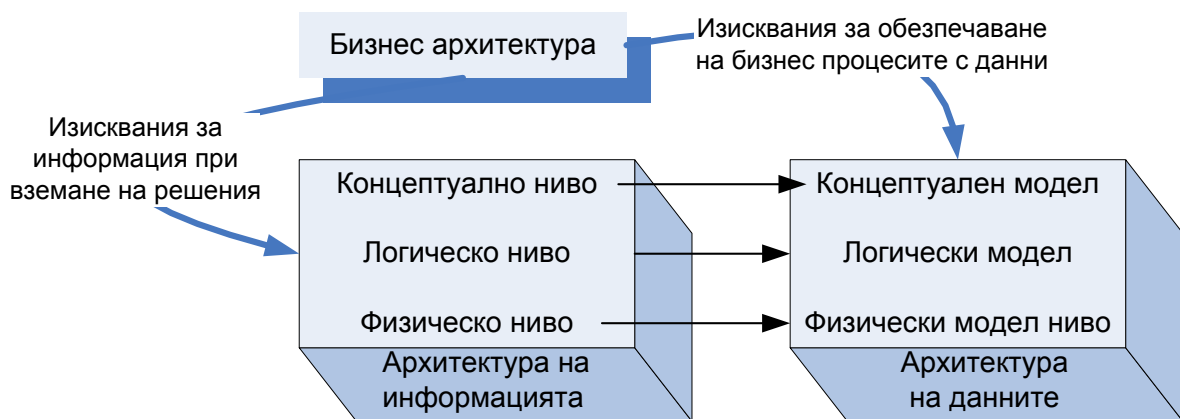
Архитектурата е фундаментална организация на системи, реализирана чрез нейните компоненти, връзките помежду им, и с обкръжаващата среда, както и принципите, определящи нейното проектиране и развитие [18, 19] .

Архитектура на предприятието е процес на моделиране и документиране на всички аспекти на организацията, за да се гарантира, че услуги, процеси, приложения, информация, данни, технологии, места, хора, събития и срокове са съобразени с целите и задачите. Тя осигурява синхронизация на развитието на бизнеса и информационните технологии (ИТ), като включва четири взаимнозависими архитектури [4, 13, 18]:

- **архитектура на бизнеса** – описва, как работи бизнеса от гледна точка на целите, стратегията, бизнес средата и е насочена към подобряване на производителността му. Основни компоненти се явяват бизнес процеси, организация и заинтересовани страни;
- **архитектура на данните** – тя е фундамента, върху който се изграждат останалите архитектури. Описва структурата на данните, източниците им, ролята при въвеждане, обработка и съхранение;
- **архитектура на приложенията** – характеризира информационните системи, които се използват за управление на данните и поддръжката на бизнес процесите, техния състав и структура, както и предоставяните от тях услуги;
- **архитектура на инфраструктурата** – включва средствата и платформите, върху които се разполагат останалите архитектури. Осигурява надеждни ИТ услуги, предоставяни в

рамките на цялата организация и координирани централизирано.

Настоящият доклад има за цел да представи ролята, мястото и задачите на архитектурата на данни в архитектурата на предприятието, като ги разгледа през призмата на архитектурата на информацията и изискванията на бизнес процесите (фиг.1) [18].



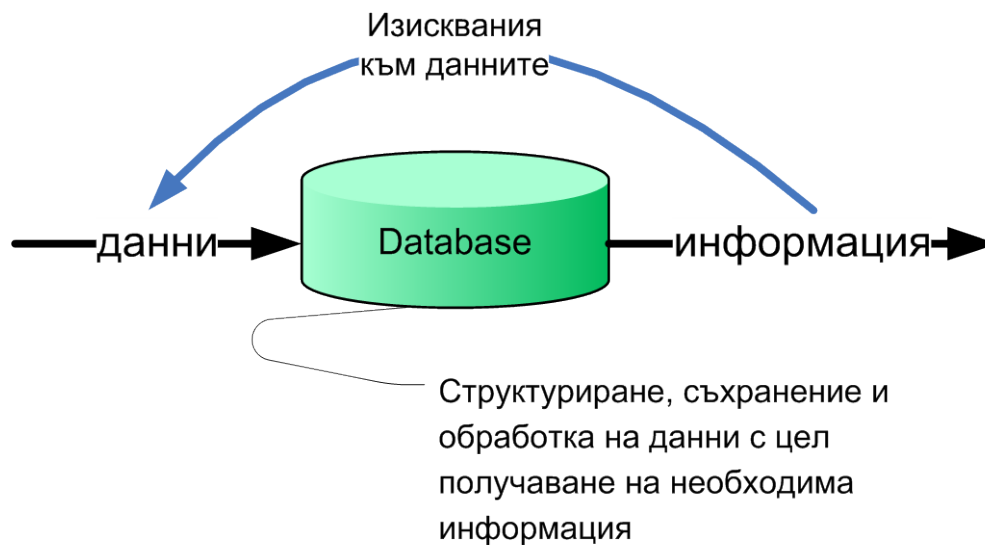
Фиг. 1. Архитектура на информацията и изискванията на бизнес процесите.

