

Data Engineering & Management

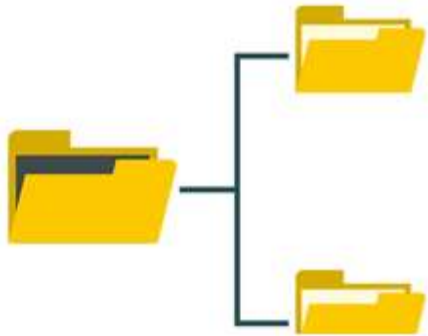
1



DB Architecture for an IS

2

File System - Designed for and associated with app. Pgms.



- Issues

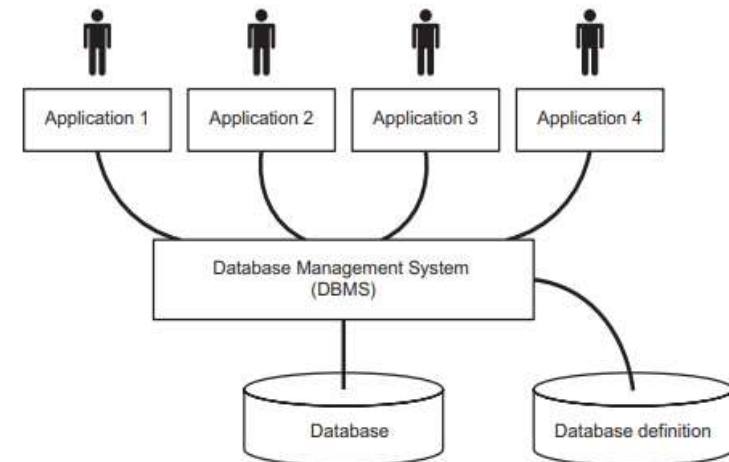
- Sequential & Direct access

- Location depends on Data structure

DB Approach - Two data store

DB and DB defn. - Schema – Set of Data defn.

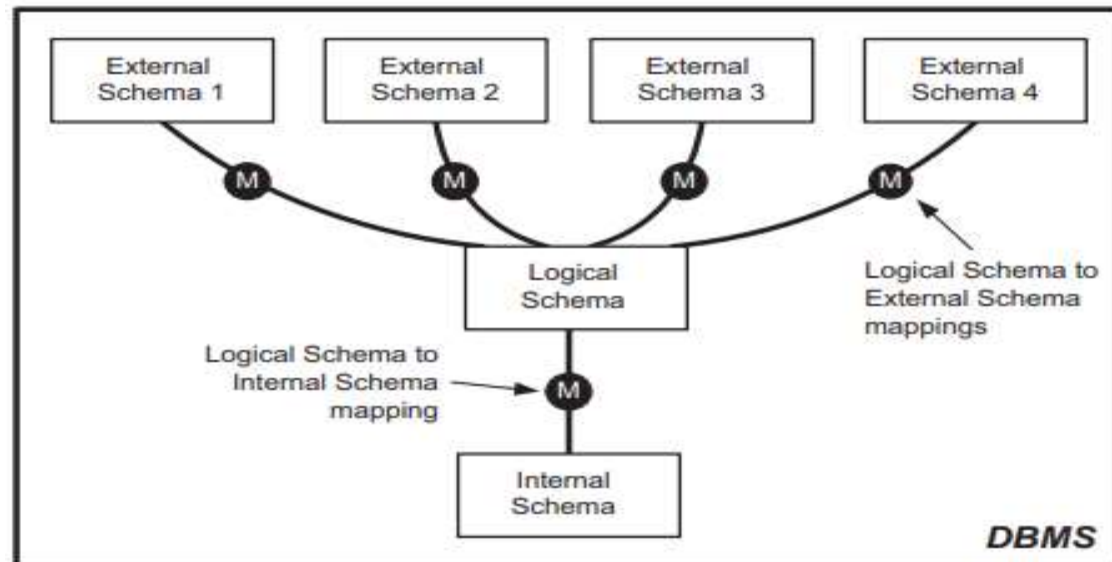
- ❖ Data Independence Integration & sharing of data
- ❖ Consistency of data Data redundancy
- ❖ Integrity controls Access and responsibility
- ❖ Ease development & maintenance



The three-level schema architecture

3

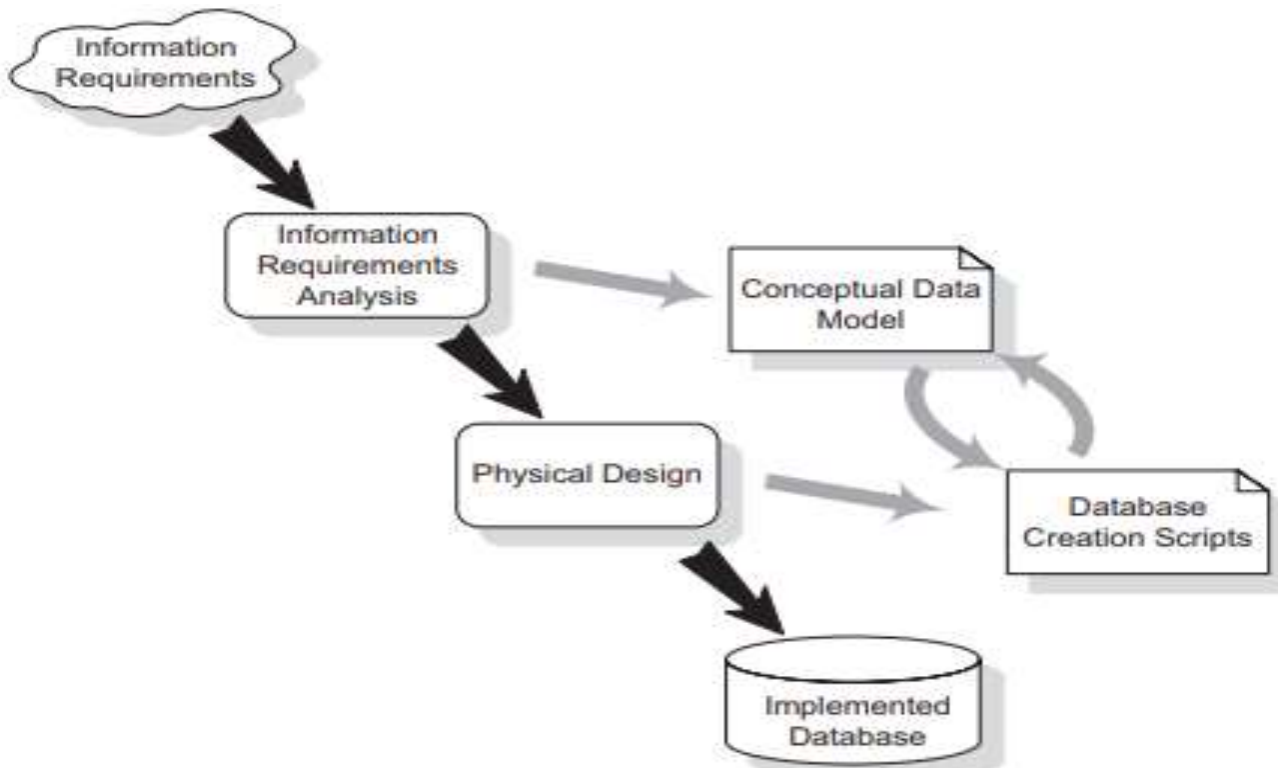
- Proposed by the ANSI Standards Planning and Requirements Committee (SPARC) in 1975
- The schema at the **logical level** is the central, and main, component of the architecture. It defines the properties of all the data.
- The schema at the **internal level** defines how the database is physically stored in files and how these files are accessed.
- Each schema at the **external level** defines the data required to support one or more user processes



DATABASE DEVELOPMENT PROCESS

4

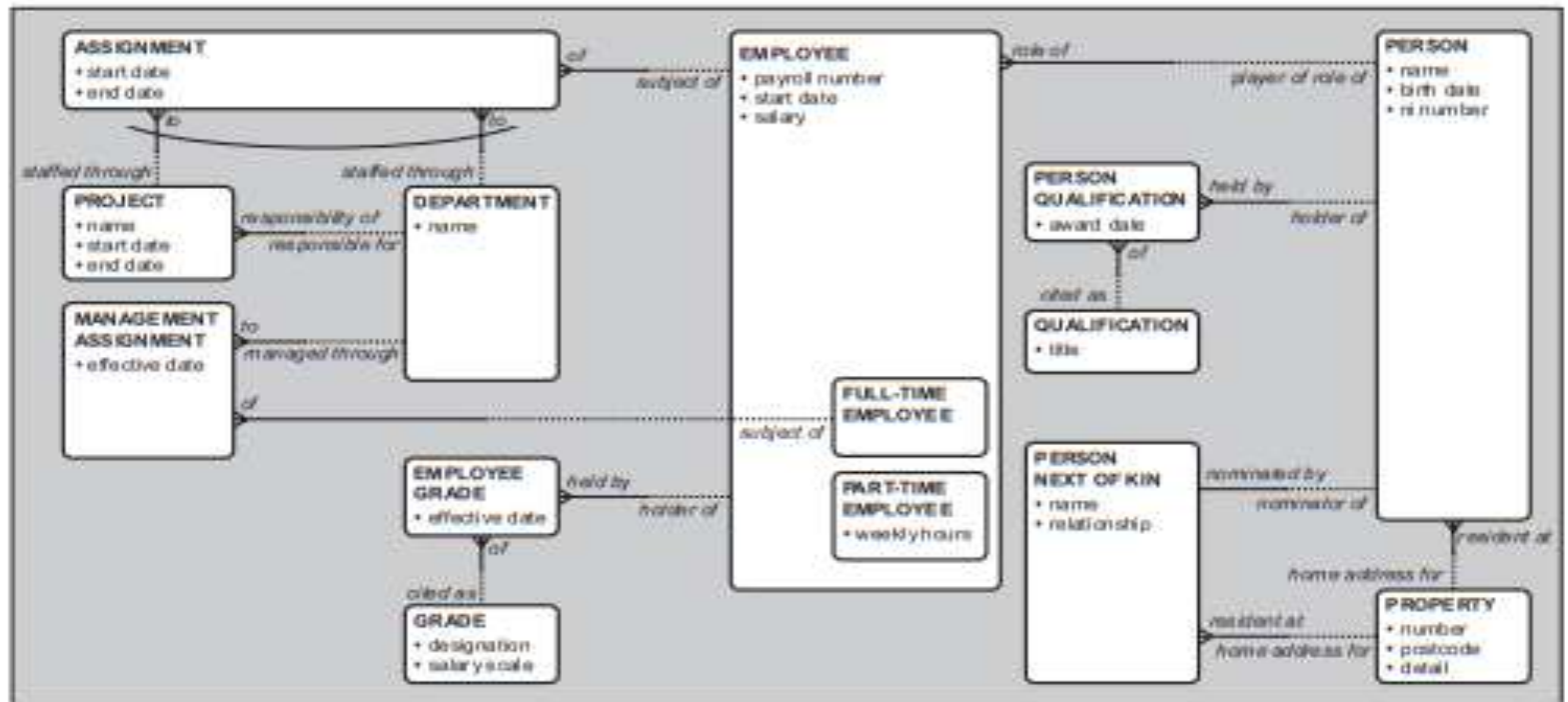
- IS meets requirements – help business process



Conceptual Model

5

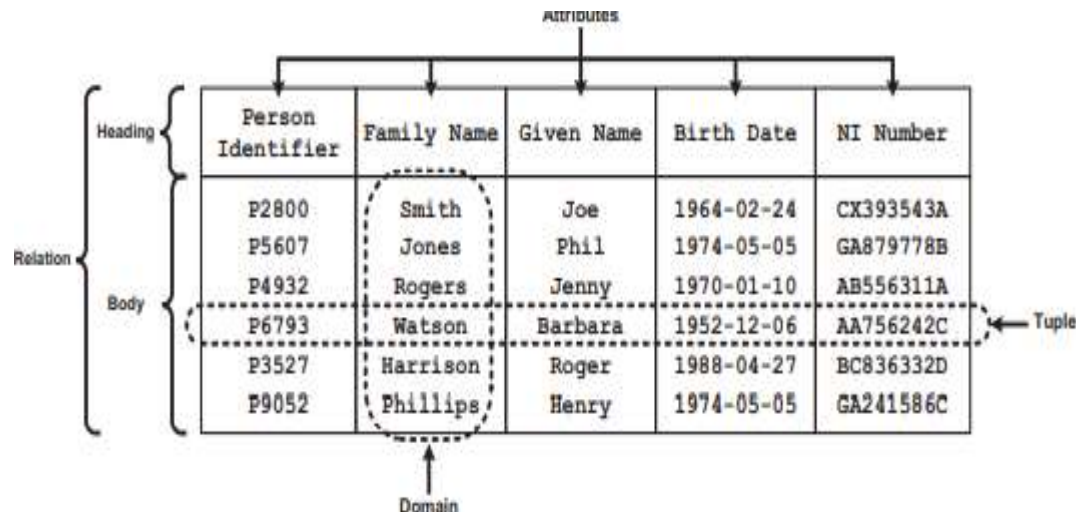
- UML notation – Ellis Barker
- ER Modeling – Richard Barker 1990
- Cons – Covers data requirement our business area & perspective single project – No formal documentation



Relational Model

6

- Relational data analysis is based upon the relational model of data proposed by Edgar F. Codd (1970)
- A relation has two parts – a heading and a body.
 - ▣ The heading consists of a set of attributes
 - ▣ body is a set of elements that are called tuples. Tuple is a set of attribute-name:attribute-value pairs.



Normalization

7

- A large database defined as a single relation may result in data duplication. This repetition of data may result various issues:
- What is Normalization?
 - ▣ Normalization is the process of organizing the data in the database.
 - ▣ Normalization is used to minimize the redundancy from a relation or set of relations. It is also used to eliminate undesirable characteristics like Insertion, Update, and Deletion Anomalies.
 - ▣ Normalization divides the larger table into smaller and links them using relationships.
 - ▣ The normal form is used to reduce redundancy from the database table.

- Why do we need Normalization?
 - ▣ The main reason for normalizing the relations is removing these anomalies.
- **Insertion Anomaly:** Insertion Anomaly refers to when one cannot insert a new tuple into a relationship due to lack of data.
- **Deletion Anomaly:** The delete anomaly refers to the situation where the deletion of data results in the unintended loss of some other important data.
- **Updation Anomaly:** The update anomaly is when an update of a single data value requires multiple rows of data to be updated.

		1NF	2NF	3NF	4NF	5NF
Decomposition of Relation	Conditions	R	R ₁₁ R ₁₂	R ₂₁ R ₂₂ R ₂₃	R ₃₁ R ₃₂ R ₃₃ R ₃₄	R ₄₁ R ₄₂ R ₄₃ R ₄₄ R ₄₅
		Eliminate Repeating Groups	Eliminate Partial Functional Dependency	Eliminate Transitive Dependency	Eliminate Multi-values Dependency	Eliminate Join Dependency

1NF

10

EMPLOYEE table:

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385, 9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389, 8589830302	Punjab

The decomposition of the EMPLOYEE table into 1NF has been shown below:

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385	UP
14	John	9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389	Punjab

In the 2NF, relational must be in 1NF.

In the second normal form, all non-key attributes are fully functional dependent on the primary key

11

TEACHER table

TEACHER_ID	SUBJECT	TEACHER_AGE
25	Chemistry	30
25	Biology	30
47	English	35
83	Math	38
83	Computer	38

TEACHER_DETAIL table:

TEACHER_ID	TEACHER_AGE
25	30
47	35
83	38

TEACHER_SUBJECT table:

TEACHER_ID	SUBJECT
25	Chemistry
25	Biology
47	English
83	Math
83	Computer

EMPLOYEE_DETAIL table:

EMP_ID	EMP_NAME	EMP_ZIP	EMP_STATE	EMP_CITY
222	Harry	201010	UP	Noida
333	Stephan	02228	US	Boston
444	Lan	60007	US	Chicago
555	Katharine	06389	UK	Norwich
666	John	462007	MP	Bhopal

Super key in the table above:

{EMP_ID}, {EMP_ID, EMP_NAME}, {EMP_ID, EMP_NAME, EMP_ZIP}....so on

Candidate key: {EMP_ID}

Non-prime attributes: In the given table, all attributes except EMP_ID are non-prime.

EMPLOYEE table:

EMP_ID	EMP_NAME	EMP_ZIP
222	Harry	201010
333	Stephan	02228
444	Lan	60007
555	Katharine	06389
666	John	462007

EMPLOYEE_ZIP table:

EMP_ZIP	EMP_STATE	EMP_CITY
201010	UP	Noida
02228	US	Boston
60007	US	Chicago

BCNF

13

- A table is in BCNF if every functional dependency $X \rightarrow Y$, X is the super key of the table.
For BCNF, the table should be in 3NF, and for every FD, LHS is super key.
- **Example:** Let's assume there is a company where employees work in more than one department.

