

Agile BI – The Future of BI

Mihaela MUNTEAN, Traian SURCEL
 Department of Economic Informatics and Cybernetics
 Academy of Economic Studies, Bucharest, Romania
 munteanm@ie.ase.ro, tsurcel@ase.ro

In a rapidly changing economy, Business Intelligence solutions have to become more agile. This paper attempts to discuss some questions which help in creating an agile BI solution such as: What is Agile? Why agile is so well suited for BI? Which are the key elements that promote an agile BI solution? Also, this paper briefly looks at technologies that can be used for enabling an agile BI solution.

Keywords: *Agile Business Intelligence, Agile Business Analytics, Agile Development Methodologies, In-Memory Bi Approaches, Data Virtualization Server*

1 Introduction

Business Intelligence (BI) was defined in different ways. The Data-Warehousing Institute has defined Business Intelligence as “the tools, technologies and processes required to turn data into information and information into knowledge and plans that optimize business actions” [6]. Turban has defined BI as “a broad category of applications and techniques for gathering, storing, analyzing and providing access to data to help enterprise user make better business and strategic decisions.” [23]. The range of capabilities that can be defined as business intelligence is very broad. The

spectrum of BI technologies is presented in [16].

Most enterprises have hundreds of internal and external data sources such as: databases, e-mail archives, file systems, spreadsheets, digital images, audio files and more. 80% of the organizational data are unstructured and semi-structured data.

Traditional Business Intelligence systems use a small fraction of all the data available. Also, traditional BI systems use only structured data. The core components of a traditional BI architecture are: ETL tools, an enterprise data warehouse with metadata repository and business analytics (Figure 1).

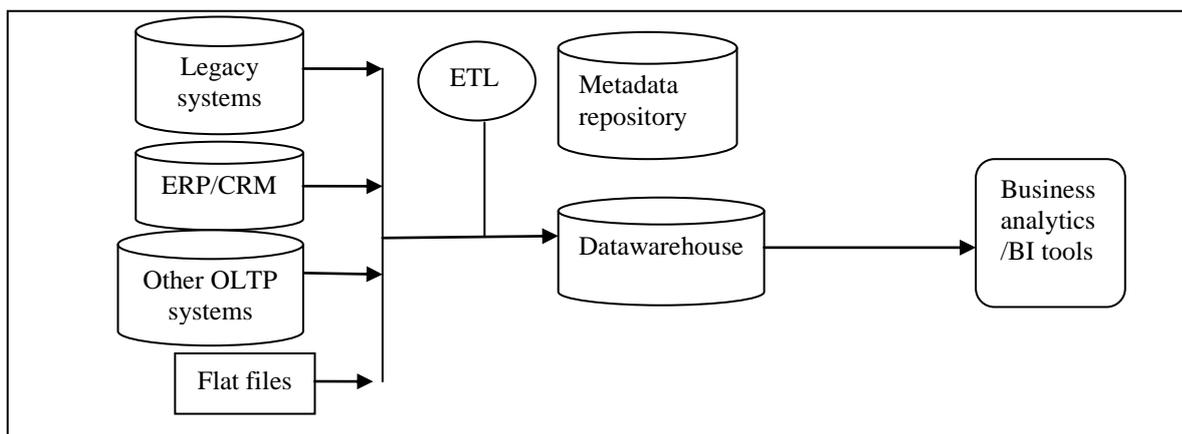


Fig. 1. A traditional BI architecture

Traditional BI systems use ETL tools for extracting data from multiple sources and temporarily storing those datasets at a staging area. Organizations use data warehouses to aggregate cleaned and structured data.

Business analytics/BI tools include enterprise reporting tools, ad hoc query tools, statistical analysis tools, OLAP tools, spatial-OLAP analysis tools, dashboards, scorecards and advanced analytics. Advanced analytics

typically refer to: data mining tools, text mining tools, predictive analytics, artificial intelligence, and natural language processing. But this architecture is unable to get adapted to change. A study by Aberdeen Group [24] showed that *“this style of BI is predominantly controlled, driven and delivered by corporate IT. Often, only static views of data are available and any changes or enhancements must be made by the IT organization”*. These

characteristics are in contradiction with frequently changing business requirements and *“big data”*. *Big data* typically refers to the following types of data: semi-structured data (XML and similar standards), unstructured data, Web data (social data, Web logs) and real-time data (event data, spatial data, machine-generated data). Table 1 briefly summarizes the main disadvantages of traditional BI systems.

Table 1. The main disadvantages of traditional BI systems

Disadvantages	Problems
huge amount of duplicate data	- every change already made requires an extra change of duplicate data - data inconsistencies - data quality risks
use different tools for different tasks	- non-shared metadata specifications - inconsistent results
rigid relational or multidimensional data models	- limited flexibility to changing - limited support for analysis on unstructured and external data
waterfall approach	- the long development lifecycle and less visibility to user - users are not involved in the development cycles - inflexible to analytical requirements modifications - testing at the end of the development cycle

How to eliminate these problems? By building an agile BI solution. A study by TDWI Research [22] showed that many traditional business intelligence systems are not agile:

- *“33% of the organizations needed more than three months to add a new data source to an existing business intelligence system”*
- *“developing a complex report or dashboard with about 20 dimensions, 12 measures and 6 user access rules took on average 7 weeks in 2011”*

The next section presents briefly the concept of agile BI and the key elements that together promote an agile BI solution.