АРХИТЕКТУРА НА ИНФОРМАЦИЯТА

Архитектурата на информацията като част от системната архитектура не включва създаването на бази данни или моделите на всички данни използвани в организацията. По същество става дума за модели за по-общо описание на информацията, необходима на бизнеса, както и за политики и правила за работа с нея.

Архитектура на информацията е процес на организация и представяне на значима за потребителите информация в

интуитивно разбираема форма, с използване на съответни средства за систематизация.



Фиг. 2. Архитектура на информацията.

Тя включва визия, принципи, модели и стандарти, които осигуряват процеса по създаване, използване и поддържане на информация, необходима за реализиране дейността на организацията. Архитектурата на информацията описва начините за:

- бързо и ефективно вземане на решения, чрез използване на ИТ;
- разпространение на информация, вътре и извън организацията;
- технологиите и методите за работа с различни типове информация:
 - структурирана информация (релационни и обектни модели);
 - полуструктурирана информация (XML стандарти);

- неструктурирана информация във форма на текст, графика, изображения, съпроводена от описателни данни (метаданни и каталози).

Модели на информацията

Цел на разработката на модели на информацията се явява графичното представяне на потребностите на организацията и отделните бизнес процеси от информация. Архитектурните модели на информацията описват:

- процесите по обработка на информацията;
- основните информационни обекти, свързани с бизнес събитията;
- информационните потоци;
- принципите за управление на информацията.

От процеса на разработка на архитектурата на информация могат да се получат следните резултати:

- документирано описание на съществуващите източници на данни;
- модели на данните, описващи тяхното използване от бизнес потребителите;
- описание на съществуващите и планирани информационни потоци, включващо съответните интерфейси, алгоритми на преобразуване на данните, както и услуги свързани с предаване на данни;
- описание на решенията по управлението на съхранение на данните – от каталози до хранилища на данни;

- използвани технологии и средства за преобразуване и управление на данните.

Нива на абстракция при архитектурата на информация [1,18]

На концептуално ниво

Разработват се модели от високо ниво, описващи информационните потоци между функционалните подразделения в най-общ вид. Тези потоци се разглеждат на бизнес ниво без да се уточнява тяхната практическа реализация.

На логическо ниво

Моделите на информация и данни описват изискванията към информацията във форма и термини, разбираеми от бизнес потребителите. На това ниво се осигуряват средства за:

- откриване, анализ, определяне, стандартизация и нормализация на отношенията между бизнес процесите и приложните системи;
- идентификация на потоците от информация и съответстващите елементи на данните, които са необходими на организацията.
- констатиране на логическите факти, касаещи процесите, информационните потоци и елементи на данните, които организацията е длъжна да поддържа за изпълнение на бизнес операциите;

На физическо ниво

Описват се конкретните обработки на данните, за получаване на информация при вземане на решения.

АРХИТЕКТУРА НА ДАННИТЕ

В информационните технологии, архитектурата на данни се състои от модели, политики, правила и стандарти, които регламентират кои данни се събират и как се съхраняват, подредени, интегрирани и готови за използване в информационните системи и в организациите. Може да се направи сравнение с дизайна на къща, където всички описания на структурата на къщата (от избора на материали, размери и стил на стаите и покрива) са структурирани в план [5, 6, 7].

Архитектурата на данните се явява базов компонент на архитектурата на предприятието. Тя е основата, върху която се изграждат останалите архитектури. При разработката трябва да бъде даден отговор на следните въпроси [3, 9, 10] :

- Каква е структурата на данните?
- За какви бизнес процеси са предназначени данните?
- Кой е собственик на данните?
- Кой отговаря за данните?
- Кои са потребителите на данните?
- Какви обеми заемат данните и колко бързо нарастват?

