

# INTRODUCTION TO BUSINESS INTELLIGENCE

What to consider implementing a Data Warehouse and Business Intelligence

## Summary:

This note gives some overall high-level introduction to Business Intelligence and some advices from a user perspective in implementing Business Intelligence in a company. What to look for when using an external Business Intelligence Consultancy company.

Version 1.0

## Table of content

Table of content .....	2
What is business Intelligence .....	3
General definitions .....	3
Business Intelligence components .....	4
1 - Multidimensional analysis .....	4
2 - Reporting .....	6
3 - Data mining .....	6
4 - Financial consolidation and budgeting .....	6
5 - Key Performance Indicators .....	6
Business Intelligence Consulting Companies .....	7
Consulting companies tries to differentiate themselves .....	7
Performance Management .....	7
Master Data Management .....	8
Inconsistent Master Data .....	9
Master Data Management and Slowly Changing Dimensions .....	9
Quality Management and Risk Management .....	10
Conclusion – BI consultants .....	10
Business requirement and what you get .....	11
Typical 'non-data warehouse' system solution .....	11
Building data Warehouses to support analysis and reporting .....	12
What is the real business need .....	13
Master Data Management .....	13
Business need: Multidimensional analysis or fixed reports .....	14
What to look for by choosing consultancy company .....	16
Don't just accept the 'full standard package methodology' .....	16
ETL process .....	17
Presentation level 1 process – Creation of data marts .....	18
Presentation level 2 process – Creation of cube(s) .....	18
Alternative processes .....	18

# What is business Intelligence

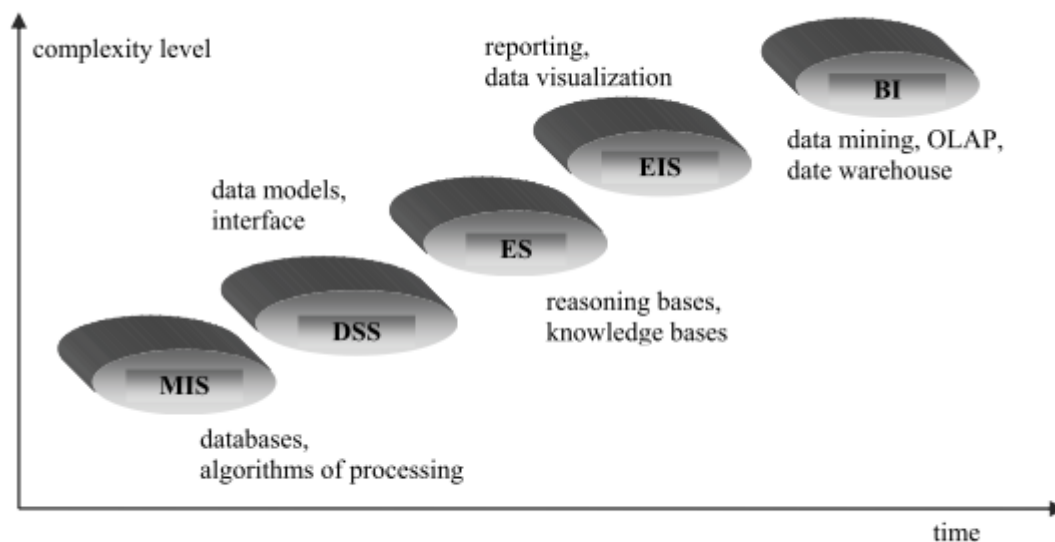
## General definitions

First a public definitions from the internet:

*"Business intelligence (BI) is a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information. It allows business users to make informed business decisions with real-time data that can put a company ahead of its competitors. Traditionally, core features like reporting and analytics have been the focus of BI technology choices, but as those features get commoditized, a whole new set of possibilities has emerged. Forrester's BI research shows that the technology is evolving and that enterprises on the cutting edge of these new trends can gain competitive advantage in their industries."*

Source: [http://www.forrester.com/rb/Research/topic\\_overview\\_business\\_intelligence/q/id/39218/t/2](http://www.forrester.com/rb/Research/topic_overview_business_intelligence/q/id/39218/t/2)

There can be found many different definitions and views upon the term "Business Intelligence" but most of them describes broadly that the purpose is to provide meaningful information for decision making. In late 1970'es and in the 1980'es it was more common to use the term "Decision Support Systems" and at the end it will all come down to the same purpose: To be able to present meaningful information for decision making. See the evolvement of technologies in the figure below:



**Figure 1. Development of management information systems.**

Source: (Olszak, & Ziemba, 2004).

Source: <http://ijikm.org/Volume2/IJIKMv2p135-148Olszak184.pdf>

## Business Intelligence components

In most cases Business Intelligence involves the use of multidimensional analysis and reporting. By providing multidimensional analysis and reporting the company often builds a Company Data Warehouse to assemble the needed data.

However the term Business Intelligence covers at least the following five components that will be explained in more detail in the following:

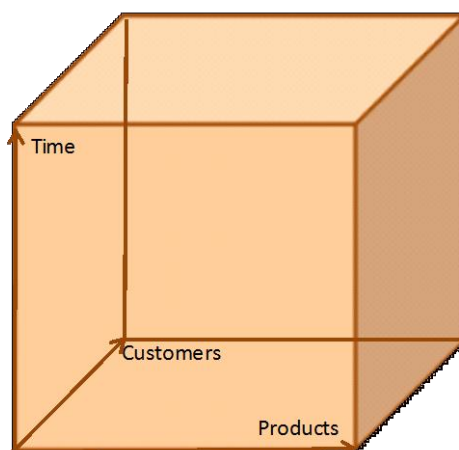
1. Multidimensional analysis
2. Reporting
3. Data mining
4. Financial consolidation and budgeting
5. Key Performance Indicators

### 1 - Multidimensional analysis

This area covers the possibility to slice-and-dice data (factual information) in many dimensions. For many this is known as pivoting data.

A pivot table is something that you can build and use in most spreadsheets including Excel. In Excel you can summarize data in a pivot table on many levels on each dimension and you can have a table of rows and columns of summarized data in both dimensions as well as filtering the shown data. Dimensions are often a hierarchy with multiple levels (product groups down to product level eg.)

