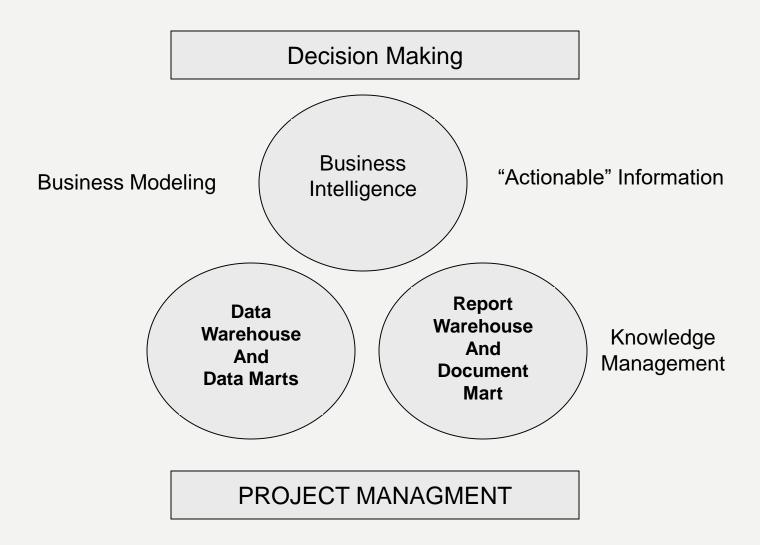
BUSINESS ANALYTICS

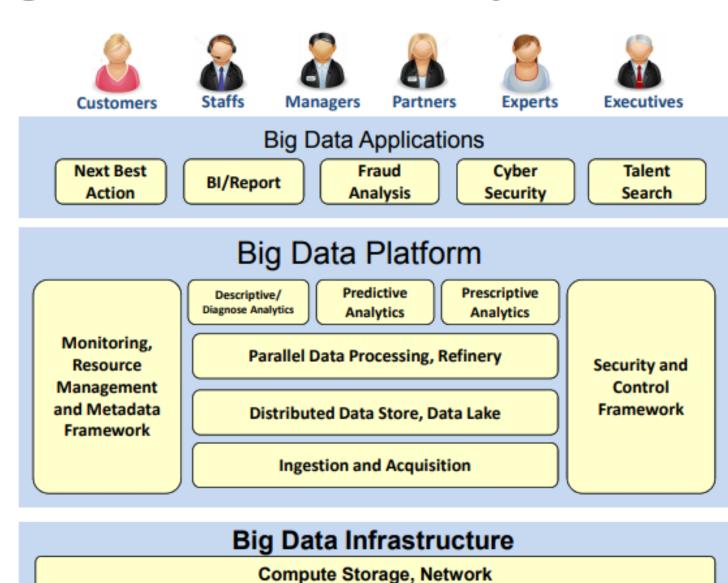
WHAT IS BUSINESS INTELLIGENCE?

- Bl(Business Intelligence) is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions. It is a suite of software and services to transform data into actionable intelligence and knowledge.
- BI has a direct impact on organization's strategic, tactical and operational business decisions. BI supports fact-based decision making using historical data rather than assumptions and gut feeling.
- BI tools perform data analysis and create reports, summaries, dashboards, maps, graphs, and charts to provide users with detailed intelligence about the nature of the business.

Business Analytics



Big Data for Business Analytics Platform



How is data analytics transforming the world?

The analytics play a vital role in every kind of business because it takes every organization to the next level by reducing major stress in business. There are so many <u>analytics tools</u> has been introduced in the market to analyze data rapidly where this leads to tackle so many trolls in business without relying on others. The data Analytics lets you track business such as your store and generates reports that will help you with current updates on your business.

It also derive the right insights of your business and able to visualize effectively with <u>search based analytics</u>. It gives the ability to analyze all the disparate data sources that companies deal with every day and now the executives iterate and drive solutions in seconds instead of letting months or weeks, Through this insight, businesses may be able to gain an edge over their rivals and make superior business decisions and thus changes the world in all way.

- I) Health department: Big data and analytics provides a significant support in the field or area of health care. For eg: U have a huge data of the patients of a particular country or state. Now by the proper analysis of this huge amount of data u can classify the kind of diseases, health problems and the percentage of it in that particular area. Based on your results, a proper healthcare plan for that area can be summoned and prepared. Seeing the business perspective, a particular medical company can provide specific medics to that area based on the analysis.
- 2) Crime: Getting the data of how many prisoners are there in a particular area, the number of crimes committed, cases registered, weapons confesticated and drug trafficking. Analysing all these data and than planning a different model for different regions. The security departments can use this analysis to deploy a team with the particular skills for that location.

- 3) Science: scientists are using big data analytics to predict the upcoming events in nature like, earthquakes, tornados, floods, rain, draught, Particular weather data of any area is collected and than by the use of analytics, a result is generated to know about the near future happenings.
- 4) Business: This is the area where big data and analytics is playing the most significant and profitable role. Analysing the market trend based on sales data, demand data etc and than setting a road map for the future of the company. Stock market analysis (predictive analysis) helps investors to judge that particular stock in which they can invest or withdraw. Big data analytics help in improving the customer satisfaction policies for various companies and brands like Mac D, Wallmart and various other brands

Business Intelligence Are So Important For Businesses

While most things in life are anything but predictable, the cost of implementing and maintaining technology, especially business intelligence analytics.

Businesses stay in line with leading-edge capabilities without significant capital and head count. Each month, a monthly subscription is paid to access the exact analytics capabilities they need now – with the flexibility to introduce or remove organizations, users, or capabilities later as the company evolves.

Decision makers can discover, predict, visualize, collaborate, model, and manage data – all in one place and with a common logic. With this infrastructure, businesses can take advantage of the latest digital innovations in a seamless, familiar user experience. This strategy not only reduces the cost of maintaining analytics, but also increase ROI.

Guaranteed storage in a data center can be protected against a range of ransomware and viruses. As analytics become a natural part of running all aspects of the business, decision makers can take advantage of a growing, yet secure, collection of data to adopt predictive analytics and cross-functional forecasting.

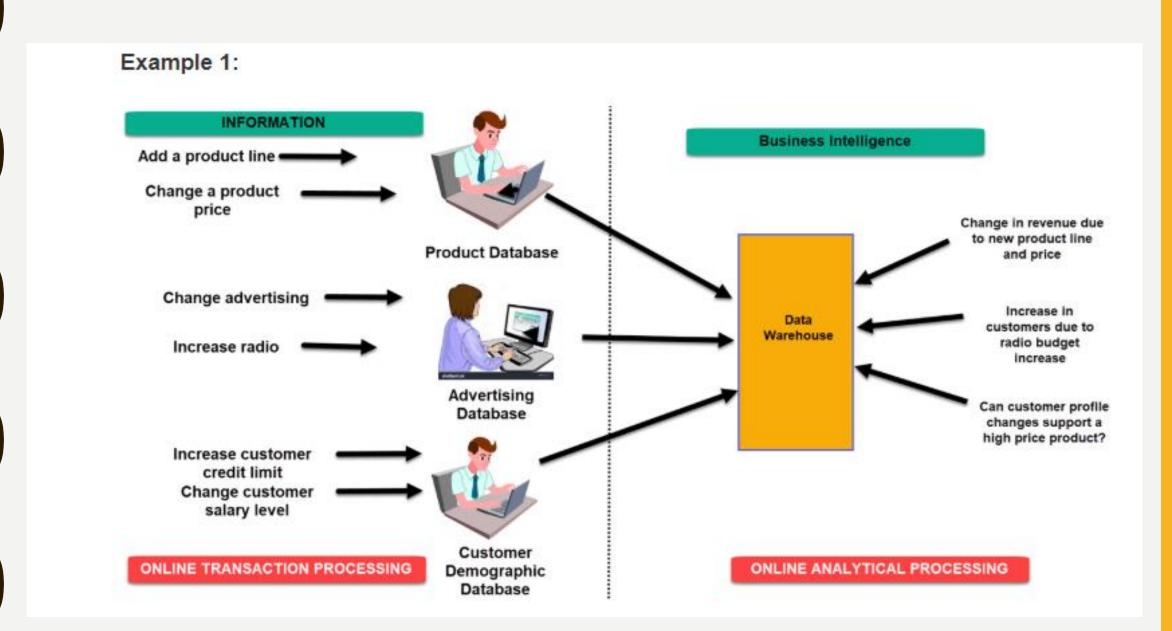
WHY IS BI IMPORTANT?

- Measurement: creating KPI (Key Performance Indicators) based on historic data
- Identify and set benchmarks for varied processes.
- With BI systems organizations can identify market trends and spot business problems that need to be addressed.
- BI helps on data visualization that enhances the data quality and thereby the quality of decision making.
- BI systems can be used not just by enterprises but SME (Small and Medium Enterprises)

HOW BUSINESS INTELLIGENCE SYSTEMS ARE IMPLEMENTED?

- **Step I**) Raw Data from corporate databases is extracted. The data could be spread across multiple systems heterogeneous systems.
- **Step 2)** The data is cleaned and transformed into the data warehouse. The table can be linked, and data cubes are formed.
- **Step 3)** Using BI system the user can ask quires, request ad-hoc reports or conduct any other analysis.