Решавани задачи от архитектурата на данните

В хода на разработка на архитектурата на данните се решават следните задачи:

- идентификация и инвентаризация на съществуващите данни, включително определяне на източниците им, процедурите за редактиране и използване, отговорност, оценка на качеството;
- съкращаване на излишеството и фрагментарността на данните с цел намаляване стойността на обслужване и повишаване на качеството на данните за сметка на

изключване на нееднозначността и противоречивостта на различни екземпляри;

- изключване на ненужните премествания и копираня на данните, свързани с наличието на наследени и остарели приложения;
- осигуряване достъпност до данните в режим близък до използване в реално време;
- интеграция на метаданните с цел цялостно представяне на данните от различни източници;
- реализиране на мерки за осигуряване на защитата на данните.

Необходимост от използване на архитектурата на данни [11, 12]

- избягване на дублиране на данни;
- систематизиране на готови архитектурни решения;
- многократно използване на проверени в практиката архитектурни решения;
- уеднаквяване на терминологията, използвана при моделиране на данните;
- добро средство за обучение в моделиране на данни.

Характеристики на архитектурата на данни

- разбираема на всички нива – всички заинтересовани лица трябва да могат да разберат представените с архитектурата на данни процеси и модели;
- гъвкава – архитектурата на данни трябва точно и стабилно да отразява целта, за която е предназначена, както и да бъде достатъчно гъвкава, за да позволява реализиране на

бъдещи допълнителни изисквания без съществени модификации;

- базирана на стандарти – трябва да са спазени всички приложими стандарти, независимо от това как се е появил на стандарта;
- за многократна употреба – архитектурата на данните трябва да позволява многократна употреба в подобни приложения със същия мащаб и функционалност.

Етапи на разработка на архитектурата на данни (модел TOGAF) [14, 16]:

- разработка на описанието на базовата архитектура на данните;
- преглеждане и проверка на принципите, еталонните модели, гледни точки и средства;
- създаване на архитектурни модели, в това число концептуални и логически модели на данните, модели на управление на данните и модели на отношения, в които бизнес функциите се съпоставят с операциите над данните (създаване, четене, обновяване и премахване);
- избор на компонентите на архитектурата на данни;
- формален анализ на контролните точки на модела на архитектурата и неговите компоненти, заедно със заинтересованите лица.
- анализ на качествените критерии (производителност, надеждност, безопасност и цялостност);
- завършване на архитектурата на данните;
- анализ на контролните точки и последствията;
- анализ на различията.

Базови архитектурни принципи на архитектурата на данните

Избягване дублиране на данните – един път съхранени, данните могат да се използват многократно в едно или няколко приложения.

Поддържане съответствие на данните – една естествена последица от премахването на дублирането е, че се намалява възможността от несъответствие при актуализиране на данните.

Независимост на данните и програмите – промените в структурата на данните или в програмните приложения не са взаимозависими.

Логически аспект на данните за потребителите и потребителските програми – представянето на данни е независимо от начина на физическото им съхраняване. Възможни са различни логически представяния на едни и същи данни за различни потребители и програми.

Обслужване на широк кръг приложения – веднъж запомнени, данните могат да се обработва по различни начини и с произволна желана форма, което разширява областта на възможните приложения.

Използване на правила и стандарти – фактът, че достъпа до данните става само чрез системата за управление на бази данни (СУБД), позволява на администратора на базата данни да изисква и следи за изпълнението на определени стандарти при представянето на данните от приложенията.

Сигурност на данните – администраторът на базата данни (БД) контролира достъпа до нея. Той може да следи за задаването на кодове за достъп на потребителите, ограничаващи ги само до онези части и функции от базата от данни, които са им разрешени.

Моделиране на архитектурата на данни

През 1975 ANSI [19] описва три вида модели на данните:

Концептуална схема – процес на изграждане модела на данните, свързан с дейността на дадена организация, който е независим от физическите фактори. Описва понятия от предметната област (домейн), техните взаимовръзки, както и налаганите ограничения на данните. Предназначението на концептуалния модел на данните е:

- детайлно изясняване на структурата, семантиката, връзките и ограниченията в БД, независимо от конкретната СУБД, която ще се използва за реализирането ѝ;
- статично описание на съдържанието на БД;
- изяснено влияние върху приложенията свързани с БД;
- улеснено общуване между участниците в проекта, благодарение на използваното диаграмно описание.

