Data Warehousing:

This article highlights the importance and intersection of data mining and data warehousing in the context of big data. In a data warehouse, data helps analysts to make informed decisions in an organization. Based on your understanding of a database warehouse and the data mining life cycle, consider a specific issue in an industry you know. How might you apply these steps to address the issue?

Databases play a critical role in almost all areas where computers are used, including education, library, science, medicine, business, law, engineering, and so on. Due to the recent developments, the storage capacity and the computing speed of the computers have increased, so now big data is being handled by the computers using different techniques handling the data. Data is a valuable asset for an organization. Data mining is the process of finding patterns in a given data set. Today, data mining is used in contexts of fraud detection, as an aid in marketing campaigns, and even by supermarkets. Data warehouse provides us generalized and consolidated data in a multidimensional view. Along with this view of data, data warehouses also supply us with online analytical processing (OLAP) tools. There is no frequent updating done in a data warehouse. Data cleaning and data transformation are important steps in improving the quality of data and data mining results. Sometimes data mining is called data or knowledge discovery.

Data Warehouse defenition:

Data warehouse was first coined by Bill Inmon in 1990. According to the Inmon, a data warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data. This data helps analysts to make informed decisions in an organization.

- A data warehouse is a database that is kept separate from the organization's operational database.
- Possesses consolidated historical data, which helps the organization to analyze its business.
- Help in the integration of diversity of application systems. Why a data warehouse is separated from operational databases?
- An operational database is constructed for well-known tasks and workloads such as searching particular records. Indexing etc.
- An operational database query allows reading and modifying operations.
- An operational database maintains current data.

Features:

- Data warehouse is subject oriented because it provides information around a subject.
- Data is integrated from heterogeneous sources.
- Data collected in a data warehouse is identified with a particular time period.
- Data is non volatile.

- A data warehouse does not require transaction processing, recovery, and concurrency controls.

Applications of Data Warehouses

A data warehouse helps business executives to organize, analyze and use their data for decision making. Data warehouses are widely used in the following fields:

1. **Financial services**: They are the economic services provides by the finance industry which encompasses a broad range of businesses that manage money and some government sponsored enterprises.
2. **Banking services**: We combine global capabilities with deep local knowledge to provide innovative products and services to meet the needs of our customers and clients.
3. **Consumer goods**: Consumer goods are goods that are ultimately consumed rather than used in the production of another good. For example, a microwave oven or a bicycle which is sold to a consumer is a final good or consumer good, whereas the components which are sold to be used in those goods are called intermediate goods.
4. **Retail sectors**: Retail is the process of selling consumer goods and/or services to customers through multiple channels of distribution to earn a profit.
5. **Controlled manufacturing**: Quality Controlled Manufacturing Inc. specializes in the precision machining of complex components and assemblies in all metals, including exotic alloys using Six Sigma methodology and lean manufacturing principles.

Types of Data Warehouses

There are three types of data warehouse applications:

1. Information processing: The data can be accessing by means of querying, basic statistical analysis, reporting using crosstabs, tables, charts or graphs.
2. Analytical processing: The data can be analyzed by means of basic OLAP operations, including slice and dice, drill down, drill up, and pivoting.

Data mining: Data mining supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classifica

3. tion, and prediction.

Example: **Facebook**

Example of data warehousing that everyone can relate to is what Facebook does. Facebook gathers all of your data – your friends, your likes, and so on – and then stores that data into one central repository. For many reasons, they want to make sure that you see the most relevant ads that you're most likely to click on, they want to make sure that every friends that they suggest are the most relevant to you, keep in mind that this is the data mining phase in which meaningful data and patterns are extracted from the aggregated data

Data warehousing problem solving and solutions:

1 Missing values in data sources

2 Additional columns

3 Use of different representation formats in data sources

4 Non-Compliance of data in data sources with the Standards

5 Failure to update all replicas of data causes DQ Problems.

6 Approximations or surrogates used in data

7 Different encoding formats (ASCII, EBCDIC,….)

8 Lack of business ownership, policy and planning of the entire

enterprise data contribute to data quality problems.

Fig:1 Types of data sources Organizations extract data from.

Causes of Data Quality Issues at Data Profiling

When possible candidate data sources are identified and finalized data profiling comes in play immediately.

Data profiling is the examination and assessment of your source systems' data quality,integrity and consistency

sometimes also called as source systems analysis

Table 2:causes of data quality at Data profiling

Sr.No CAUSES OF DATA QUALITY PROBLEMS AT DATA

PROFILING

1 Insufficient range and distribution of values or threshold

analysis for required fields.

2 Unreliable and incomplete metadata of data source

3 UserGeneratedSQLqueriesforthedataprofilingpurpose leave

the data quality problems.

4 Inability of evaluation of inconsistent business processes

during data profiling cause data quality problems.

5 Inability of evaluation of data structure, data values and

data relationships before data integration ,propagates poor

data quality

6 Inability of integration between Data profiling and ETL

causes Data quality problem

7 Inappropriate selection of Automated profiling tool cause data

quality issues

8 Insufficient data content analysis against external reference

data causes data quality problems.

9 Insufficient structural analysis of the data sources in the

profiling stage.

10 Insufficient Pattern analysis for given fields within each data

store

Data Quality issue at D ata Staging ETL(Extra action), Transformation and Loading)

One consideration is whether data cleansing is most appropriate a the source system, during the ETL process,

at the staging database,or within the data warehouse..A data cleaning process is executed in the data staging

area in order to improve the accuracy of the data warehouse. The data staging area is the place where all

grooming' is done on data after it is culled from the source systems. Staging and ETL phase is considered to

be most crucial stage of data warehousing where maximum responsibility of data quality efforts

resides.It is a prime location for validating data quality from source or auditing and tracking down data issues.

