

LECTURE 3

Introduction to information architecture and design –part v

Learning Goals

Types of information architecture and design

1. Financial
2. Telecommunication
3. Insurance
4. Human Resource

Types and Typical Applications of information architecture and design

Financial information architecture and design

A. First data warehouse that an organization builds. This is appealing because:

1. Nerve center, easy to get attention.
2. In most organizations, smallest data set.
3. Touches all aspects of an organization, with a common denomination i.e. money.
4. Inherent structure of data directly influenced by the day-to-day activities of financial processing.
5. Financial data warehouses suffer from an anomaly of inability to match balanced accounts, due to many legitimate reasons. Financial data warehouses are often the first data warehouse that an organization builds. This is appealing because:

Financial data is ALWAYS at the nerve center of the organization. Therefore, getting attention drawn to a well-built financial data warehouse is a very easy thing to do. In most organizations (but not all), financial data represents one of the smallest volumes of data that exist. Finance cuts across all of the aspects of the corporate data and has a common denominator i.e. Money. Financial data inherently has a structure of data directly influenced by the day-to-day activities of financial processing and the list goes on. For these reasons, finance becomes a very good target for beginning to build the corporate data warehouse.

Financial data warehouses have some severe drawbacks that are not found elsewhere. When purists use a financial data warehouse, they are overjoyed to find that it is almost impossible to

reconcile down to the rupee. Again and again they will report that they saw a report yesterday that stated the balance to be Rs. 12,345,678 but when did the same report today the using our accounting MIS system the balance was Rs. 13,245,678. Hence one cannot trust the new system. Note that it is unlikely that the financial data warehouse would balance to the last rupee, such a notion is completely wrong for a number of reasons. The accounting periods may Be different in different operational systems, but in a data warehouse, it is the corporate calendar.

Telecommunication

- Dominated by sheer volume of data
- Many ways to accommodate call level detail:
- Only a few months of call level detail,
- Storing lots of call level detail scattered over different storage media,
- Storing only selective call level detail, etc.
- Unfortunately, for many kinds of processing, working at an aggregate

Level is simply not possible.

Telecommunications Data Warehouse

The telecommunications data warehouse is dominated by the sheer volume of data generated at the call level subject area. There are many other types of data in this environment as well. No other data warehouse environment is dominated by the size of one of the subject areas as call level detail dominates a telecommunications data warehouse.

There are many ways that call level detail can be accommodated:

1. Only a few months of call level detail,
2. Storing lots of call level detail scattered over different storage media,
3. Summarizing or aggregating call level detail,
4. Storing only selective call level detail, and so forth.

Unfortunately, for many kinds of processing, the only way it can be accommodated is by working at the call level detail. Working at a summary level or an aggregate level simply is not a possibility.

Insurance

1. Insurance data warehouses are similar to other data warehouses BUT with a few exceptions Stored data that is very, very old, used for actuarial processing.
2. Typical business may change dramatically over last 40-50 years, but not
3. Insurance. Therefore, if someone says that the insurance business has not changed much over the last 40-50 years would be telling the truth.
4. In retailing or telecomm there are a few important dates, but in the insurance
5. Environment there are many dates of many kinds. In the retail sales environment there are a few dates such as: sale date, stocking date, and perhaps manufacturing date. In the banking environment there is the transaction date. However, in the insurance environment there are dates everywhere of every kind.
6. Long operational business cycles, in years. Thus the operating speed is different.
7. Consider a bank, you insert the ATM card and draw the money instantly, or deposit a check and it is cashed at most in a week. You go to store and purchase an item, and instantly make a telephone call. However, in an insurance environment a claim is made and it may take several years before it is settled.
8. Transactions are not gathered and processed, but are in kind of “frozen” state.
9. Thus a very unique approach of design & implementation.

Differences between different types of data warehouses

1. Financial data warehouses, often the first a corporation builds. However, it will
2. Not reconcile down to the last rupee with the existing operational financial environment.
3. Insurance data warehouses are similar to other data warehouses with a few
4. Exceptions: such as the length of time that insurance data warehouses exists, in terms of the dates found in the business, and in terms of the operational business cycle.
5. The human resources data warehouse is different from others because this without
6. Doubt has only one major subject area (yes you guessed it right!).
7. The telecommunications data warehouse is dominated by the sheer volume of data
8. Generated at one subject area.

Indeed there are many types of data warehouses, each with their own peculiarities and all of them

cannot be discussed here.

