Business Intelligence and Data Warehousing

Learning Unit 1

Business Intelligence and Data Warehousing

INTRODUCTION

This learning unit introduces this course with an overview of Business Intelligence. We begin with a short, gentle, readable book about the topic: *Business Intelligence en datawarehousing*.

> LEARNING OBJECTIVES
> After studying this learning unit, you should be able to study the Sharda book.

Study hint
The workload is about 7 to 8 hours.

STUDY CORE

1 Reading Assignment: Business Intelligence en datawarehousing

Reading assignment

Read most of the course textbook *Business Intelligence en datawarehousing* by Karien Verhagen.
The chapters you should read are

1 Bestuurlijke informatie: de vraag
2 Business Intelligence en de bedrijfsprocessen
3 Business Intelligence en datawarehousing
4 De voorkant van de BI-omgeving
5 De achterkant van de BI-omgeving
6 De binnenkant: de opslagstructuur
8 De zijkant van de BI-omgeving
9 ETL (algemeen)
12 Project- en programma-aanpak

The goal here is to get a comfortable introduction to the topic of Business Intelligence as a whole. This will provide a broad conceptual foundation for more rigorous study of more specific topics in the rest of this course.

Corporate social responsibility

Learning Unit 2

Corporate Social Responsibility

INTRODUCTION

This learning unit will give you the highlights of sustainability of and by IT. During task 1 you will describe an organization suitable for the Audit for Business Intelligence. As one of the items in that description you investigate if the company has a policy about sustainability, reports on sustainability and makes use of an energy efficient data center.

LEARNING OBJECTIVES
After studying this learning unit, you should be aware that
- the intensive use of IT such as during BI comes with environmental impacts
- when you make choices, you have to consider the balance between people, planet and profit
- IT can also be an enabler to make processes more sustainable.

Study hints
The workload is about 1 hour.
In this workload you are not required to read the references.

STUDY CORE

1 Introduction to sustainable IT

It has been long assumed that the IT industry does not affect the environment. Recently, however, people became aware of the high energy consumption (comparable to the global aviation industry [1]) and the depletion of scarce metals the use of IT equipment brings. Future IT professionals should understand the impact of the life cycle of IT equipment on the environment and also be aware of how IT can be used to enhance sustainability and reinforce the green economy.

2 What is Green IT?

Brundtland report: sustainable development

In the literature both terms 'green' and 'sustainable' are used. The concept *sustainable development* was introduced in the *Brundtland report* in 1987: 'Sustainable development is a pattern of economic growth in which resource use aims to meet human needs while preserving the environment so that these needs can be met not only in the present, but also for generations to come'.
Sustainable development has four dimensions: here and elsewhere in the world; now and in the future. In the nineties of the last century, organizations introduced the concept corporate social responsibility in which the preservation and improvement of three kinds of capital are important: the social, ecological and economic capital.

Social The *social* capital is about the well-being of employees, a company's own employees as well as the employees in the company's supply chain.

Ecological The *ecological* capital is formed by the natural resources such as minerals, forests, rivers and a clean environment.

Economic The economic capital is how profitable the company is.

'Triple P': People, Planet, Profit These three kinds of capital are referred to as *'triple P': People, Planet, Profit*.

In this course we conform to reference [1] and use 'Green' to describe those situations where the impact on the environment is considered. We use 'Sustainable' for situations which affect the 'triple P' People, Planet, Profit, derived from the Brundtland definition and relating to corporate social responsibility.

Equilibrium Sustainability is the act of striving for *equilibrium* in the triangle People, Planet, Profit with the target of less 'waste'. Balance with respect to Profit and People in IT means having no waste of money or working hours by failure of IT-projects; promotion of code reuse and code without failures. Balance between Profit and Planet means, among other things, energy efficiency of hardware and software and careful e-waste. Balance between People and Planet means an attitude of respect for the planet and knowledge of sustainability of and by IT.