2 Agile BI

Agile means the ability to be adaptable. Agile BI was defined in different ways. The Forrester Research defines agile BI as *“an approach that combines processes, methodologies, organizational structure, tools and technologies that enable strategic,*

tactical and operational decision –makers to be more flexible and more responsive to the fast pace of changes to business and regulatory requirements” [9]. According to Data Warehousing Institute agile BI *“addresses a broad need to enable flexibility by accelerating the time it takes to deliver value with BI projects. It can include technology deployment options such as self-service BI, cloud-based BI, and data discovery dashboards that allow users to begin working with data more rapidly and adjust to changing needs.”* [tdwi.org/portals/agile-bi.aspx].

In conclusion, an agile BI solution should provide access to accurate information in the right format to the right person at the right time. Below it identifies the key components that together promote an agile BI solution (Figure 2):

1. Agile development;
2. Agile business analytics;
3. Agile information infrastructure.

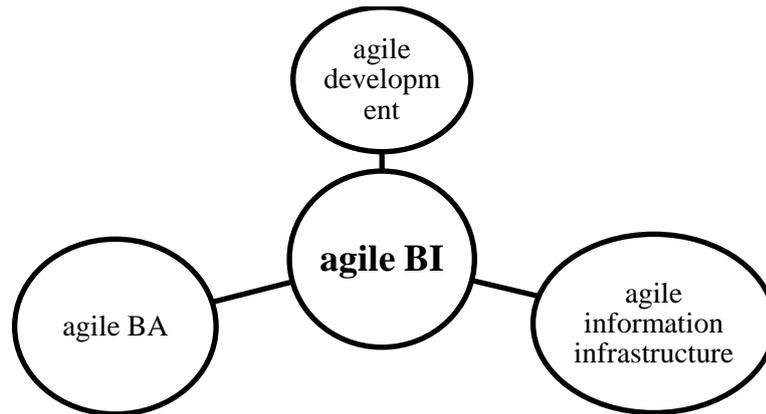


Fig. 2. The key components that promote an agile BI solution

2.1 Agile Development

An agile BI solution must be implemented quickly. According to Forrester Research the purpose of agile BI solution: “*is to 1) get the development done faster and 2) react more quickly to changing business requirements*” [8].

Two distinct approaches are relevant in the context of development of BI solutions: waterfall development and agile development. A waterfall approach is illustrated in Figure 3 (adapted from [7], [20]).

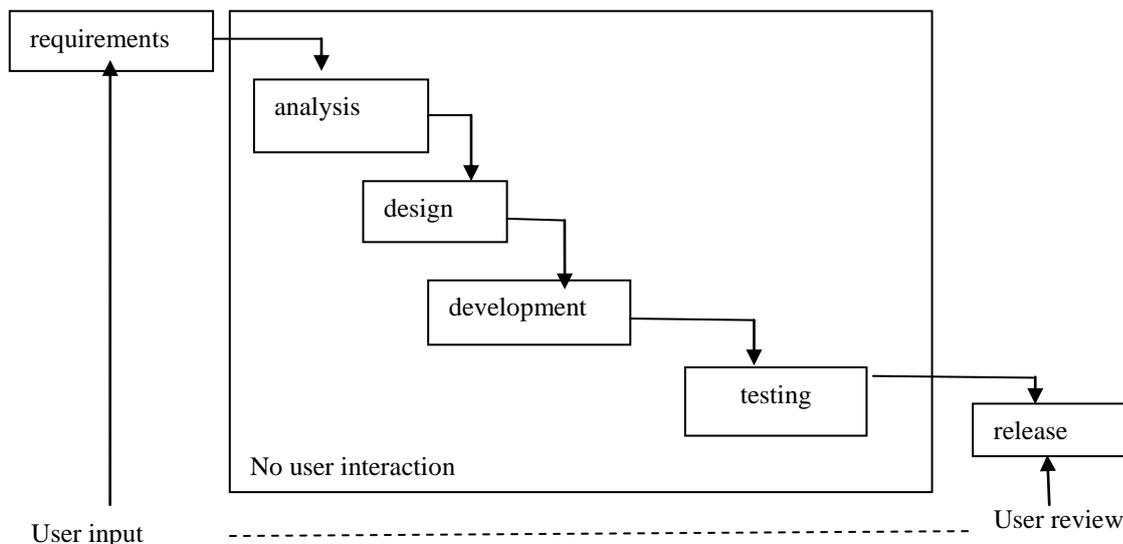


Fig. 3. Waterfall approach to the development of BI solution

But waterfall approach is poorly suited for BI. The main problems of this approach are:

- the long times between the system request and the delivery of the BI solution;
- users are not involved in the analysis phase, design phase, development phase and testing phase;
- it is inflexible to analytical requirements changes;
- testing at the end of the development life cycle.