EMPLOYEE table:

EMP_ID	EMP_COUNTRY	EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
264	India	Designing	D394	283
264	India	Testing	D394	300
364	UK	Stores	D283	232
364	UK	Developing	D283	549

BCNF

14

- **Example:** Let's assume there is a company where employees work in more than one department.

EMPLOYEE table:

EMP_ID	EMP_COUNTRY	EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
264	India	Designing	D394	283
264	India	Testing	D394	300
364	UK	Stores	D283	232
364	UK	Developing	D283	549

- In the above table Functional dependencies are as follows:

$EMP_ID \rightarrow EMP_COUNTRY$

$EMP_DEPT \rightarrow \{DEPT_TYPE, EMP_DEPT_NO\}$

Candidate key: {EMP-ID, EMP-DEPT}

BCNF

15

- The table is not in BCNF because neither EMP_DEPT nor EMP_ID alone are keys.
- To convert the given table into BCNF, we decompose it into three tables:
 - **EMP_COUNTRY** table:
 - **EMP_DEPT** table:
 - **EMP_DEPT_MAPPING** table:
- **EMPLOYEE** table:

EMP_ID	EMP_COUNTRY	EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
264	India	Designing	D394	283
264	India	Testing	D394	300
364	UK	Stores	D283	232
364	UK	Developing	D283	549

BCNF

16

EMP_COUNTRY table:

EMP_ID	EMP_COUNTRY
264	India
264	India

EMP_DEPT_MAPPING table:

EMP_ID	EMP_DEPT
D394	283
D394	300
D283	232
D283	549

EMP_DEPT table:

EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
Designing	D394	283
Testing	D394	300
Stores	D283	232
Developing	D283	549

Functional dependencies:

$EMP_ID \rightarrow EMP_COUNTRY$

$EMP_DEPT \rightarrow \{DEPT_TYPE, EMP_DEPT_NO\}$

Candidate keys:

For the first table: EMP_ID

For the second table: EMP_DEPT

For the third table: {EMP_ID, EMP_DEPT}

THE ROLES OF A DATA MODEL

17

- Conceptual data models can be developed to understand the information requirements of a business and to form the basis for a physical database design to support the business.
- Understanding information requirements
 - ▣ Aim is highlight the inconsistencies
- The basis for physical database design
 - ▣ A conceptual data model, that is a data model that has been developed to understand information requirements
- Physical database design
 - ▣ Once the conceptual data model is complete it provides the start point for the design of the physical database

THE ROLES OF A DATA MODEL

18

- 'Physical database design' process has two stages:
 - ▣ First-cut database design
 - ▣ Optimised database design.
- Understanding information requirements
 - ▣ Aim is highlight the inconsistencies
- The basis for physical database design
 - ▣ A conceptual data model, that is a data model that has been developed to understand information requirements
- Physical database design
 - ▣ Once the conceptual data model is complete it provides the start point for the design of the physical database

THE ROLES OF A DATA MODEL

19

- First-cut database design
 - ▣ The aim is to use the conceptual constructs of the logical-level schema of the target database management system to develop a design that matches the conceptual data model as closely as possible.
 - ▣ Each entity type conceptual data model becomes a table, with each of the attributes of the entity type becoming a column of that table. If the foreign keys needed add the column if not exists
 - ▣ Naming of tables and column relates to schema
 - ▣ There has to be an explicit primary key declaration for each table
 - ▣ Another important element of the first-cut database design is the specification of the file storage

THE ROLES OF A DATA MODEL

20

- Optimised database design stage
 - ▣ The first-cut design gives a database design that closely resembles the conceptual model
 - ▣ Such a design should be robust, easy to understand and meet all the requirements
 - ▣ Enhance or move away from the first-cut design to improve the performance of the DB, there are two stages
 - Make use of the built-in facilities of the database management system;
 - Compromise on the design of the logical schema.
 - ▣ Two facilities provided by most database management systems are
 - The ability to cluster data and the ability to create indexes.
 - Data clustering means arranging data
 - Indexing provide another way to access

Data Management

21

- Data management is a corporate service which helps with the provision of information services by controlling or co-ordinating the definitions and usage of reliable and relevant data.
- The Problems Encountered Without Data Management
 - ▣ The information systems within the enterprise cannot be interfaced.
 - ▣ Data is not shared between the information systems. Even if it is possible
 - ▣ Communication breaks down and information gets lost.
 - ▣ Data is unnecessarily transcribed and rekeyed.
 - ▣ The wheel keeps being reinvented.
 - ▣ The competitive edge of the organization is reduced.
 - ▣ Frustration sets in

Data management responsibilities

22

- Strategically supports the corporate definition, management and use of business data
- Operationally supports the development and maintenance of computerised information systems
- Key areas of responsibilities
 - ▣ Achieving recognition of data
 - ▣ Improving the quality of data
 - ▣ Facilitating information share
 - ▣ Making various level of management
 - ▣ Achieving single source of reference data to support
- To fulfil the above responsibilities, the data management function needs to identify the specific activities that it needs to carry out and then obtain sufficient resources to perform the activities.

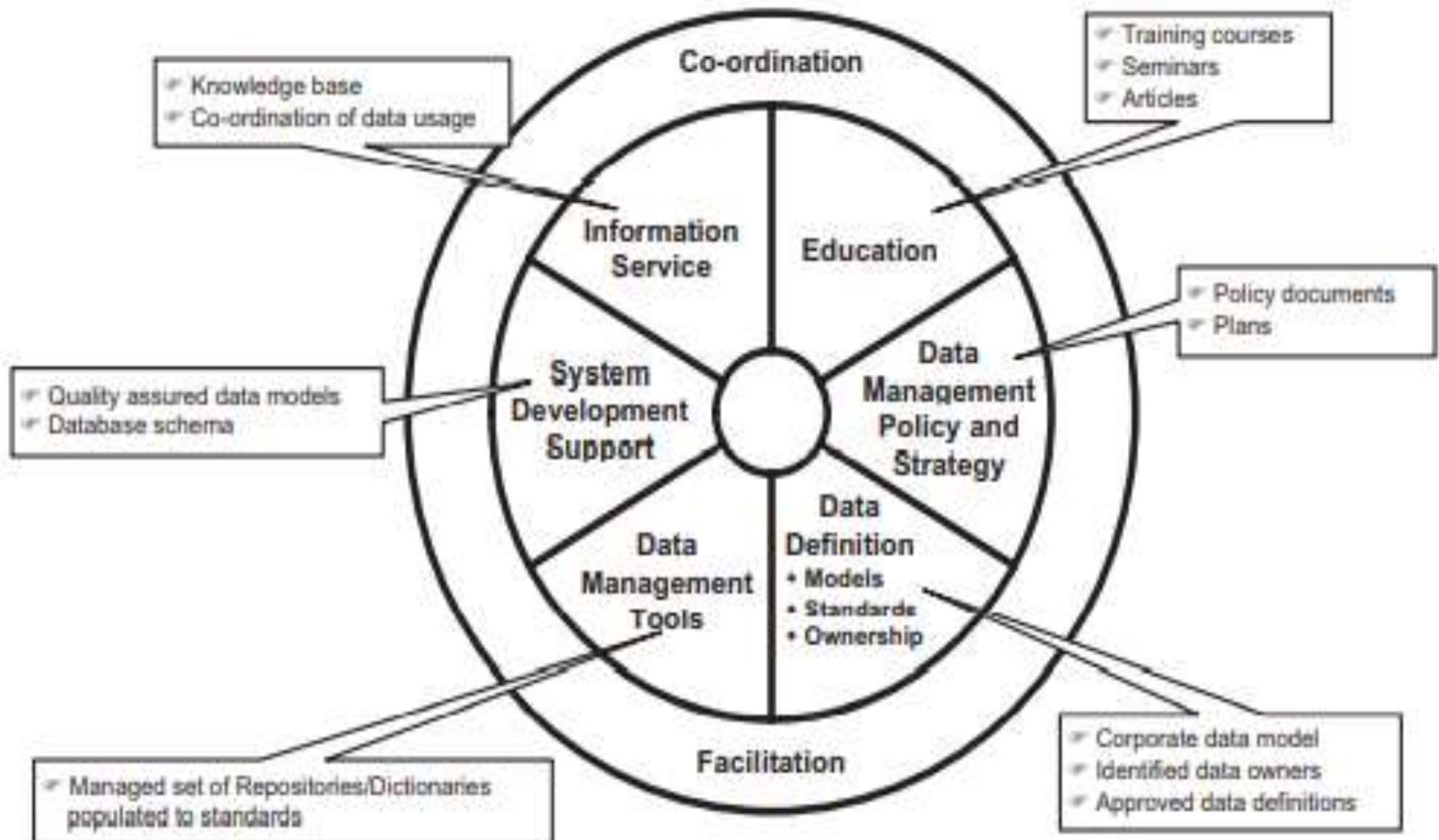
Data management activities

23

- Educate the importance of data mgmt
- Develop the corporate data definitions
- The functional mgmt ensure all activities are coordinated & facilitated
- The IT or IS department is often cast in the role of 'advocate' for the creation and implementation of a data management function.

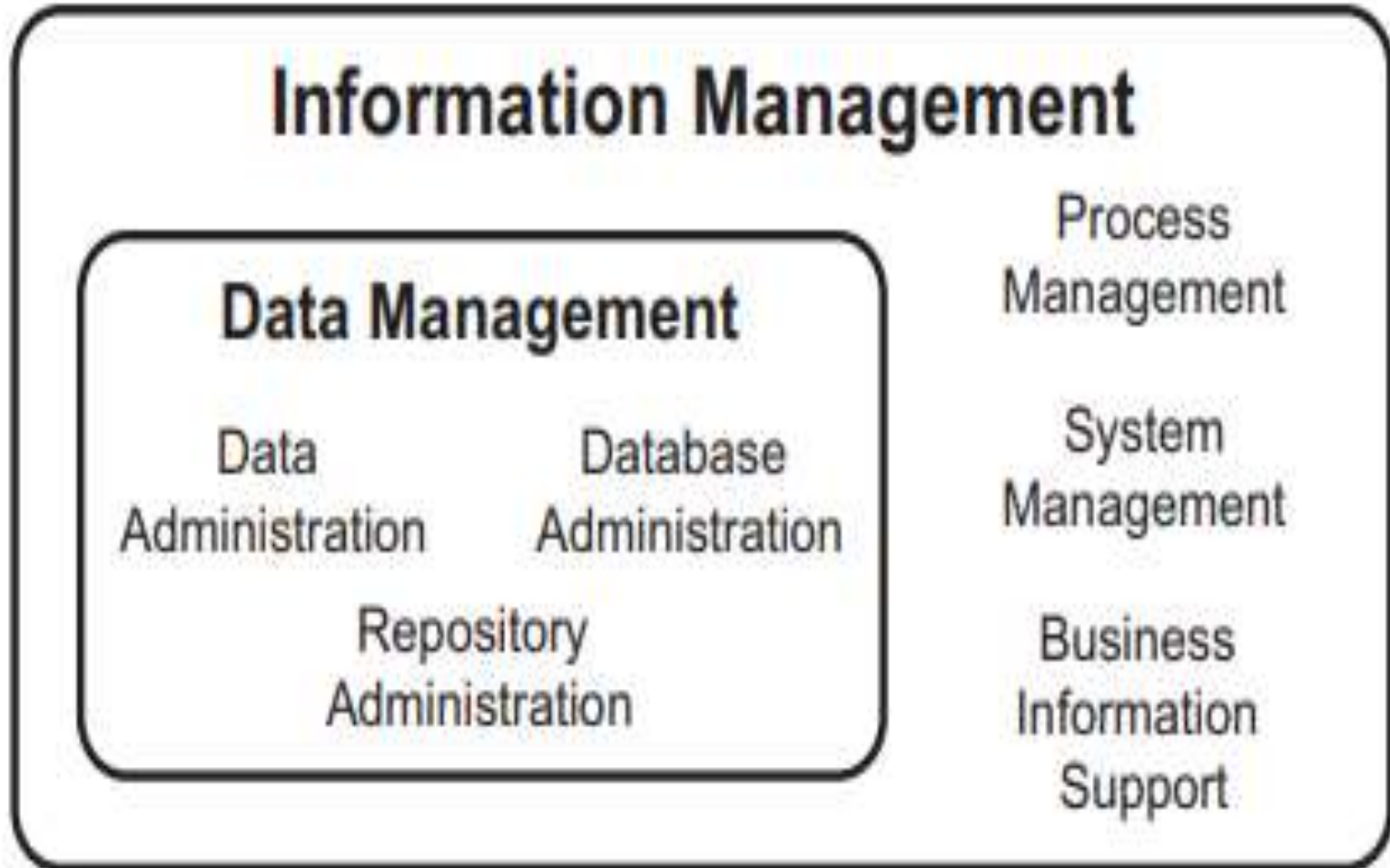
Data management deliverables

24



Roles within data management

25



Roles within data management

26

□ Information Management

- ▣ Process Mgmt – Looks at the processes that use data;
- ▣ System management -which is the management of the computer systems that support the business processes;
- ▣ Business information support, which provides a service to the business users to enable them to exploit the information available.

□ Data Management

- ▣ Data administration – Mechanism of definitions, QC & accessibility
- ▣ Database administration - is concerned with the management and control of the software used to access physical data.
- ▣ Repository administration - is concerned with the management and control of the software in which 'information about information' is stored, manipulated and defined.

Benefits of data management

27

- Benefits related to cost savings
- Business related benefits
 - ▣ Increased availability of information
 - ▣ Improvement in data quality
 - ▣ Improvement in overall efficiency & effectiveness
 - ▣ Helps to improve the customer service & competition edge
- Information related benefits
 - ▣ Reuse of Information and data analysis products
 - ▣ Increase productivity in Sys. Development & leading to cost saving

The relationship between data management and enterprise architecture

28

- US Department of Defense Architecture Framework (DoDAF)
- UK Ministry of Defence Architecture Framework (MODAF)
 - ▣ DoDAF (dodcio.defense.gov/dodaf20.aspx) and MODAF (www.modaf.com) are specifically targeted at the defence communities of the respective nations.
- Open Group Architecture Framework (TOGAF)
 - ▣ TOGAF (www.opengroup.org/togaf) is an enterprise architecture framework, developed by members of The Open Group Architecture Forum (www.opengroup.org/architecture), that uses models at four levels: Business, Application, Data and Technology.
- Framework for Enterprise Architecture developed by John Zachman
 - ▣ Consists of six column & Six rows

The relationship between data management and enterprise architecture

29

□ Framework for Enterprise Architecture developed by John Zachman

□ Consists of six column –

- What – Inventory sets
- How – Process flow
- Where – Distribution network
- Who – Responsibility assignments
- When – Timing cycle
- Why – Motivation intentions

□ Six rows

- The ‘executive perspective’ – the view of the business context planners
- The ‘business management perspective’ – the view of the business concept owners
- The ‘architect perspective’ – the view of the business logic designers
- The ‘engineer perspective’ – the view of the business physics builders
- The ‘technician perspective’ – the view of the business component implementers
- The sixth row represents the functioning enterprise

Corporate Data Model

30

- Why? - Relates similar to Conceptual data model
 - ▣ It comprises
 - Business interest – Entity
 - Info about the things- Attributes
- Is the corporate model used as
 - ▣ Business model
 - ▣ DB design model
 - ▣ Interface design model
- Nature of Corporate data model
 - ▣ Scope – Beyond the single Information System
 - ▣ Support all future systems

DEVELOP A CORPORATE DATA MODEL

31

- Three approach
 - ▣ Attribute-trawling approach (Bottom approach)
 - ▣ Joining multiple project model or Area model
 - ▣ Top down approach
- Attribute trawling approach
 - ▣ Studying all existing information
 - ▣ Collecting data definitions
 - ▣ Sorting and Reuse the definitions
- Problems
 - ▣ Some Business processes not supported by IS
 - ▣ IS will not meet the requirements
 - ▣ It is unclear the data definitions are analyzed and compared

DEVELOP A CORPORATE DATA MODEL

32

- Joining multiple project model or Area model
 - ▣ It involves the independent modeling of information or data requirements of the separate business areas within the enterprise.
 - ▣ These models are o form an enterprise-wide corporate data model.
 - ▣ However, this approach often fails, even when a common set of modelling standards is used.
 - The reason for these failures is exactly because the models are developed independently of each other.
- Top down approach
 - ▣ Building top-down implies the development of a single conceptual data model that, from its inception, is intended to cover the complete information and data requirements of enterprise
 - The best approach is to build the corporate data model 'topdown', starting from a core or framework model that represents the major objects and Concepts of business

