Week One

Analysis of data in 3 overlapping disciplines; data analytics, data mining, and business intelligence.

Data Analytics: Further than just visualization of data, the goal of analytics is to support a decision or prove a hypothesis by using quantitative techniques. This confirmatory approach will validate or summarize information to provide the answer.

Data Mining: Describes examining data sets to identify undiscovered patterns and uncover hidden relationships. This exploratory approach may use sophisticated models to determine its patterns.

Business Intelligence: Describes data analysis efforts that is focused on answering analytical questions about the business, supports business management processes, or provides views of the business whether it be enterprise or departmental.

We tie Business Intelligence (BI) with a Data Warehouse (DW) solution.

Transaction Data:

Records business events e.g., retail purchases, call detail records, bank deposits/withdrawals, insurance claims, stock trades/quotes, etc.

Usually recorded along with date and timestamp for each transaction.

Considered raw data or detailed view of data as analysis may be done at a point in time versus looking at each transaction.

Snapshot Data:

Records current or past ‘state’ of a business entity or relationship e.g., customer, account or measures of metric values at a certain point in time.

Unlike transaction data, a query will typically want to access only one “time instance” of the snapshot data e.g., balance on the account for month end close.

Multiple snapshots can be used for trending or constructing averages over time.

Two philosophies:

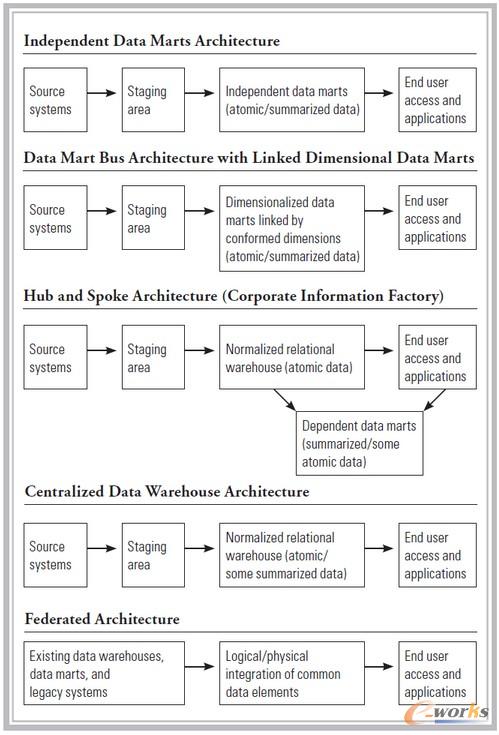
1.Bill Inmon claimed that this uniform structure is in a 3NF\*. Analysis is done off of dimensionally structured data marts\*\* that is produced from this 3NF data warehouse. More of a “top down” approach.

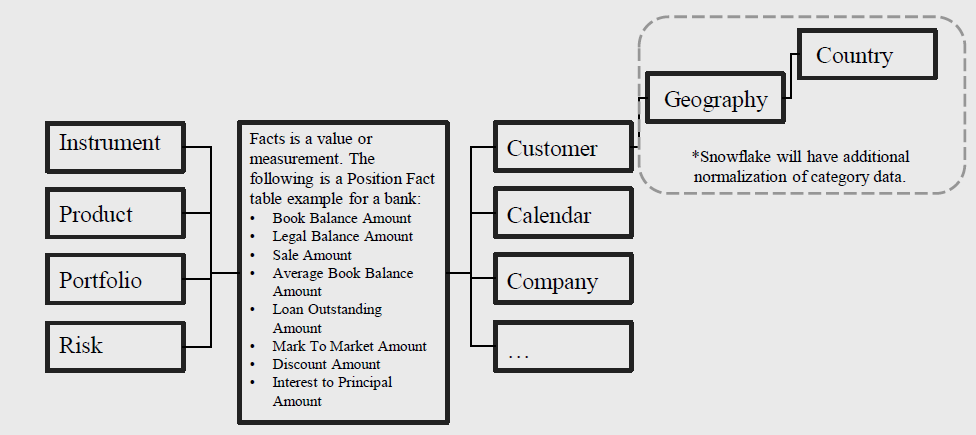
2.Ralph Kimball claimed that this uniform structure is a conglomeration of departmental data marts that may share data through an information bus. Hence, a data warehouse itself may also have a dimensional structure. More of a “bottom up” approach.