3 Environmental problems

Earth: closed system Even as early as 1972, the report of the 'Club of Rome' recognized that the earth should be 'handled with care' because the *earth is a closed system* in which the cycles should not be disturbed. The environmental policy in the Netherlands reacted by a subdivision of the problems into eight themes [2] to preserve the environment:

– climate change
– waste
– hazardous waste
– exhaustion
– disturbance
– aridity
– acid rain
– eutrophication.

The order of this list reflects the impact of IT on the environment.

3.1 CLIMATE CHANGE

Greenhouse gases: enhanced greenhouse effect *Greenhouse gases* trap the heat in the atmosphere and man-made greenhouse gases are responsible for the *enhanced greenhouse effect* which is supposed to be the cause of climate change. They are: *water vapor* – due to natural evaporation of water and industrial processes; *carbon dioxide* – produced by burning natural or synthetic organic substances (e.g. fossil fuels, trees, etc.); *methane* – emitted by natural sources such as wetlands, livestock, decay of organic waste, etc.; *nitrous oxide*–direct effect of human activities such as agriculture, fossil fuel combustion, wastewater management, and industrial processes; *fluorinated gases* (also known as ozone depleting substances) -emitted from a variety of industrial processes.

In the IT industry, the indirect production of carbon dioxide (CO_2) is the most important impact factor on the environment, caused by the enormous energy consumption. Because most energy comes from the combustion of fossil fuels, energy consumption is related to the weight of CO_2 formed or the carbon footprint. But also when renewable energy is used (and no net CO_2 is formed), people should be aware that energy is never free!

A very simple, but chemically correct carbon cycle is shown in figure 2.1.

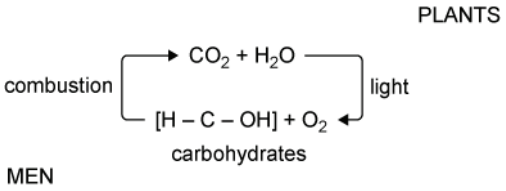

FIGURE 2.1 Carbon cycle

Plants can convert carbon dioxide and water by the energy from sunlight to oxygen, O_2, and carbohydrates (schematically represented by [H-C-OH]). Men consume these plants and metabolize the carbohydrates into carbon dioxide and water. Incompletely decomposed organisms left underground for a long time turn into fossil fuels. These consist of hydrocarbons, which are combusted to carbon dioxide. If more carbon dioxide is formed then used the cycle is no longer closed and the content of carbon dioxide increases. This causes the enhanced greenhouse effect, which in turn causes climate change.

3.2 WASTE

Discarded electronic devices

E-waste is defined as 'any refuse created by *discarded electronic devices* and components as well as substances involved in their manufacture or use'. Ways to reduce e-waste are: recovery, reuse and recycling.

3.3 HAZARDOUS WASTE

Batteries

Examples of hazardous waste include most *batteries* that are used in computer equipment. These contain hazardous compounds that are harmful to the environment. When batteries are incinerated, the metals they contain pollute the atmosphere and the incineration residues pollute the soil. However, when batteries end up in landfills, the metals can leach into the soil and water. Thus, the chemical substances in batteries can cause atmospheric, soil, and water pollution.

Nuclear waste

Nuclear waste from nuclear power plants is another example. Most nuclear waste remains radioactive for hundreds of years.

3.4 EXHAUSTION

Depletion of scarce materials

Exhaustion means the depletion of substances. An IT-related example is the *depletion of scarce materials* such as rare metals that are used in touchscreens. The indium used is difficult to come by, expensive, and difficult to handle. In addition, indium is politically difficult to obtain because 97% of rare earths (to which indium belongs) are found in China.

Deforestation

Another IT-related example is *deforestation*, since paper made from wood deforestation can be diminished by making fewer printouts.

3.5 DISTURBANCE

Water temperature

Under disturbance are categorized: heating (of water), noise and noxious odors. An IT-related example is the disturbance (rise) of the *water temperature* when water is used as cooling agent for data centers and then not brought back to room temperature before it is discharged into surface water.