Examples of Business Intelligence System used in Practice



- In an Online Transaction Processing (OLTP) system information that could be fed into product database could be
 - add a product line
 - change a product price

Correspondingly, in a Business Intelligence system query that would be executed for the product subject area could be did the addition of new product line or change in product price increase revenues

In an advertising database of OLTP system query that could be executed

- Changed in advertisement options
- Increase radio budget
- Correspondingly, in BI system query that could be executed would be how many new clients added due to change in radio budget
- In OLTP system dealing with customer demographic data bases data that could be fed would be
 - increase customer credit limit
 - change in customer salary level
- Correspondingly in the OLAP system query that could be executed would be can customer profile changes support support higher product price

FOUR TYPES OF BI USERS

- I.The Professional Data Analyst:
- The data analyst is a statistician who always needs to drill deep down into data. BI system helps them to get fresh insights to develop unique business strategies.
- 2.The IT users:
- The IT user also plays a dominant role in maintaining the BI infrastructure.
- 3. The head of the company:
- CEO or CXO can increase the profit of their business by improving operational efficiency in their business.
- 4.The Business Users"
- Business intelligence users can be found from across the organization. There are mainly two types of business users
- Casual business intelligence user
- The power user.
- The difference between both of them is that a power user has the capability of working with complex data sets, while the casual user need will make him use dashboards to evaluate predefined sets of data.

ADVANTAGES OF BUSINESS INTELLIGENCE

- I. Boost productivity
- With a BI program, It is possible for businesses to create reports with a single click thus saves lots of time and resources. It also allows employees to be more productive on their tasks.
- 2.To improve visibility
- Bl also helps to improve the visibility of these processes and make it possible to identify any areas which need attention.
- 3. Fix Accountability
- BI system assigns accountability in the organization as there must be someone who should own accountability and ownership for the organization's performance against its set goals.

4. It gives a bird's eye view:

- BI system also helps organizations as decision makers get an overall bird's eye view through typical BI features like dashboards and scorecards.
- 5. It streamlines business processes:
- BI takes out all complexity associated with business processes. It also automates analytics by offering predictive analysis, computer modeling, benchmarking and other methodologies.
- 6. It allows for easy analytics.
- BI software has democratized its usage, allowing even nontechnical or non-analysts users to collect and process data quickly. This also allows putting the power of analytics from the hand's many people.

BI SYSTEM DISADVANTAGES

I. Cost:

• Business intelligence can prove costly for small as well as for medium-sized enterprises. The use of such type of system may be expensive for routine business transactions.

2. Complexity:

• Another drawback of BI is its complexity in implementation of datawarehouse. It can be so complex that it can make business techniques rigid to deal with.

3. Limited use

• Like all improved technologies, BI was first established keeping in consideration the buying competence of rich firms. Therefore, BI system is yet not affordable for many small and medium size companies.

4. Time Consuming Implementation

• It takes almost one and half year for data warehousing system to be completely implemented. Therefore, it is a time-consuming process.

TRENDS IN BUSINESS INTELLIGENCE

- Artificial Intelligence: Gartner' report indicates that Al and machine learning now take on complex tasks done by human intelligence. This capability is being leveraged to come up with real-time data analysis and dashboard reporting.
- Collaborative BI: BI software combined with collaboration tools, including social media, and other latest technologies enhance the working and sharing by teams for collaborative decision making.
- **Embedded BI:** Embedded BI allows the integration of BI software or some of its features into another business application for enhancing and extending it's reporting functionality.
- Cloud Analytics: Bl applications will be soon offered in the cloud, and more businesses will be shifting to this technology. As per their predictions within a couple of years, the spending on cloud-based analytics will grow 4.5 times faster.

UNIT 1

WHAT IS ANALYTICS?

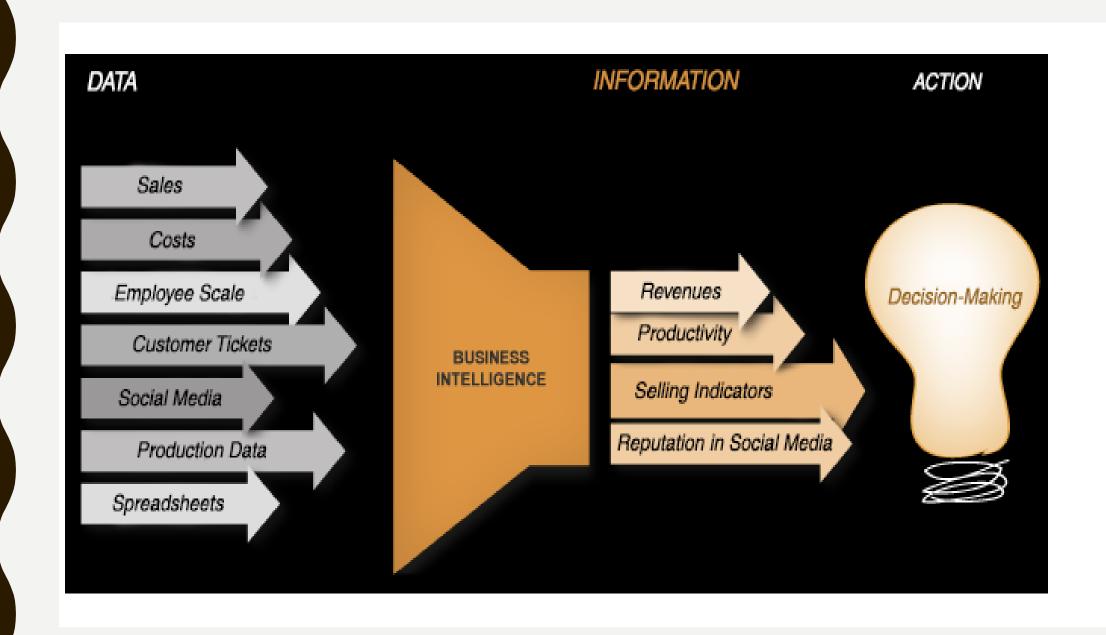
- Business analytics (BA) is the practice of <u>iterative</u>, methodical exploration of an organization's data, with an emphasis on statistical analysis. Business analytics is used by companies committed to data-driven decision-making.
- BA is used to gain insights that inform business decisions and can be used to automate and optimize business processes. Data-driven companies treat their data as a corporate asset and leverage it for a competitive advantage. Successful business analytics depends on data quality, skilled analysts who understand the technologies and the business, and an organizational commitment to data-driven decision-making.

BUSINESS ANALYTICS EXAMPLES

- Business analytics techniques break down into two main areas. The first is basic <u>business</u> <u>intelligence</u>. This involves examining historical data to get a sense of how a business department, team or staff member performed over a particular time. This is a mature practice that most enterprises are fairly accomplished at using.
- The second area of business analytics involves deeper <u>statistical analysis</u>. This may mean doing <u>predictive analytics</u> by applying statistical <u>algorithms</u> to historical data to make a prediction about future performance of a product, service or website design change. Or, it could mean using other <u>advanced analytics</u> techniques, like cluster analysis, to group customers based on similarities across several data points.

HOW BUSINESS INTELLIGENCE WORKS?

- Through a data capture process, the information is acquired from any system and recorded in a database specifically modelled for your business.
- With this modeling, known as Dimensional Modeling, it is possible to work with a large mass of data keeping a high performance.



WHY BUSINESS ANALYTICS HAS BECOME IMPORTANT IN VARIOUS INDUSTRIES

- Companies have widely embraced the use of analytics to streamline operations and improve processes. But <u>implementing analytics data</u> that informs intelligent and effective business decisions is not as easy as a snap of the fingers.
- In a survey conducted by Bloomberg Business week Research Services, nearly 97% of respondents reported their companies have adopted analytics. The three most sought-after goals were the ability to reduce costs, increase profitability and improve risk management. However, many organizations struggle with making sure the <u>data is accurate and consistent</u>.
- Analytics data is everywhere and sorting through it to find what is useful and pertinent to your business is a necessary skill to be effective in the current marketplace. These days, analytics is being used to determine everything from Supreme Court case outcomes to personalized marketing efforts. The challenge is to understand how analytics can help your business and begin to address any issues you believe are most important to short- and long-term success.

Analyzing Data to Identify Business Opportunities

- Analyzing data more often than not increases efficiency, but also helps <u>identify new business</u> <u>opportunities</u> that may have been otherwise overlooked, such as untapped customer segments. In doing so, the potential for growth and profitability becomes endless and more intelligence based.
- Many professionals can discern short-term trends, but are less proficient at predicting obstacles that plague their business down the road. Computer models based on data analytics help companies see shifts in what customers buy and give a clear picture of what products should be highlighted or updated. Whether it's a production concern, a customer service issue or a deficiency among your employees, analytics can help to highlight key areas of concern when it comes to your venture's ability to make a profit.
- Big data has also been <u>used as an HR tool</u> to recruit perspective job candidates. Collecting data from many different sources allows companies to assess a candidate's skills and traits to help determine how they could fit into the corporate culture and workplace.

Using Analytics to Prevent Shipping Breakdowns

- Shipping companies are tasked with the logistics challenge of delivering millions of packages each day. Many have turned to analytics to maximize the performance and reliability of their vehicles. By looking at sensor data from each vehicle within a shipping fleet, companies can keep track of the state of the parts in the vehicle and determine what parts may prove to be problematic.
- Addressing problem areas before they become major issues makes it possible for companies to ensure their vehicles stay on the road and don't interrupt the flow of business reducing driver downtime, overall maintenance cost and customer dissatisfaction. By incorporating analytics into their approach to mechanical maintenance, the shipping industry has made itself more efficient.

Better Targeting Customers with Business Analytics

- An analysis by McKinsey & Company showed that using data to make better marketing decisions can increase marketing productivity by 15-20%. A good example of this is retail giant Target's "pregnancy prediction score." Target assigns a score based on a customer's purchases that indicate the possibility of a pregnancy; the retailer uses purchase data to determine the types of coupons and special discounts Target would send to a customer's email address.
- There is a ton of information companies can use for predictive analytics that help streamline a customer's experience with a brand. Finding the right tools to examine your customer's buying and Internet browsing habits, and implementing them to provide reliable and actionable intelligence can activate buyer instincts and embed your brand into customers' minds.