Typical Applications of DWH

There are, and there can be many applications of a data warehouse. It is not possible to discuss all of them. Some representative applications are listed to be discussed as follows:

1. Fraud detection.
2. Profitability analysis.
3. Direct mail/database marketing.
4. Credit risk prediction.
5. Customer retention modeling.
6. Yield management.
7. Inventory management.

ROI on any one of these applications can justify hardware/software and consultancy costs in most organizations.

Fraud detection

1. By observing data usage patterns.
2. People have typical purchase patterns.
3. Deviation from patterns.
4. Certain cities notorious for fraud.
5. Certain items bought by stolen cards.
6. Similar behavior for stolen phone cards.

You can look at patterns of data usage and behavior of customer and detect fraud very easily. Let's suppose you are a Credit Card Company. You know that Danyal as a customer has certain behaviors. He tends to buy airplane tickets, make hotel reservations, rents cars using his credit card. When he goes shopping, it is usually only on a weekend and only on certain stores that he goes shopping and buys sports goods and action clothing. And then all of sudden you see very different behaviors coming in, such as credit card transactions in Kenya. Kenya by the way has very high stolen credit card rate. And the patterns of buying behavior are stereos, and CDs and

TVs i.e. electronic goods that are typically the things that people buy on stolen credit cards. You can notice and say “look, this is not airplane or hotel reservations, this does not look like Danyal”. It is not in the city where Danyal lives or goes to very often. Although he was there a week ago, so maybe this is how you can tell that when and where his credit card was stolen. And the purchases were made after he left the city. Hence “high probability of fraud”.

So I can predict based on the behavior of the individuals, when things out of the typical routine happen. Then I can stop the fraud happening even before it has spread. That’s the example in credit card but you can do the same thing with telephone cards. I am sure you have used these telephone cards. You punch in your number and you get access to telephone network. You find people at airports or railway stations who make their living by stealing the telephone card codes. They stay near the phones at the air port, they act like they are talking on the phone, but they are really looking over your shoulder when you are punching in the code. And then they memorize the code and then go sell it on the street to people who make long distance phone calls to all their friends in Brazil and Australia all these expensive places. You can look at the calling patterns and say look “Danyal makes regular calls to areas like say Rawalpindi and to friends in Kilifi” and all of sudden we see all these phone calls to Toba Taik Singh or Bhawalpur, meaning there is something wrong here.

Profitability Analysis

1. Banks know if they are profitable or not.
2. Don’t know which customers are profitable.
3. Typically more than 50% are NOT profitable.
4. Don’t know which one?
5. Balance is not enough, transactional behavior is the key.
6. Restructure products and pricing strategies.
7. Life-time profitability models (next 3-5 years).

It is another example. If you go to an average bank, to Kenya or anywhere in the world, the bank will know if they are profitable or not. But they don’t know which customers are profitable or not, and at most banks this is true, especially true for retail banks. These banks do business with the consumers, such that more than 50 % of the customers are unprofitable. In other word the

bank is losing money on 50% of their customers. But they don't know which 50%? That's the problem. So the idea behind profitability analysis is to do the analysis to figure out which customers are profitable and which customers are not profitable. And then based on that they can restructure their product offering and their pricing strategies to do more profitable business. I used a banking example but same is true for Telecommunications Company or any other consumer oriented business that requires access to detail data. It is not sufficient to just know the account balance, but also transactional behavior and so on to do profitability analysis. And once I know profitability retrospectively, I can also build predictive models to understand what it's going to be prospectively. So I can build what's called lifetime value models to using the historical data to predict what the future profitability will be for the next 3 or 5 years.

Direct mail marketing

1. Targeted marketing.
2. Offering high bandwidth package NOT to all users.
3. Know from call detail records of web surfing.
4. Saves marketing expense, saving pennies.
5. Knowing your customers better.

Assume you are a telecommunications company, and you have an offering that you can make such as to give value added Internet access through mobile phones. You can send a letter or an SMS to every customer in the country and tell them about this value added Internet access service. But 95% of that mail will be completely wasted. Because a small percentage of customers will be interested in such as service. So the question here is how you your message to the right people in a meaningful way. If you have the call detailed records, you can tell very easily which of your customer's access the Internet and which of them just make missed calls. Because the patters of the call detail records of Internet access are completely different from making voice calls and making missed calls. They have different durations; they have different data bandwidth requirement and so on. So by having access to detail data, you can target those people that are better to get a second account, or better to get a value added connection/service.