CORPORATE DATA MODEL PRINCIPLES

33

- Develop the model 'top-down'
- Give primacy to the core business
- Cover the whole enterprise
- Future-proof the model
- Develop co-operatively
- Gain consensus, not perfection

DATA NAMING CONVENTIONS

34

- The purpose of a data naming convention
 - ▣ To provide consistent, unique and meaningful names for all existing and new items within the enterprise's common data resource.
 - ▣ A consistent approach to data naming should be applied across the enterprise to help achieve unambiguous understanding of data.
- A typical data naming convention
 - ▣ Either single words or a number of words, in a precise and predefined manner.
 - Mandatory prime term that provides the context of the data, which normally means the entity type or table holding the 'data item';
 - One or more optional modifier terms that are used to make the meaning of the data explicit;
 - A mandatory class term that indicates the 'class' of the data.
- Two problems : Over prescriptive & May not deliver what expected

DATA DEFINITIONS & NAMING

35

A data definition with validation criteria and valid operations

Name or Label	Salary
Synonyms	Rumuneration
Significance Statement	has significance as the annual reward, expressed in whole pounds Sterling, paid to an employee of the company before the addition of any extra payments for overtime and/or performance bonuses and before the application of any statutory deductions
Format	Currency
Value List	Not applicable
Validation Criteria	> 0
Valid Operations	multiply by number; answer is currency divide by number; answer is currency add currency; answer is currency subtract currency; answer is currency divide by currency; answer is ratio
Ownership	HR Director
Users	HR Management System, Payroll System
Source	Interview Deputy HR Manager, 15 Sep 06
Comments	[None]
Date created	15 Sep 06
Author	K F Gordon
Date last updated	26 Feb 13

Data Quality

36

- Quality – Fitness of use
- Data quality – Satisfy the requirements of intended use
- Poor data quality – Issues leads to
 - ▣ Rework
 - ▣ Implement new systems
 - ▣ Delay in delivery
 - ▣ Lost of customers
 - ▣ Supply chain problem

Data Quality

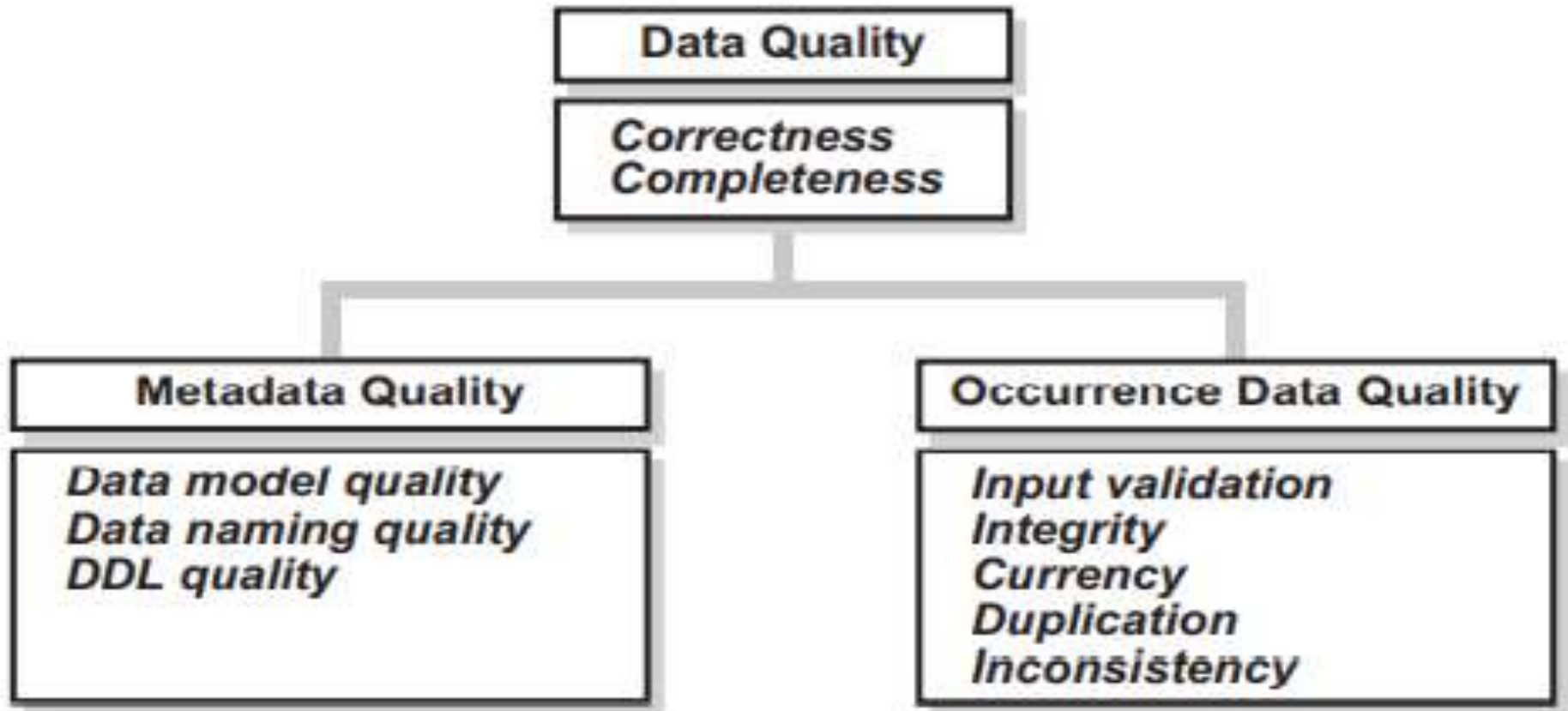
37

- Causes of Poor data quality
 - ▣ Inappropriate schemas in DB
 - ▣ Data entry errors
 - ▣ Data decay (destroyed over time)
 - ▣ Data corruption
 - ▣ Lack of understanding

Dimension of Data Quality

38

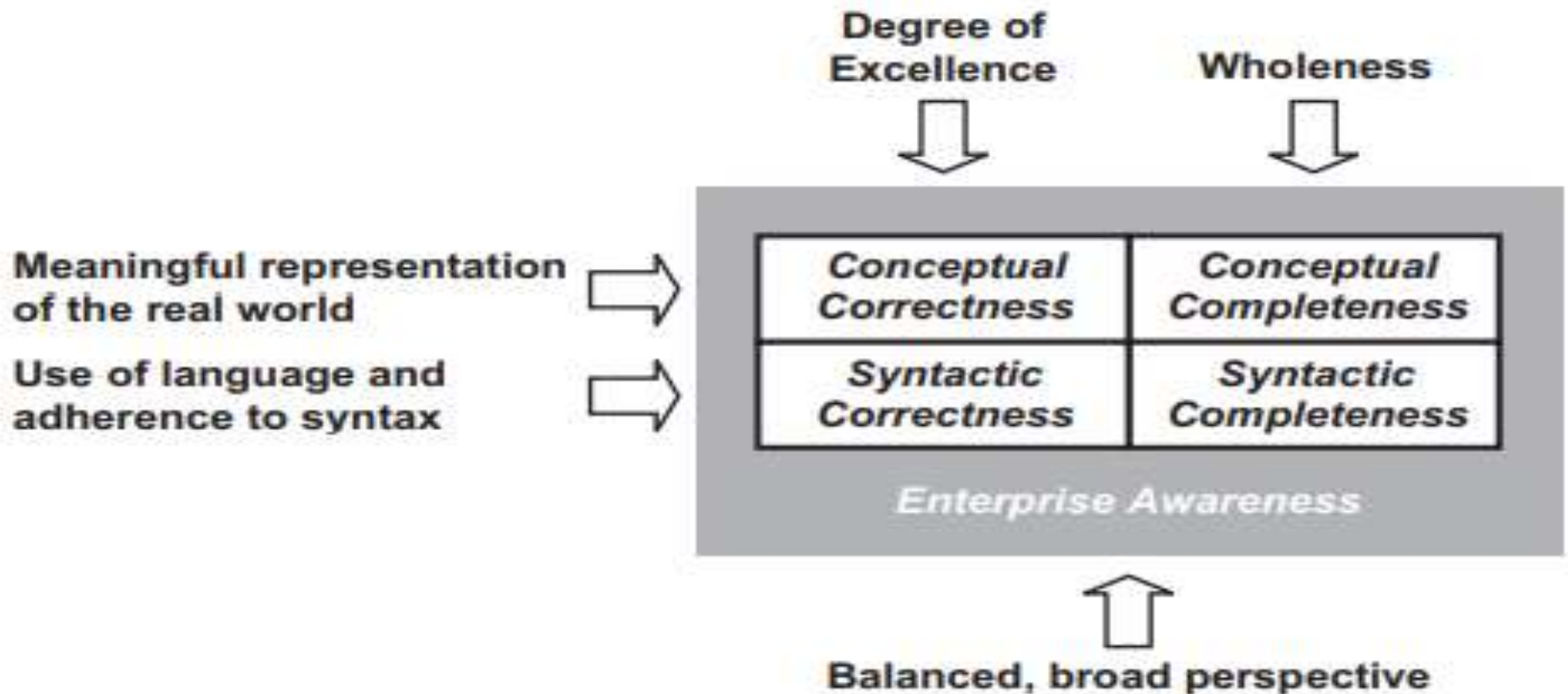
- Completeness – Assesses - data reflect to real world
- Correctness –Assesses – Data has constraints & validation



Data model Quality

39

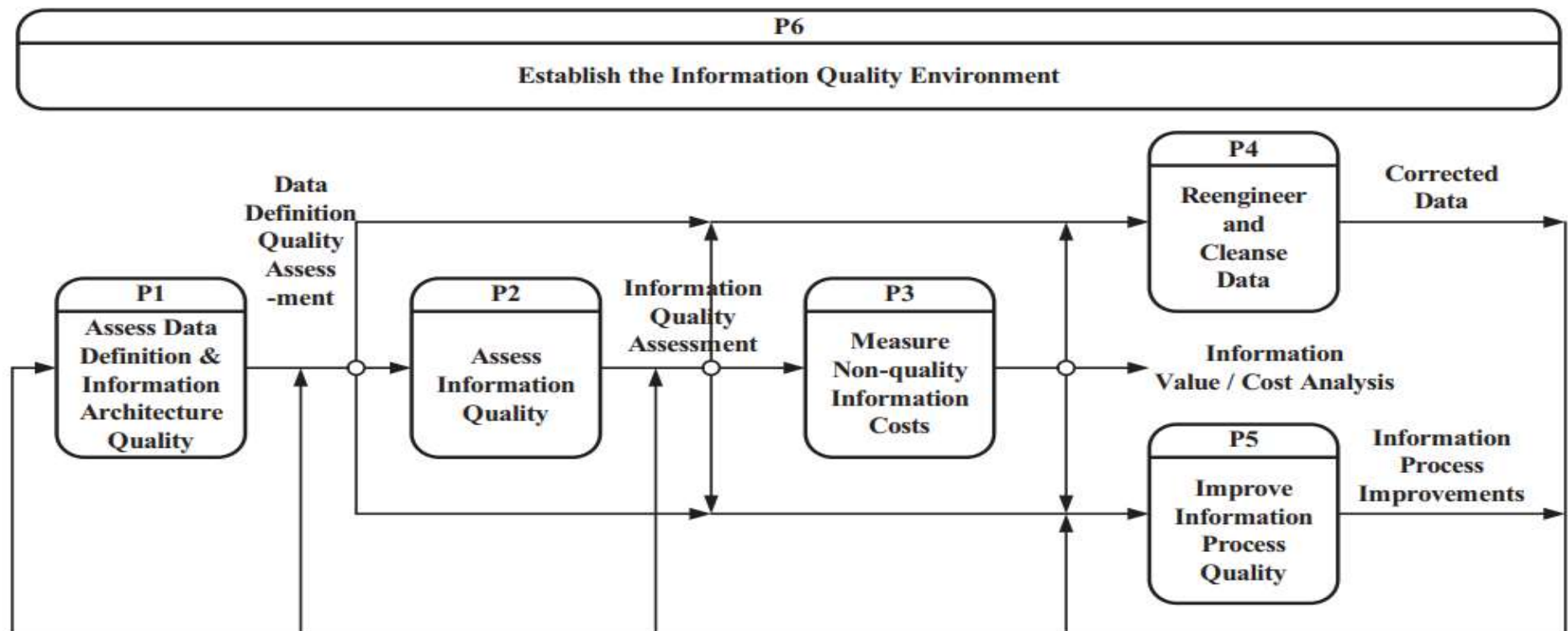
- Assesses – Quality of conceptual data model
- Approaches – Qualitative & Quantitative



Improving Data Quality

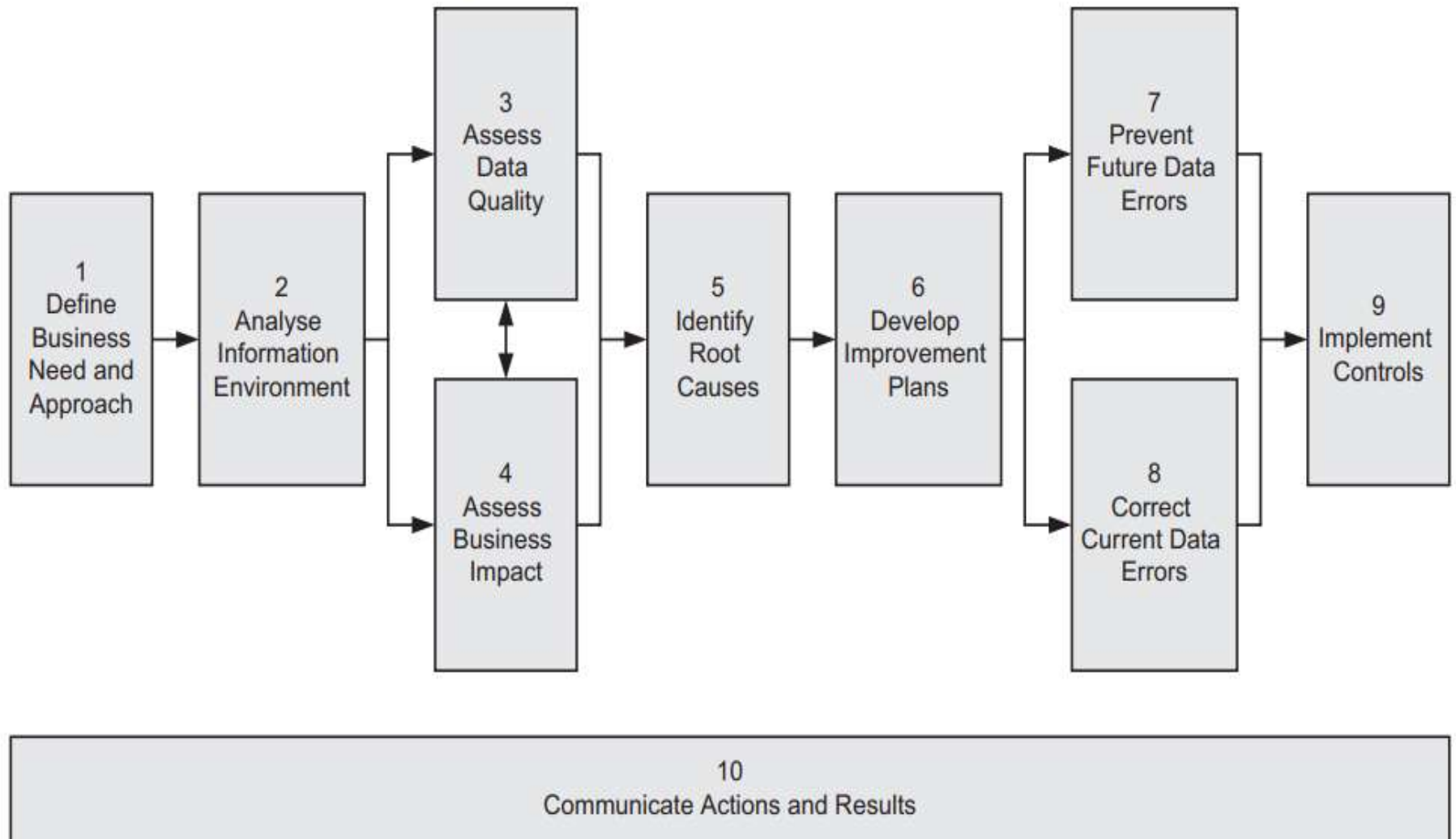
40

- To achieve enduring quality – Enterprise implement
 - ▣ Set of Procedures & Culture
- Larry English – TQdM