**比尔·恩门**的“Building the Data Warehouse”主张建立数据仓库时采用自上而下（DWDM）方式，以第3范式进行数据仓库模型设计，而他生活上的好朋友Ralph Kimball在“The DataWarehouse Toolkit”则是主张自下而上（DMDW）的方式，力推数据集市建设，以致他们的FANS吵闹得差点打了起来，直至恩门推出新的BI架构CIF（Corporation information factory），把Kimball的数据集市包括了进来才算平息。

**自上而下（DWDM）方式**：是基于业务数据驱动模式建设的方法，从统一标准化的数据仓库（DW）到数据集市（DM）建设过程，此方法建设周期长，效率低，投资大。注重底层数据模型、统一标准化、元数据管理等等。  
**自下而上（DMDW）方式**：与DWDM方式相反，以业务需求驱动模式，从某公司业务部门需求出发，先建设单一的数据集市（DM），逐步发展全公司各业务部门数据集市大集中管理，发展到数据仓库(DW)统一标准化建设。本方法在国内大多数企业引用，见效快，周期短，利用率高，分阶段实施。面向用户，满足业务需求为目的，从设计上弱化数据模型，对未来的扩展性需求不能把握，方便对接决策系统、ODS系统、数据仓库系统。  
**两者相结合，“齐头并进”建设：**  
以数据为驱动的系统分析(自上而下)和以业务为驱动的系统分析(自下而上)同时进行。以业务驱动分析为主线，以数据驱动分析为辅。

Bill Inmon的集线器架构/企业信息工厂架构（Hub and Spoke / CIF – Corporate Information Factory）与 Ralph Kimball的数据集市/数据仓库总线架构（Data Mart Bus Architecture/Data Warehouse Bus Architecture）则是DW架构的争论焦点。





Data Warehouse

Source of consistent, integrated, enterprise-wide data

Mechanism providing the analytical and decision support needs of the enterprise

Data is at multiple levels of granularity, including transaction-level and summarization

Data is typically retained for 3-7 years

Data may be sourced from the Operational environment and/or the ODS

Data Mart

Mechanism providing the analytical and decision support needs of a business function

Data is highly summarized and is usually specific to the business function

Data is typically retained for 3-7 years

Data may be sourced from the ODS and/or the Data Warehouse

Operational Data Store (ODS)

Mechanism enabling the collection, cleansing, and integration of operational data for population in either a Data Warehouse or a Data Mart

Data is typically at the transactional level of detail

Data may be retained for short time spans (90-120 days)

**People Management:**

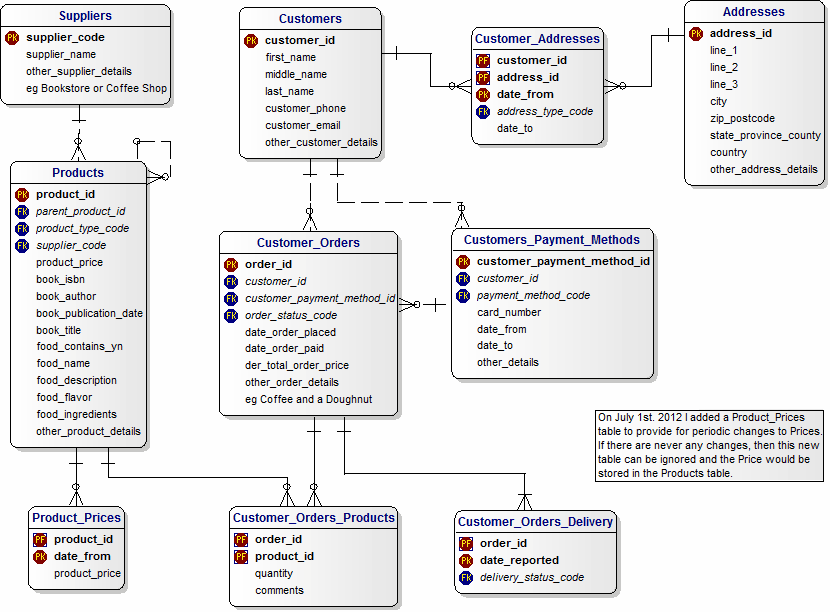
Stakeholder Readiness - *understanding opinions*