From a more technical perspective most database suppliers and Business Intelligence suppliers support multidimensional analysis with a special multidimensional database. A multidimensional database is often also called a CUBE. A Cube is often presented as 3-dimensional data information storage like in the example below:

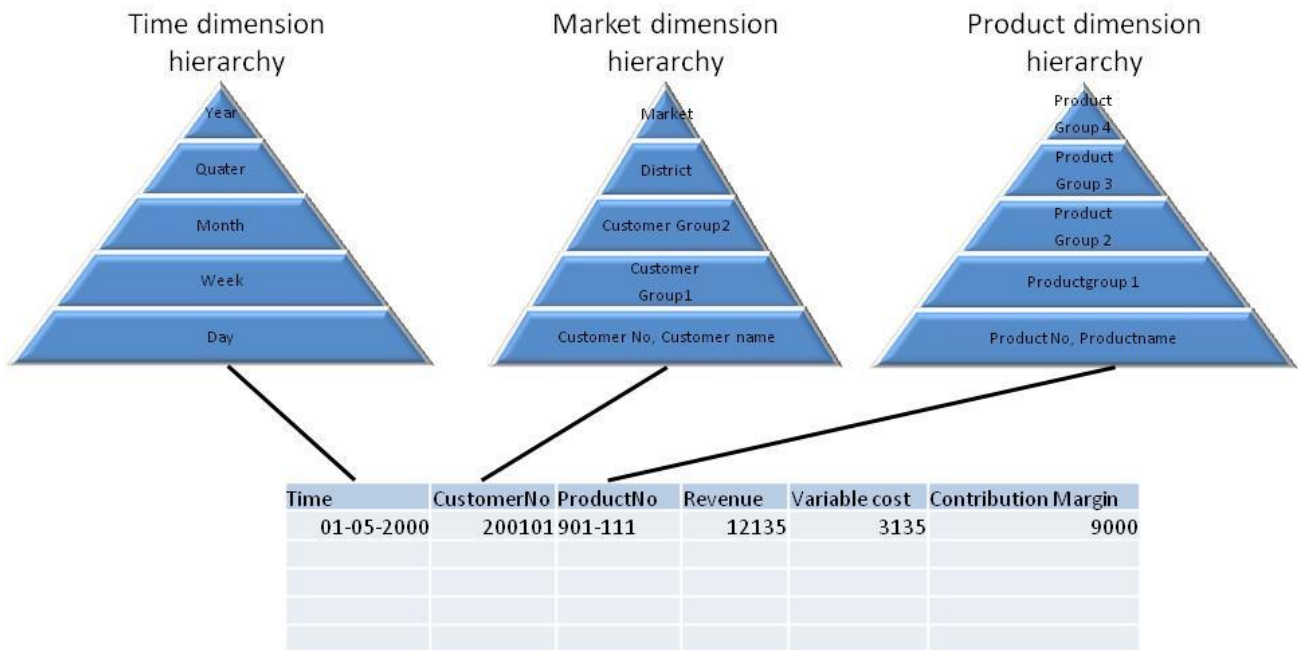


In the figure above this cube has Time, Customers, and Products in its 3 dimensions. However a multidimensional CUBE can normally contain many more dimensions than 3. However as it is difficult to represent a 4+ multidimensional database in a figure then this is normally shown as a cube in 3 dimensions even if the multidimensional database can have 4+ dimensions.

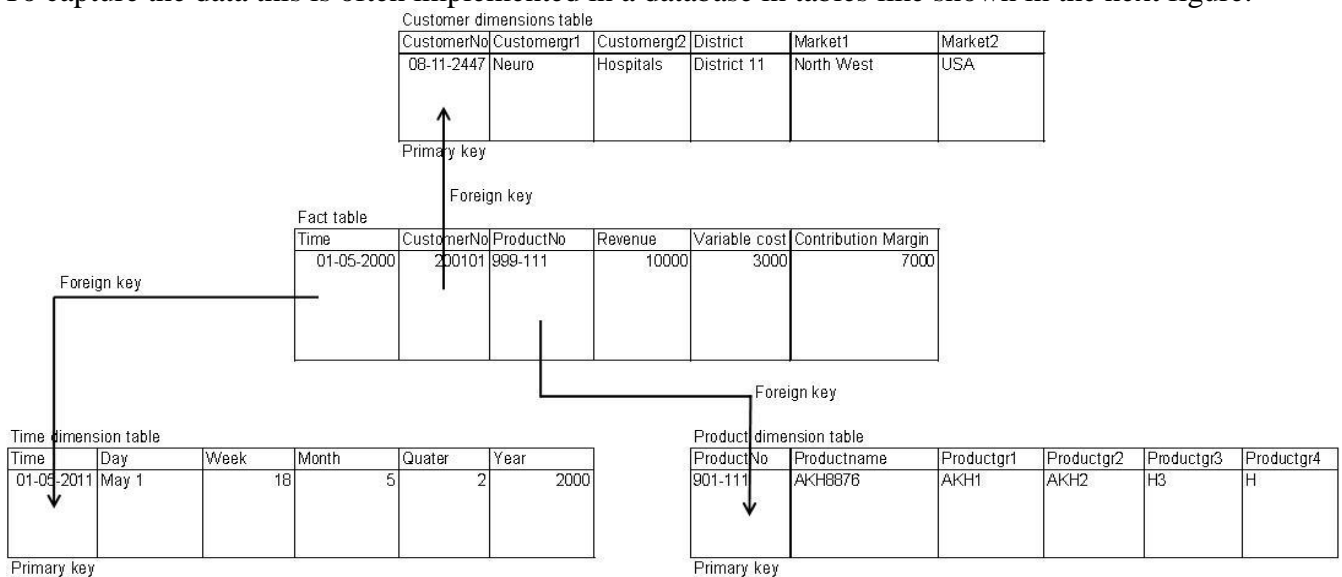
For now let us continue this introduction with an example of the above mentioned 3 dimensions. Many companies has a need to analyze data in many dimensions as mentioned here but it could also be other types of data like purchasing information where it was not a customer dimension but a Supplier dimension. It can also be manufacturing data and much more.