So a different approach is needed to make BI applications more flexible and able to react faster to changing business requirements. The way to achieve agility in BI development is the usage of agile development methodologies. Agile development methodologies refer to a group of software development methodologies based on the following characteristics: collaboration between cross functional teams, iterative development and tolerance for changes [1], [3], [4]. There are different agile

development methodologies such as: Scrum, Extreme Programming, Crystal, Dynamic Systems Development, Lean and others. Many of agile software development characteristics can be applied to BI projects from team structure, project management, BI system design, BI system development and analytical techniques [5]. The most popular agile development methodologies for BI projects are: Scrum, Extreme Scoping and Agile Data Warehousing.

The main concepts of Scrum [21] are: *user story*, *product backlog*, *sprint backlog*, *sprint* and *daily scrum*. In this methodology the BI requirements are divided into small “*user stories*”. An agile BI project consists of a collection of “*user stories*”. Each story is then designed, developed, tested and released to the users. One sprint is a full life cycle of understanding the BI requirements, analysis, design, development and user testing. Each sprint lasts for 1-2 weeks. Users are involved in sprint steps. User stories need to be categorized in one of two ways: product backlog and sprint backlog. Sprint backlog is a list of tasks the team expects to do during

the sprint. At the end of each sprint, the business has a deliverable such as a new report or dashboard. Product backlog is a list of all requirements ordered by highest priority of what is needed. The user is responsible for ranking the features on the product backlog. Another concept is the daily scrum that is a short meeting in which every member of the team answers three questions: What did you do yesterday? What will you do until our next meeting? Do you have any problems?

Extreme Scoping and Agile Data warehousing are well suited if BI solution includes a data warehouse. Agile Data warehousing is defined as “*the application of two agile development approaches –Scrum and Extreme programming – to the specific challenges of data warehousing and BI*” [11]. Extreme scoping is an agile enterprise data warehousing approach that includes the business integration activities. Also, this approach uses agile principles. The BI solution is separated into multiple releases for iterative development [17].

Figure 4 shows a typical agile BI cycle.

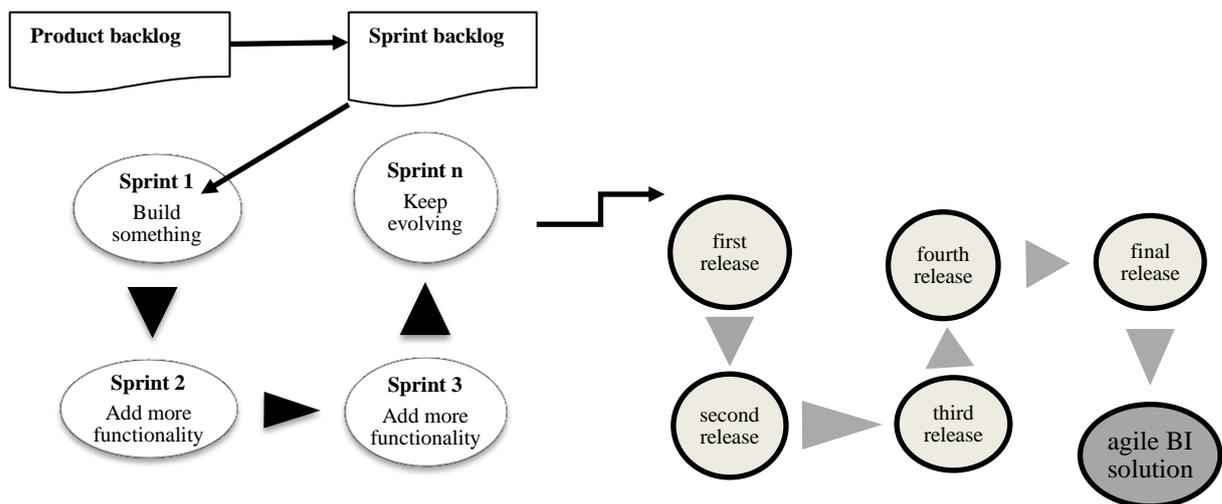


Fig. 4. A typical agile BI cycle

2.2 Agile Business Analytics (BA)

Besides developing a business intelligence system with agile design methodologies, it’s also recommended adopting agile BA [2]. Both the design methodology and the tools have to be agile. Agile BA must enable BI users to become less dependent on IT. Also,

agile BA must be easier to be used by all types of users. So agile BA should provide at least office suite integration, a business glossary and advanced visual features such as interactive dashboards and drill-down capabilities. A recent study by Forrester showed that agile BA should be integrated

with all parts of the “*information workplace*” such as: spreadsheet, presentation, word processing software, email, search portals, collaboration platforms and social communities [10]. There are several technologies to support agile BA such as: SaaS (Software-as-a-Solution) BI and in-memory BI technology. SaaS BI includes:

- packaged software-as-service BI applications that can be deployed in a cloud environment;
- SaaS BI tools that can be used to develop BI applications for deployment in a cloud computing;
- on-premises environment and data warehousing in the cloud.