Ten step processes

41



Data Accessibility

42

- Data Security - Protecting DB
- Create – Data security policy – Enforce security mechanism
 - ▣ Access control – Rely on authentication process
 - Ensure the user processes
 - Granting & Revoking privileges
 - Table privileges
 - Function & procedure privileges
 - Database object privileges
 - Discretionary & mandatory access controls
 - Granted access rights to others
 - Access rights cannot be changed by the users
 - Multilevel Security – two user get two results in same data

Data Accessibility

43

- ❑ Encryption
 - ❑ Audit trails
- ❑ DATA INTEGRITY – Protecting the DB against authorized users
- ❑ Constraints – That ensure the data consistency & complies with business rules
- ❑ Rules
 - ❑ All candidate key are unique
 - ❑ Entity integrity – Primary key Component – Not null
 - ❑ Referential Integrity – Foreign key values must be matched

Data Accessibility

44

- Data Recovery – Restoring the DB after failure
- Causes of failure
 - ▣ Transaction failure – ACID properties
 - Atomicity
 - Consistency
 - Isolation
 - Durability
 - ▣ System crashes
 - ▣ Media failures
- Recovery Mechanism
 - ▣ Physical redundancy of data
 - ▣ Back up / Disk mirroring
 - ▣ Data is held in more than one operational DB

Data Management Environment

45

- Use of Packaged Application Software
- Ready to use – Specific, but It support related operations
 - ▣ Financial Packages – Support sales, purchase, & stock
 - ▣ Human resource package – Recruitment, Training & payroll
- It appropriate for broad range of business
- It is relatively cheap & deliver speedy business
- Issues – It does not support sharing
- Need “bespoke” interface

Data Management Environment

46

- Impact on Data Management
- To be addressed - Data mgmt perspective – Packages has issues
 - ▣ Not reliable conceptual model for data sharing
 - ▣ Doesn't meet the functional & local requirements
 - ▣ How to adopt & handle new versions of packages?
- When the true costs of developing, maintaining and managing interfaces as well as the tailoring required by the business area are included in the cost–benefit analysis.
- Off-the shelf application packages often turn out to be a more expensive option than bespoke development

Distributed Data and Databases

47

- Rationale for Distributed Data
- Large enterprise – Geographically dispersed – Still use single centralized database
- Principle reasons for choose Distributed DB
 - ▣ Work is not disturbed by communication failures
 - ▣ Existing system dispersed through the enterprise with own DB
 - So Data managed as cohesive whole
- Distributed data provide greater
 - ▣ Reliability & availability with reduced response time
 - ▣ Decide – What data to be store – Where & How managed
- Top down data distribution & Bottom up integration

Distributed Data and Databases

48

- Perfect Distributed Database System
- Chris Dates – States that
 - ▣ Fundamental Principles - To the user, a distributed system should look exactly like a non-distributed system.
 - ▣ 12 Objectives

Local autonomy	Distributed query processing
No reliance on central site	Distributed transaction mgmt
Continuous operations	H/W independence
Location independence	OS Independence
Replication transparency	N/W independence
Fragmentation transparency	DB independence

Distributed Data and Databases

49

- ❑ Top down fragmentation & Partitioning
- ❑ The development of a set of local logical schemas from a global logical schema involves using the techniques of fragmentation and partitioning.
- ❑ Fragmentation is the splitting of a table (or relation) into fragments, each of which represents a distinct subset of the data required at one or more specific sites.
- ❑ Partitioning then involves combining fragments
- ❑ So that all the fragments required at a particular site are grouped into a partition.
- ❑ Fragmentation – Vertical, Horizontal & Hybrid

Distributed Data and Databases

50

- ❑ Top down fragmentation & Partitioning
- ❑ The development of a set of local logical schemas from a global logical schema involves using the techniques of fragmentation and partitioning.
- ❑ Fragmentation is the splitting of a table (or relation) into fragments, each of which represents a distinct subset of the data required at one or more specific sites.
- ❑ Partitioning then involves combining fragments
- ❑ So that all the fragments required at a particular site are grouped into a partition.
- ❑ Fragmentation – Vertical, Horizontal & Hybrid

Vertical Fragmentation – Primary key in all fragments

51

payroll_number	surname	first_name	birth_date	salary	department	grade
AY334	Watson	Barbara	1952-12-06	20340	Finance	4
AY478	Wilson	John	1953-07-03	13436	Production	5
BZ987	Smith	Joe	1964-02-24	35625	HQ	1
CA446	Jones	Phil	1974-05-05	27750	Production	2
CX137	Rogers	Jenny	1970-01-10	27750	Finance	2
DJ777	Phillips	Henry	1974-05-05	22570	Finance	3
EX115	Thompson	Brian	1979-06-11	21785	Production	3
FJ678	Harrison	Roger	1988-04-27	14300	Finance	4
FL233	Smith	Jane	1989-08-25	12725	Production	5

payroll_number	birth_date	salary
AY334	1952-12-06	20340
AY478	1953-07-03	13436
BZ987	1964-02-24	35625
CA446	1974-05-05	27750
CX137	1970-01-10	27750
DJ777	1974-05-05	22570
EX115	1979-06-11	21785
FJ678	1988-04-27	14300
FL233	1989-08-25	12725

payroll_number	surname	first_name	department	grade
AY334	Watson	Barbara	Finance	4
AY478	Wilson	John	Production	5
BZ987	Smith	Joe	HQ	1
CA446	Jones	Phil	Production	2
CX137	Rogers	Jenny	Finance	2
DJ777	Phillips	Henry	Finance	3
EX115	Thompson	Brian	Production	3
FJ678	Harrison	Roger	Finance	4
FL233	Smith	Jane	Production	5

Figure 13.2 An example of hybrid fragmentation

payroll_number	surname	first_name	birth_date	salary	department	grade
AY334	Watson	Barbara	1952-12-06	20340	Finance	4
AY478	Wilson	John	1953-07-03	13436	Production	5
BZ987	Smith	Joe	1964-02-24	35625	HQ	1
CA446	Jones	Phil	1974-05-05	27750	Production	2
CX137	Rogers	Jenny	1970-01-10	27750	Finance	2
DJ777	Phillips	Henry	1974-05-05	22570	Finance	3
EX115	Thompson	Brian	1979-06-11	21785	Production	3
FJ678	Harrison	Roger	1988-04-27	14300	Finance	4
FL233	Smith	Jane	1989-08-25	12725	Production	5

payroll_number	birth_date	salary
AY334	1952-12-06	20340
AY478	1953-07-03	13436
BZ987	1964-02-24	35625
CA446	1974-05-05	27750
CX137	1970-01-10	27750
DJ777	1974-05-05	22570
EX115	1979-06-11	21785
FJ678	1988-04-27	14300
FL233	1989-08-25	12725

payroll_number	surname	first_name	department	grade
BZ987	Smith	Joe	HQ	1

payroll_number	surname	first_name	department	grade
AY334	Watson	Barbara	Finance	4
CX137	Rogers	Jenny	Finance	2
DJ777	Phillips	Henry	Finance	3
FJ678	Harrison	Roger	Finance	4

payroll_number	surname	first_name	department	grade
AY478	Wilson	John	Production	5
CA446	Jones	Phil	Production	2
EX115	Thompson	Brian	Production	3
FL233	Smith	Jane	Production	5

Distributed Data and Databases

53

- Horizontal fragmentation involves the assignment of the rows of the table to one (and only one) fragment.
- Hybrid fragmentation is a combination of the vertical and horizontal fragmentation where a vertical fragment is further fragmented horizontally or a horizontal fragment is further fragmented vertically.
- Any fragmentation must be lossless and disjoint.
- Lossless means that no data must be value (other than the primary key values in the case of vertical fragmentation) appears in only one fragment.

Distributed Data and Databases

54

- All of these fragments can be grouped into a single partition to be allocated to the database. This partition happens to be equivalent to the original table, but that is coincidence.
- The aim of fragmentation and partitioning, and any associated decisions about replication, is to organise data so that it is placed closest to where it needs to be used.

Distributed Data and Databases

55

- ❑ Bottom Up integration
- ❑ Unfortunately there is no commonly agreed process for integrating a number of local logical schemas into a single global logical schema.
- ❑ The task is the equivalent of developing a corporate data model by joining project or area models.
- ❑ The development of a global logical schema through the direct combination of a number of local logical schemas is equally likely to fail.
- ❑ The formats used for common data items were different in global and local schemas

Distributed Data and Databases

56

- The situation can be eased if there is a common understanding of the global data requirements that the global logical schema is to represent.
- Then it is a case of identifying how each of the individual data requirements is handled in each of the local logical schemas and arranging a mapping to a common representation of the requirement in the global logical schema.
- An integration project of this kind is more likely to succeed if there is an existing data management initiative and a corporate data model within the organisation.

Management of Replication

57

- Some or all of the data may be replicated in a distributed data system.
- Replication is duplication of data and one of the fundamental principles of database design is that duplication of data should be avoided.
- Replication is acceptable because it is managed duplication;
- Replication implies that we have management procedures in place to ensure that the collective distributed database remains consistent.

Management of Replication

58

- Replication may be
 - ▣ Full or partial;
 - ▣ Synchronous or asynchronous;
 - ▣ Subject to master–slave or update-anywhere updating.

- Full replication is where the entire body of data, including all updates, is copied to every database instance within the distributed system. In partial replication, only data that it is deemed necessary to hold at a site is copied to the database supporting that site.

Management of Replication

59

- Synchronous or asynchronous;
 - Synchronous replication is where all the copies of the data are held in a strictly consistent state
 - Asynchronous: Individual databases are allowed to become inconsistent for a time with the intention that all sites become consistent eventually.
 - The time that databases are inconsistent may be very short, fractions of a second say, but in an environment where data is being constantly updated the databases may never actually be totally consistent.

Management of Replication

60

- Subject to master–slave or update-anywhere updating.
 - Master–slave updating is where one site, the master, has control of all updating.
 - Any site that wishes to update data informs the master, which then propagates the update to every site, including the site that originated the update.
 - All sites then commit the update to their database. Update-anywhere is a form of updating that does away with the reliance on a central site, the master.
 - Each site is now aware of the location of all the replicas and passes updates to sites with replicas as necessary.

Business Intelligence

61

- Business intelligence is the name that has been given to the set of 'techniques' that is used to transform raw data into information that can be used to inform high-level decision making.
- This set of techniques enables decision makers to take historical, current or predictive views of the business.
- The techniques include
 - ▣ Data warehousing,
 - ▣ Data mining and
 - ▣ OnLine Analytical Processing (OLAP).

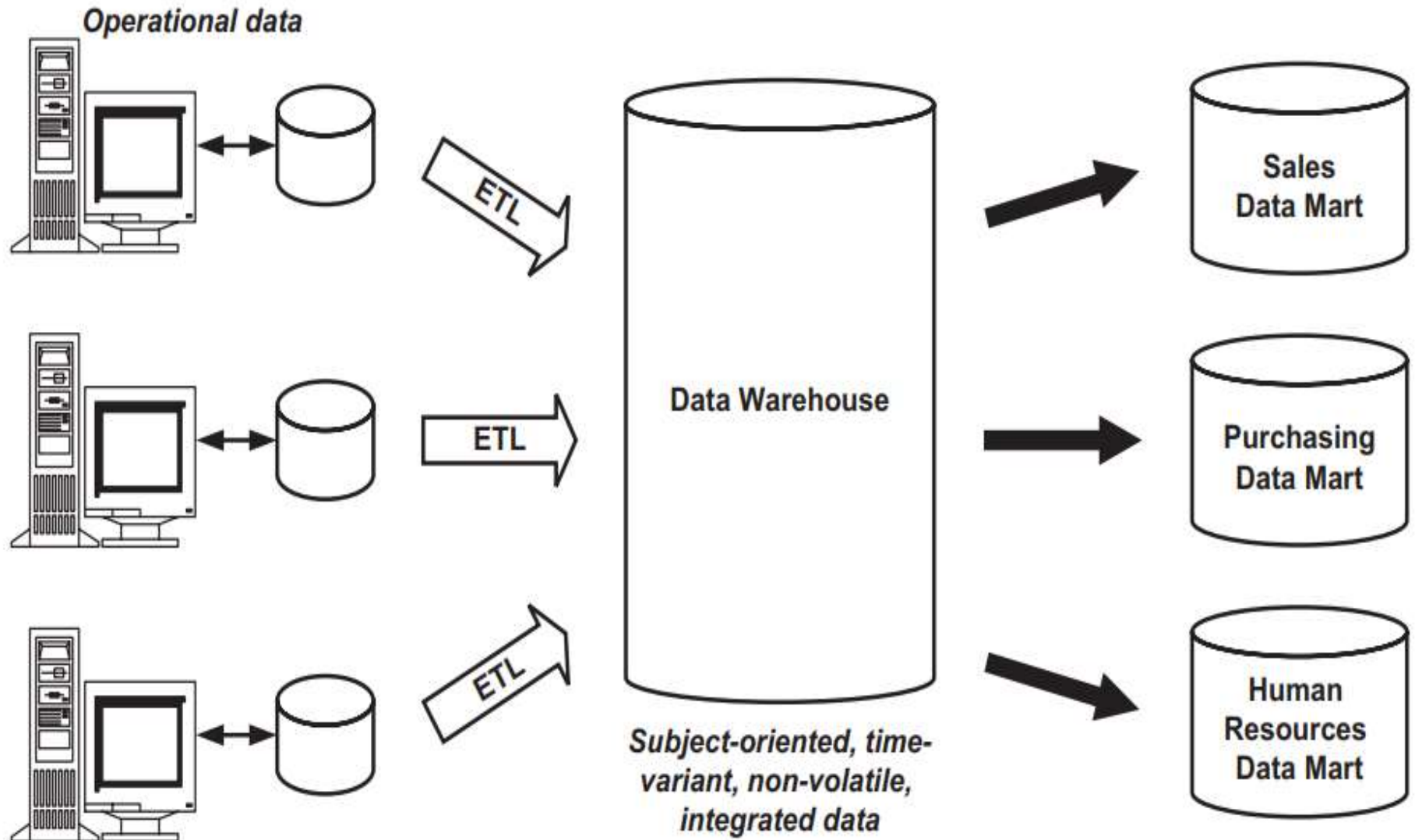
Business Intelligence – Data warehouse

62

- ❑ A data warehouse is not just a large collection of data.
- ❑ At its simplest it is a copy of transactional data – that is, the data created in the operational systems used by the enterprise.
- ❑ This copied data is specifically structured to make it easy for the data to be queried and analyzed.
- ❑ The data is copied using ETL process
 - ❑ Extraction
 - ❑ Translation
 - ❑ Loading

Data warehouse architecture

63



Business Intelligence – Data warehouse

64

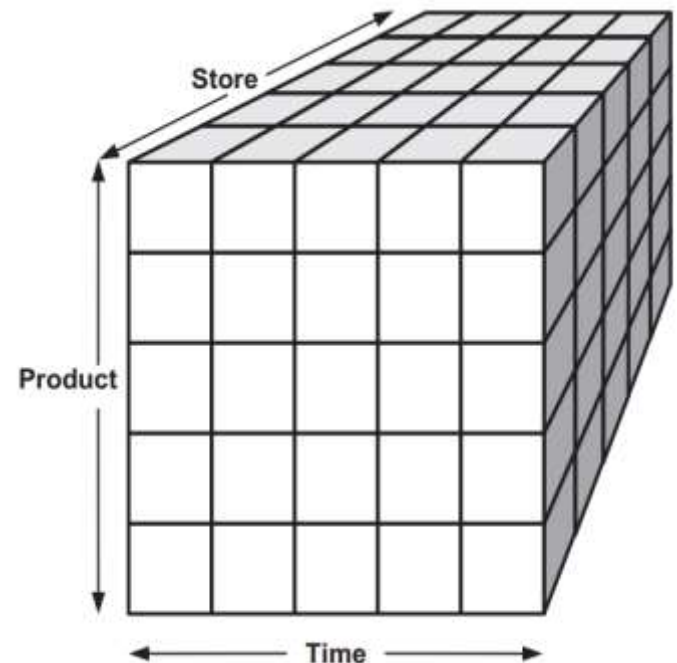
- The data warehouse contains the data that is
 - Subject oriented
 - Time-variant
 - Non-volatile
 - Integrated

Business intelligence –

The multidimensional model of data

65

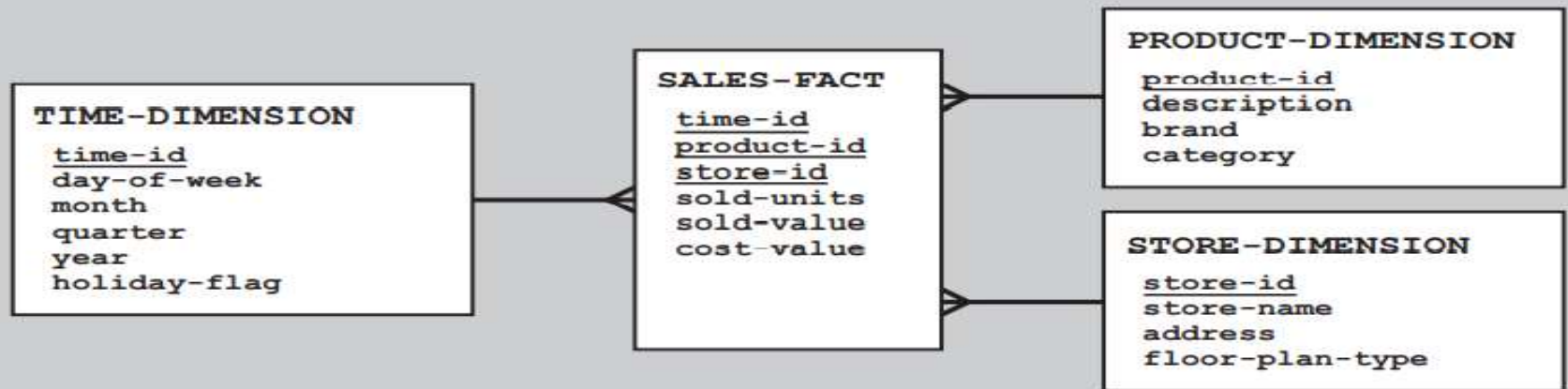
- ❑ The data warehouse is easily visualize the data as a cube by 3/4/5/more dimensions
- ❑ Cubes with more than three dimensions are sometimes known as ‘hypercubes’.
- ❑ Once structured in this way the data can be:
 - ❑ Used with standard reporting tools
 - ❑ Queried using OLAP techniques
 - ❑ Subject to data mining.