3.6 ARIDITY

The proper humidity for soil is important not only for the growth of vegetation but also for the stability of buildings. A modern way to cool and heat buildings (and also to cool data centers) is through the use of thermal energy storage. Using this technique, one must note the amount of water (aridity), the temperature of the groundwater (disturbance), and the possible pollution if the installations fail (hazardous waste).

3.7 ACID RAIN

Acid rain is caused by sulphur dioxide, SO_2, nitrous gases, NO_x, and ammonia, NH_3. The first two originate from combustion from brown coal and incomplete combustion, the third from animal farming. Thus, the first two can be related to IT use.

3.8 EUTROPHICATION

Eutrophication is caused by the addition of artificial or natural substances, mainly phosphates, through detergents, fertilizers, or sewage, or nitrates to an aquatic system. One result is the 'bloom' or great increase of phytoplankton in the water as a response to increased levels of nutrients. This problem is not IT-related.

3.9 OTHER METHODS TO CATEGORIZE ENVIRONMENTAL PROBLEMS

Apart from the above mentioned thematically classification, human interventions into the environment can also be classified according to the type [3, 4]: pollution, exhaustion or degradation.

Pollution

Pollution is the introduction in the environment of chemical substances of physical phenomena in amounts that are larger than what is naturally present such that it causes damage.

Exhaustion

Exhaustion means that the exploitation of a component is faster or larger than can be renewed.

Degradation

Degradation is a change in structure of the environment by which the landscape and/or the ecological equilibrium is disturbed and thereby the quality of the environment is diminished.

3.10 SOLVING ENVIRONMENTAL PROBLEMS

Behavior
Technology

Environmental problems can be solved by a combination of *behavior* and *technology*. Examples of technology are energy saving, dematerialization (use as little material as possible) and closed loops.

The first step in the change of behavior is to make people aware of the problem so that IT professionals become aware of their impact on the environment.
Examples of technologies used in green IT are energy efficient hardware, software and services, new cooling techniques for data centers and techniques for non-carbon energy production. Material loops are closed when equipment can be recycled by separating it into components that can be reused in new equipment. A paperless office is an example of dematerialization.

4 Greening from and by IT

The approaches can be grouped into two categories:

Greening of IT

– *Greening of IT* – aims to mitigate the environmental impact of ICT itself. This encompasses energy efficient and environmental sustainable designs, operations, use and disposal of ICT equipment, infrastructure and systems. Sustainability becomes a *quality requirement*.

Quality requirement

Greening by IT

– *Greening by IT* –aims to harness IT (via ICT-enabled solutions) to mitigate the environmental impact of other sectors. In SMARTer2020 [1], the named sectors are: power, transportation, manufacturing, agricultural, building, service and consumer. This addresses applying ICT to create energy-efficient and environmental sustainable operations, processes, practices, etc. Traditional business intelligence can play an important role to control these smart processes.
Greening of IT is a subject of scientific literature and whitepapers since about 2007; greening by IT since about 2009.

5 Sustainable IT and Business Intelligence

On applying Business Intelligence sustainability *of* and *by* IT can both be important.
BI and especially Data discovery BI (Sharda book chapters 4 and 5) use much data center capacity, so it goes hand in hand with high energy consumption. Questions that may be asked related to sustainability *of* IT, are: 'Is the application indeed necessary?' and 'Is the energy consumption of the data center as low as possible?'
An example of sustainability *by* IT is the application of traditional BI (Sharda book chapter 3) for data center infrastructure management [5].

6 Environmental management and BI

Corporate social responsibility is based on 'triple P' and can be associated with the introduction of an environmental management system [6]. The components of which are:
– policy for sustainability by the board
– sustainability plan (which activities)
– integration of sustainability in the company operations
– measurement and registration
– reports (internal and external)
– training and education of personal
– inspections (internal)
– regular audits of the environmental management system.
Measurement and reporting can be supported by BI [7].

7 References

[1] Pattinson, C. and Kor, A. (2014) Chapter, 1 Introduction to green IT, in Pattinson, C., Slaidins, I. and Counotte, A. (eds.), *Green Sustainable Data Centres*, http://portal.ou.nl/web/green-sustainable-data-centres, and references therein.