Organizational Readiness – *understanding situational challenges*

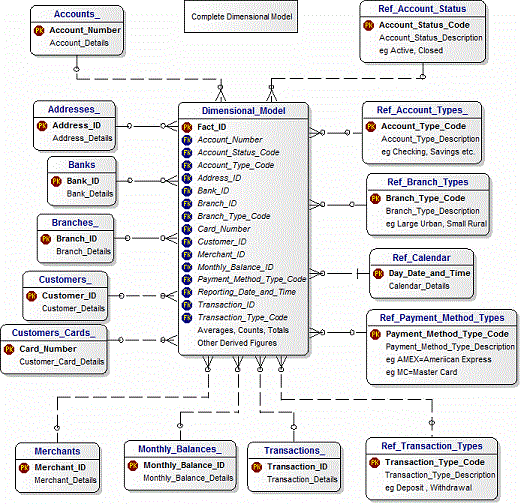
Financial Readiness

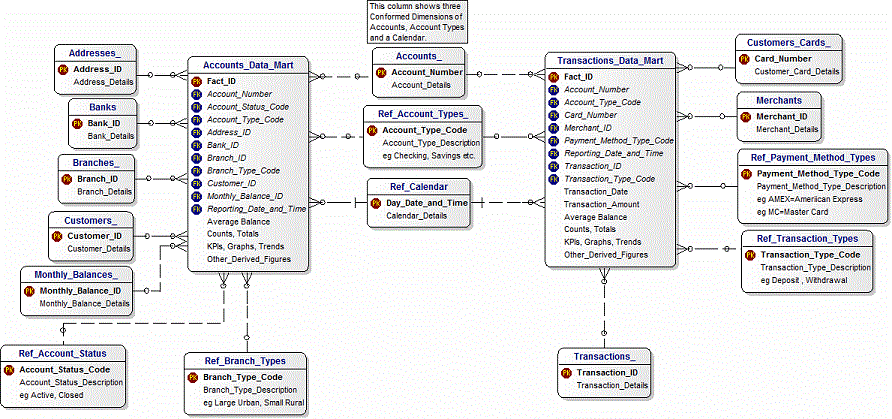
Data & Technology Readiness

**Sample retail data model:**



**Sample Bank data model:**





Data Analysis Methodology:

A.Inspect - All data is inspected and “cleansed”

B.Transform - Data is standardized, improved or derived e.g., profitability, scores

C.Integrate - All of the required data is in one logical structure e.g., transaction, position, account, customer, household, product, instrument, branch

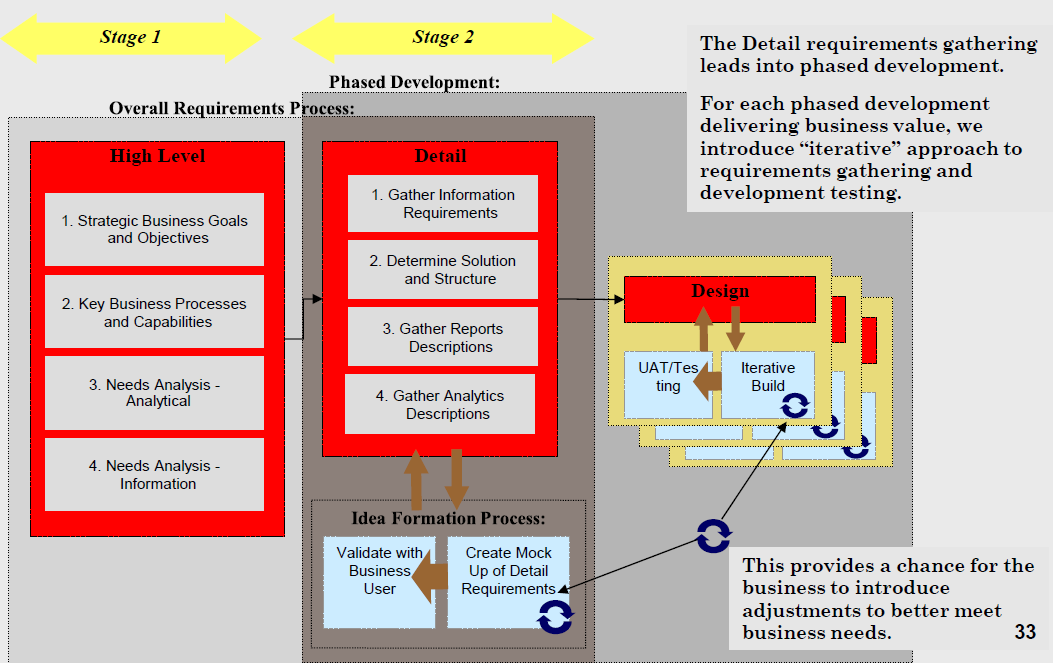
D.Organize BI - Data is stored dimensionally