THE HISTORY OF ANALYTICS AND HOW IT HAS CHANGED TODAY

- Ten years ago, Jeanne Harris and I published the book <u>Competing on Analytics</u>, and we've just finished updating it for publication in September. One major reason for the update is that analytical technology has changed dramatically over the last decade; the sections we wrote on those topics have become woefully out of date. So revising our book offered us a chance to take stock of 10 years of change in analytics.
- Of course, not everything is different. Some technologies from a decade ago are still in broad use, and I'll describe them here too. There has been even more stability in analytical leadership, change management, and culture, and in many cases those remain the toughest problems to address. But we're here to talk about technology. Here's a brief summary of what's changed in the past decade.

- The last decade, of course, was the era of big data. New data sources such as online clickstreams required a variety of new hardware offerings on premise and in the cloud, primarily involving distributed computing spreading analytical calculations across multiple commodity servers or specialized data appliances. Such machines often analyze data "in memory," which can dramatically accelerate times-to-answer. Cloud-based analytics made it possible for organizations to acquire massive amounts of computing power for short periods at low cost. Even small businesses could get in on the act, and big companies began using these tools not just for big data but also for traditional small, structured data.
- Along with the hardware advances, the need to store and process big data in new ways led to
 a whole constellation of open source software, such as Hadoop and scripting languages.
 Hadoop is used to store and do basic processing on big data, and it's typically more than an
 order of magnitude cheaper than a data warehouse for similar volumes of data. Today many
 organizations are employing Hadoop-based data lakes to store different types of data in their
 original formats until they need to be structured and analyzed.

- Since much of big data is relatively unstructured, data scientists created ways to make it structured and ready for statistical analysis, with new (and old) scripting languages like Pig, Hive, and Python. More-specialized open source tools, such as Spark for streaming data and R for statistics, have also gained substantial popularity. The process of acquiring and using open source software is a major change in itself for established businesses.
- The technologies I've mentioned for analytics thus far are primarily separate from other types of systems, but many organizations today want and need to integrate analytics with their production applications. They might draw from CRM systems to evaluate the lifetime value of a customer, for example, or optimize pricing based on supply chain systems about available inventory. In order to integrate with these systems, a component-based or "microservices" approach to analytical technology can be very helpful. This involves small bits of code or an API call being embedded into a system to deliver a small, contained analytical result; open source software has abetted this trend.

- This embedded approach is now used to facilitate "analytics at the edge" or "streaming analytics." Small analytical programs running on a local microprocessor, for example, might be able to analyze data coming from drill bit sensors in an oil well drill and tell the bit whether to speed up or slow down. With internet of things data becoming popular in many industries, analyzing data near the source will become increasingly important, particularly in remote geographies where telecommunications constraints might limit centralization of data.
- Another key change in the analytics technology landscape involves autonomous analytics a form of artificial intelligence or cognitive technology. Analytics in the past were created for human decision makers, who considered the output and made the final decision. But machine learning technologies can take the next step and actually make the decision or adopt the recommended action. Most cognitive technologies are statistics-based at their core, and they can dramatically improve the productivity and effectiveness of data analysis.

- Of course, as is often the case with information technology, the previous analytical technologies haven't gone away after all, mainframes are still humming away in many companies. Firms still use statistics packages, spreadsheets, data warehouses and marts, visual analytics, and business intelligence tools. Most large organizations are beginning to explore open source software, but they still use substantial numbers of proprietary analytics tools as well.
- It's often the case, for example, that it's easier to acquire specialized analytics solutions say, for anti-money laundering analysis in a bank than to build your own with open source. In data storage there are similar open/proprietary combinations. Structured data in rows and columns requiring security and access controls can remain in data warehouses, while unstructured/prestructured data resides in a data lake. Of course, the open source software is free, but the people who can work with open source tools may be more expensive than those who are capable with proprietary technologies.

- The change in analytics technologies has been rapid and broad. There's no doubt that the current array of analytical technologies is more powerful and less expensive than the previous generation. It enables companies to store and analyze both far more data and many different types of it. Analyses and recommendations come much faster, approaching real time in many cases. In short, all analytical boats have risen.
- However, these new tools are also more complex and in many cases require higher levels of expertise to work with. As analytics has grown in importance over the last decade, the commitments that organizations must make to excel with it have also grown. Because so many companies have realized that analytics are critical to their business success, new technologies haven't necessarily made it easier to become and remain an analytical competitor. Using state-of-the-art analytical technologies is a prerequisite for success, but their widespread availability puts an increasing premium on nontechnical factors like analytical leadership, culture, and strategy.

STEPS FOR ANALYZING UNSTRUCTURED DATA

- Data analysis is becoming an important part of businesses growth. It is important for businesses to understand structured and unstructured data in order to make a right decision for their businesses to grow.
- Below are 10 steps to follow that will help analyze unstructured data for successful business enterprises.

I. Decide on a Data Source

• It's very important to understand the source of data that is beneficial for your small business enterprise. You may use one or more data source to collect the information that is relevant to your business. Collecting data from random sources is never a good idea because you might corrupt the data or even lose some. Hence it's recommended to survey the relevant data source before you start collecting data. There are some online big data development tools that you can use to collect the data.

2. Manage Your Unstructured Data Search

• Collected data will vary in usage if it's structured or unstructured. Finding and collecting data is only one step; structuring your unstructured data search and making it useful is entirely another thing. The second step is as important as collecting the data but can have a negative impact on your clients and your own business if not managed properly. Invest in a good business management tool before you have too much unstructured data.

3. Eliminating Useless Data

• After collection and structuring the data comes the third step of eliminating data. Although most data is going to only further your company's growth, sometimes it can also be detrimental. If your unstructured data takes up too much space on your businesses hard drives, storage, or backups, this may affect your business' ability to strive. This reduces further confusion and saves you from wasting your time on data that are not beneficial.

4. Prepare Data for Storage

• Preparing data means to remove all the whitespace, formatting issues, etc. from the data. Now when you have all the data, no matter useful for the business or not, you can start making a stack of useful data and indexing unstructured data once the data is prepared.

• 5. Decide the Technology for Data Stack and Storage

• After the elimination of useless data, *stacking* your data is the ideal next step. Be sure to use the latest technology to save and stack data so that it is easy for you and your employees who are also working with data to fetch the most important and mandatory data in no time. Also, ensure that you have a maintained and updated data backup and recovery service.

6. Keep All the Data Until It Is Stored

• Seems obvious, but always make sure you save data — whether it is structured or unstructured — before deleting anything! Recent natural disasters around the globe have proven that a current and updated data backup recovery system is essential and necessary, especially during times of crisis. You may not know that all of your data is about to get deleted. So, think ahead and save your work often.

7. Retrieve Useful Information

• After a proper data backup, you can recover data. This step is useful because you will need to retrieve data after converting unstructured information as well.

8. Ontology Evaluation

• It's good if you can show a relationship between the source of information and the data extracted. This will help you in providing useful insights in regards to the organization of data. Your company will need to be able to explain the steps and processes you took, so keep a record in order to recognize patterns and keep consistent with the process.

9. Record Statistics

• Once you have made the unstructured data search into the structured data through all the steps mentioned above, it's time to create statistics. Classify and segment the data for easy use and study in order to create a great flow for future use.

I0.Analyze the Data

• This is the last step of indexing unstructured data. After all the raw data are structured, it comes the time to analyze and make decisions that are relevant and beneficial for the business. Indexing also helps your small business make consistent patterns for future use.

HOW ANALYTICS IS MAKING THE WORLD SMARTER

Citizen feedback

Alberta Parks manages about 250 campgrounds and 14,000 campsites that receive more than 1.8 million overnight visitors every year. From May to mid-October, the Parks department receives 15,000 customer survey replies. With the help of software that analyzes text, the department doesn't have to wait until a post-season quiet time to sift through the surveys. It is able to quickly identify trends in the feedback comments and make adjustments dynamically to improve the camping experience for visitors.

Criminal justice and public safety

One of the UK's largest police forces deployed an intelligence analytics platform across the entire enterprise. The mission-critical system contains 12 million documents, 9 million structured records and provides real-time intelligence 24 hours a day. It is used by more than 40,000 officers and police personnel daily and can be accessed securely by other government agencies. The platform provides integration with confidential, highly specialized and highly secure protected units. With the intelligence management system in place, information can be acted upon in real time to protect the public around the clock.

Child well-being

In New Zealand, the Ministry of Social Development is using analytics as a transformation tool that helps struggling young people create a better future. This is a perfect example of inclusive growth: It helps the individual, the society and the economy alike. Better targeting empowers welfare beneficiaries with confidence and life skills, and reduces the cycle of long-term benefit dependency.

Economic development

Local government agencies in the Campania region around Naples, Italy, rely on analytics to evaluate and respond to more than 7,000 applications for regional projects; verify compliance with tender specifications; and manage the financing phases and progress toward completion – all while operating with maximum transparency and speed. In the past few years, officials have transitioned from manual processes to an integrated monitoring system (IMS) that manages the financing of its scientific research initiatives.

Education

The Fox Chapel Area School District outside Pittsburgh has earned numerous accolades, with each school in the district winning the US Department of Education's coveted Blue Ribbon award. Yet there was one group of students the district wanted to serve better: those with learning disabilities. By using Pennsylvania's Value-Added Assessment System (PVAAS), the district was able to better track year-to-year growth of all students. The result: A decade ago, learning-disabled 11th graders showed 14 percent proficiency in math and 29 percent proficiency in reading on state-mandated tests. Today, the proficiency level is 69 percent for both subjects.