Credit risk prediction

1. Who should get a loan?
2. Customer segregation i.e. stable vs. rolling.
3. Qualitative decision making NOT subjective.
4. Different interest rates for different customers.
5. Do not subsidize bad customer on the basis of good.

Credit risk prediction is another area where data warehouse is used quite frequently. Let's say I and my friend both apply for home loan to construct a house. How would the bank determine which one of those should get a loan? So typically they will say, ok I have been employed for 5 years, have good salary, married, have children etc etc and then based on the context of application then they decide yes we think you are a low risk we will give you the loan. Or say no you are a high risk because you had a job only for one year, your salary is not high, and you have lots of change of jobs recently, so you know we don't feel so good about the risk we are taking, so we won't give you the loan. So typically today it is based on what is called qualitative decision-making. Basically in the offices these are just references say OK Saeed is a nice person or is related to me or is my neighbor so we give him the loan. And I don't really know Waleed over here so we don't give him a loan. This is not the best way to give fair loan to people and is not the best way to manage risk in the business. So with credit risk, the idea is to build quantitative models, which predict risk based on historical data. And not only I use the historical data to predict the risk, if I get more sophisticated, what I can do is called risk based pricing. So today lets say you offer me (Saeed) a loan, and also to the other applicant i.e. Waleed also loan, and both are given the loan at the same interest rates. However, if Waleed is a lower risk than Saeed then why should both pay the same interest rates? If Waleed is a low risk, then he should get a lower interest rate otherwise the bank is subsidizing the higher risk loan. So what's happening is that you are using very precise credit prediction models in order to adjust the interest to give lower interest to people who are at lower risk of default. And again it should use some data mining algorithms to predict the risk, which is called the risk index on every individual consumer. So that every consumer gets a loan at an interest that is appropriate to their risk level.

Yield Management

1. Works for fixed inventory businesses.
2. The price of item suddenly goes to zero.
3. Item prices vary for varying customers.
4. Example: Air Lines, Hotels, Car rentals etc.
5. Price of (say) Air Ticket depends on:
 - i. How much in advance ticket was bought?
 - ii. How many vacant seats were present?
 - iii. How profitable is the customer?
 - iv. Ticket is one-way or return?

Yield management works, for example for sophisticated airlines which use a variable pricing mechanism for the seats. Say you are flying on a plane; 90% of the time the person sitting next to you pays a different price for the ticket as compared to yours. The seats are basically same, yet you pay a different price for the ticket, why is it that? Because the airline used a sophisticated method to see how much in advance the person bought the ticket, and how many seats were available on the plane at the time the person bought the ticket. Or did they buy a round ticket or one-way ticket; there are all kinds of sophisticated ways to do the pricing. Yield management helps decide what should be the price at any time for that airline seat. Yield management is important for almost any type of fixed inventory business. For example, once the plane leaves, the price of the seat if it is empty or full means nothing any more. So the goal of the airline is to fill every seat with the maximum possible price for that seat. The same is true for hotels. For example a three star hotel had an empty room last night. If they can get Rs. 1500 for that night it is better than having nothing at all. So someone who booked the room for example as a group would have paid lower price per room as compared to someone who is booking only one room. The price is different, because the hotels use yield management to fill out their hotel rooms as profitably as they can. This is applicable to car rentals also.

Agriculture Systems

1. Agri and related data collected for decades.
2. Metrological data consists of 50+ attributes.

3. Decision making based on expert judgment.
4. Lack of integration results in underutilization.
5. What is required, in which amount and when?

Each year different government departments and agencies in Pakistan create tens of thousands of digital and non digital files from thousands of pest-scouting surveys yield surveys, metrological data collection, river flows etc. This data collection has been going on for decades. The data collected has never been compiled, standardized and integrated to give a complete picture. Thus the lack of data integration and standardization contributes to an under-utilization of historical data, and inevitably results in an inability to perform any scientific predictive analysis for effective decision support and policy making. An Agriculture data warehouse is the answer, as processing 100+ variables by experts for large historical data is not possible. This has repeatedly resulted in tragic outcomes.

Although the benefits are many, but are not without some major issues or problems, some of them are as follows:

Major Issues

1. Unavailability of data and in an illegible format and form.
2. Most of the detail data is not digitized; therefore it cannot be readily used by computers.
3. Dispersion of data across multiple and geographically displaced organizations.
4. Even though they said organizations are required to share the data with common
5. Citizens, no official data sharing mechanism is currently in place.

Until the requisite training is provided, government employees and farmers are not equipped with enough technical and literate skills to use latest information technology tools to use the data warehouse