The primary goal of the in-memory BI technology is to eliminate traditional disk-based BI solutions which are relational or OLAP-based. In-memory BI technology can save significant development time by eliminating the need to store pre-calculated data in OLAP cubes or aggregate relational tables.

The common characteristics of in-memory BI approaches are: easy to use, visual interface, dashboards, self-service, in memory processing, speed of response, low costs, and quickly to deploy.

We see that in-memory BI technology has the potential to help BI systems to become more agile. Table 3 presents a SWOT analysis for implementation of in-memory BI solutions.

Table 3. A SWOT analysis for implementation of in-memory BI solutions

S (STRENGTHS)	W (Weaknesses)
<ul style="list-style-type: none"> • in-memory processing • faster speed of response, rapid access to reports, analysis and business metrics • improving self service through analytic flexibility • allows companies to integrate data from transactional systems, external data sources, spreadsheets or data warehouses • quickly to deploy • eliminates the need to store pre-calculated data in OLAP cubes or aggregate relational tables • visual interface, dashboards • ease of use for end users • low costs 	<ul style="list-style-type: none"> • a limited metadata management • limited by physical memory • data quality • limited ETL • sometimes requires multidimensional data modeling
O (Opportunities)	T (Threats)
<ul style="list-style-type: none"> • allows for real time business intelligence without a DW • eliminates the need for a pre-built OLAP cube or data mart • cloud computing 	<ul style="list-style-type: none"> • not a real-time analysis because data is analyzed in memory, not in the data store.

There are different in-memory BI approaches such as: IBM Cognos TM1, MicroStrategy, Microsoft PowerPivot and QlikView. For example, QlikView originally called “QuikView” as in “Quality, Understanding, Interaction, Knowledge” uses AQL

(associative query logic) technology. QlikView holds all data in memory with every association between data points defined [19]. Table 4 presents few in-memory BI approaches with their characteristics.

Table 4. In-memory BI approaches

Approach	Characteristics	Examples
In-memory OLAP	<ul style="list-style-type: none"> - MOLAP cube loaded entirely in memory - accessible by MDX tools - requires multidimensional data modeling - limited by physical memory 	IBM Cognos- Applix(TM1) Actuate BIRT
In-memory ROLAP	<ul style="list-style-type: none"> -ROLAP metadata loaded in memory -requires multidimensional data modeling 	MicroStrategy

	-not limited by physical memory	
A columnar database	-load and store data in a columnar database -less modeling required than an OLAP based solution -limited by physical memory	Tableau Software
In memory spreadsheet	-spreadsheet loaded into memory -no modeling required -access by third-party tools.	Microsoft PowerPivot
In memory “associative” data model	-loads and store all data in an “associative” data model (array) that runs in memory. -all joins and calculations are made in real time. -less modeling required than an OLAP based solution -limited by physical memory -some scripting required to load the data	QlikView

2.3 Agile Information Infrastructure

True agility is reached by making all parts of a BI system agile. An agile BI solution can be seen as consisting of two layers: *an agile information infrastructure layer* and *an agile analytic layer* (figure 5).

Information infrastructure addresses how the data architecture and data integration infrastructure ensure agility to react to changing business requirements. An agile information infrastructure must be able to extract and combine data from any data sources, internal and external sources including relational, semi-structured XML, multidimensional and “Big Data”. How to get an agile information infrastructure? By using data virtualization.

Data virtualization is “*the process of offering data consumers a data access interface that hides the technical aspects of data stores, such as location, storage structure, API, access language, and storage technology*” [12]. According to [13], [14] data virtualization “*is the technology that offers data consumers a unified, abstracted, and encapsulated view for querying and manipulating data stored in a heterogeneous set of data stores. Data virtualization means on-demand data transformation, on-demand data integration, and on-demand data cleansing*”.

Data virtualization can be implemented in many ways such as: using a data

virtualization server or placing data sources in the cloud.

Traditional BI systems use ETL tools for extracting data from multiple sources and temporarily storing those datasets at a staging area. As opposed to ETL tools, data virtualization server:

- allows the source data to remain in their original locations;
- eliminates staging of the data;
- abstracts source data, resolving structural and semantic issues;
- generates business views and/or data services that provide data required.

A business view is conceptually equivalent to a relational view. The views can read data from multiple data sources including: relational databases, multidimensional databases, text files, XML documents, spreadsheets, HTML pages, NoSQL databases, and so on. Applications access source data through the business views/data services using different interfaces such as: JDBC with SQL, ODBC with SQL, SOAP/XML and MDX.

Data virtualization server offers:

- data modeling capabilities;
- data profiling capabilities;
- data transformation capabilities;
- on-demand data integration capabilities which result in more agile BI systems. [15], [18].