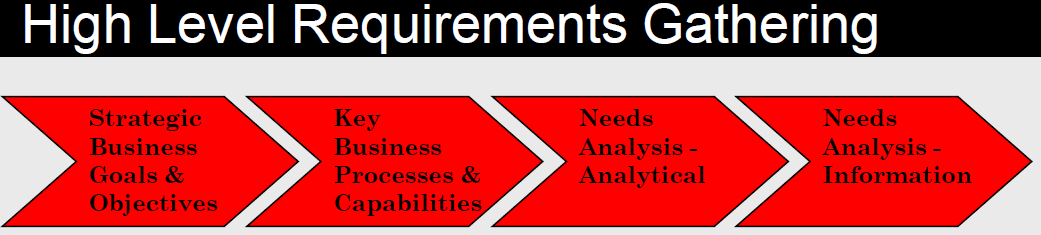
E.Organize A/DM – Data is stored in modeling specific data structure

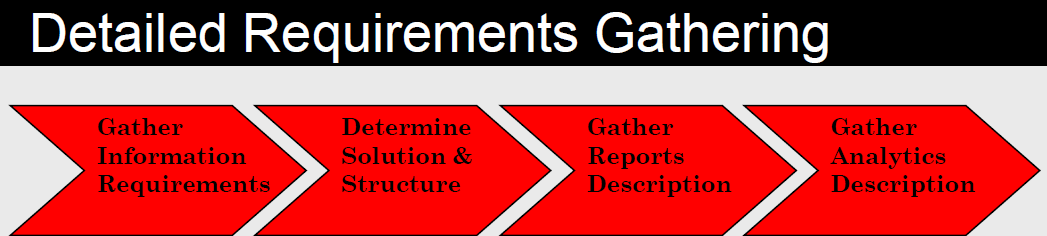
F.Explore - Search for anticipated relationships, unanticipated trends and anomalies:

G.Document - Data definitions and transformation rules are documented and accessible

Requirements Gathering Process:







Excel Basics - cheat sheet

**Navigation/Selection:**

a)CTRL+Tab – rotate through open excel workbooks (ALT+Tab flips through open application)

b)CTRL+PageUp/PageDown – move left/right inside workbook – between sheets

c)CTRL+Arrow – goes to end of cells of “continuous formatting”

d)CTRL+Home – top-left corner of worksheet (unless “freeze panes” is activated)

e)CTRL+End – bottom-right of worksheet

f)Holding shift key + arrow/mouse click (or CTRL+arrow) – selects continuous cells

g)Holding down CTRL key + mouse click – selects multiple cells (copying and pasting will eliminate any un-selected cells between the copied/cut selections)