Another typical characteristic with a dimension is that most dimensions have some hierarchy (like Customer, Customer Parent, City, County, Region, Country, Group of Countries e.g.

See an example of data we want to be able to analyze (in the table) by pivoting these in many hierarchic dimensions:



To capture the data this is often implemented in a database in tables like shown in the next figure:



How this concept is implemented in Data Warehouses and the technical aspect of this is not the scope of this note but will be discussed in much more detail in another note covering Data Warehouse concepts and Slowly Changing Dimensions.

## 2 - Reporting

Most companies have a need for different type of reports. In many cases hundreds of different type of reports and occasionally often more.

Business Intelligence software often has comprehensive reporting tools available that can extract and present data in many different media types (like over an internal Web page/Intra net, Internet (to customers), Excel file format, PDF format e.g. In many cases these reporting facilities will be controlled by parameters that can be chosen real time and present a report that has been run directly against data (often a Data Warehouse or multidimensional data).

## 3 - Data mining

*“Data mining, a branch of computer science, is the process of extracting patterns from large data sets by combining methods from statistics and artificial intelligence with database management.”*

Source: [http://en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining)

In many cases Business Intelligence also covers some functionality to perform Data Mining on the company data. As the definition describes above the purpose is to find (so far unknown) patterns from large data sets.

In general to use the methods to find some “new information” in regards of already collected data and with this new information of “patterns” to be able to make “informed business decisions”.

## 4 - Financial consolidation and budgeting

In many cases the Business Intelligence area also covers systems and functionality for Groups to perform financial group consolidation and budgeting.

This area covers the whole aspect of elimination of inter-company transactions within a group to present the financial figures as these inter-company transactions had never happened and that the group is presented as one entity.

In some cases this area within consulting companies also covers some elements of Performance Management and metrics like Key Performance Indicators.

## 5 - Key Performance Indicators

This term is mainly seen in the context of Balance Scorecard developed by Robert S. Kaplan and David P. Norton.

It is not in scope of this note to introduce the concepts of Balance Scorecard but Key Performance Indicators are metrics that you measure periodically (weekly, Monthly, Yearly or other timeframes) and where you often keeps track of historical trends as well as a goal or target.

In some cases Business Intelligence also supports setting goals for specific metrics in the data set as well as calculating and presenting trends.

## Business Intelligence Consulting Companies

When a company needs to collect and present certain information it often turns to a consulting company for help. When we talk about information analysis and reporting the company often tries to find some specialists within this area and this will often be a company specialized within Business Intelligence.

### Consulting companies tries to differentiate themselves

As it could be seen from the first figure shown in this note Information Systems has developed over time from late 1970'es until into 201x.

While Business Intelligence has been 'a hot word' for nearly 15 years now some consultancy companies tries to differentiate them compared to the competitors. This is often done by trying to tell the customer that the company has a 'broader scope' or 'more holistic' scope.

They will typically use a 'broader view' and that can be with using some terms like:

- Performance Management
- Master Data Management
- Quality Management
- Risk Management

### Performance Management

There are many definitions of the term Performance Management and within Business Intelligence Consulting companies this is often covering a broader group of technologies like Consolidation Software, Balance Scorecard Software including Key Performance Indicators.

Performance Management is in some HR related consulting companies mainly defined within the scope of organizational performance and individual goals and metrics in this area and how to handle this. In other definitions this is mainly the overall process of handling the performance of the company in many other aspects including a Balance Scorecard view – and covers both financial performance as well as organizational performance.

In the figure below found on the Internet searching for the term "Performance Management" this can be seen as mainly a Balance Scorecard inspired definition:



Personally I see Performance Management as a general concept of how to manage a company. The term: Performance Management does not cover specific tools like Consolidation or Key Performance Indicators but is a process of 2 closed loops:

- Strategic loop: Mission->Vision->Strategy->Strategy execution/Goals, Follow-up, Adjust
- Yearly loop: Budget/Projects/KPI's, Implement actions, Follow up, Adjust goals



However many Business Intelligence consulting companies will provide consultants that has knowledge of financial consolidation and elimination and/or Balance Scorecard and will call this Performance Management and having this as their overall umbrella for their Business Intelligence services.

## Master Data Management

Master Data Management is the concept of having control of the companies Master Data often defined as the Customer Master file, Suppliers Master File, Items Master File, General Ledger e.g.



Master data are often used as dimensions in a multidimensional database (see previously about multidimensional analysis).

Some Business Intelligence consulting companies have used this area as their overall umbrella concept for implementing Business Intelligence.

Master Data Management is not an umbrella concept but it is however crucial for a successful implementation of a Business Intelligence solution. If master data that needs to be used in the dimensions and reporting are corrupt then the solution will suffer from this and will typically not be used because you can't 'trust' the data.

### Inconsistent Master Data

Many companies would say that they have very good Master Data but in most cases this is far from the truth. By implementing a Business Intelligence solution it is often shown that the company Master Data is very poor bringing the implementation project into huge problems.

Many ERP (Enterprise Resource Planning) systems that should be expected to have a consistent management of Master Data do not provide this management at all. This means that in some ERP systems you are able to record transactions where the transactions are stamped with certain Master Data and then you later on change or even deleting the Master Data without any update or handling of the underlying transactions having stamped Master Data as "orphan data" compared to the current Master Data files.

Two examples where a company needs to have clear procedures and knowledge of how data are recorded in the companies transactional databases are:

- Customer data (and same issue with Supplier data):
  - o Change of customer data like the address (what if the customer changes his address)
  - o Two customers merge (what will we do with the historical data)
  - o Customer groups (new grouping structure. What about historical data)
- Item data:
  - o One new Item developed and takes over from an old Item
  - o Item groups (new grouping structure. What about historical data)

In addition to this many companies end up having the same customer created two or even multiple times and this also goes for suppliers and items.