Business intelligence – A relational schema for a data warehouse

66

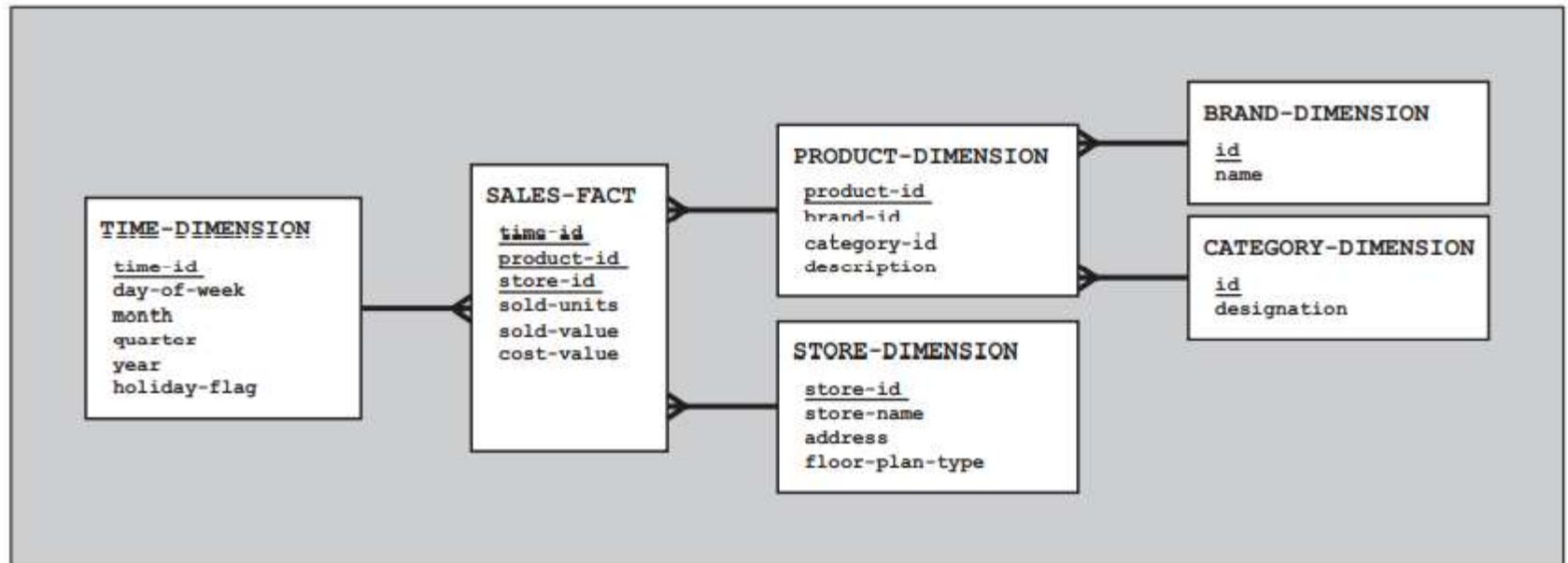
- The multidimensional model of data, is possible to emulate the multidimensional view of data using a relational database management system.
- Our three-dimensional cube (product, store and time) could be represented using the star schema



Business intelligence – A relational schema for a data warehouse

67

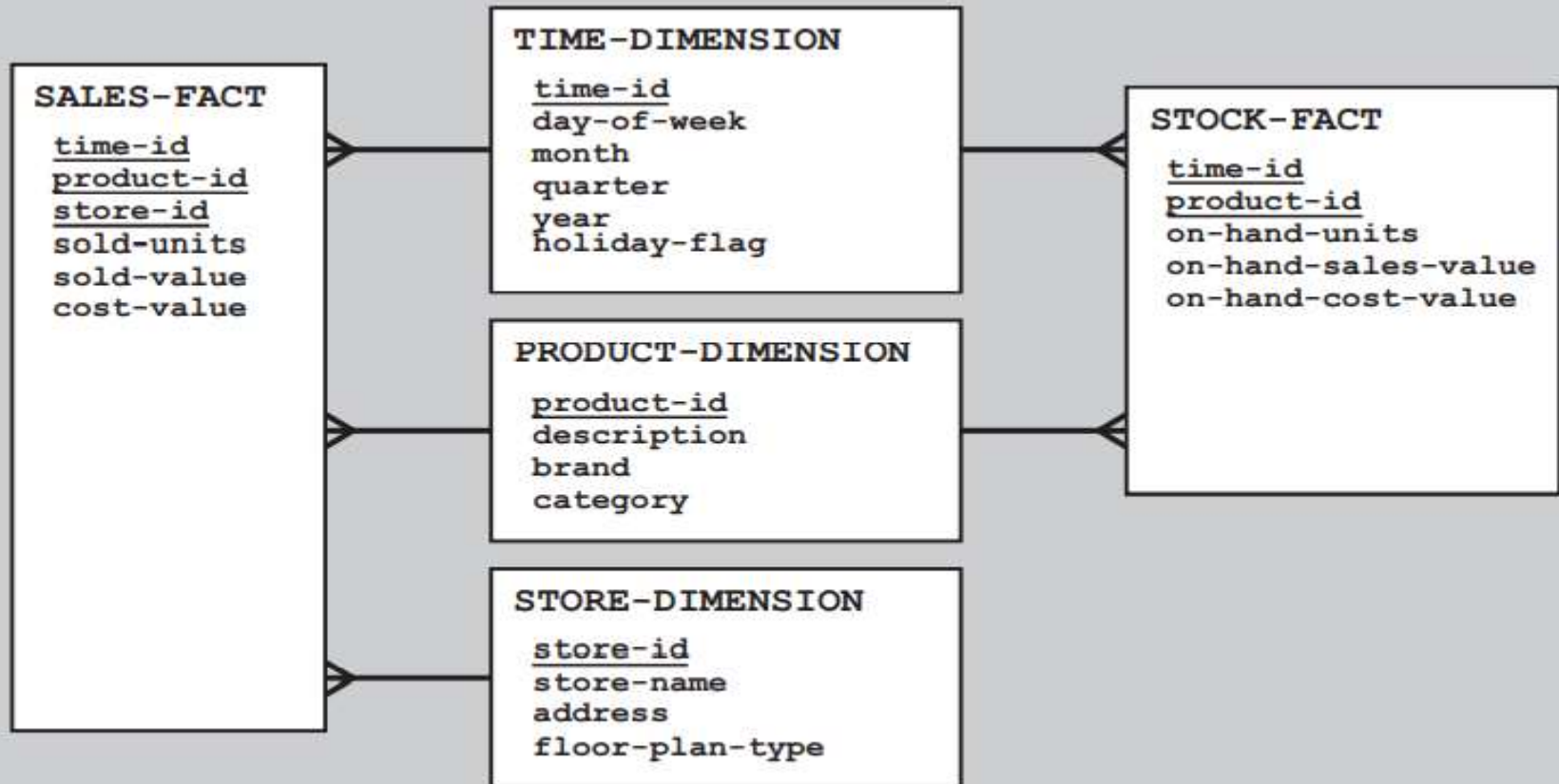
- ❑ Variations on Star schema: Snowflake and Galaxy
- ❑ Snowflake schema – introduces the degree of normalization in to dimensions



Business intelligence – A relational schema for a data warehouse

68

- Galaxy schema – Schema where two or more fact tables share one or more of the dimensions.



CRM

69

- ❑ Orchestrate your business and get your team on the same page with a central repository of customer data.
- ❑ Business success – Depends on Customers
 - ❑ Harder economic fundamentals,
 - ❑ Increasing competition,
 - ❑ Stricter regulations,
 - ❑ Digital disrupters,
 - ❑ Demanding customers,
 - ❑ Mobility, and
 - ❑ Price sensitivity
- ❑ Shifting the power from Companies to Customer

CRM

70

- ❑ Where do they want to go with it?
- ❑ How do they view the outcome?
- ❑ What do they expect from the system?
- ❑ What is the role they play in this journey?

CRM - Pillars

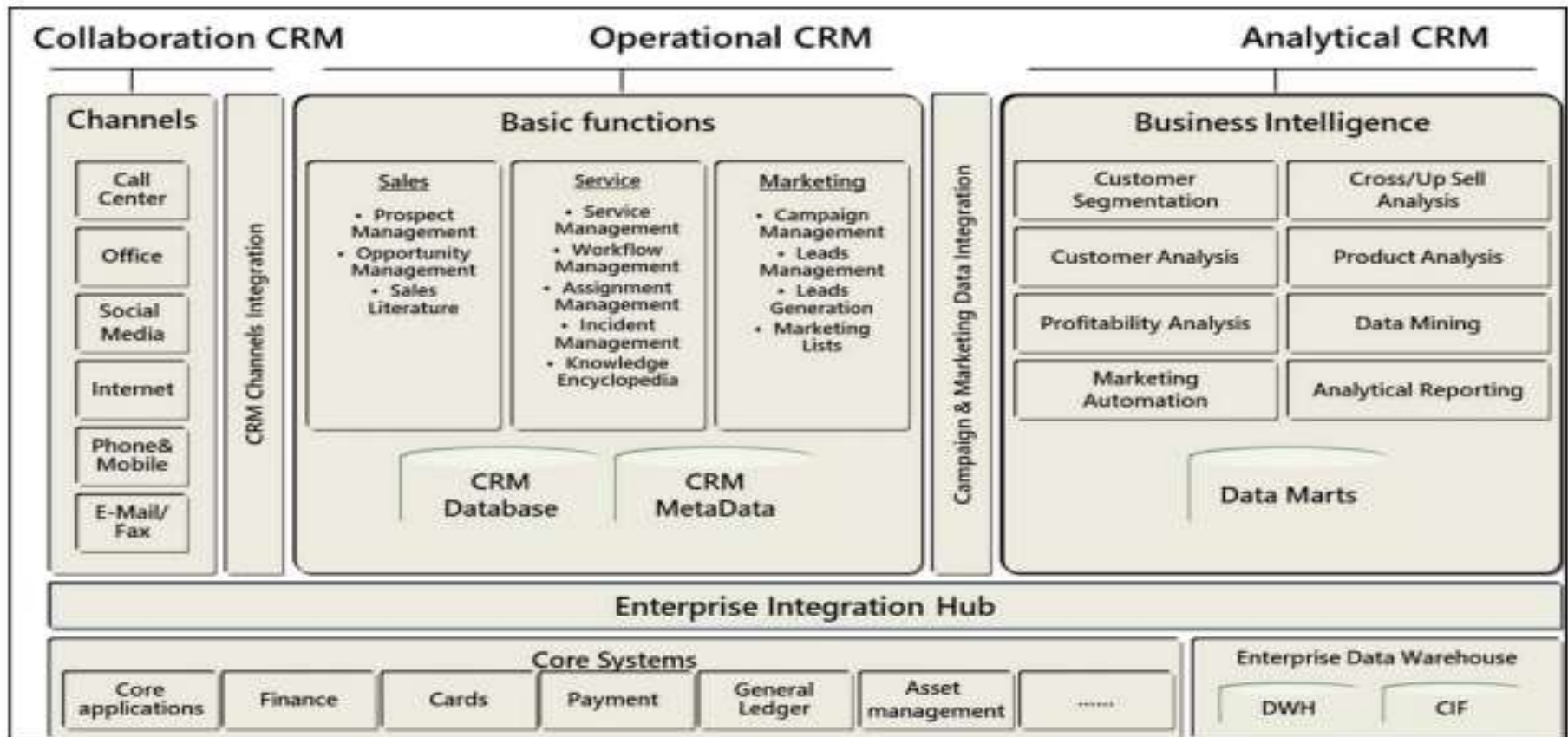
71

- For implementation and they deliver value to the business:
- **Operational CRM:** The operational CRM is all about marketing, sales, and services functionalities.
- **Analytical CRM:** The analytical CRM will use the data collected from the operational CRM and provide users and business leaders with data about business performance as they require
- **Collaboration CRM:** The collaboration CRM will provide the technology to integrate all kinds of communication channels and frontends with core CRM for both internal and external users

Pillars of CRM

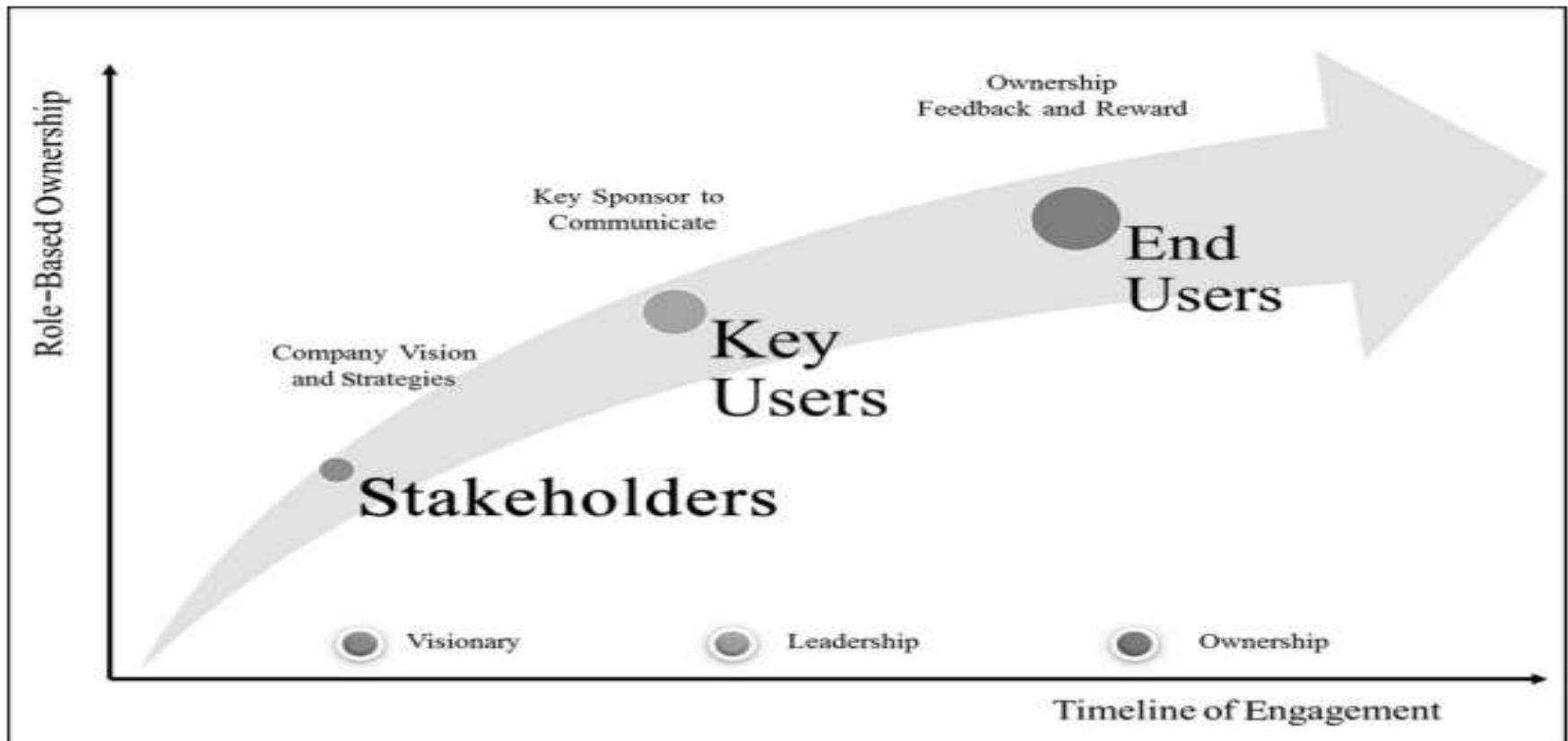
72

- Able to create a comprehensive view of your business and manage clients' communication over all your channels.