h)CTRL+[ (or CTRL+]) – navigate to cell linked to/from cell

**Shortcuts:**

a)CTRL+C = copy

b)CTRL+X = cut

c)CTRL+V = paste

d)CTRP+P = print

e)CTRL+S = save

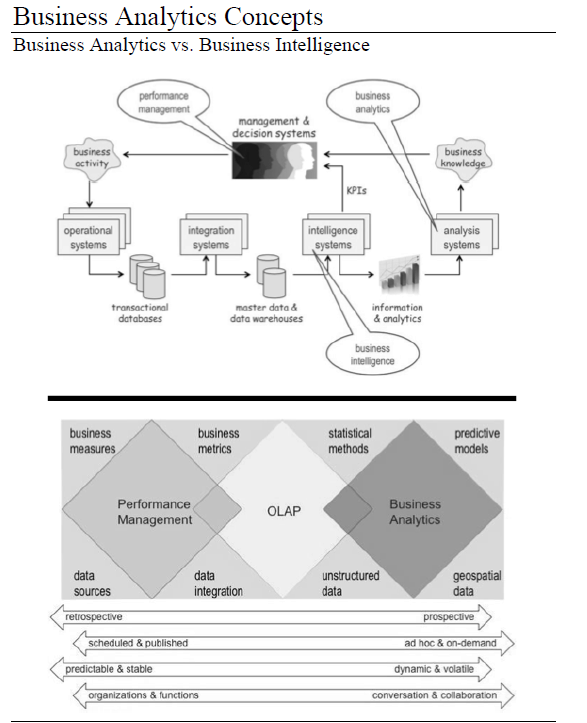
f)CTRL+Z = undo

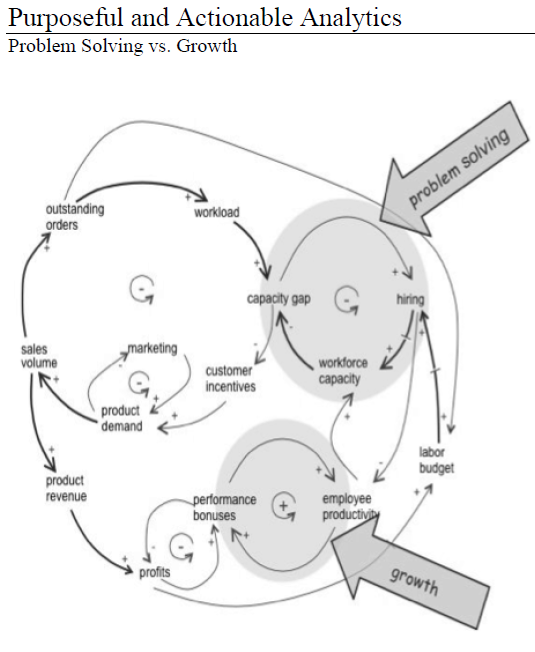
g)CTRL+Y = repeat (also F4 for multiple repeats)