When a company has to implement a new ERP system and have to convert Customers, Suppliers and Items to the new system it often shows how miserable the Master Data actually is.

While I do not think that Master Data Management is a "broader" or more "holistic" concept than Business Intelligence then Master Data Management and a needed 'clean-up' of existing Master Data, as well as keeping Master Data clean in the future is fundamental for implementing a successful Business Intelligence project.

### Master Data Management and Slowly Changing Dimensions

Master Data Management has a close relationship to what within Business Intelligence is called 'Slowly Changing Dimensions'. As described earlier in this note a dimension is often a hierarchy of Master Data and Master Data seldom keeps the same over time. Sometimes a Customer changes the

Customer Name, phone number, Address and Zip code, or as described earlier merge with another customer.

An item can be used for a period of time but are then replaced with a new version that has a new item number.

The terms slowly changing dimensions holds concepts to handle these cases. One solution could be just to update the dimensions in the multidimensional database with the new information, but in that case the historical (and correct information) would be lost (like a customer that moves to another address and thereby to another zip code – the revenue to this customer is not suddenly increased in the new zip-code).

There are many different ways of handling this historical dimensional information while still having the updated information but it is out of scope for this note. See another note regarding Data Warehouse concepts and Slowly Changing Dimensions for details.

## Quality Management and Risk Management

Another area where Business Intelligence consultancy companies try to ‘broaden’ and have a “holistic” view is bringing either Quality Management or Risk Management into the framework as the umbrella terms of Business Intelligence.

This is often by providing IT technical solutions to handle Quality Management (providing controlled documents and work-flow of approval e.g.) and/or Risk Management (providing creation of risks; assessment; probabilities; risk action plans e.g.).

However these concepts are not some overall concepts ‘above’ business intelligence but rather other specific disciplines that a company needs to handle in many cases.

## Conclusion – BI consultants

If you have an information business need (see next section), then do not get trapped by ‘smart concepts’.

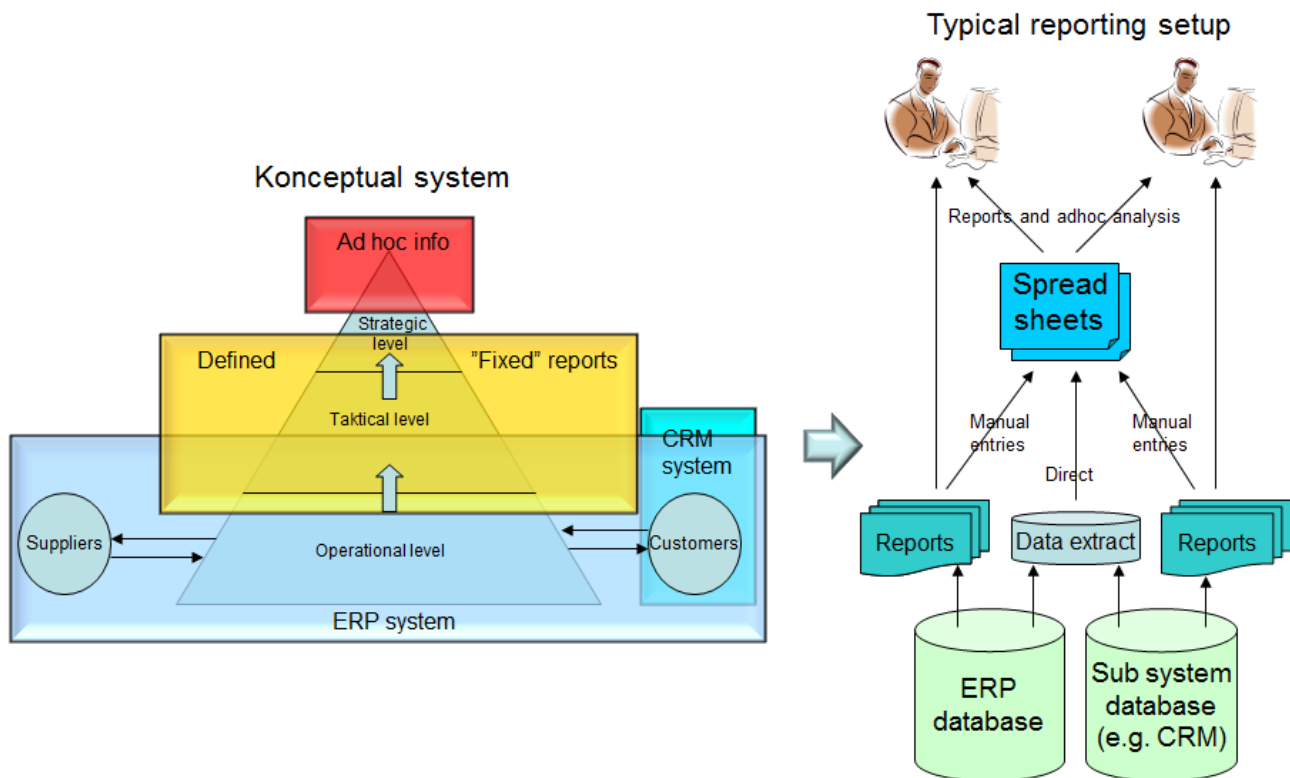
If you need someone to help out with ‘Performance Management’ then describe what this concept covers.

Do you need specific help to implement Balance Scorecard and identify a performance metrics system, or do you need help to improve the organizational style to move towards a more performance oriented culture. In some cases you need consultant’s that has specific knowledge of IT solutions that can help you and in other cases you need specialists in change management.

You need to clearly define what your objective and your goal is.