[2] Nelissen, N.J.M. in Glasbergen, P. (ed) (1994) *Milieubeleid, een beleidswetenschappelijke inleiding*, chapter 3: Het themagerichte milieubeleid.'s Gravenhage,VUGA uitgeverij b.v.

[3] Cörvers, R.J.M. et al. (2004). *Milieuproblemen en duurzame ontwikkeling*, part 1, page 22.. Heerlen, Open Universiteit.

[4] Boersema, J.J. e.a. (ed) (1986). *Basisboek milieukunde. chapter 1.* . Meppel, Boom b.v.

[5] Dalton, D. (2014) Chapter 6, Data Centre Infrastructure Management, in Pattinson, C., Slaidins, I. and Counotte, A. (eds.), *Green Sustainable Data Centres*, http://portal.ou.nl/web/green-sustainable-data-centres, and references therein

[6] Glasbergen, P. e.a. (1991), *Milieumanagement in bedrijven*. Heerlen, Open Universiteit Nederland.

[7] Carreira, P. and Silva, C.A. (2014), Chapter 8, Greening by IT, in Pattinson, C., Slaidins, I. and Counotte, A. (eds.), *Green Sustainable Data Centres*, http://portal.ou.nl/web/green-sustainable-data-centres, and references therein.

Assignment Overview

Learning Unit 3

Assignment Overview

INTRODUCTION

In this learning unit you browse through the assignment bundle. The intention is that, before you start with the Sharda book, you first have an impression of the final goal of the course: the Audit of the present BI of an organization and recommendations for future BI, based on the knowledge of this course. In learning unit 10 and during the first face to face tutoring session you will learn more about the assignment.

> LEARNING OBJECTIVES
> After studying this learning unit, you are aware of which questions you should study the Sharda book.

Study hint
The workload is 1 hour

STUDY CORE

1 **Browse Assignment: Information Audit for Business Intelligence (IA4BI)**

Browse assignment

Browse the Assignment bundle: *Audit for Business Intelligence (IA4BI)*

Business Intelligence Overview

Learning Unit 4

Business Intelligence Overview

INTRODUCTION

This learning unit gets us started with this course's main textbook: *"An Overview of Business Intelligence, Analytics, and Decision Support"*.

> LEARNING OBJECTIVES
> Here we follow the learning objectives of the Sharda book.

Study hints
You will experience that the book provides:
– Learning objectives
– Overview of the chapter
– Opening vignette
– Content
– (Section) review questions
– Application cases
– Resources, links, and Teradata University network connections
– Chapter highlights
– Key terms
– Question for discussions
– Exercises
– References.

Study You should *study* the Learning objectives, Content and the Chapter highlights.

Read You should *read* the Opening vignette and the Application cases.

Be aware You should *be aware* that the book provides the other items in case you need some details.

The workload is about 2 to 3 hours.

STUDY CORE

1 Reading Assignment: Business Intelligence Overview

Reading assignment Study the first chapter *An Overview of Business Intelligence, Analytics, and Decision Support* from the Sharda book *Business Intelligence: A Managerial Perspective on Analytics* to get a broad view of Business Intelligence from the perspective of this course's main textbook.

Data Warehousing

Learning Unit 5

Data Warehousing

INTRODUCTION

The previous learning unit provided a general introduction to Business Intelligence as a whole. Now we focus on one of the core topics in BI: data warehousing. With this topic, we learn how to manage the vast amount of data and data sources that Business Intelligence utilized. Then the rest of this course describes different aspects of exploiting the enormous repositories of information that data warehouses provide.

LEARNING OBJECTIVES
Here we follow the learning objectives of the Sharda book.

Study hint
The workload is about 4 to 5 hours.

STUDY CORE

1 Reading Assignment: Data Warehousing

Reading assignment

Read chapter 2 *Data Warehousing* from the Sharda textbook to become familiar with this fundamental topic in Business Intelligence.