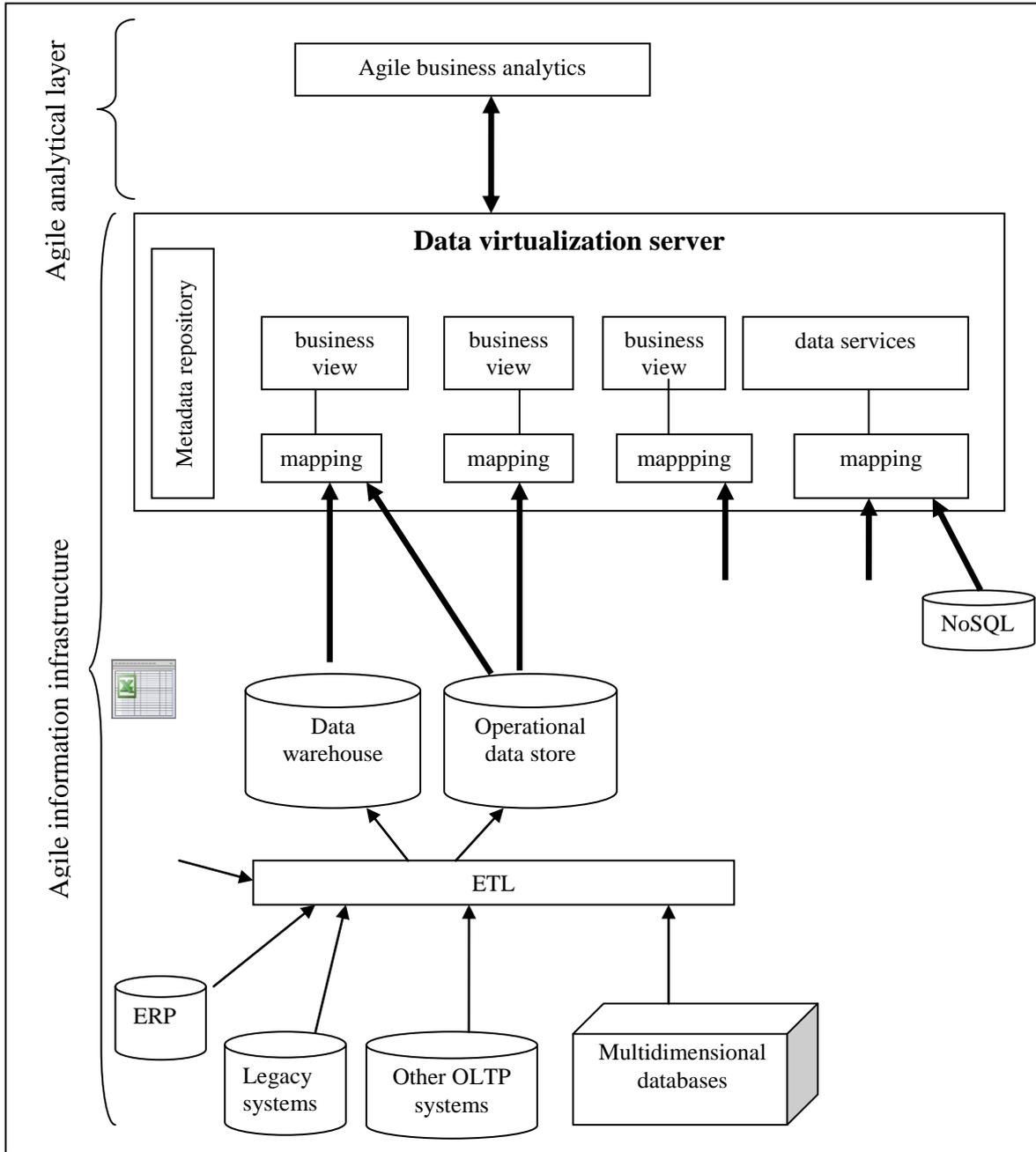


Fig. 5. A high level architecture of an agile BI system

Many data virtualization servers are currently available such as: Composite Information Server, Denodo Platform, IBM InfoSphere Federation Server, Informatica Data Services, and so on.

A high-level architecture of an agile BI system is illustrated in Figure 5. This architecture is based on data virtualization server. Data virtualization server examines

the data source structure and the resulting metadata is stored in metadata repository. Then you can create business views or data services using the data source's metadata

In conclusion, an agile BI solution requires: an agile development methodology, agile BA and an agile information infrastructure (Figure 6). Also, Figure 6 briefly summarize strengths of an agile BI solution.