## Business requirement and what you get

### Typical 'non-data warehouse' system solution



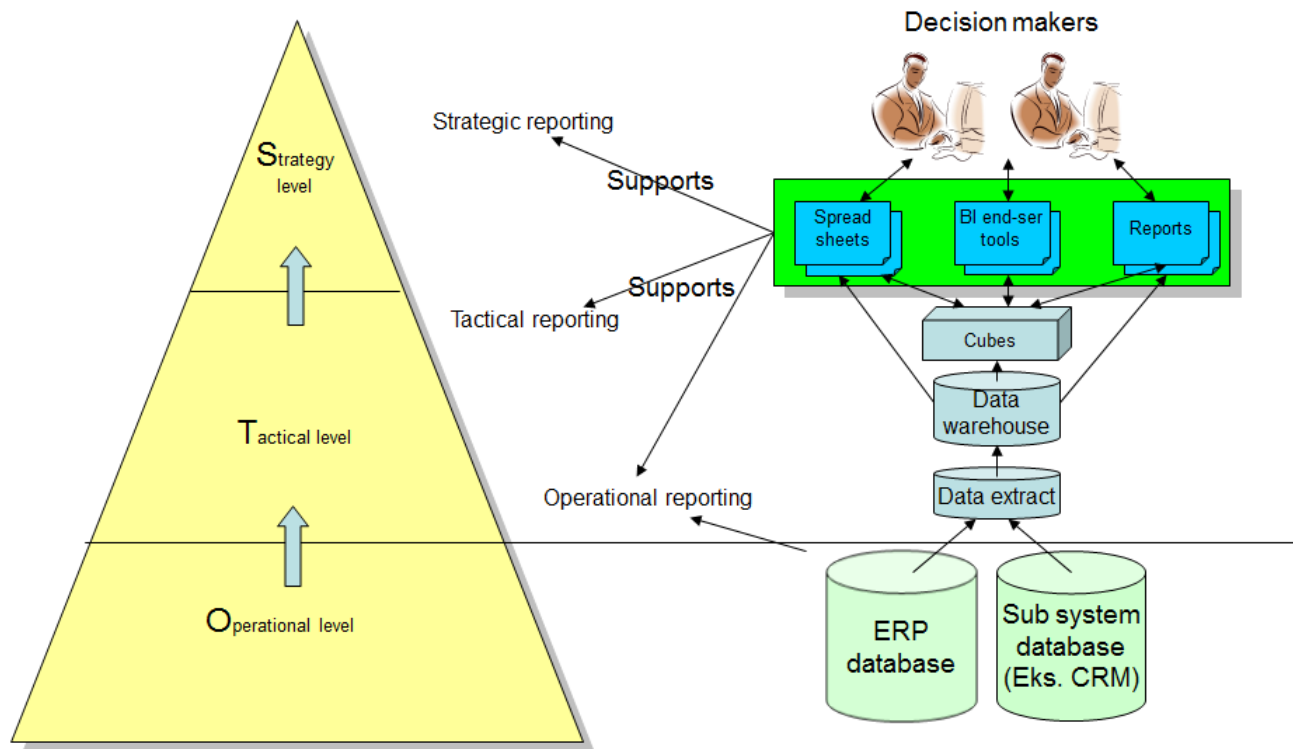
In the model above you see that a company typically have an integrated ERP system that handles total transactional data from customer orders, deliveries, invoicing, customer payments, production (bill of materials, materials planning e.g.) over to procurement from suppliers, warehouse management and much more. Likewise many companies can have a lot of other transactional sub systems like CRM (if not integrated part of the ERP system) and many other sub systems.

In this setup data are typically extracted from the underlying ERP database and sub system databases as mostly fixed reports and often to a minor degree some defined data extracts. Often most of the reporting to tactical and strategic levels is done by using spreadsheet models taking the transactional data and making some defined reporting.

However the transactional systems and databases are built to handle the day-to-day business requirements based on often huge amounts of daily transactions and the data are stored to handle this requirement and not a reporting requirement.

Some systems have seen the need to provide data and information to tactical and strategic levels and have made some effort to create some "built-in" functionality of Business Intelligence functionality directly in the databases by creating some pre-defined summary data, but as each company have its own uniqueness this will not – actually not that often – solve the reporting requirements the company have.

## Building data Warehouses to support analysis and reporting



By introducing a data warehouse as described above you change the reporting model from the typical “old” scenario where you report directly out of the transactional data, but instead extract the transactional data into one or more “reporting databases” where you hold the information – often in a more reporting friendly structure. Information can in some cases be summarized data as the reporting need in some cases do not need the underlying individual transactions but can be some information sums.

Benefits from this approach over the previous approach are like:

- Data reporting do not access (and slow down) the transactional systems by running many reports.
- Information can be validated, summarized or structured to easy reporting (see multi-dimensional reporting described earlier).
- Only “one truth” holding the same information from all underlying transactional systems.
- You discover your problems with the validity of your data (both Master Data and transactional data).

However this approach has also some disadvantages:

- You need to invest in additional Data Warehouse and BI/Cube servers
- You need to invest in the extract and transformation from transactional databases and into a company data warehouse. Often you need consultants to do this work.
- You need to maintain the systems when you change systems and databases over time.

## What is the real business need

If you want to move from the old reporting concept and introducing some kind of a Data Warehouse and multi-dimensional cubes and reporting you need to make your needs clear.

In some cases you might need multi-dimensional functionality but not always. This question is one of the most important questions you can ask about Business Intelligence and this will be discussed in much more detail later. But another question almost as important is the company Master Data management.

## Master Data Management

It has already been discussed in a previous section about the problem with inconsistent master data as well as the problems when master data over time change or gets updated.

Before you go for a Data Warehouse you need to have an idea of the quality of your data in the transactional systems. Often there are 2 methods/events when a company gets a true sanity check of its master data: When you introduce a new ERP system and need the old data converted and when you have to setup a new Data Warehouse.

Company after company has tried to implement new ERP systems and have got a real eye-opener when looking at its master data. Analyzing e.g. the Customer master table gives examples where the same customer has been created multiple times and used multiple times (often with slightly different names). This can be because of a problem identifying the existing customer or because the customer has changed, merged e.g. and you want to keep the historical transactions on the “old” customer.