Citizen services

In France, job seekers who collect unemployment benefits are receiving assistance that is customized to their unique situations. Analytics helps empower local service branches to design personalized pathways to employment, helping the branches customize for citizen needs while still meeting statewide quality and consistency standards and goals.

Fraud detection

In Los Angeles County, the Department of Public Social Services (DPSS) offers a range of programs to alleviate hardship and promote health, personal responsibility and economic independence. Across the county's many communities, DPSS offers temporary financial assistance, employment services, free/low-cost health insurance, food benefits, in-home services for the elderly and disabled, and other financial assistance. To support program integrity efforts in the CalWORKs Stage I Child Care Program, LA County turned to SAS® Analytics to identify potential fraud, enhance investigations and prevent improper payments. By doing so, it has helped the most vulnerable members of the community while protecting millions in taxpayer dollars.

Health care

In Wake County, North Carolina, cardiac arrest victims have a better chance at survival thanks to new, analytics-driven recommendations from the county's Emergency Medical Services (EMS). Based on an analysis of 20 years of data about cardiac arrest patients, Wake County EMS changed its recommendations for how long to conduct CPR from 25 minutes (the industry standard) to 60 minutes or more – if EMTs see cardiac activity during CPR. A study found that using the new CPR guidelines saved 100 people in the first year.

Natural resources and conservation

Cary, NC, helps citizens monitor water usage and saves millions of dollars in the process with an automated water meter system. Aquastar replaced monthly manual readings with a wireless IoT system that collects hourly meter data, allowing the town to operate more efficiently and provide citizens with in-depth water usage reports powered by SAS Analytics. Cary replaced approximately 60,000 residential and commercial water meters with new, state-of-the-art meters and radio transmitters. Through operational savings, Aquastar will save the utility more than \$10 million above the cost of the project.

Transportation

The resundsbron between Denmark and Sweden is a five-mile expanse of bridge and tunnel that connects the two countries and two major metropolitan areas: Copenhagen in Denmark and Malmö in Sweden. With the help of analytics, it also connects its "customers" with destination ideas based on their unique likes and dislikes – the bridge's way of driving the revenue-generating traffic it needs to support itself. Using SAS® Customer Intelligence, the bridge's marketing and customerservice teams share meaningful, relevant offers unique to each of the 180,000 travelers who own toll passes. Open rates for the weekly emails are 30 percent.

• In this way analytics serving the greater good in cities around the world. Analytics can help prioritize resources in disaster relief efforts, measure the results of innovative new programs to benefit foster children, and help conserve dwindling energy reserves.

WHERE THE FUTURE OF ANALYTICS LIES

- The era of big data has witnessed a paradigm shift into analytics. Today, it's no longer sufficient to simply gather data from social media, IoT, and wearable devices, and be unable to manage or filter it. It is more about delivering the right data to the right person, at the right time.
- This trend is growing crucial as data is multiplying every day and pouring in from various devices and smart machines including wearables, electronic gadgets, and other devices. Such factors call for the treatment of vast pools of structured and unstructured data with care and precision. This is precisely where invisible analytics come in.

Big Data was the Past; 2015 is the Start Point to Take Analytics to the Next Level

- By far, big data has remained as an enabler of the new wave of analytics solutions. However, the challenge for big data analytics lies in the traditional hardware storage capacity and processing rates that execrably lag during operations, thus becoming inefficient in supporting the demand for handling large amounts of data.
- As we look into the future, more products and technologies are leaning towards what can be possibly done with the large amounts of data that is already present, without the need of harvesting more data. This time, analytics will be the enabler. Market experts predict that analytics will become deeply, but invisibly embedded everywhere. The increasing invisibility in analytics is in the same breath as the growth of the volume of data and the rising trend toward embedded business intelligence (BI).
- Pervasive BI is gaining immense traction these days. The adoption rates of BI hover at approximately 30 percent in a typical business or enterprise environment.

• The Future is All about Advanced, Pervasive, and Invisible Analytics

- Analytics will continue to grow due to the Internet of Things (IoT), creating large pools of data. Analytics will be deeply embedded and virtually invisible in the coming years. It will be the major highlight in the future as the volume of data generated by various embedded systems is rapidly increasing. Every application will need to be an analytic application and the value will be in the answers, not the data.
- With the help of advanced analytics techniques such as natural language processing, data mining, text analytics, statistics, predictive analytics, and machine learning, organizations will utilize big data analytics and similar analytical tools to gain deeper insights and make significantly better business decisions.

HOW BUSINESS ANALYTICS CAN HELP TURN DATA INTO INSIGHT



"Let's shrink Big Data into Small Data ... and hope it magically becomes Great Data."

• Being able to collect the right data is one thing, but making it extremely useful requires a different skill- and mindset.

• Based on surveying 49 Analytics experts here we have compiled a list of the top 10 strategies to turn data into actionable insights.

I. Measure the right things

- You can't optimize what you don't measure. There is not a one-size-fits-all solution. Every business is different and should be treated in a unique way.
- Let's assume you run an ecommerce site. In this case you probably want to know:
- Which channels drive the most conversions?
- What are your leaking buckets (places where people leave your website)?
- Whether people use multiple devices before purchasing your products?
- What are the look-to-buy ratios for your individual products and product categories?
- What landing pages need to be improved and in which channel?
- You need to think in advance about what is important for your business. It all starts with a <u>clear</u> and concise measurement plan for your business

2.Ask the right questions to stakeholders

- Go the extra mile to answer all of your stakeholderis questions. This means tapping into the stakeholderis aspirations and challenges by asking the right questions.
- It's a must to formulate a clear business question before you start your next data analysis. You can easily waste hours of your time by getting lost in your data. And coming up with "insights" that are already known or not deemed important can be highly frustrating as well!

3. Use segmentation to drive action

- Go for segmentation if you want to take action on your data! By grouping visitors that have some attributes in common, you can start digging deeper. Choosing which segments to study depends on the business question you are trying to answer.
- Identifying segments will greatly enhance your understanding of how your customers behave. You can use this information for setting up an optimization plan.



Digital analytics tools like Google Analytics come with a lot of <u>built-in segments</u> and provide you with all the freedom to customize them to your needs.

• 4. Use clear visualizations to convey your message

- The way in which you present your data will make a huge difference in the outcome. Do you remember these presentations that only include numbers and words? And this compared to clear visualizations that promote cognition instead of confusion.
- It's important to <u>articulate your specific data story</u> with as much what, how and why behind it. This will turn your data into insights and profitable business decisions.

5. Discover the context of your data set

- Everyone has got data and their own personal data driven insight (opinion). In most cases superior understanding of context leads to the best decisions.
- Make sure to establish context for the data you are seeing. What do these numbers mean? Are they important? Does it really affect the business? And how is the data collected?
- Data without context isn't that meaningful and can actually lead to bad business decisions because of interpreting it in the wrong way!

6. Build a solid optimization plan

• Use the "Define Measure Analyze Improve Control" process (<u>DMAIC</u>) to improve your business. It's one of the Six Sigma concepts you can directly apply in your situation.



- **Define** the problem or hypothesis, stakeholders and scope of analysis.
- Measure relevant data and conduct basic analysis to spot anomalies.
- Analyze correlations and patterns, put your statistics and visualization skills to work.
- Improvement based on insights and showing several options to explore.
- Control the change by deploying (A/B) tests and monitoring KPIs.

7. Construct a great hypothesis

- A clearly articulated hypothesis is the start of any "analysis". And the hypothesis should have the potential to drive action. Come up with your hypothesis by completing two fill-in-the-blank statements: I) I believe ______. (this is the hypothesis), and 2) If I am right, we will _____. (this is the qualification).
- Formulating these statements isn't always easy, but you will save a ton of time wandering through your data and coming up with interesting but not actionable findings.

8. Integrate data sources

• The integration of data sources leads to better and faster business decisions. Think about integrating traditional databases with big data solutions (like Hadoop). Mining through and connecting all your sources will enhance your customer understanding and can deliver great insights.

9. Break down organizational silos

• A healthy organization is the foundation of everything. Communication instead of confrontation. Inspire, motivate and be curious about the data and the possibilities it has for your organization. Treat any obstacles first and improve the communication between the business and analytics leaders.

• 10. Don't forget to hire smart people

• Tools can collect data, but people - who understand the business - build insights. Smart people are required to find the useful data, translate it into data-driven stories of useful knowledge - the insights. It's a team effort where combining internal business experts with external analysts might be your best bet.

HOW PREDICTIVE ANALYTICS IS TRANSFORMING ALL TYPES OF ORGANIZATIONS

- Manufacturing Machine Learning in a Predictive Environment
- Machine learning -- a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed -- is part of nearly every aspect of our day-to-day lives. Take credit card fraud alerts, for example. Machines learn your regular activity over time, so when a suspicious transaction is made with your credit card, the credit company immediately puts a hold on your account.
- Data is the theme behind this growing trend in technology. Simply put, there's more data available than ever before, so there's enough data to fuel predictive algorithms.
- Modern organizations are adjusting business practices in this predictive environment to adapt to changing demands. For manufacturers specifically, there are several areas where machine learning could positively impact their bottom line as well as the overall customer experience with predictive analytics.

• Modern organizations are adjusting business practices in this predictive environment to adapt to changing demands. For manufacturers specifically, there are several areas where machine learning could positively impact their bottom line as well as the overall customer experience with predictive analytics.