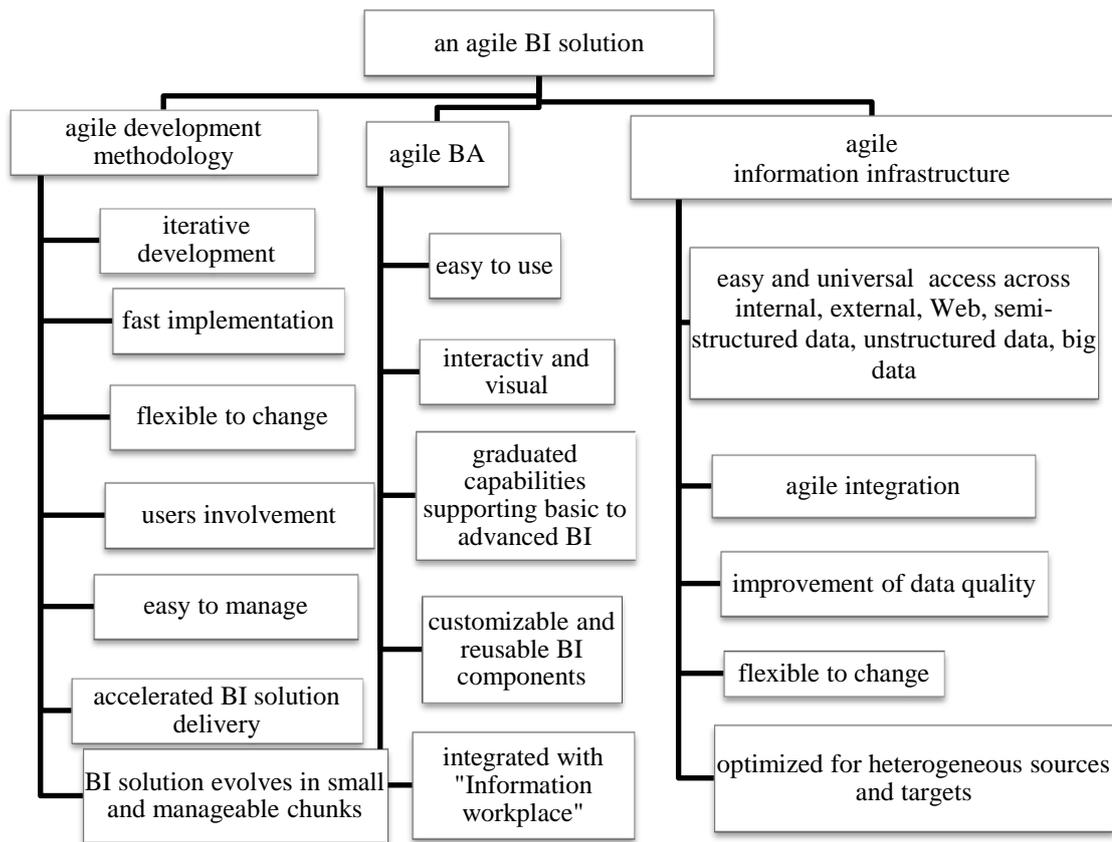


Fig. 6. The key components that promote an agile BI solution

The Table 5 presents a comparative analysis between traditional BI solution and agile BI solution using the following criteria: business requirements, integration approach, data

timeline, data refresh, information delivery, data source format, development methodology, development cycle and BA.

Table 5. Traditional BI versus agile BI

Criteria	Traditional BI	agile BI
Business requirements	- the customer knows what he needs - well defined - not change significantly	- the customer discovers during the project what he needs - change frequently
Integration approach	- ETL tools - moves/copies data from data sources to stage area - replicated data	- data virtualization - data remains stored at the source and a conceptual view is materialized on demand
Data timeline	historical data	on real-time data
Data refresh	end on day /end of last load	on real-time/near real time
Information delivery	takes too long to deliver	faster
Data source format	- structured data (relational databases), - limited semi-structured data, - Excel files, - multidimensional databases	- structured data, - semi-structured data - unstructured data, Big data
Development methodology	waterfall methodology	agile development methodologies
Development cycle	- too slow - too inflexible for BI - during the project nothing changes	- faster - there is a lot of changes during the project
Type of Business analytics	traditional BA	agile BA

Also Figure 7 shows a traditional BI workflow versus an agile BI workflow. In a traditional BI workflow the business analyst gathers user analytics requirements. He decides how to interpret requirements and then delivers them to the data warehouse architect. The data warehouse architect defines cubes, facts, dimensions and granularity of facts/dimensions. Then the data warehouse modeler decides how the dimensions and facts should be integrated into data warehouse. Also, he develops data models for staging area, warehouse database and cubes. The ETL developer develops ETL code to load data. Then, the BI developer

develops cubes and dashboards. The Data quality analyst verifies the quality of dashboards. Then dashboards are published to server and the customer uses these dashboards.

The BI consultant advises customers in the fields of information management and in selecting the most suitable BI solution. He is the first point of contact for the customer. He has experience with modern business intelligence techniques, data modeling, ETL tools, software development cycles, and so on. In short a BI consultant is responsible for the requirements gathering, the design and the development of BI solutions.