Some of the identified area from literature review are shown in Table

auses of Data Quality Issues at Data Profiling

When possible candidate data sources are identified and finalized data profiling comes in play immediately.

Data profiling is the examination and assessment of your source systems' data quality,integrity and consistency

sometimes also called as source systems analysis

Table 2:causes of data quality at Data profiling

Sr.No CAUSES OF DATA QUALITY PROBLEMS AT DATA

PROFILING

1 Insufficient range and distribution of values or threshold

analysis for required fields.

2 Unreliable and incomplete metadata of data source

3 UserGeneratedSQLqueriesforthedataprofilingpurpose leave

the data quality problems.

4 Inability of evaluation of inconsistent business processes

during data profiling cause data quality problems.

5 Inability of evaluation of data structure, data values and

data relationships before data integration ,propagates poor

data quality

6 Inability of integration between Data profiling and ETL

causes Data quality problem

7 Inappropriate selection of Automated profiling tool cause data

quality issues

8 Insufficient data content analysis against external reference

data causes data quality problems.

9 Insufficient structural analysis of the data sources in the

profiling stage.

10 Insufficient Pattern analysis for given fields within each data

store

Data Quality issue at D ata Staging ETL(Extra action), Transformation and Loading)

One consideration is whether data cleansing is most appropriate a the source system, during the ETL process,

at the staging database,or within the data warehouse..A data cleaning process is executed in the data staging

area in order to improve the accuracy of the data warehouse. The data staging area is the place where all

grooming' is done on data after it is culled from the source systems. Staging and ETL phase is considered to

be most crucial stage of data warehousing where maximum responsibility of data quality efforts resides.It is a prime location for validating data quality from source.

What is Design Thinking?

Let's start with a definition. Design thinking, consistent with IDEO's CEO Tim Brown, is:

"a human-centred approach to innovation that pulls from the designer's toolkit to integrate the requirements of individuals , the chances of technology, and therefore the requirements for business success."

Let's break that down a touch . Design thinking is:

- Human-centered: you've got to place people at the guts of your design
- Approach to innovation: it's a strategy and set of guiding rules
- Designer's toolkit: all the tools, tips, and tricks that designers use to bring ideas and ideas to plug.
- Integrate people, technology, and business: it's to compile the top user (people) using the technology that exists now / within the near future, and style must do that while meeting business requirements.

With all that in mind, we define design thinking as:

A specific methodology/approach for designing new products, services, and experiences that put the user at the guts of the method , using every trick within the book to tie together the user and technology while achieving business requirements.

Design thinking is about taking the approach a designer would — that specialize in the user, incorporating technology, considering the intersection of feasibility, desirability, and viability — outside of a design context. This suggests tinkering and twiddling with a thought , reviewing and failing all the time, and taking note of the people you're designing for and staying laser-focused on their needs.

Design Thinking Process

So that's the ethos behind design thinking. Next, we'd like to know how it's actually done. The method could vary counting on what model you consult, but design thinking generally fits into this framework.

1. Empathize. Work to know your audience – who your audience is, what are their pains, problems, attitudes, and what they need to accomplish at an emotional level beyond your product.

2. Define. Define the matter that you're getting to solve within the context of the pains and challenges you uncovered in your empathy discovery. This results in a problem/challenge statement.
3. Ideate. Come up with a bunch of various ways to unravel the matter .
4. Prototype. Turn those ideas into rock bottom possible fidelity tests that you simply can execute and still get clean data. This stage usually cuts the ideas down as problems are discovered.
5. Test. Get your prototypes ahead of users and see what they assert .

Why Business are Embracing Design Thinking

Over the last 20 years, companies became far, much more customer-centric. As an example , it's hard to imagine a sales team being run today like in Glengarry Glen Ross. Sales are a consultative process now. Products are built with countless revisions of UX testing baked directly into the method .

For sure, every now and again, a Steve Jobs visionary comes and shifts the paradigm of innovation. However, overwhelmingly, organizations are currently within the business of intimately knowing their customers and building and selling things their customers want.

Also, let's be honest — knowing your customer isn't a replacement idea. Organizations have always tried their best to know who's buying them and make products that their customers would also like.

Design thinking, with its customer centricity, may be a natural slot in the new business paradigm.

Why Use Data within the Design Process?

To recap, design thinking is thinking sort of a designer, putting your customer/user first, and solving non-design problems with a design toolkit and methodology.

Now, to data.

The emergence of the web and subsequent massive data collection and storage mean that now, businesses can know their customers better than they ever thought possible. This drives up product quality, successively driving up customer expectation during a vicious feedback circuit .

Now, organizations are expected to understand and anticipate user needs at a private level.

Moreover, you can't deliver that quality with design thinking alone.

Data, especially big data about your users' actual behaviour, can assist you understand where you would like to focus, how users are using your product/service "in the wild", and exactly what pieces are driving people to like your organization.

The acute example of this type of data-ization is product-led growth (PLG), where the merchandise is so focused and attuned to solving users' problems that it literally sells itself.

However, albeit you're not a VC-backed PLG machine like Slack or Dropbox, you'll still use data to know your customers' pains better, and fuel design-thinking-led development .

Conclusion

Design thinking doesn't need to be something that only creative agencies use. Anyone can use the framework to drive business objectives.

Understanding your customer, their pains, and iterating to unravel them is usually getting to be a sound business strategy. And by overlaying the tactic with hard data, you'll drive more meaningful insights, faster and more effectively than ever before.

So what's your next step? Are you able to embrace design thinking and achieve your growth goals?