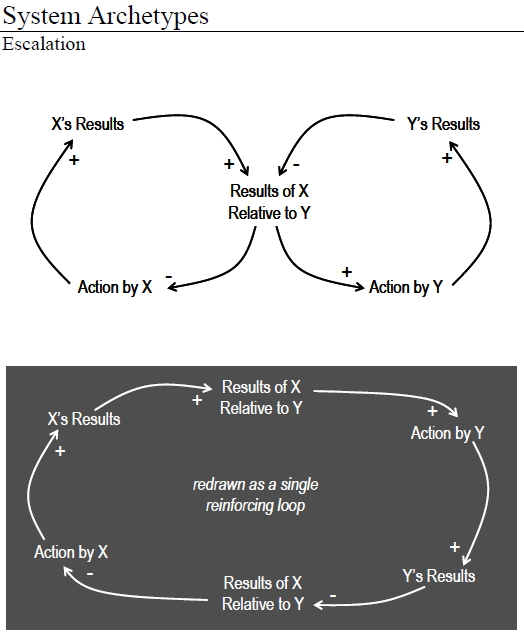
h)CTRL+A = select all cells

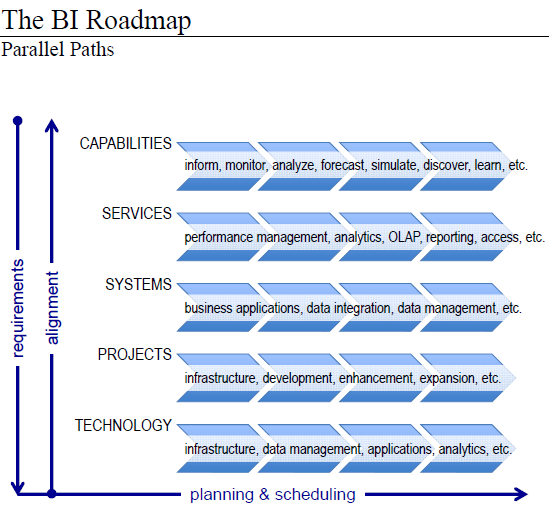
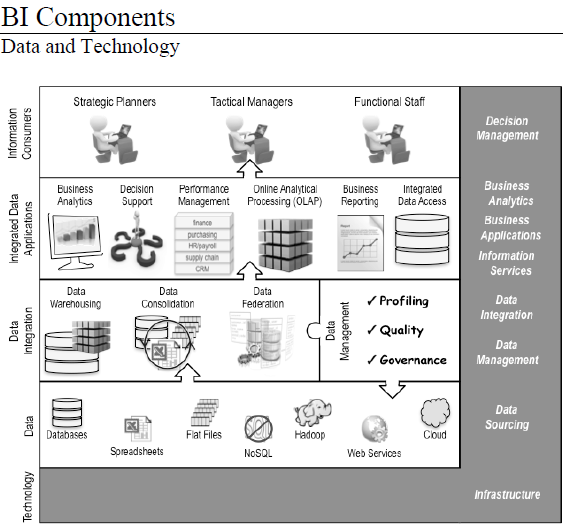
i)F4 = repeat and quick absolute reference toggle (when in cell)

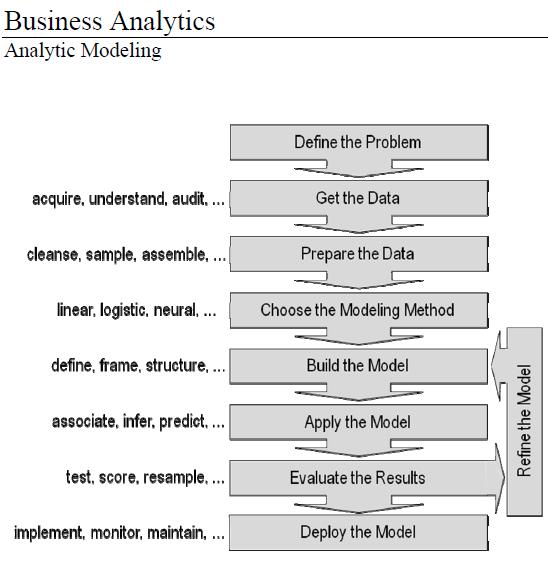
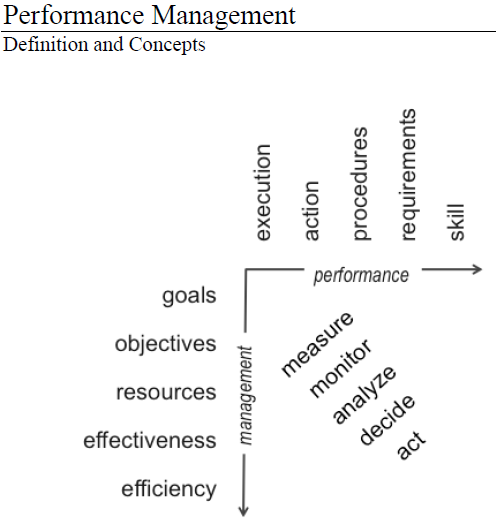
j)CTRL+1 = cell formatting menu