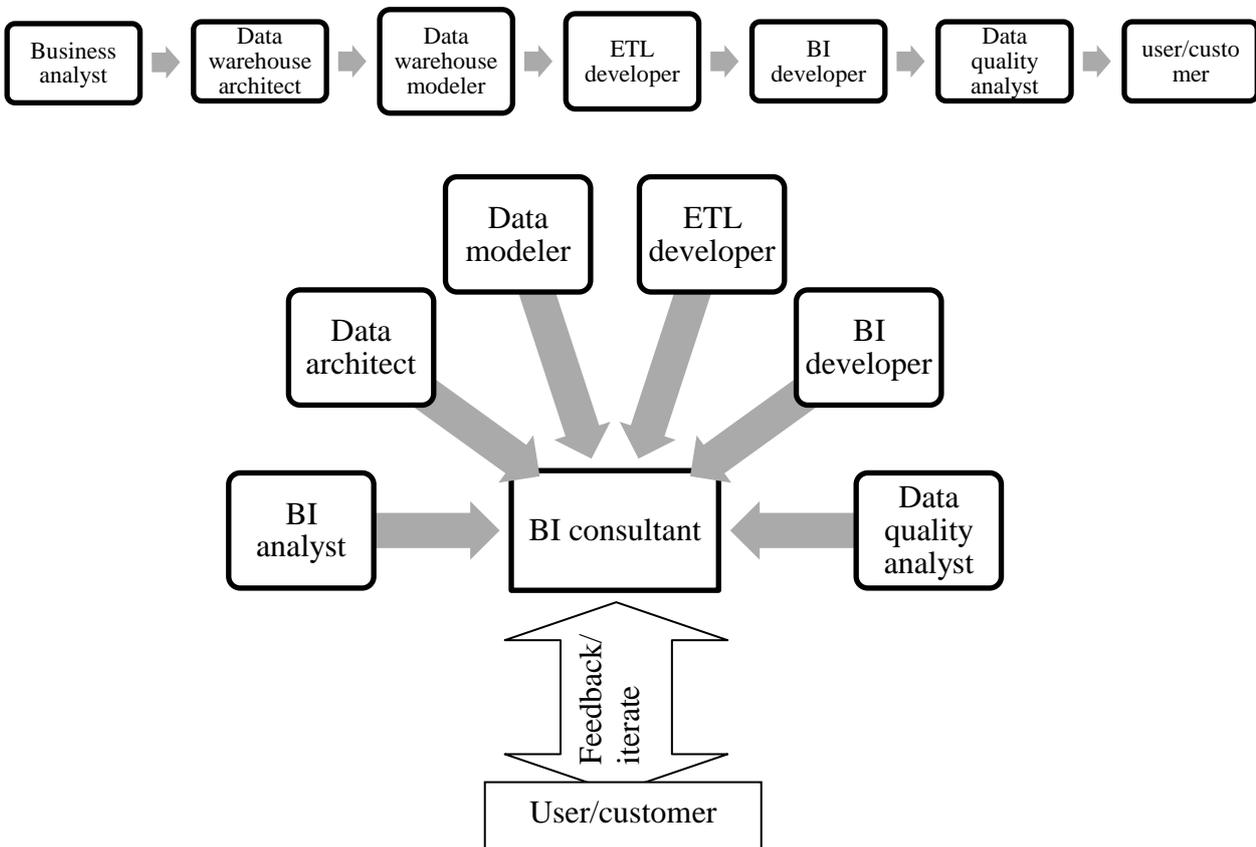


Fig. 7. A traditional BI workflow versus an agile BI workflow

4 Conclusion

In conclusion, the main reasons for implementing agile BI are:

- constantly changing business requirements;
- inability of IT to meet business user demands;
- slow access to information.

Agile BI solutions enable organizations to adapt to changing market conditions. This paper has identified the key elements that together promote an agile BI solution. There are plenty of technologies that can make an agile BI. This paper briefly looked at technologies that can be used for enabling an agile BI solution such as: in-memory

technology and data virtualization. Deploying of both technologies in a BI system results in an agile BI architecture.