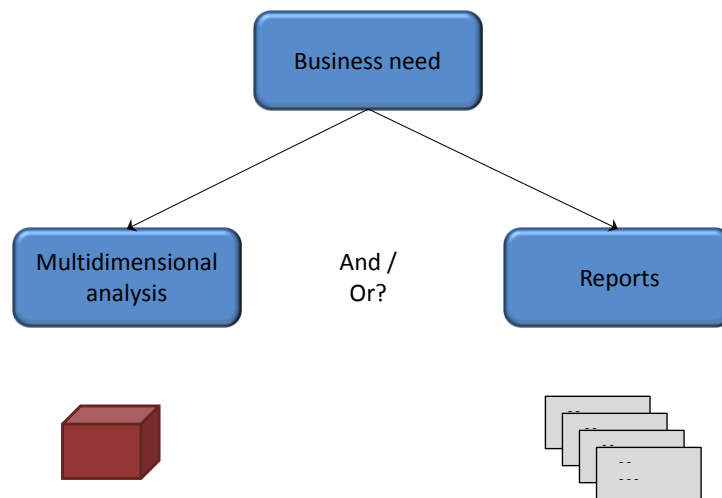
Whatever reason there can be then it is often a problem because the company does not have a Master Data policy and strict management of creation, change/update and deletion of Master Data.

Even if the company does not have valid existing Master Data then before starting a Warehouse project the company needs to define how to ‘scrub and clean’ the existing data as well – and very important – as create systems to make sure that master data are valid for the future.

Often a Data Warehouse and BI project can get out of control and be very expensive if the company has not made the prep work.

Another area you need to think about before introducing a Data Warehouse is how you want to handle changes of Master Data. Will you just take the latest updated master data (ignoring historical info (like change in address, business name e.g.)), or do you need to keep track of historical values of the master data while also being able to report and analyze on the latest information.

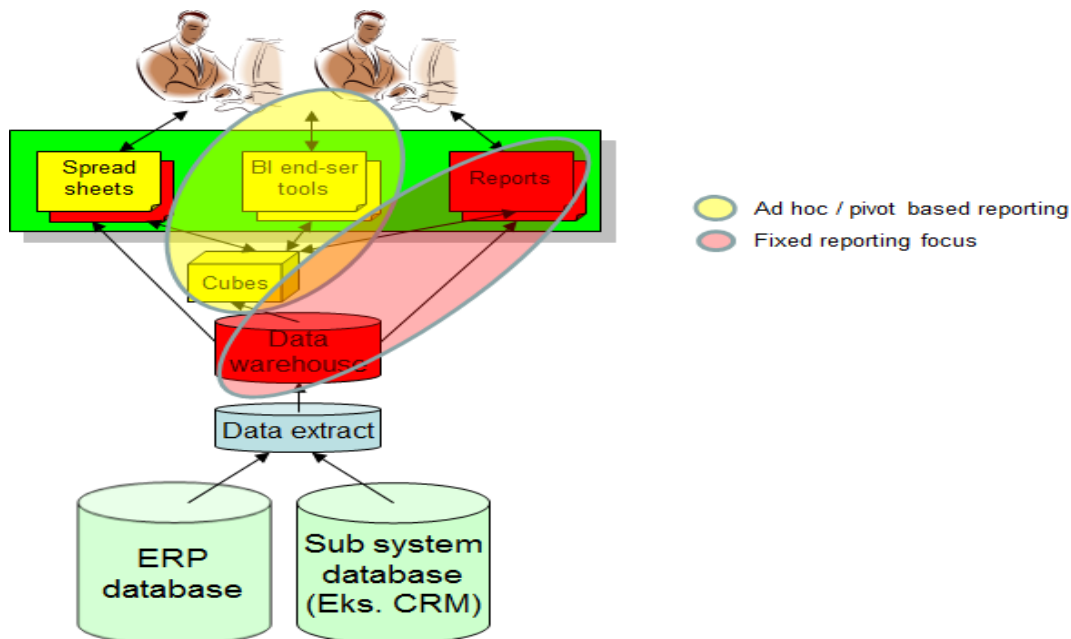
## Business need: Multidimensional analysis or fixed reports



Even a more important question introducing Business Intelligence is about the business needs. You need to make sure what your real business needs are.

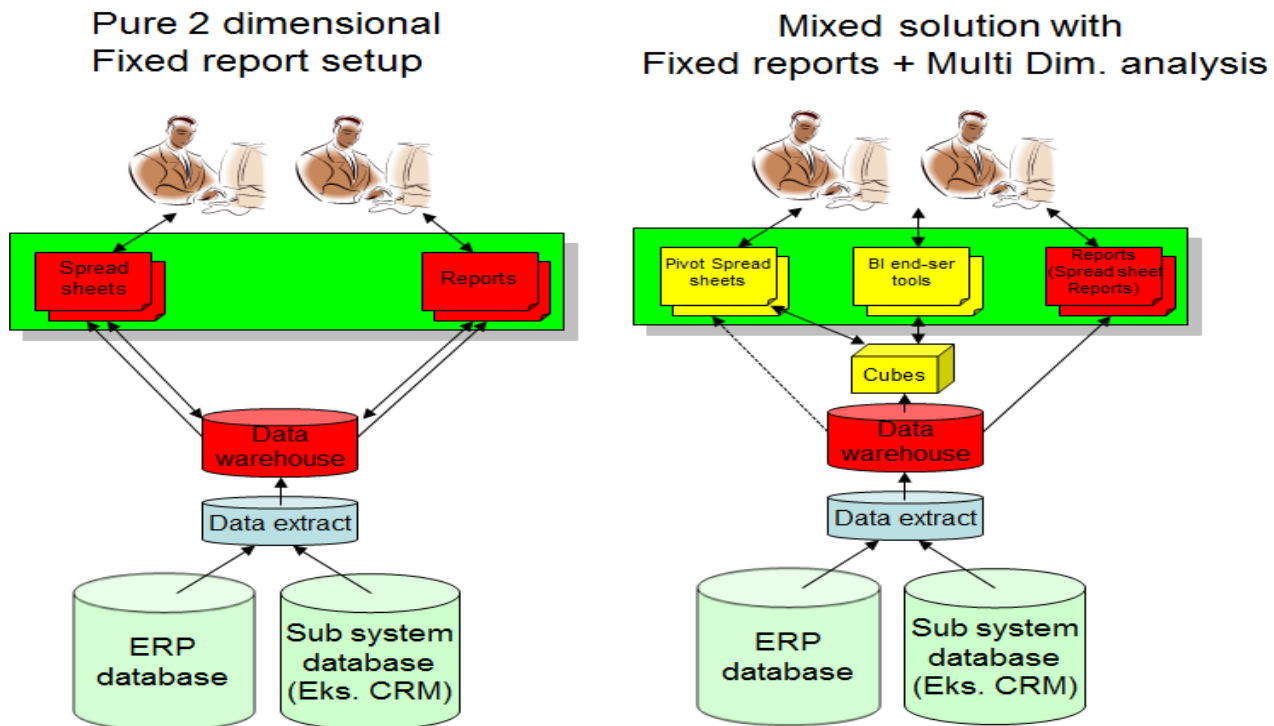
A multi-dimensional reporting setup with cubes to be able to make pivot analysis are extremely helpful for making reporting **of ad-hoc queries against known factual transactional data summarized against multiple dimensions. However this solution is NOT optimized for 2 dimensional fixed report demands.**