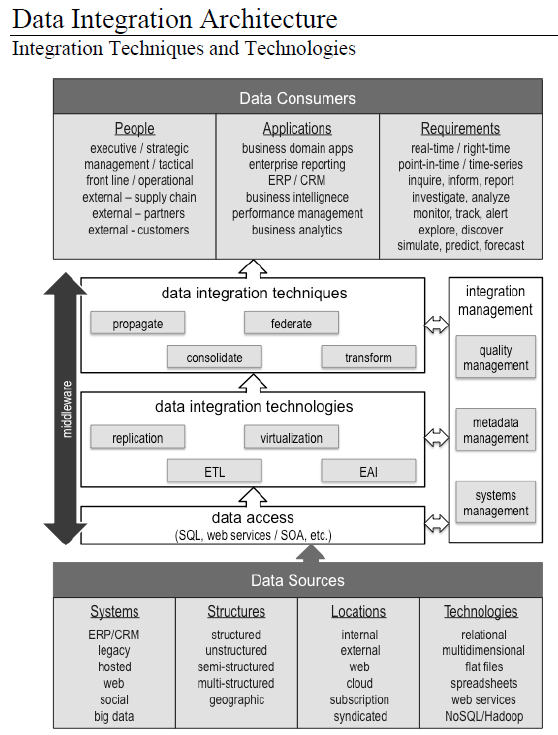


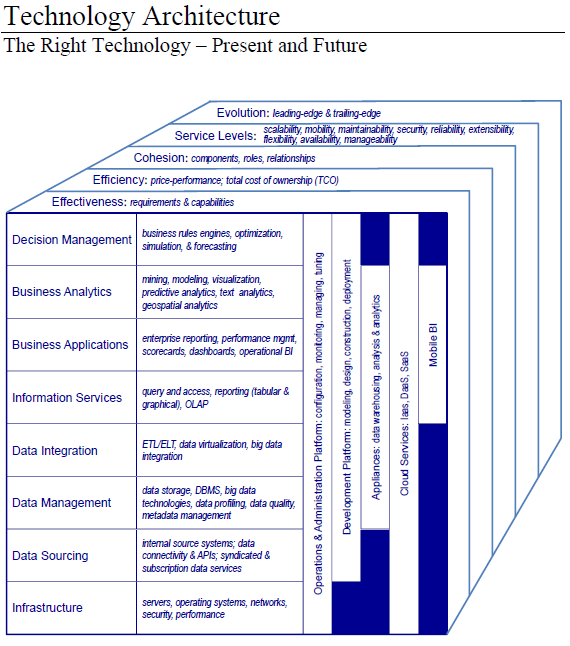


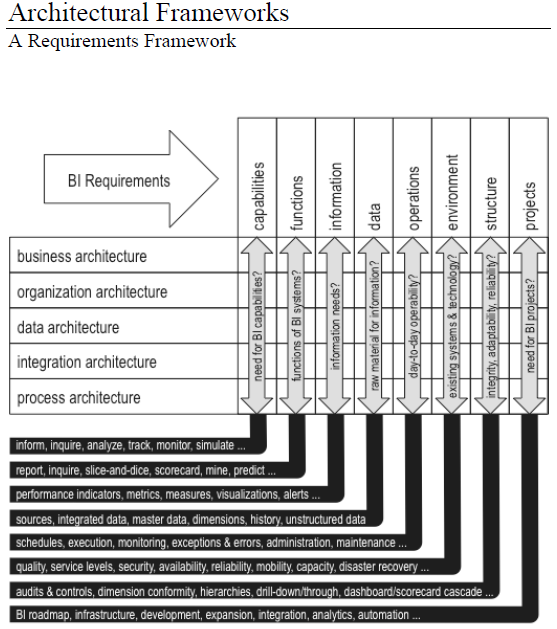




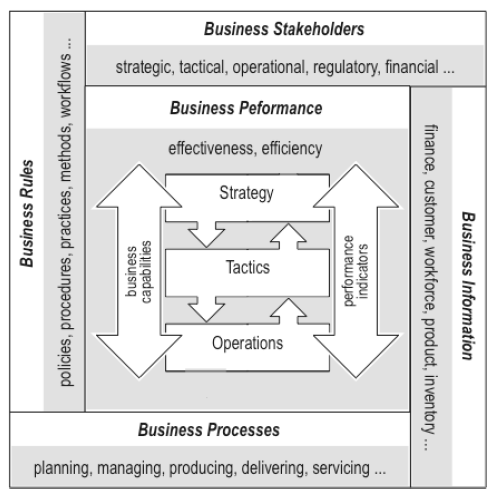








**Business Architecture Concepts**



**Business Performance**