Логическа схема – описва данните чрез средствата на конкретната СУБД в два етапа:

- трансформиране концептуалната схема в схема на релационната БД.
- отразяване специфичните особености на избраната за реализация СУБД върху концептуалния модел.

Физическата схема – описва физически средства, използвани за съхраняване на данни. Процес на избор на специфични структури за съхранение и пътища за достъп до файловете, изграждащи БД. Това осигурява ефективността на работещите с БД приложения. Всяка СУБД предлага различни файлови организации и пътища за

достъп: индексирание, клъстериране на свързани записи чрез указатели, хеширане и т.н.

Според ANSI, този подход позволява три гледни точки да бъдат относително независими една от друга. Технологията за съхранение на данни може да се промени, без да се засяга логическата или концептуалната схема.

В следващата таблица могат да се видят ключовите разлики между схемите за моделиране на данни.

Концептуален модел на данните	Логически модел на данните	Физически модел на данните
Включва високо ниво на конструкции на данни	Включва обекти (таблици), атрибути (колони/полета) и взаимоотношения (ключове)	Включва таблици, колони, ключове, типове данни, правила за проверка, тригери, съхранени процедури, домейни и ограничения за достъп
Нетехнически имена, директори и мениджъри на всички нива, така че може да разбере БД на архитектурното описание	Използва бизнес имена за лица и атрибути	Използва по-определено и по-малко родово конкретни имена за таблици и колони, като съкратените имена на колони, които са ограничени от СУБД и всяка компания определени стандарти
Използва общи от високо ниво данни и конструкции, като архитектурните описания са създадени в технически термини	Независима е от технологията (платформа, СУБД)	Включва първични ключове и показатели за бърз достъп до данните.
Не може да се нормализира	Нормализира се до четвърта нормална форма(4NF)	Може да бъде де-нормализирани. Не отговарят на изискванията, въз основа на

		естеството на БД Ако естеството на БД е онлайн обработка на транзакции (OLTP) или оперативни данни се съхранява (OPB) обикновено не е де-нормализирани. Де-нормализирането е обща в Datawarehouses.
--	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Табл. 1. Ключови разлики между схемите за моделиране на данни.

Нива на моделиране на данните, според модела на Захман

Нива	Модел	Значение	Професия	Данни
Първо	Бизнес модел	Основни понятия определящи бизнеса	Плановик	Списък важни понятия и обекти
Второ	Концептуален модел	Определяне в термини на бизнеса структурата на организацията, ключовите и спомагателни бизнес процеси .	Собственик	Концептуален модел данни
Трето	Логически модел	Бизнес процесите се описват в термините на информационните системи . Включват се различни типове данни, правилата за преобразуване и обработката им с цел изпълнение на бизнес функциите описани на горното ниво	Архитект	Логически модел данни
Четвърто	Технологичен модел	Осъществява се привързване на данните и операциите над тях с избраната технология на реализация - СУБД , средства за работа с неструктурирани данни и т.н.	Проектант	Физически модел данните
Пето	Детайлна реализация	Описват се конкретните модели оборудване, топология на мрежата , тип и версия на СУБД, програмен код и т.н. Много от дейностите на това ниво се изпълняват от външни изпълнители.	Разработчик	Описание структура данни
Шесто	Работеща система	Представят се инструкции за работа със системата, фактически бази данни , предлагани услуги за управление на ИТ и т.н.	Сервиз	Данни

На фигура 3 е представен модел на Захман [18].

Фиг.3. Нива на моделиране на данните, според модела на Захман.

Характеристики за качеството на архитектурата на данните [15]

Концептуална коректност – предполага, че архитектурата на данните точно отразява бизнес обектите, представляващи интерес

за предприятието. Постигането на концептуалната коректност зависи от способността да се представи структурирана информация за бизнес средата на семантичен език, който дава смислено и точно представяне на реалния свят.

Концептуална пълнота – предполага, че модела на данните съдържа обекти, напълно покриващи обхвата на бизнес домейна, който модел цели да представи. Представата за качеството на модела на данни е тясно свързана с външни фактори, като например държавни и правни мандати, финансови ограничения и заинтересованите изисквания.