Pillars of CRM - Level of Engagement

73

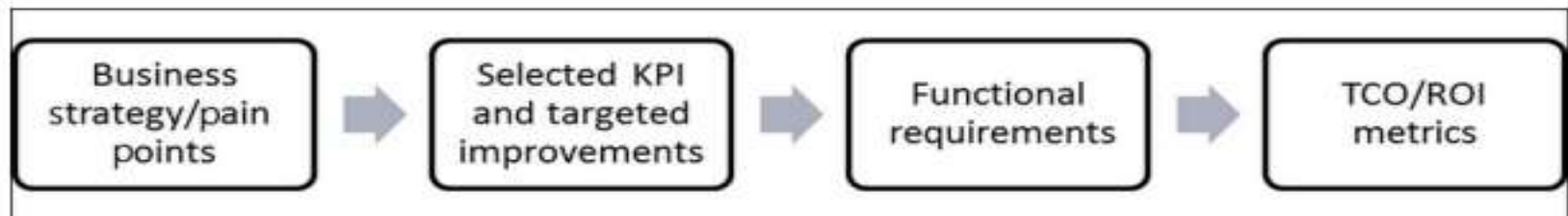


- Helps to identify the metrics for KPI & Total Cost of Ownership and Return on Investment (TCO/ROI) of your project.

Pillars of CRM - Level of Engagement

74

□ Steps defining TCO/ROI metrics



- CRM projects is building a comprehensive 360- degree client view that will be a supporting pillar for sales, marketing, and services processes automation.
- It is also important to support and services processes automation. It is also important to support managing the business and resources more efficiently based on insightful decisions, at a reduced cost of operation.

KYC – Know Your Customer

75

- A common definition of a single customer view is an aggregated and holistic presentation of client data that's being held by a company or by its business units.
- This data could come from multiple sources,
- Related terms for single customer view
 - ▣ 360-degree client view
 - ▣ Client single view (CSV)
 - ▣ Single customer view (SCV)
 - ▣ Single customer repository

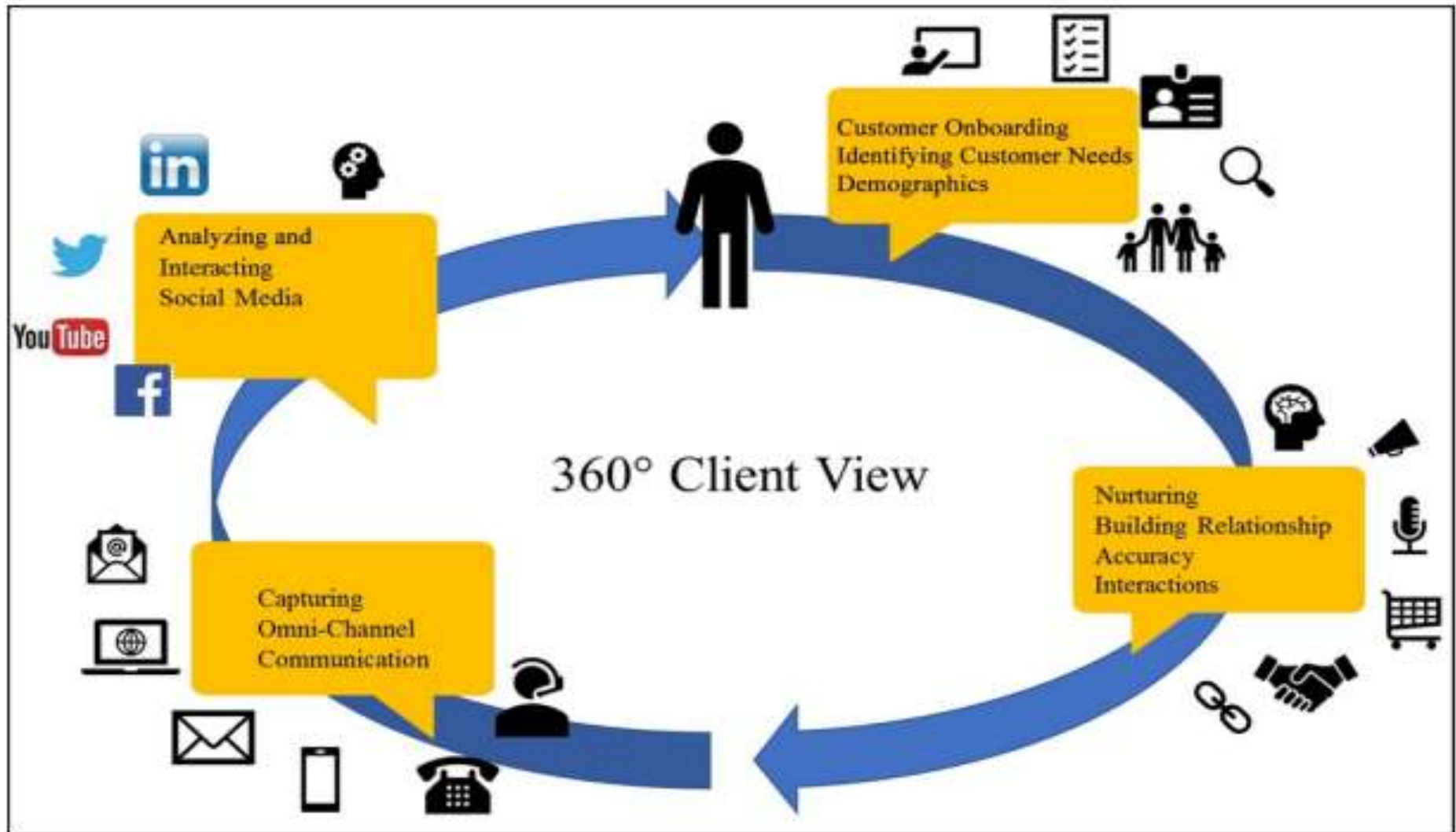
KYC – 360-degree client view

76

- The 360-degree client view will not only be the tool that your company uses to record and assist the customer across all your channels with the best possible customer experience, but it will be the strategic approach for your business to streamline processes and reduce operational costs.
- Type of client data needed in order for the business unit to be more efficient.
- customers (role-based). A discussion about how much information is Additionally, it's important to define who should be entitled to see what type of data (entitlement) from customers (role-based).

KYC – 360-degree client view

77



KYC – 360-degree client view

78

- Building a 360-degree client view could have many more faces, such as the size of the organization, the role of the user, the processes involved, the business unit requirements, the nature of the industry, and the client segment, to name a few.
- Multiple parameters for 360-degree client view
 - ▣ Volume, Segment, Size, and the nature of your business.
- Build trust relationship –
 - ▣ Manage the wealth of customers
 - ▣ Advise the future planning
 - ▣ Customer Life Cycle Management
 - ▣ Customer Health Index – Empower your employees

Utilizing AI and ML in CRM Strategy

79

- AI and ML support CRM processes to enrich traditional business applications
- Digital Business
 - ▣ 2013 @ 5% , 2022 @ 85%
 - ▣ Nine out of 10 young customers
 - ▣ Technologies



Evolution of AI

80

- ❑ In 1940, AI was key component in Science fiction films as a story telling
- ❑ Algorithms and theories rooted from 1950's
- ❑ 1980's the first machine with AI Systems were developed
- ❑ A subset of AI is ML. In the 1990s, ML was reorganized as a separate field from AI and from there it began to flourish.
- ❑ It certainly benefited from the increasing availability of digitized information, and the adoption of the internet in order to exchange information and data

Evolution of AI

81

- The development of ML shifted away from the traditional approaches it had inherited from AI and more toward methods and models borrowed from statistics and probability theory.
- ML changed its desire of achieving comprehensive AI to tackling solvable problems of a practical nature using data.
- The algorithms of ML can leverage mathematical models from sample data and, in turn, make predictions or decisions without being explicitly programmed to do so.

The Current state of AI

82

- AI today is used in almost all fields and industries, including consumer, healthcare, finance, government, and automotive.
 - ▣ Search engines: Such as Microsoft Bing or Google search.
 - ▣ Autonomous vehicles: Such as airplanes and self-driving cars.
 - ▣ Medical diagnosis: Such as predicting the chances of you having cancer
 - ▣ Playing games: Such as chess or Go.
 - ▣ Online assistants: Such as Apple's Siri or Amazon's Alexa.
 - ▣ Creating art: Such as poetry.
 - ▣ Speech recognition: Such as speech-to-text software.
 - ▣ Image and pattern recognition: Such as spam filtering and targeting online advertisements.

The Current state of AI

83

- One of the first mass AI-driven business applications ever was the AutoCorrect feature in Microsoft Office, which was first introduced by MS office in 1993

Big Data + Affordable HPC + Deep Neural Networks = Progress in AI

- Text analysis and understanding
- Image understanding
- Speech recognition
- Robotics
- Deep learning (ML)
- IoT

The Teaming up AI with People

84

- AI will expand functional areas of our daily lives.
- In terms of CRM, the opportunities AI presents are endless: from operational efficiencies to improved customer satisfaction and increased productivity and revenue.
- High priorities in regards to leveraging AI in business:
 - ▣ Marketing
 - ▣ Social media
 - ▣ Chat bots
 - ▣ Operational efficiency
 - ▣ Logistics
 - ▣ Supply chain

The Teaming up AI with People

85

- AI algorithms are good for repetitive tasks that are well defined, such as
 - ▣ Customer service,
 - ▣ Processing a large amount of data,
 - ▣ Finding patterns within that data, and
 - ▣ statistical reasoning for decision making, such as in
 - Business intelligence, reporting, and forecasting.
- Teaming up AI with real people in order to allow them to collaborate and work together, using their abilities and their strengths to achieve the maximum return for businesses.

Applying AI to CRM Solution

86

- How we could apply all these technologies into our CRM design.
- How we could set the stage for collaboration between people and AI systems in order to achieve more.
- How can we leverage these technologies for improving processes in sales, marketing, and customer service?
- Actually, CRM vendors are already infusing AI into their cloud applications. For example, Microsoft Dynamics 365

Applying AI to CRM Solution

87

- By mixing AI and CRM, you will gain
 - ▣ Key performance analytics,
 - ▣ Coupled with easy,
 - ▣ Natural language queries,
 - ▣ In order to respond quickly with the uncovered information and trends in data.
- By proactively analyzing data using ML models, you can gain insights into your performance and KPIs.
- These are elements can provide better visibility into your business forecast, actuals, pipeline, leads, and activities.

The ethical aspects of AI

88

- One important aspect of AI development: the ethical questions around it.
- A key principle here is that AI will partly be used to augment human intelligence, but this is not always the case, nor should this be done in an uncontrolled way.
- Augmenting humans is not the sole problem for the field of AI, must also understand the impact of this space.
- Make thoughtful and transparent decisions to mitigate the potential risks from leveraging AI.
- Engage both employees and customers in the stages of the processes and create transparency in all these discussions.

The ethical aspects of AI

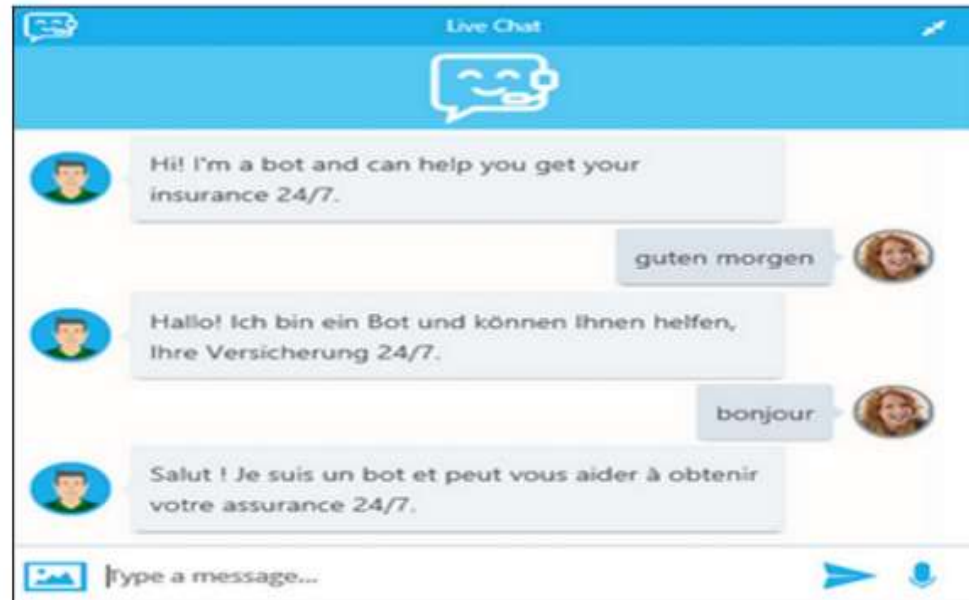
89

- ❑ 2018 Facebook/Cambridge Analytica scandal, where it was revealed that millions of records of user data had been misused in the run-up to the U.S. election and the UK's Brexit referendum with social marketing techniques that were based on AI.
- ❑ The ethical design and human control factor have to be a part of every AI and big data project today.
- ❑ There's no doubt across the industry today that AI, ML, and data modeling have to be transparent for all the people involved, whether that is the company or the customer.
- ❑ Ethics committees that review the use of AI, when designing CRM solution

An example of AI in CRM Processes

90

- More and more organizations starting to realize the power of AI when combined in a CRM platform.
- Four design technologies
 - ▣ Bot Framework (or conversation agents)
 - ▣ Language Understanding Intelligent Service (LUIS)
 - ▣ Cognitive Services
 - ▣ Translation Services
- Insurance case



The Deployment of a CRM platform

91

- Deployment option – Microsoft Dynamics 365, Salesforce.com,...
- A mixture of the on-premise and cloud deployment options, referred as Hybrid options
 - ▣ such as Infrastructure as a Service (IaaS) or Software as a Service (SaaS) deployment
- Deployment factors
 - ▣ Business requirement & Fit gap
 - ▣ Cost-value matrix
 - ▣ User acceptance
 - ▣ Technical consideration
 - ▣ Explicit & Implicit costs
 - ▣ Regulatory Compliance