Forecasting the Predictive Analytics Future

- In the wake of these advancements, which industries are fertile ground for transformation through predictive analytics? The Internet of Things (IoT) and predictive analytics can deliver after-sales service value across innumerous industries, but they are most attractive in industries with part-intensive products that are field or depot repairable.
- As technology platforms mature and the cost of implementing IoT and predictive analytics solutions decreases over time, manufacturers will begin to accelerate their pace of turning predictive analytics into strategic advantage. Below are three different areas of opportunity for predictive analytics to take hold.

Area #1: Predictive Demand Forecast

- Efficient demand forecasting, which predicts future demand for products and parts based on past events and prevailing trends, is a key component of after-sales service success. With an accurate picture of demand, manufacturers can improve service after the initial sale of a product without having to raise costs.
- Predictive technologies such as machine learning and cloud-based inventory management solutions eliminate overstocking and enable warehouses to work with each other -- as opposed to each operating in individual silos -- to meet demand. The ultimate results are high service-parts fill rates, high levels of product uptime with minimal risk, and increased customer service levels.

Area #2: Predictive Pricing

• Many manufacturers today rely on pricing practices of the past, such as cost-plus models and Excel spreadsheets, to price service parts. Unfortunately, this leads to parts and products being sold at different prices in different locations -- creating poor customer experiences as well as a missed profit opportunity for manufacturers. When using predictive algorithms to set prices for service parts, manufacturers need to be mindful of the different factors that can affect sales, including part location, seasonality, weather, and demand. With predictive capabilities, manufacturers can incorporate all these factors and others to automatically adjust prices based on what the market will bear.

Area #3: Predictive Maintenance

- The break-fix service model is reactive and inefficient. Fifty percent of service attempts fail because the needed service part was not available when needed, which results in extended product downtime, lost revenue for the product owner, and an unhappy customer.
- With IoT and predictive analytics, smart parts and sensors will detect when a part is about to fail so the manufacturer can determine when and where parts are needed, proactively routing them to a dealer or repair center instead of storing them in stocking locations around the world. This enables manufacturers to reduce excess inventory and costs, improve part fill rates, avoid the cost and disruption of unscheduled downtime, and, ultimately, maximize customer loyalty.

Optimizing After-Sales Service with Predictive Analytics

- It's no secret that the revenue and margins generated on a manufacturer's service side of its business are attractive, especially when compared to the product side, making optimizing the service side increasingly important to manufacturers' long-term financial performance.
- Combined with the near-term financial benefits associated with service parts inventory optimization, this is driving companies across multiple industries to explore opportunities to increase the financial value from their service organizations. Predictive analytics is key to aftersales service organizations' success.

HOW ANALYTICS SUPPORTS RETAIL COMPANIES



I. Going omnichannel

- Retail started to move beyond brick-and-mortar stores long ago. However, many retailers still consider making the first step in this direction. If you are among those, you need a proper data analytics solution.
- Let's say, a consumer electronics retailer runs both physical stores and an online one. The retailer's brick-and-mortar stores are showing brilliant sales, while their online store is terribly lagging behind. Naturally, the retailer is not happy with the format that does not bring sales. However, is their e-store really useless and they are just wasting money to keep it running?
- After scrutinizing <u>customer big data</u> for both channels, the retailer may find out the following: 75% of website visitors are surfing the online catalog to compare product features and finish their purchases in one of the physical stores. In this case, if the retailer abandons an online store, they may also lose the majority of customers who would prefer shopping with another retailer in a way that they find convenient.

• 2. Creating unique customer experience

- Customers make purchases in brick-and-mortar and online stores, take part in loyalty programs, create shopping lists and place orders in the apps. In a word, they interact with a retailer in different ways and expect them to offer a personalized approach in return. As a retailer, you must understand how precious customer data is, so you should be striving to repay your customers with targeted marketing campaigns and product offers.
- Let's assume, you are a drugstore retailer. You collect and analyze customer data to understand their behavior and preferences. Your analytical system knows that Customer A used to visit your store once a month to buy 3 packs of nappies, a washing powder of brand X and a dishwashing liquid of brand Y. But this month, Customer A didn't appear. To encourage him or her to visit your store, you may then send them a 5% coupon for their favorite washing powder brand.
- As a real-life example of data analytics in action, we've chosen Nordstrom with their initiative Reserve Online & Try In Store. Their app users can select items and book them for trying on in a particular store and at a convenient time. Nordstrom has gone further: the retailer can recognize if a customer is passing by their store and kindly invite them to come in. How can anyone stay indifferent when they receive the message: Hello from Nordstrom. It looks like you are nearby, and your item is ready to try! To crown it all, once a customer is in the store, he or she will find their name on the door of a fitting room. Here is the personalized approach in action.

3. Dynamic pricing to stay competitive

- Once competitive intelligence was a challenge for brick-and-mortar retailers. Naturally, the rivals were unwilling to share any information, and price monitoring at the competitor's stores was time-consuming, prone to mistakes and exhausting. In the era of ecommerce, retailers can benefit from new approaches to competitive intelligence. By definition, online stores are publicly available, and a lot of information such as product details, promo offers, prices and category hierarchy is always at hand. This all made dynamic pricing possible, when the system is able to scan competitors' prices in real time, run a complex analysis very quickly and change a retailer's prices automatically based on the defined rules. Now, if an ecommerce retailer wants to be 5% cheaper than their competitors are, with big data analytics they are able to do this.
- Brick-and-mortar retailers can also rely on big data analytics, though some time ago they experienced some implementation limitations, as store employees had to replace price tags manually, which was time-consuming. With the invention of electronic price tags, dynamic pricing has become available to brick-and-mortar retailers, too. In no time, the system can change prices based on analyzing the sales, stock, competitors' prices, customer demand, shelf life, etc.

4. Building effective relationships with suppliers

- An industry benchmark, Walmart laid the rules on how to collaborate with the suppliers. And this policy can add \$1 billion to Walmart's revenue. The retailer has implemented a scoring system to assess the suppliers according to *On-Time*, *In-Full* principle. In simple words, either delivery is late or early, the supplier should pay a fine. The same scenario works if the supplier delivers timely, but the quantity of goods is wrong, their quality is poor, or packaging is damaged.
- If you are to follow Walmart's best practice, you can also think about tuning your data analytics system to differentiate between strategic and non-strategic suppliers, critical and non-critical product categories. Additionally, you can set different thresholds. For instance, the critical number of troublesome deliveries may be 10%.
- With such a scoring system, a retailer will easily identify whether a supplier is reliable or not. Besides, this approach will contribute to a more efficient inventory management. So, a retailer's persistent headache caused by over-stocks and out-of-stocks should finally subside.

THE DIFFERENCES BETWEEN BUSINESS INTELLIGENCE AND BUSINESS ANALYTICS

• Business Intelligence (BI) and Business Analytics are similar, though they are not exactly the same. Business Intelligence involves the process of collecting data from all sources and preparing it for Business Analytics. Business Intelligence is more of a first step for companies to take when they need the ability to make data-driven decisions. Business Analytics, on the other hand, is the analysis of the answers provided by Business Intelligence. While Business Intelligence answers what happened, Business Analytics answers why it happened and whether it will happen again. Business Intelligence includes reporting, automated monitoring and alerting, dashboards, scorecards, and ad hoc query; Business Analytics, in contrast, includes statistical and quantitative analysis, data mining, predictive modeling, and multivariate testing.

THE IMPORTANCE OF BUSINESS ANALYTICS

- Business analytics is a methodology or tool to make a sound commercial decision. Hence it impacts functioning of the whole organization. Therefore, business analytics can help improve profitability of the business, increase market share and revenue and provide better return to a shareholder.
- Facilitates better understanding of available primary and secondary data, which again affect operational efficiency of several departments.
- Provides a competitive advantage to companies. In this digital age flow of information is almost equal to all the players. It is how this information is utilized makes the company competitive. Business analytics combines available data with various well thought models to improve business decisions.
- Converts available data into valuable information. This information can be presented in any required format, comfortable to the decision maker.

HOW BIG DATA AND ANALYTICS CAN HELP IN UNDERSTANDING CONSUMER/CUSTOMER BEHAVIOR

- Customer behavior analytics is about understanding how your customers act across each channel and interaction point digital or non-digital and what influences their actions. It gives you a way to implement what I like to call, the "four rights" talk to the right audience, through the right channel, with the right message, at the right time.
- Understanding customer behavior can help your organization in more ways than you think. The entire customer lifecycle can be optimized using behavior analytics:
- Customer acquisition: Marketing will target high-value customer segments identified by behavior analytics and study behavior patterns to determine the best potential offers.
- **Customer engagement**: Behavior patterns will be used to generate personalized next-best, cross-sell and up-sell offers, while behavioral customer segmentation will be used for more general customer marketing offers.
- **Customer retention**: Behavior patterns will be used to detect possible customer churn and generate next-best retention offers.

- The value of customer behavior analytics can be measured in a number of key metrics:
- Increased customer acquisition and conversion rates
- Lower cost of acquisition
- Larger average sale on initial purchases
- Increased number of purchases per customer
- Larger order sizes on repeat purchases
- Lower cost per sale
- Increased lifetime value of customers
- Higher customer retention/Reduced churn
- Lower cost of service
- These are all metrics every organization strives to improve and can be dramatically impacted by customer behavior analytics

EXPLAIN HOW ANALYTICS CAN HELP MANAGE ASSETS

- Improved customer satisfaction Some critical goals to enhance customer satisfaction include managing asset risks, avoiding unexpected outages, and maintaining critical assets before failure strikes. Consumers expect to be notified about planned outages well in advance to plan their utility consumption. However, it is not possible to notify them when it comes to unexpected equipment failure. Adding predictive intelligence capabilities to the existing systems can help Utilities to control & avoid asset failures, outages, and penalties. Therefore, planning & prioritizing asset maintenance activities and informing customers about it will help achieve improved customer satisfaction.
- Reduced asset maintenance costs Predictive intelligence system provides an accurate prediction of failures, events & outcomes. This would help avoid significant unexpected failure costs including lost revenue, expense of field asset in service, disposal of damaged utility asset, collateral damage cost, regulatory penalty, intangible costs, etc. Thus, eliminating the overhead of unplanned asset maintenance and reducing fixed costs in millions of dollars.
- Improved safety & compliance system Integrating data from various sources such as SCADA, CIS, EAM-GIS, weather channels and online monitoring systems with Predictive asset analytics system enables utilities to address possible safety risks. They can quickly take any appropriate operational action and mitigate safety risks.