References

- [1] S. Ambler, "Agile database techniques: effective strategies for the agile software developer", Wiley Application Development, published by Wiley Publishing, Inc, 2003, pp. 3-18
- [2] K. Collier, "Agile Analytics. A value-driven approach to business intelligence and data warehousing", Agile Software Development Series, Pearson Education, Inc, 2012, pp: 3-23
- [3] K. Beck, J. Grenning, et.al, "Principles behind the Agile Manifesto", 2001, available on-line at: www.agilemanifesto.org/principles.html
- [4] K. Beck, J. Grenning, et.al, "Manifesto for Agile software development", 2001, Available on-line at: www.agilemanifesto.org
- [5] L. Corr, "Agile datawarehouse design; collaborative dimensional modeling, from whiteboard to star schema", published by DecisionOnePress, Burwood House, 2011, pp: 3-26
- [6] W. Eckerson, "Predictive Analytics. Extending the Value of Your Data Warehousing Investment", TDWI Best Practices Report, 2007, Available on-line at http://www.sas.com/events/cm/174390/assets/102892_0107.pdf
- [7] W. Eckerson, "The secrets of creating an Agile adaptable BI environment", 2010, TDWI Education, Available on-line at: <http://tdwi.org/webcasts/2007/11/the-secrets-of-building-an-agile-adaptable-bi-environment.aspx>
- [8] B. Evelson, "Agile BI out of the box", pp. 1-18, 2010, Available on-line at: <https://www.wherescape.com/getattachment/resource-library/white-papers/agile-bi-out-of-the-box/Agile-BI-out-of-the-box.pdf>
- [9] B. Evelson, "Trends 2011 and beyond; Business intelligence", 2011, Available on-line at: [http://www.mxisoft.com/Portals/53068/docs/Forrester %20trends_2011_and_beyond_business_intelligence%5B1%5D.pdf](http://www.mxisoft.com/Portals/53068/docs/Forrester%20trends_2011_and_beyond_business_intelligence%5B1%5D.pdf)
- [10] B. Evelson, "The future of BI", pp: 1-8, 2012, available on-line at: http://www.hostanalytics.com/sites/default/files/The_Future_Of_BI-Forrester.pdf
- [11] R. Hughes, "Agile data warehousing: Delivering world-class business intelligence systems using Scrum and XP", publisher: iUniverse, 2008, pp: 1-23
- [12] R. F. van der Lans, "What is data virtualization?", 2011, Available on-line at <http://www.irmuk.co.uk/articles/Rick%20van%20der%20Lans%20-%20What%20is%20Data%20Virtualization.pdf>
- [13] R. F. van der Lans, "Data Virtualization for business intelligence systems: revolutionizing data integration for data warehouses", the Morgan Kaufmann series on business intelligence, Elsevier, 2012, pp: 1-106
- [14] R. F. van der Lans, "Data Virtualization for business intelligence agility", 2012, available on-line at http://purl.manticoretechnology.com/ImgHost/582/12917/2012/resources/whitepapers/Whitepaper_DVAgility_RickVDL_February%202012.pdf
- [15] D. Loshin, "Effecting data quality improvement through data virtualization", 2010, available on-line at <http://dataqualitybook.com/kii-content/DataQualityDataVirtualization.pdf>
- [16] M. Muntean, „A Business Intelligence framework for universities”, *The tenth International Conference on Informatics in Economy - Education, Research & Business Technologies, (IE2011)*, Department of Informatics in Economy, Faculty of Cybernetics, Statistics and Informatics in Economy, Academy of Economic Studies, București, 2011, "Education, Research & Business Technologies", pp. 1-6
- [17] L. Moss, "An agile Approach to enterprise data warehousing and Business Intelligence", pp: 1-9, 2012,

- available on-line at http://www.technologytransfer.eu/article/104/2012/9/An_Agile_Approach_to_Enterprise_Data_Warehousing_and_Business_Intelligence.html.
- [18] P. Russom, "Data Integration for Real-Time data warehousing and data virtualization, TDWI Checklist report, TDWI Research, 2010, Available on-line at: http://www.bleent.com/wp-content/uploads/2011/01/7239_tdwi_di_rt_dw_wp.pdf
- [19] R.L. Sallam, et.al., "Magic Quadrant for Business Intelligence Platforms", Gartner RAS Core Research Note, 2012, available on line at <http://businessintelligence.info/docs/estudios/Magic-Quadrant-for-Business-Intelligence-Platforms-2012.pdf>
- [20] R. Sabherwal, I. B. Fernandez, "Business Intelligence, chapter 8 : Development of Business Intelligence", John Wiley & Sons, 2010, pp. 218-243
- [21] J. Sutherland, "Scrum handbook", Scrum Training Institute Press, available on-line at: <http://jeffsutherland.com/scrumhandbook.pdf>, 2010, pp 1-67
- [22] TDWI Research, "2011 TDWI BI Benchmark Reports: Organizational and Performance metrics for Business Intelligence teams", 2011, Available on-line at: <http://tdwi.org/research/2011/09/2011-tdwi-bi-benchmark-report.pdf>
- [23] E. Turban, R. Sharda, D. Delen, "Decision Support Systems and Intelligent Systems", 8th edition, Prentice Hall, 2006
- [24] D. White, "Agile BI: Completing Traditional BI to address the shrinking Decision-window", 2011, Aberdeen Group, Available on-line at: <http://www.blockconsulting.net/PSBDGroup/Aberdeen-agile-BI-shrink-decision-window.pdf>



Mihaela MUNTEAN is associate professor in Economic Informatics Department, Faculty of Economic Cybernetics, Statistics and Informatics, Academy of Economic Studies of Bucharest. She received her doctoral degree in Economics in 2003. Since 1997 she is teaching in Academy of Economic Studies, in Economic Informatics Department. She is interested in Databases, Information Technology & Communication, OLAP technology, Business Intelligence Systems and Economic Information Systems Design.

She published over 50 articles in journals, over 30 scientific papers presented at national and international conferences, symposiums and workshops and she was member over nine research projects. She is the author of two books and she is coauthor of seven books.



Traian SURCEL is Professor at Academy of Economic Studies Bucharest, Faculty of Economic Cybernetics, Statistics and Informatics, Department of Informatics in Economy, PhD in Economic Cybernetics from 1987. He coordinates the Fundamentals of IT&C for Business Management professors group and also PC Laboratories for Faculty of, Marketing, Commerce and International Business and Economics. He is Internal Auditor for the ASE Bucharest. His main research areas are: ERP and Information System and

Database analyze and design, IT Systems Audit, e-Learning applied methodology, IT&C for Communications and Business Management.