Factors Influencing Vendor Selection

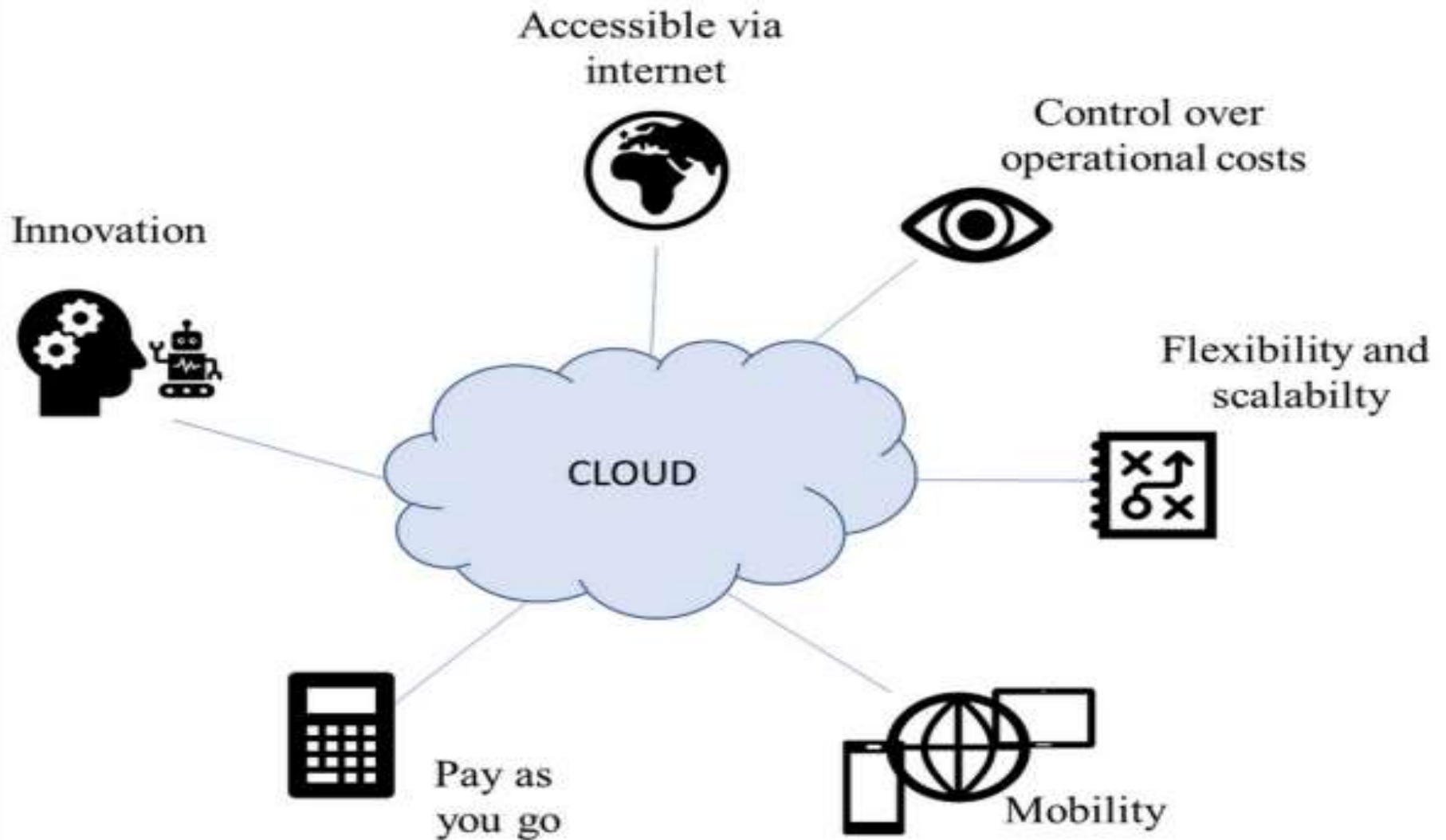
92

- Vendor selection factors
 - ▣ Capabilities of the Platform
 - ▣ Cost of the system
 - ▣ Complexity & Feasibility for Customization
 - ▣ Portability & deployment limitations
 - ▣ No. of International data centers
 - ▣ Training & Skills required

Factors Influencing Vendor Selection

93

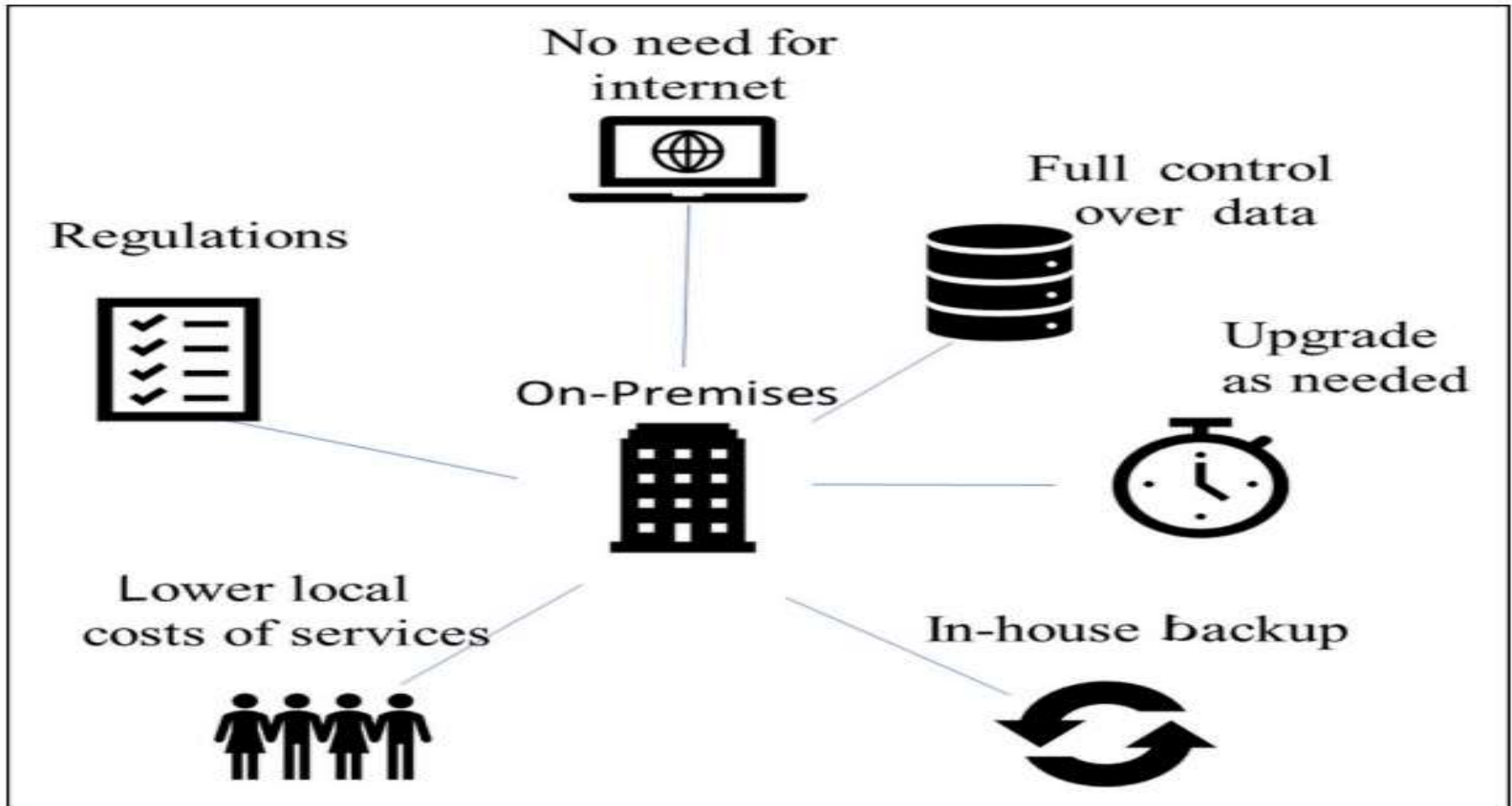
- Cost and Complexity
 - ▣ The initial costs: License, infrastructure, and so on.
 - ▣ The implementation time: It needs to be short.
 - ▣ The degree of customizations: Must be lower than 5%.
 - ▣ The control over the data and the solution: Including backups & updates
 - ▣ The on going maintenance costs.
- Cloud deployment (Amazon Elastic Compute Cloud, MS Azure)
 - ▣ Solution capabilities
 - ▣ Collaboration with internal IT
 - ▣ Regulation
 - ▣ Speed, Security & Portability



Factors Influencing Vendor Selection

95

□ On-premise deployment



Mixed Deployment

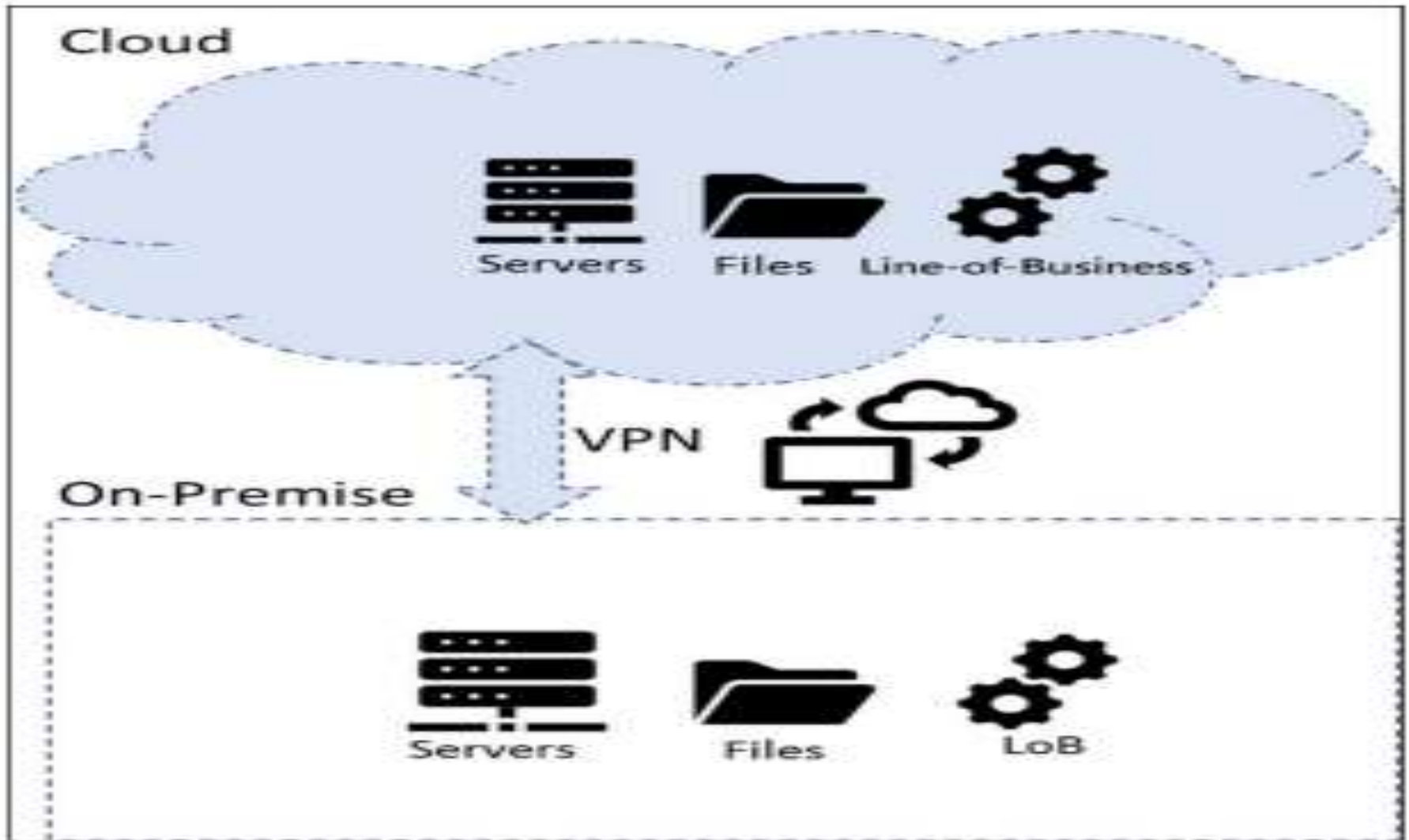
96

□ IaaS, SaaS & PaaS

Responsibility	On-Prem	IaaS	PaaS	SaaS
Data Classification and Accountability	Customer	Customer	Customer	Customer
Client and end-point Protection	Customer	Customer	Customer	Shared
Identity and Access Management	Customer	Customer	Shared	Shared
Application level Controls	Customer	Customer	Shared	Shared
Network Controls	Customer	Shared	Shared	Shared
Host Infrastructure	Customer	Shared	Shared	Shared
Physical Security	Customer	Shared	Shared	Shared
	Customer	Cloud Provider	Shared	

Hybrid Development

97



Dynamics365	On-Premise	Cloud/Online	Hybrid	Partner Hosted
Releases/Year	One release	Two releases (April and October)	Mix	Mix
Upgrade/Update	By admin	Continuous Automatic Updates and Rollups. Automatic Upgrade after approvals	Mix	By admin
Skype (business), Yammer, Exchange, SharePoint	Yes	Yes	Yes	Yes
Office 365, Power BI, ML, Bots, Social,	No	Yes	No	Yes (federated)
Authentication	Active directory	Dynamics365 Authentication and Azure AD-Synch,	AD, ADFS, Azure Synch	ADFS, Synch Azure
Large Data Integration	Faster with low latency	Latency optimized by bulk API, or Azure-Server	Mix	Very fast
Mobile for Smart Phone and Tablet	Yes limited	Yes (online and offline)	Yes	Yes

CRM Differentiators

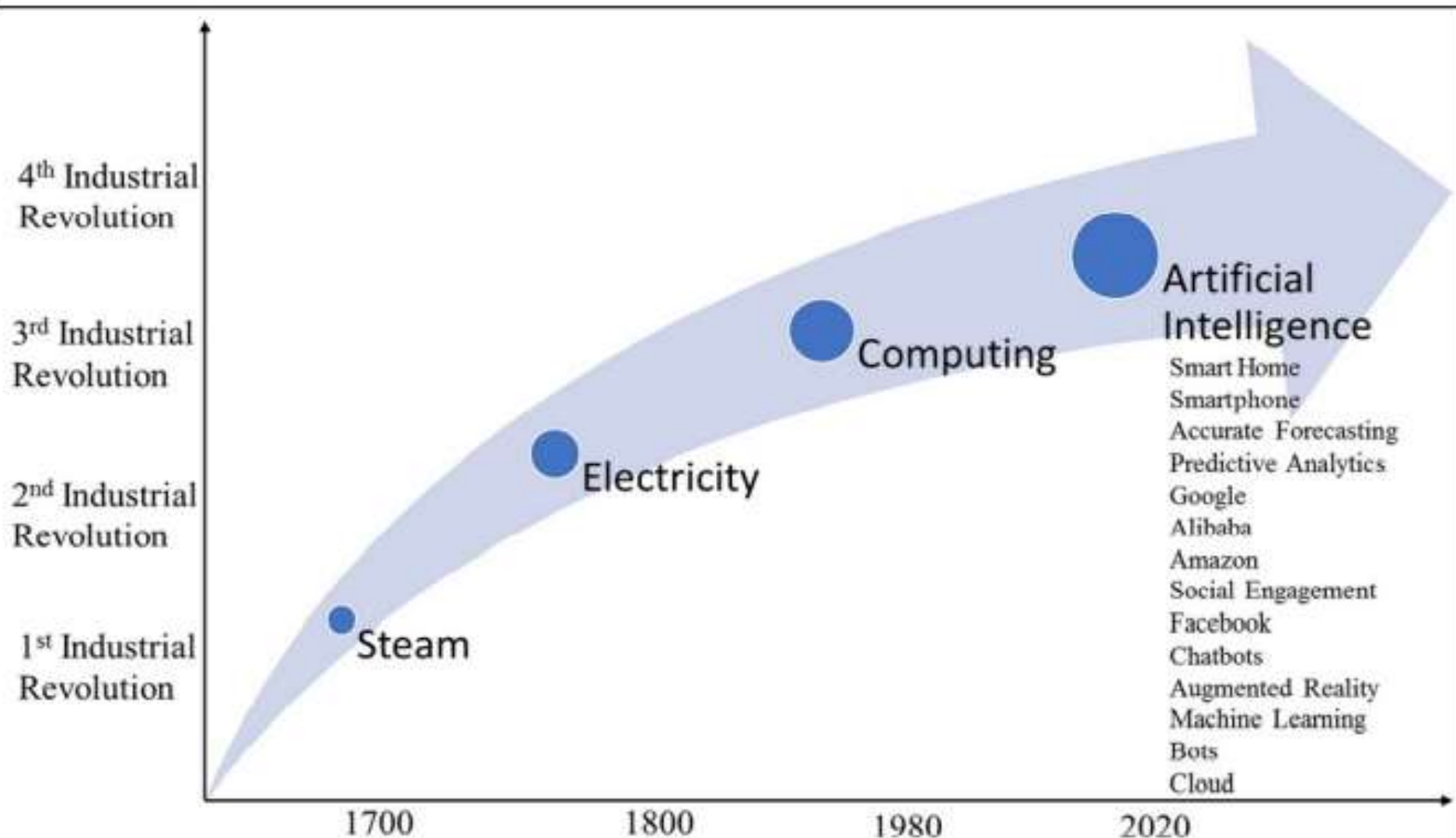
Feature list – Eco System

99

- Instead of CRM elements – The ECO system support Business
- Innovation in CRM technologies
- Industry 4.0
 - ▣ AI and Cognitive Service – Has great impact in CRM platforms
- Fourth Industrial Revelation – Two main driver for CRM
 - ▣ Digitization, Collaborate & Connect with each other
 - ▣ Massive Innovations Technology