- Optimization of field force Remote Asset Management & proactive maintenance systems can help utilities to keep a watch on the health of field assets at pre-scheduled intervals. This avoids the reallocation of field-force from existing location to restore the encountered outage and a complete reorganizing of other planned asset maintenance activities.
- Challenges in Implementing Predictive Asset Analytics
 As utilities leverage predictive analytics for optimizing asset maintenance, they must deal with the following challenges:

• [An] encompassing and multidimensional field that uses mathematics, statistics, predictive modeling and machine-learning techniques to find meaningful patterns and knowledge in recorded data."

- Analytics uses the scientific method where an analyst makes hypotheses and uses the Analytics Tools to test their premises.
- Analytics is often broken down into three components:

- <u>Descriptive Analytics</u>: Covers the more standard BI activities aimed at understanding activities and opportunities from historical data. It answers the questions:
 - What Happened?
 - Why did it Happen?
- <u>Predictive Analytics</u>: Provides insight based on modeling. It answers the question: What is likely to happen? It is sometimes considered a subfield of supervised learning that is <u>rooted in</u> statistics.
- <u>Prescriptive Analytics</u>: Has its basis in scenarios and asks: What should we do to make things happen? It takes Predictive Analytics a step further to define actions that will affect outcomes, rather than just predicting the outcomes from actions that have occurred.

Descriptive Analytics Predictive Analytics Prescriptive Analytics What could happen in the future based on previous trends and patterns?" What has happened? What should a business do?

Descriptive Analytics

Business Intelligence and Data mining

"The simplest class of analytics, one that allows you to condense big data into smaller, more useful nuggets of information."- Dr. Michael Wu



90% of organizations use descriptive analytics.



Analyses the data coming in real-time and historical data for insights on how to approach the future.



Most of the social analytics are descriptive analytics.

Predictive Analytics

Forecasting

"Predictive analytics can only forecast what might happen in the future, because all predictive analytics are probabilistic in nature."- Dr. Michael Wu



Analytics is the next step of data reduction.



Predictive analytics provides answers to questions that cannot be answered by BI. Predictive analytics can be further categorized as -



What will happen next if <condition>?

Predictive Modelling



Why this actually happened? Root Cause Analysis



Identifying correlated data

Data Mining



What if the existing trends continue? Forecasting ?

What could happen?

Monte-Carlo Simulation



When should an action be invoked to correct a process. Pattern Indentification and Alerts

Prescriptive Analytics

Simulation and Optimization

Prescriptive analytics is an advanced analytics concept based on -



Optimization that helps achieve the best outcomes.



Stochastic optimization that helps understand how to achieve the best outcome and identify data uncertainties to make better decisions.



Used in producing the credit score which helps financial institutions decide the probability of a customer paying credit bills on time.



Aurora Health Care system saved \$6 million annually by using prescriptive analysis to reduce readmission rates by 10%.

BUSINESSES USE ANALYTICS TO:

- Find and act on business opportunities that may be discovered through datasets.
- Improve process.
- Grow revenue.
- Be proactive on risk management.
- Derive a <u>competitive advantage</u>.
- Develop new services.

STEPS TO APPLYING DATA ANALYTICS

- Consistently seen across available literature are five common steps to applying data analytics:
- Define your Objective
- Understand Your Data Source
- Prepare Your Data
- Analyze Data
- Report on Results

STEP 1 - DEFINE YOUR OBJECTIVE

- What are you trying to achieve?
- What could the result look like?
- Before you even start looking at the data, figure out what it is you think the data could tell you. Imagine what the results could be. For example, if you are going to look at a listing of transactions from an accounts payables system, you might come up with objectives like: look for transactions that are not supported by purchase orders, or look at the days of the week the transactions posted. The result could show that not all transactions were supported, or we may find transactions posted on the weekend when no one should be working. The ACFE and The IIA both offer courses on data analytics theory that provide examples on the types of analytics and when to use each one. Starting with proper and realistic goals can reduce wasted time as you move through the process.

STEP 2 - UNDERSTAND YOUR DATA SOURCE

- What information do I need?
- Can I get the data myself, or do I need to ask an IT resource?
- Depending on data availability, you may need to combine information from multiple sources. Many times the, type of testing you can perform will be dictated by the information that is available to you. For example, let's say you are planning to look for transactions that were posted on the weekend when no one is supposed to be working. At a minimum, the data you need for this test must include a date/time stamp on each transaction. You will probably want more than just the date/time stamp. You will also want to know who processed the transaction, what was the dollar amount, and was there a supervisor approval.
- Your next question will be about actually getting the data. If you have the required access and sufficient training on the systems involved, you may be able to get the information on your own. Realistically, most IT departments do not want us all pulling our own information. It is perfectly acceptable to ask for help. Have the IT department extract the information you need. When involving someone else in this process, make sure you give them clear instructions on the data to pull and set scoping parameters. If you ask for all the data, you may end up with 20 years' worth of transactions when you only wanted all of the transaction from this year.

STEP 3 - PREPARE YOUR DATA

- Does the data need to be cleansed?
- Does the data need to be normalized?
- Data preparation includes many different aspects, so we will focus on two of the broadest, most encompassing points: cleansing the data, and normalizing the data. Cleansing data addresses the quality of the information, while normalization eliminates redundancies.
- Cleansing the data is especially true when the information is coming from multiple sources. Sometimes you will have a column of text in a spreadsheet, but some of the cells also have numbers, or spaces in front of the letters, or symbols in the data. Cleansing the data will remove all of the unrecognizable information from the cells.
- Normalizing the data is very closely related to cleansing. Normalizing looks for different version of the same data entry. For example, you may have the last name O'Brien in your data six different ways: O'Brien, O_Brien, Obrien, O_brien, O_brien, OBrien. These are probably all the same name, just input into systems in different ways. Normalizing converts all of the variations into one format. If you don't cleanse and normalize the data, the output will either produce an error, or the results will be unusable or unreliable.

STEP 4 - ANALYZE DATA

- What tests can I run on the data?
- Is help available to understand results?
- At this point, you will have come up with the objective, pulled the data, and spent some time cleansing and normalizing the information, and now it's time to run the test. Your data analytics tools will help you summarize the information. Again, look to the professional organizations like The IIA and the ACFE for training if you are not sure which test to run.
- Once you run the tests, you may or may not understand the results. Your best resource to understand the results will likely be the people you are auditing. If it is appropriate, you should take the results back to the data owners for help understanding the output. Keep in mind, this may not always be appropriate, especially if this is for a fraud examination.

STEP 5 - REPORT RESULTS

- Ask the following questions:
- Will management understand the results?
- Can you represent the results visually?
- Avoid presenting management with tables full of numbers. We need to effectively communicate results without lengthy explanations. Use charts and graphs with simple notes.
- One of the most important factors in helping management focus and understand the results is the amount of information we present. We must be careful not to overwhelm management with endless information. Always present summary information and provide any details as an appendix to the summary. Remember that these are busy people with a very limited amount of time to dedicate to your data, so be succinct and provide your reports well in advance of any meetings.
- In line with our basic audit process, we should provide more risk information to management. You should feel free to have open, risk-based discussions with them as it relates to the results.
- Instead of getting mired in the details, where applicable, provide more trending information related to audit results. Showing trends is more illustrative of the organization's overall status. Examples could be trending by types of findings, results by business unit, or trends in the status of different data classifications. However, if we are dealing with a fraud examination, the exact opposite may be preferable. When we must present the details, be sure to make a clear, unbiased presentation of the test results.

COGNOS

IBM

WHAT IS DATA WAREHOUSING?

- A data warehousing is a technique for collecting and managing data from varied sources to provide meaningful business insights. It is a blend of technologies and components which allows the strategic use of data.
- It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

- The decision support database (Data Warehouse) is maintained separately from the organization's operational database. However, the data warehouse is not a product but an environment. It is an architectural construct of an information system which provides users with current and historical decision support information which is difficult to access or present in the traditional operational data store.
- The data warehouse is the core of the BI system which is built for data analysis and reporting.

- Data warehouse system is also known by the following name:
- Decision Support System (DSS)
- Executive Information System
- Management Information System
- Business Intelligence Solution
- Analytic Application
- Data Warehouse



HOW DATA WAREHOUSE WORKS?

- Data Warehouse works as a central repository where information arrives from one or more data sources. Data flows into a data warehouse from the transactional system and other relational databases.
- Data may be:
- Structured
- Semi-structured
- Unstructured data

- The data is processed, transformed, and ingested so that users can access the processed data in the Data Warehouse through Business Intelligence tools, SQL clients, and spreadsheets. A data warehouse merges information coming from different sources into one comprehensive database.
- By merging all of this information in one place, an organization can analyze its customers more holistically. This helps to ensure that it has considered all the information available. Data warehousing makes data mining possible. Data mining is looking for patterns in the data that may lead to higher sales and profits.