Техническа коректност – означава, че обектите, които се съдържат в модел на данните не нарушават установените правила на синтаксиса на даден език. Синтактичната коректност означава, че включените в модела на данните фигури, линии и символи се използват по предназначение и че модела се придържа към общоприети практики на избраната методология.

Техническа пълнота – означава, че всички необходими за създаване на модела на данните, компоненти, елементи и детайли са представени на съответните нива на детайлност.

Интеграция – предполага, че модела на данните е балансиран с другите елементи на архитектурата на предприятието. Архитектурата на данните е свързана и синхронизирана с бизнеса, услугите и техническите компоненти на архитектурата на предприятието.

ЗАКЛЮЧЕНИЕ

Представеният доклад дава най-обща представа за същността, използваните принципи и характеристиките на една много обсъждана в ИТ средите тема, каквато е архитектурата на данните.

Направени са изводи за фундаменталната роля и мястото на архитектурата на данни в архитектурата на предприятието. Докладът може да служи като ръководен материал за ИТ отдели, които започват да реализират проекти в областта на синхронизацията на бизнеса и информационните технологии.

ЛИТЕРАТУРА

1. Милев. В, Многомерност на архитектурното описание, Научна конференция с международно участие “Предизвикателства пред висшето образование и научните изследвания в условията на криза”, Бургас 2010.
2. В. Милев, Клачев С., Стефанова М., Архитектурния подход, основа за синхронизация на бизнеса и информационните технологии, Научна конференция с международно участие “ Mathtech2010”- Шумен 2010
3. Codd, E.F., The Relational Model for Database Management. Addison Wesley Publishing Company. ,1990
4. Dalin, A.B., A.I., Sjularenko , Enterprise Architecture, <http://www.intuit.ru/department/itmngt/entarc/0/>.,2007
5. Fowler, M. Foemmel, E.Hieatt, R.Mrr, R.Stafford , Patterns of Enterprise Application Architecture, Addison Wesley., 2002
6. Ganesan E., R. Paturi Building Blocks for Enterprise Business Architecture, Labs Briefings, <http://www.infosys.com/offerings/it-services/architecture-services/white-papers/documents/enterprise-business-architecture.pdf>. 2008
7. Goldfedder, Joy of Patterns , Using Patterns for Enterprise Development, Addison Wesley.,2001
8. H.Tujarov, Kalchev S., Milev V. Methods for creation of Enterprise Architecture, International Conference on Information Technologies (InfoTech-2011), 15-16 September 2011 Bulgaria, 2011.
9. H. Tujarov, Kalchev S., Milev V. Using the „Building Blocks" for modeling data, ISK'2010, Varna,2010
10. H.Tujarov, Kalchev S., Milev V., Component model of the enterprise architecture, Proceedings of the International Conference on Information Technologies (InfoTech-2010) 16-18 September 2010, Bulgaria.
11. Lewis, G., Comella-Dorda, S., Place, P., Plakosh, D., & Seacord, R., Enterprise Information System Data Architecture Guide Carnegie Mellon University.,2001
12. Milev V. The “Building Blocks” Concept Using for Modeling Data, International Conference on Information Technologies (InfoTech-2011), 15-16 September 2011 Bulgaria, 2011.
13. Minoli, Enterprise Architecture A to Z, Taylor & Francis Group, LLC, Op't Land, E.Poper. M. Waage. J.Cloo, Enterprise architecture ,Springer,2008
14. Raynard Boyce, TOGAF The Open Group Architecture Framework 100 Success Secrets: 100 Most Asked Questions –

- The Missing TOGAF Guide on How to Achieve and Sustain Superior Enterprise Architecture Execution, 2008
15. Schekkerman, How To Survive In The Jungle Of-Enterprise Architecture Frameworks, second edition Trafford, 2004
 16. Temnenko V., To be or not to be TOGAF: distribution of Enterprise Architecture abroad, 2007
 17. Zachman A. John ,The zachman framework for enterprise architecture: A primer for enterprise engineering and manufacturing,
http://zachmaninternational.com/2/Zachman_Framework.asp,2003
 18. Тужаров,Хр., Ст. Калчев, В. Милев, Архитектура на предприятие, <http://www.tuj.asenevtsi.com/EA/EA33.htm>,2010
 19. <http://standards.ieee.org/findstds/standard/1471-2000.html>