Industry 4.0

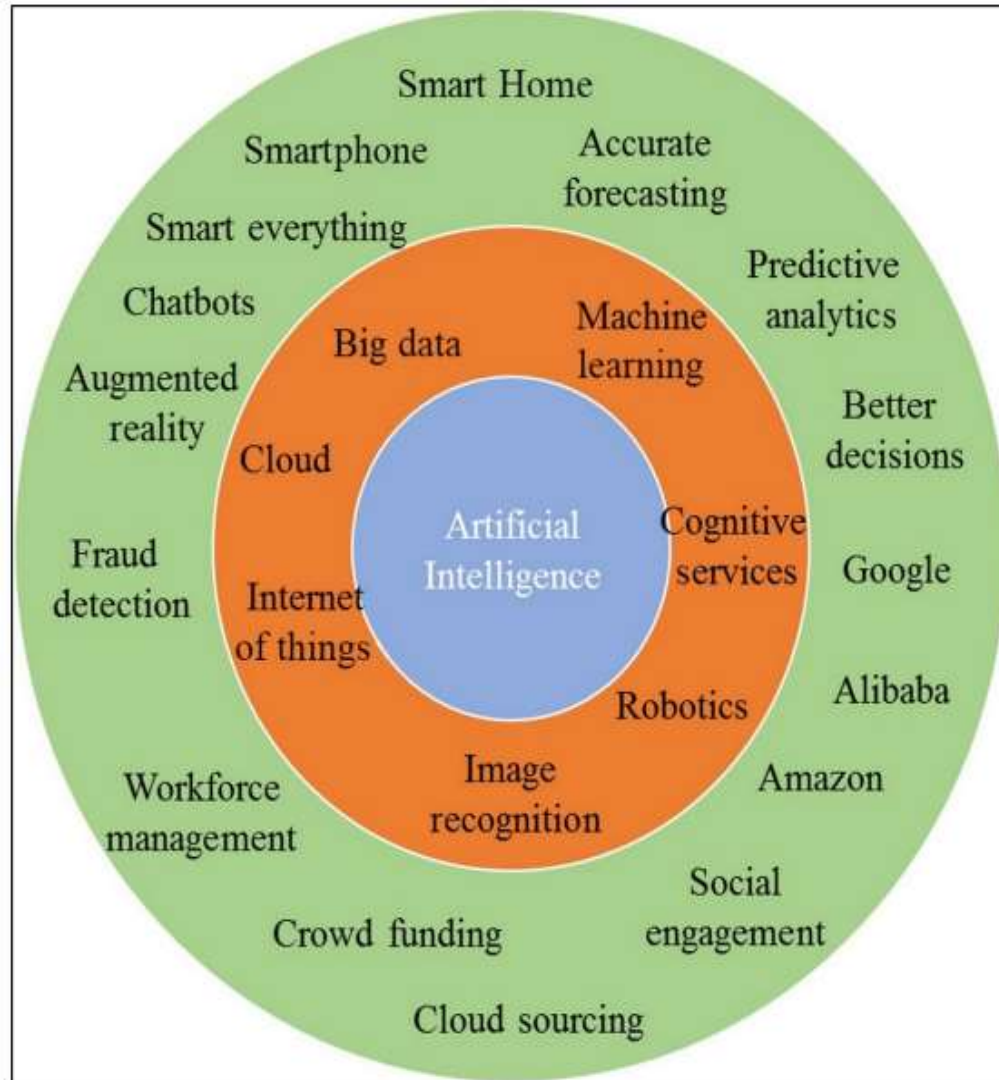
100



AI and Smart Cloud

101

- Revolution driven by AI
 - ▣ Advances in Computer Power
 - ▣ Memory size



To Cloud or Not to Cloud

102

- What options are there?
- What are the best practices?
- Should all applications move to the cloud or only some of them?
- What are the pros and cons of private cloud versus public cloud
- The company wanted to understand the integration and deployment options.
- This element was mainly focused on how to set up a hybrid application infrastructure in order to benefit from cloud services and on-premise.
- My general advice is that if you move to the cloud, then you don't want to move all your business applications to the cloud at once.

To Cloud or Not to Cloud

103

- Total ownership/return on investment (TCO/ROI) analysis on all three
 - ▣ 1. Stay where you are, on-premise, and improve what you have.
 - ▣ 2. Move from on-premise and adapt to the cloud gradually.
 - ▣ 3. Move to the cloud entirely with your core applications.
- **Leveraging smart cloud into CRM**
- Business scenarios for CRM applications
 - ▣ Routing cases by sentiment with cognitive services text analytics.
 - ▣ The intelligent agent assistant, namely bots in the agent desktop
 - ▣ Leveraging ML and AI to automatically extract metadata from audio and video files

Big data

104

- Big data, as a topic, is the new science for dealing with datasets that are too large or complex to be dealt with using traditional data processing applications.
- It's the science of leveraging AI and ML to analyze systematically extracted information from a variety of sources



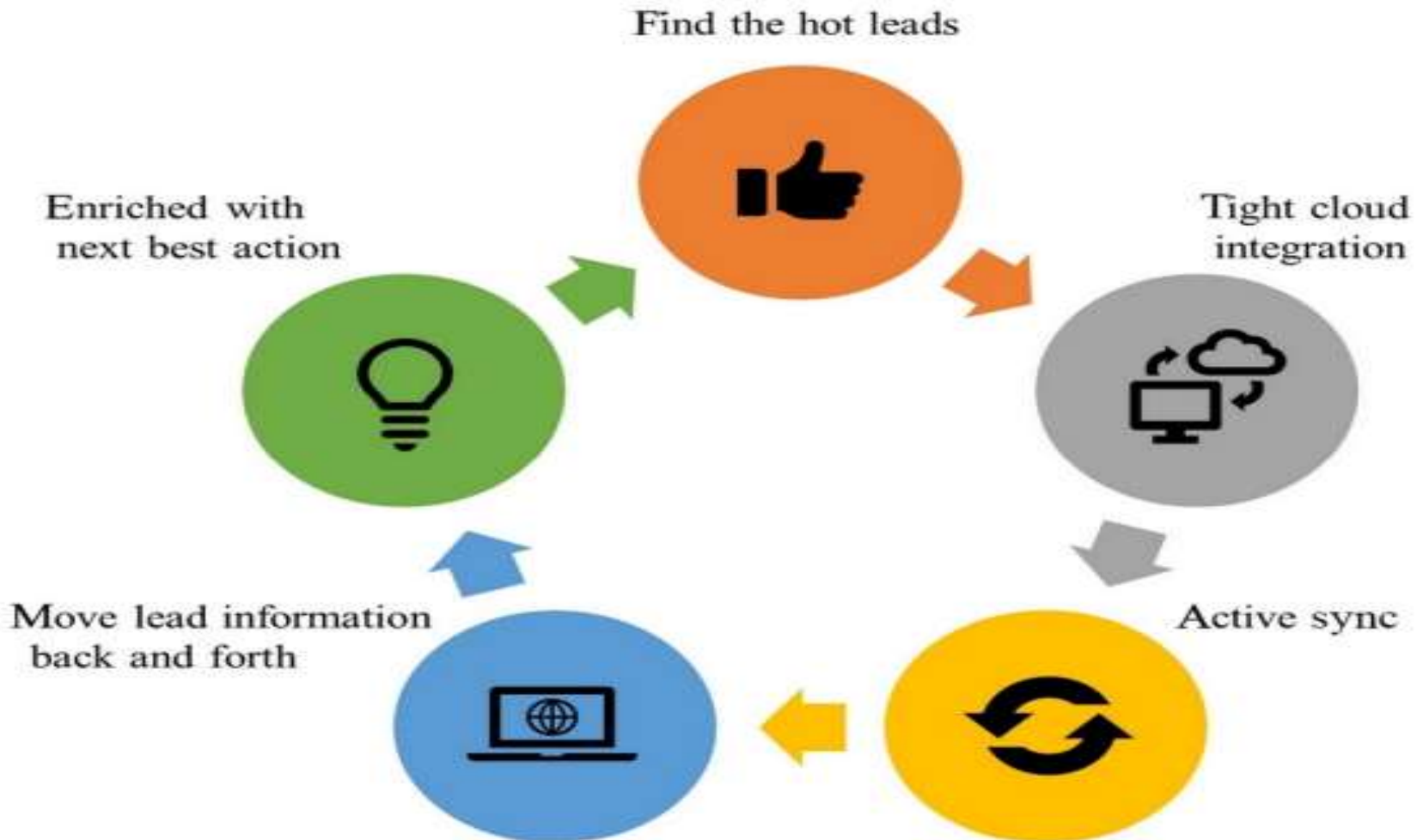
Social selling and advertising

105

- Social selling is the process of developing sales over social network platforms and it is an integrated part of the modern sales process.
- Social selling techniques include branding and interacting directly with potential buyers and customers, which will help you to develop leads and opportunities with integrated social listening in your CRM strategies.
- The concept is gaining huge popularity in a variety of industries, such as financial advisory services, the automotive industry, and consumer products.
- Social selling is the main source of income for a selection of social media platforms.

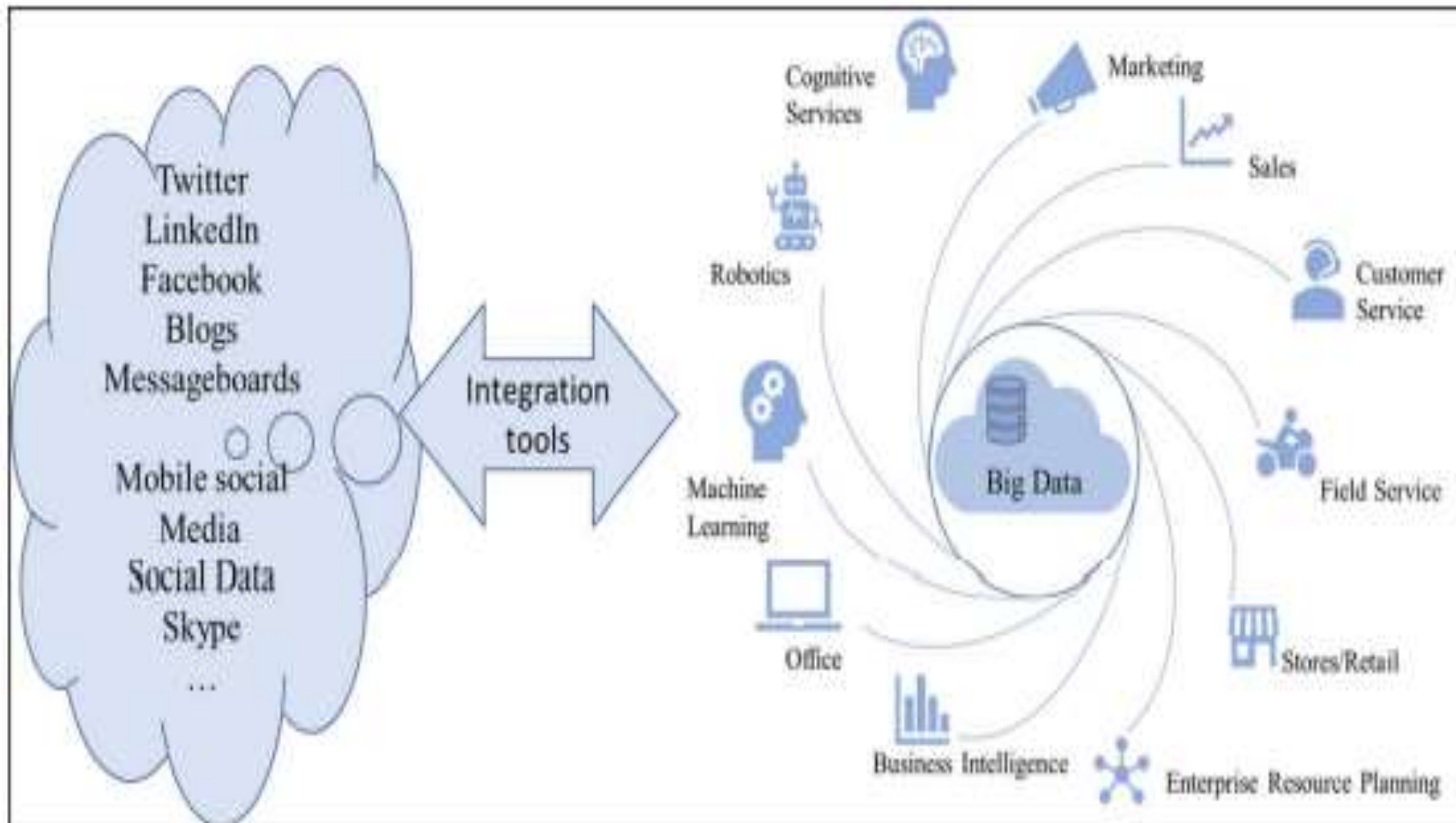
Social Selling Process

106



Social Selling Process

107



Implementation tools

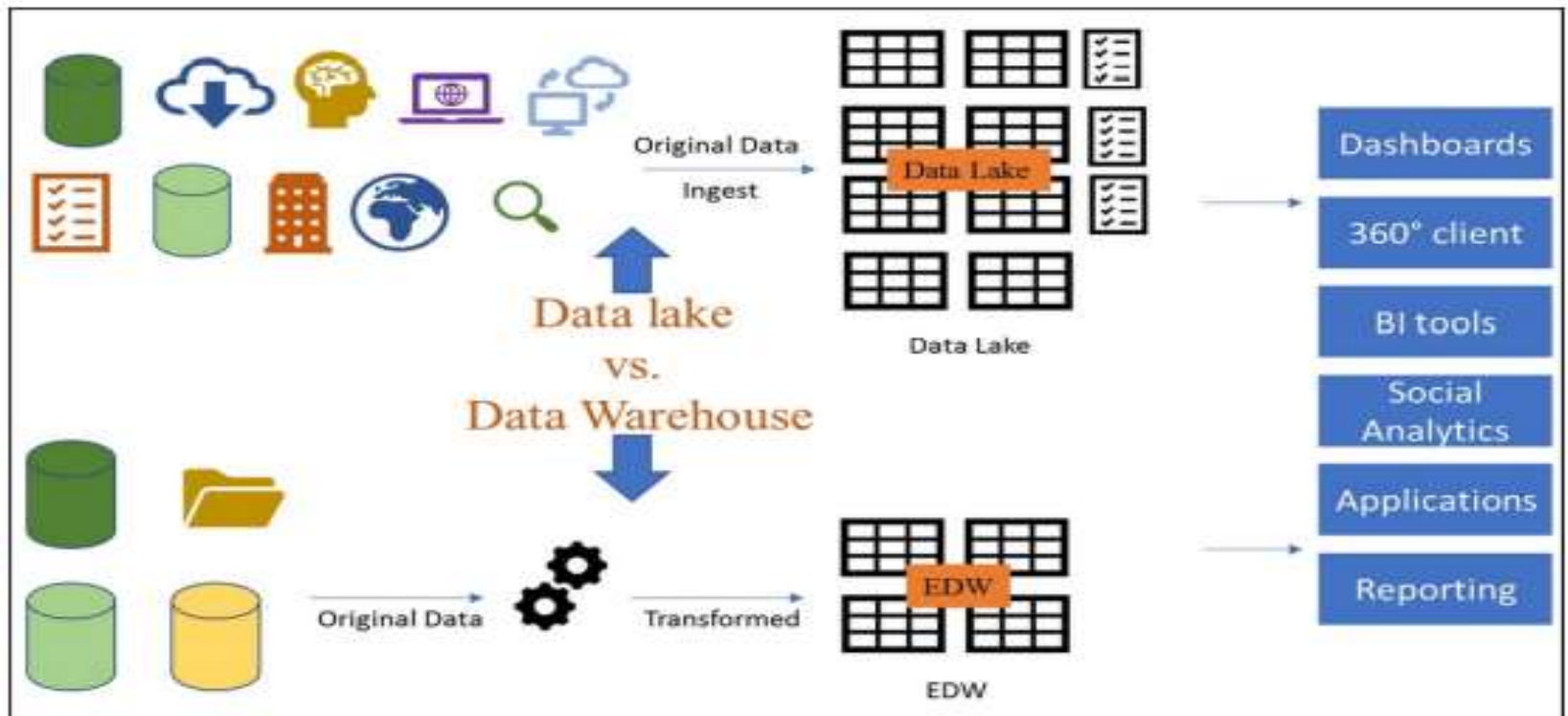
108

- Challenge is create platform that support entire business applications processes
- Microsoft Power Platform
- Components
 - ▣ Power BI
 - ▣ Power Apps
 - ▣ Flow
 - ▣ Common Data model
- Power BI is a business analytics service from Microsoft. It provides simple-to-use business intelligence and interactive visualization tools that allow users to create their own dashboards and reports on the fly
- PowerApps is an application development environment that

Implementation tools

109

- PowerApps is an application development environment that provides services, connectors, and a data platform.
 - ▣ Share point, Excel, MS 365 etc...



Sustainable Platform

110

- In designing a sustainable CRM strategy, it is legitimate to ask how CRM platforms today will differ tomorrow.
- The challenge is twofold;
 - ▣ Need to integrate and leverage big data across the organization
 - ▣ Need to expand your applications into the cloud seamlessly.
- CRM platform must leverage cloud intelligent services (smart cloud) to expand and drive business outcomes.
- Platform that enables you to greatly increase your reach and to serve more customers, while reducing operational costs.
- Your CRM will have to leverage the cloud and XaaS (virtually everything as a service) in the smart cloud to scale and support customer interaction channels.