TYPES OF DATA WAREHOUSE

• Three main types of Data Warehouses are:

1. Enterprise Data Warehouse:

Enterprise Data Warehouse is a centralized warehouse. It provides decision support service across the enterprise. It offers a unified approach for organizing and representing data. It also provide the ability to classify data according to the subject and give access according to those divisions.

2. Operational Data Store:

Operational Data Store, which is also called ODS, are nothing but data store required when neither Data warehouse nor OLTP systems support organizations reporting needs. In ODS, Data warehouse is refreshed in real time. Hence, it is widely preferred for routine activities like storing records of the Employees.

3. Data Mart:

 A data mart is a subset of the data warehouse. It specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.

•

COMPONENTS OF DATA WAREHOUSE

- Four components of Data Warehouses are:
- Load manager: Load manager is also called the front component. It performs with all the operations associated with the extraction and load of data into the warehouse. These operations include transformations to prepare the data for entering into the Data warehouse.
- Warehouse Manager: Warehouse manager performs operations associated with the management of the data in the warehouse. It performs operations like analysis of data to ensure consistency, creation of indexes and views, generation of denormalization and aggregations, transformation and merging of source data and archiving and baking-up data.
- **Query Manager:** Query manager is also known as backend component. It performs all the operation operations related to the management of user queries. The operations of this Data warehouse components are direct queries to the appropriate tables for scheduling the execution of queries.
- End-user access tools:
- This is categorized into five different groups like 1. Data Reporting 2. Query Tools 3. Application development tools 4. EIS tools, 5. OLAP tools and data mining tools.

WHO NEEDS DATA WAREHOUSE?

- Data warehouse is needed for all types of users like:
- Decision makers who rely on mass amount of data
- Users who use customized, complex processes to obtain information from multiple data sources.
- It is also used by the people who want simple technology to access the data
- It also essential for those people who want a systematic approach for making decisions.
- If the user wants fast performance on a huge amount of data which is a necessity for reports, grids or charts, then Data warehouse proves useful.
- Data warehouse is a first step If you want to discover 'hidden patterns' of data-flows and groupings.

WHAT ARE THE MOST COMMON SECTORS WHERE DATA WAREHOUSE IS USED ?

Airline:

• In the Airline system, it is used for operation purpose like crew assignment, analyses of route profitability, frequent flyer program promotions, etc.

Banking:

• It is widely used in the banking sector to manage the resources available on desk effectively. Few banks also used for the market research, performance analysis of the product and operations.

Healthcare:

 Healthcare sector also used Data warehouse to strategize and predict outcomes, generate patient's treatment reports, share data with tie-in insurance companies, medical aid services, etc.

Public sector:

 In the public sector, data warehouse is used for intelligence gathering. It helps government agencies to maintain and analyze tax records, health policy records, for every individual.

Investment and Insurance sector:

 In this sector, the warehouses are primarily used to analyze data patterns, customer trends, and to track market movements.

Telecommunication:

 A data warehouse is used in this sector for product promotions, sales decisions and to make distribution decisions.

Hospitality Industry:

 This Industry utilizes warehouse services to design as well as estimate their advertising and promotion campaigns where they want to target clients based on their feedback and travel patterns.

ADVANTAGES & DISADVANTAGES OF DATAWAREHOUSE

- Advantages of Data Warehouse:
- Data warehouse allows business users to quickly access critical data from some sources all in one place.
- Data warehouse provides consistent information on various cross-functional activities. It is also supporting ad-hoc reporting and query.
- Data Warehouse helps to integrate many sources of data to reduce stress on the production system.
- Data warehouse helps to reduce total turnaround time for analysis and reporting.

- Restructuring and Integration make it easier for the user to use for reporting and analysis.
- Data warehouse allows users to access critical data from the number of sources in a single place. Therefore, it saves user's time of retrieving data from multiple sources.
- Data warehouse stores a large amount of historical data. This helps users to analyse different time periods and trends to make future predictions.

DISADVANTAGES OF DATA WAREHOUSE:

- Not an ideal option for unstructured data.
- Creation and Implementation of Data Warehouse is surely time confusing affair.
- Data Warehouse can be outdated relatively quickly
- Difficult to make changes in data types and ranges, data source schema, indexes, and queries.
- The data warehouse may seem easy, but actually, it is too complex for the average users.
- Despite best efforts at project management, data warehousing project scope will always increase.
- Sometime warehouse users will develop different business rules.
- Organisations need to spend lots of their resources for training and Implementation purpose.

DATA WAREHOUSE TOOLS

1. MarkLogic:

- MarkLogic is useful data warehousing solution that makes data integration easier and faster using an array of enterprise features. This tool helps to perform very complex search operations. It can query different types of data like documents, relationships, and metadata.
- http://developer.marklogic.com/products

2. Oracle:

- Oracle is the industry-leading database. It offers a wide range of choice of data warehouse solutions for both on-premises and in the cloud. It helps to optimize customer experiences by increasing operational efficiency.
- https://www.oracle.com/index.html

3. Amazon RedShift:

- Amazon Redshift is Data warehouse tool. It is a simple and cost-effective tool to analyze all types of data using standard SQL and existing BI tools. It also allows running complex queries against petabytes of structured data, using the technique of query optimization.
- https://aws.amazon.com/redshift/?nc2=h_m1

CHARACTERISTICS OF DATA WAREHOUSE

- A data warehouse has following characteristics:
- Subject-Oriented
- Integrated
- Time-variant
- Non-volatile

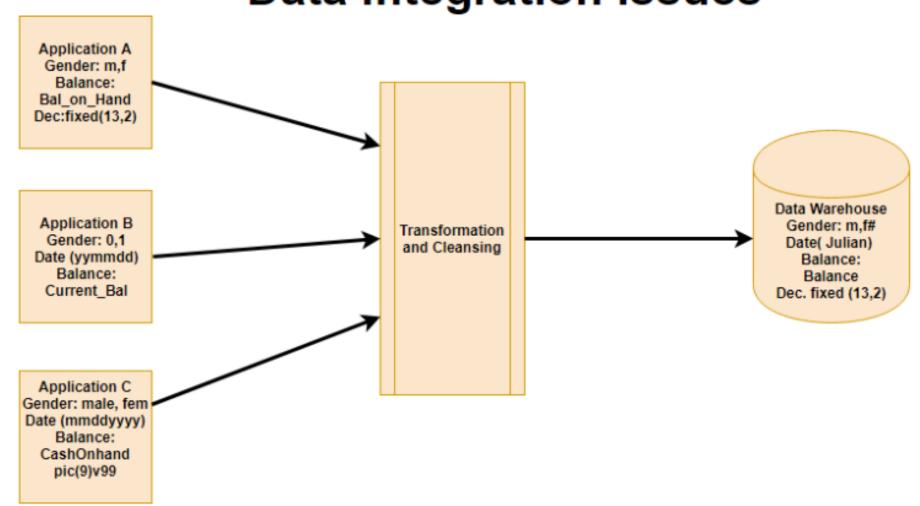
Subject-Oriented

- A data warehouse is subject oriented as it offers information regarding a theme instead of companies' ongoing operations. These subjects can be sales, marketing, distributions, etc.
- A data warehouse never focuses on the ongoing operations. Instead, it put emphasis on modeling and analysis of data for **decision making**. It also provides a simple and concise view around the specific subject by excluding data which not helpful to support the decision process.

Integrated

- In Data Warehouse, integration means the establishment of a common unit of measure for all similar data from the dissimilar database. The data also needs to be stored in the Datawarehouse in common and universally acceptable manner.
- A data warehouse is developed by integrating data from varied sources like a mainframe, relational databases, flat files, etc. Moreover, it must keep consistent naming conventions, format, and coding.
- This integration helps in effective analysis of data. Consistency in naming conventions, attribute measures, encoding structure etc. have to be ensured. Consider the following example:

Data Integration Issues



- In the above example, there are three different application labeled A, B and C. Information stored in these applications are Gender, Date, and Balance. However, each application's data is stored different way.
- In Application A gender field store logical values like M or F
- In Application B gender field is a numerical value,
- In Application C application, gender field stored in the form of a character value.
- Same is the case with Date and balance
- However, after transformation and cleaning process all this data is stored in common format in the Data Warehouse.

Time-Variant

- The time horizon for data warehouse is quite extensive compared with operational systems. The data collected in a data warehouse is recognized with a particular period and offers information from the historical point of view. It contains an element of time, explicitly or implicitly.
- One such place where Datawarehouse data display time variance is in in the structure of the record key. Every primary key contained with the DW should have either implicitly or explicitly an element of time. Like the day, week month, etc.
- Another aspect of time variance is that once data is inserted in the warehouse, it can't be updated or changed.

Non-volatile

- Data warehouse is also non-volatile means the previous data is not erased when new data is entered in it.
- Data is read-only and periodically refreshed. This also helps to analyze
 historical data and understand what & when happened. It does not require
 transaction process, recovery and concurrency control mechanisms.
- Activities like delete, update, and insert which are performed in an operational application environment are omitted in Data warehouse environment. Only two types of data operations performed in the Data Warehousing are
- Data loading
- Data access

DATA WAREHOUSE ARCHITECTURES

There are mainly three types of Datawarehouse Architectures: -

Single-tier architecture

• The objective of a single layer is to minimize the amount of data stored. This goal is to remove data redundancy. This architecture is not frequently used in practice.