You can create a 2-dimensional report out of a multi-dimensional database, but it will often be more than overkill to extract data from the data warehouse into a multi-dimensional database/cube just to extract data into basically fixed 2-dimensional reports (even if these fixed reports can be based on multiple user controlled parameters).



If we look at the same model as introduced before then in the above setup you can see that cube data and BI end user tools that can access and present pivot data (which also pivot table functionality in spreadsheets can be used) then a business need - where you need to analyze data in

many dimensions and aggregated with sums on hierarchy levels in these dimensions - would call for a solution introducing cubes and a multidimensional solution. However if the business need mainly are extract of fixed defined reports (can also be controlled with sums in many dimensions via parameters – but the layout of the presentation has a clearly defined structure) then you would NOT need the step of cubes and multi-dimensional databases but could make this happen directly from the Data Ware house as shown in the figure below:



For some this might seem just irrelevant but the cost of introducing and implementing the cube database and maintaining this is very high and if the data warehouse has been designed correctly then it can always be added later.

Creation of fixed reports based on a multi-dimensional cube are also more tricky as you can't use standard SQL language. but need some specific multi-dimensional language like MDX (**M**ulti**D**imensional **eX**pressions) and the number of consultants that can create a 2-dimensional report with this kind of tools compared to SQL programmers that can create the same output directly from the underlying Data Ware house are extremely limited. You will typically have to pay quite some more per hour having such an expert to do it from the cube database compared to a standard SQL programmer doing the same report from the underlying data ware house. And maintenance of this afterwards will be maybe even more costly.

**Conclusion:** Be aware of what type of information you need. Do not just accept a solution of multi-dimensional cube setup if your basic business need is some fixed layout reports.

And in general I will always suggest that you create fixed reporting directly from the underlying data warehouse (using standard SQL commands) and only extract multidimensional reporting from the cube database. It's only the consulting company that benefits from making it more complex (and need of specialists) by implementing 2-dimensional fixed reports from a multidimensional data base.



## What to look for by choosing consultancy company

Be aware of the following when choosing a Business Intelligence Consultancy Company:

- **Do they only cover one IT solution or many different technologies**
  - Be aware that there can be Pros and Cons using different solutions
  - Are they selling their own product (Software provider and consultants)
    - Company both develops and sell the software (be careful then)
- **What is your real business need and can the consultants solve this optimally**
  - What is your basic information need?
    - Will the suggested solution solve this need?
    - Will you get more than needed and how much will this cost?
  - Be sure that you do not pay for something you don't need
  - Have a clear definition of your need for either reports and/or multidimensional analysis.
- **Consultants competencies**
  - How many projects like yours have the named consultants implemented?
  - How much business understanding do they have?
    - Will the consultants be able to understand you business need?
    - Are they IT technology fixated or do they understand business processes
  - Will they have to use time to 'learn' that you will pay in the project?
  - e.g.
- **Contract**
  - Are the basic information need described in details?
    - Have you made sure that the expected output of the project is described in details
    - Does the contract state how to handle out of scope requests?
      - Immediate estimation of time and price needed
  - Are the contract time and material as an estimate or with a fixed expense?
    - If not a Fixed Expense contract then have at least weekly:
      - Project-to-date expense and remaining estimates
    - What happens if the consultants estimated 20-50-70% e.g. wrong?
  - Make sure that full documentation of the setup and all programs are documented
    - A complete documentation must be in place (if not you are stocked on the consultants to maintain the solution in all future even if you get dissatisfied with their performance and or prices.

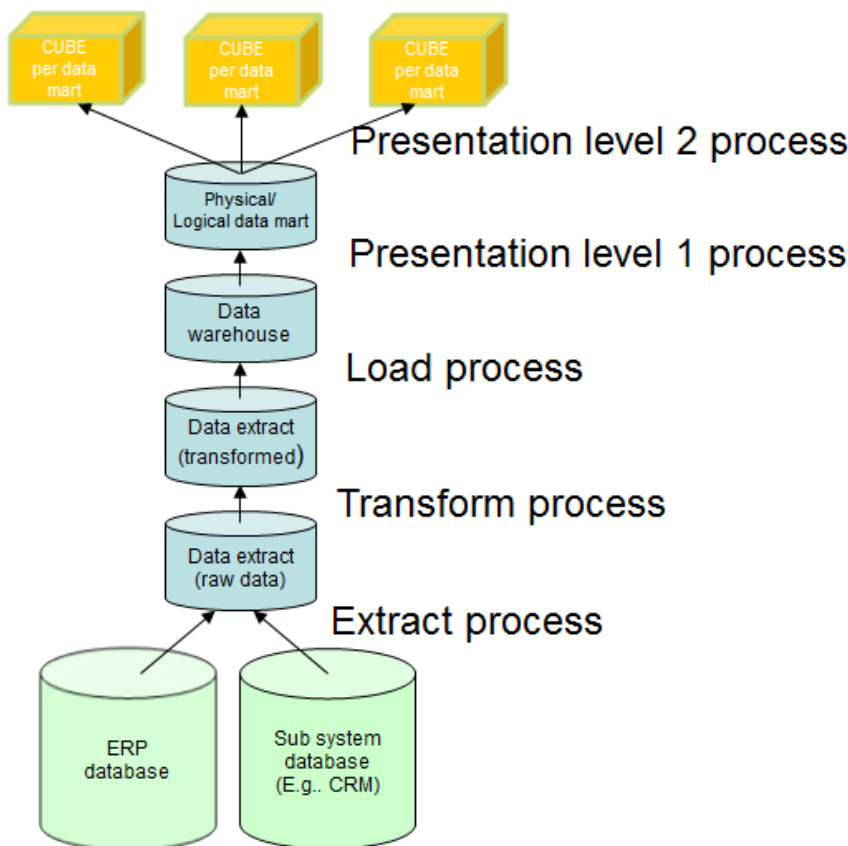
One of the most difficult areas to handle for non-business intelligence experts are how to really define and describe the business need and then being able to identify if the consultants are suggesting a solution that is specific for this need or that you 'buy the standard package'.

## Don't just accept the 'full standard package methodology'

I have worked with consultants that will only implement solutions in one way and without taking the actual business need into consideration and will follow the same BI setup process completely independently of the size of the project and the needed output.



The 'huge' total process are described in the figure below:



In the most extreme form then we have at least 5 different databases and in addition one or more cubes.

## ETL process

Within BI you can often see a term called the ETL process which stands for:

- **Extract**
- **Transform and**
- **Load**

In the figure above you extract data from the underlying transactional databases. The BI purist consultant going for the “ultimate” design will extract data in its raw form (or only transform data in the case it comes from a different database type or source where it is needed to make the ‘nearest’ data type conversion. First when all the data has been extracted to what is often called the stage area then a transformation will be done. There can be introduced many transformations like from one date format to another or calculating some numbers or conversion from one data type like a numeric float number to a big integer, or removing blank characters in text fields making sure that data do not have pre-seeding or additional blanks or zeros e.g.

## Presentation level 1 process – Creation of data marts

There are multiple definitions of what a data mart is but the most general used definition says that a data mart is a sub-set of a Data Warehouse and typically holds some specific data (typically some factual data and dimensions to be used in a cube) to be used in a specific reporting area.

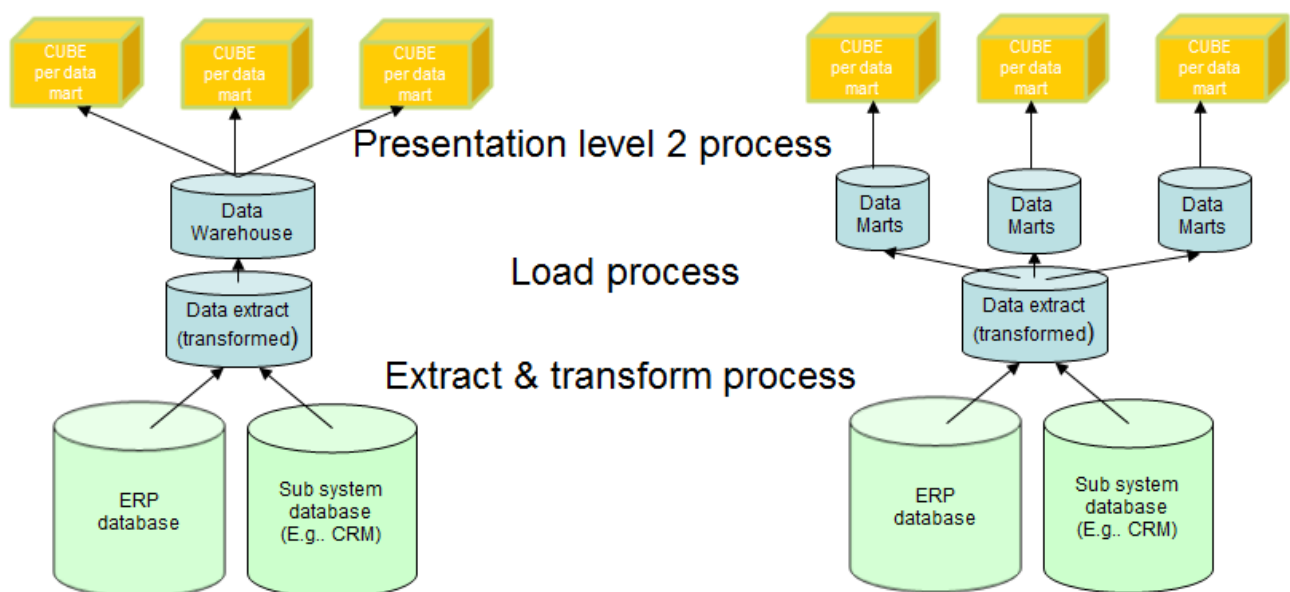
Again some purist will actually create multiple data marts based on the Data Warehouse. Another approach is to create some ‘virtual data marts’ by making data base ‘Views’ where each view is a representation of a virtual data mart).

## Presentation level 2 process – Creation of cube(s)

Last process is to create cubes based on typically the data marts or if this has not been introduced the underlying Data Warehouse.

## Alternative processes

In the figure below there are shown 2 alternative’s to the standard full process flow. In these cases the data extract process and transformation has been consolidated into one so we do not store ‘raw’ data from the underlying data sources (data is still there, so why store it one more time in ‘raw’ form, and likewise in the first one the data mart process has been skipped, and in the next example the data warehouse is actually the sum of 3 data mart databases instead.



Again it is needed to know what the business requirement is. In some cases it will make sense to have 3 data mart databases as in the example above where access rights can be granted on the database level while if the same thing should be done for the first example where we only have a data warehouse with all the data then if different users shall have different rights to view data then this security would be more complex as this would then have to be placed on table level.

Please be aware that each additional step that is introduced in a BI process has some quite considerable costs of programming/setup as well as maintenance. Often the process time will also increase.