Two-tier architecture

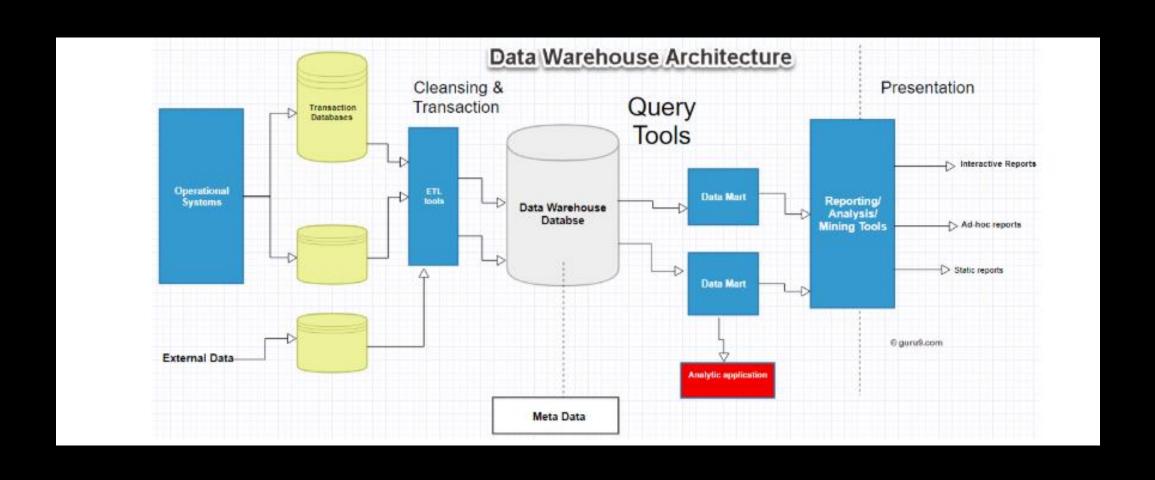
• Two-layer architecture separates physically available sources and data warehouse. This architecture is not expandable and also not supporting a large number of endusers. It also has connectivity problems because of network limitations.

Three-tier architecture

This is the most widely used architecture.

- It consists of the Top, Middle and Bottom Tier.
- Bottom Tier: The database of the Data warehouse servers as the bottom tier. It is
 usually a relational database system. Data is cleansed, transformed, and loaded
 into this layer using back-end tools.
- **Middle Tier:** The middle tier in Data warehouse is an OLAP server which is implemented using either ROLAP or MOLAP model. For a user, this application tier presents an abstracted view of the database. This layer also acts as a mediator between the end-user and the database.
- **Top-Tier:** The top tier is a front-end client layer. Top tier is the tools and API that you connect and get data out from the data warehouse. It could be Query tools, reporting tools, managed query tools, Analysis tools and Data mining tools.

DATAWAREHOUSE COMPONENTS



The data warehouse is based on an RDBMS server which is a central
information repository that is surrounded by some key components to make
the entire environment functional, manageable and accessible.

There are mainly five components of Data Warehouse:

Data Warehouse Database

• The central database is the foundation of the data warehousing environment. This database is implemented on the RDBMS technology. Although, this kind of implementation is constrained by the fact that traditional RDBMS system is optimized for transactional database processing and not for data warehousing. For instance, ad-hoc query, multi-table joins, aggregates are resource intensive and slow down performance.

Sourcing, Acquisition, Clean-up and Transformation Tools (ETL)

 The data sourcing, transformation, and migration tools are used for performing all the conversions, summarizations, and all the changes needed to transform data into a unified format in the data warehouse. They are also called Extract, Transform and Load (ETL) Tools.

Metadata

- The name Meta Data suggests some high-level technological concept. However, it
 is quite simple. Metadata is data about data which defines the data warehouse. It is
 used for building, maintaining and managing the data warehouse.
- In the Data Warehouse Architecture, meta-data plays an important role as it specifies the source, usage, values, and features of data warehouse data. It also defines how data can be changed and processed. It is closely connected to the data warehouse.

For example, a line in sales database may contain:

4030 KJ732 299.90

- This is a meaningless data until we consult the Meta that tell us it was
- Model number: 4030
- Sales Agent ID: KJ732
- Total sales amount of \$2999.90
- Therefore, Meta Data are essential ingredients in the transformation of data into knowledge.
- Metadata helps to answer the following questions
- What tables, attributes, and keys does the Data Warehouse contain?
- Where did the data come from?
- How many times do data get reloaded?
- What transformations were applied with cleansing?

Query Tools

- One of the primary objects of data warehousing is to provide information to businesses to make strategic decisions. Query tools allow users to interact with the data warehouse system.
- These tools fall into four different categories:
- Query and reporting tools
- Application Development tools
- Data mining tools
- OLAP tools

- 1. Query and reporting tools:
- Query and reporting tools can be further divided into
- Reporting tools
- Managed query tools

- **Reporting tools:** Reporting tools can be further divided into production reporting tools and desktop report writer.
- Report writers: This kind of reporting tool are tools designed for end-users for their analysis.
- Production reporting: This kind of tools allows organizations to generate regular operational reports. It also supports high volume batch jobs like printing and calculating. Some popular reporting tools are Brio, Business Objects, Oracle, PowerSoft, SAS Institute.

Managed query tools:

 This kind of access tools helps end users to resolve snags in database and SQL and database structure by inserting meta-layer between users and database.

2. Application development tools:

 Sometimes built-in graphical and analytical tools do not satisfy the analytical needs of an organization. In such cases, custom reports are developed using Application development tools.

3. Data mining tools:

 Data mining is a process of discovering meaningful new correlation, pattens, and trends by mining large amount data. Data mining tools are used to make this process automatic.

4. OLAP tools:

 These tools are based on concepts of a multidimensional database. It allows users to analyse the data using elaborate and complex multidimensional views.

DATA WAREHOUSE BUS ARCHITECTURE

- Data warehouse Bus determines the flow of data in your warehouse. The data flow in a data warehouse can be categorized as Inflow, Up flow, Down flow, Outflow and Meta flow.
- While designing a Data Bus, one needs to consider the shared dimensions, facts across data marts.

Data Marts

- A data mart is an access layer which is used to get data out to the users. It is
 presented as an option for large size data warehouse as it takes less time and
 money to build. However, there is no standard definition of a data mart is differing
 from person to person.
- In a simple word Data mart is a subsidiary of a data warehouse. The data mart is used for partition of data which is created for the specific group of users.
- Data marts could be created in the same database as the Datawarehouse or a physically separate Database.

ETL (EXTRACT, TRANSFORM, AND LOAD) PROCESS

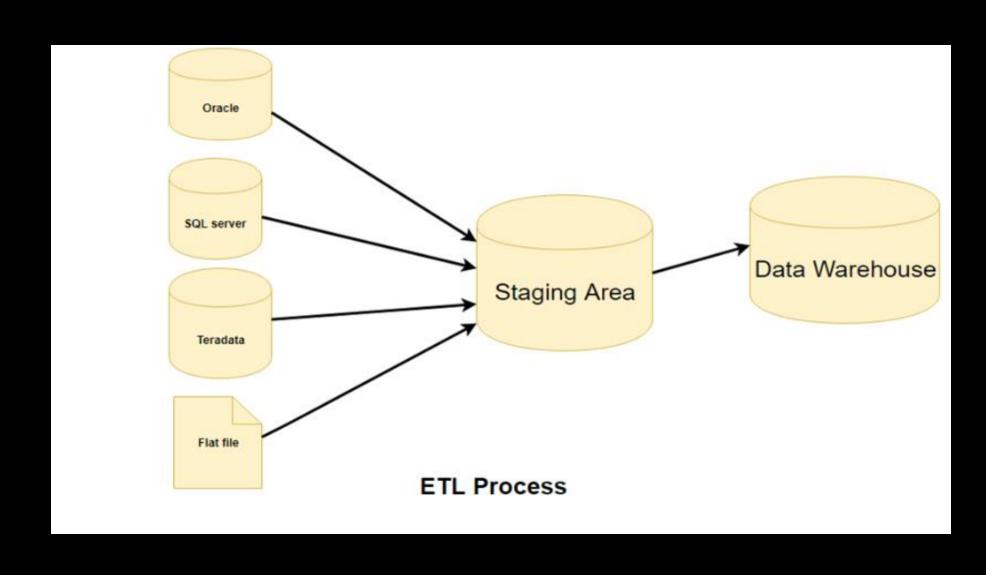
WHAT IS ETL?

- ETL is an abbreviation of Extract, Transform and Load. In this process, an ETL tool extracts the data from different RDBMS source systems then transforms the data like applying calculations, concatenations, etc. and then load the data into the Data Warehouse system.
- It's tempting to think a creating a Data warehouse is simply extracting data from multiple sources and loading into database of a Data warehouse. This is far from the truth and requires a complex ETL process. The ETL process requires active inputs from various stakeholders including developers, analysts, testers, top executives and is technically challenging.
- In order to maintain its value as a tool for decision-makers, Data warehouse system needs to change with business changes. ETL is a recurring activity (daily, weekly, monthly) of a Data warehouse system and needs to be agile, automated, and well documented.

WHY DO YOU NEED ETL?

- It helps companies to analyze their business data for taking critical business decisions.
- Transactional databases cannot answer complex business questions that can be answered by ETL.
- A Data Warehouse provides a common data repository
- ETL provides a method of moving the data from various sources into a data warehouse.
- As data sources change, the Data Warehouse will automatically update.
- Allow verification of data transformation, aggregation and calculations rules.
- ETL process allows sample data comparison between the source and the target system.
- ETL process can perform complex transformations and requires the extra area to store the data.
- ETL helps to Migrate data into a Data Warehouse. Convert to the various formats and types to adhere to one consistent system.
- ETL is a predefined process for accessing and manipulating source data into the target database.
- It helps to improve productivity because it codifies and reuses without a need for technical skills.

ETL PROCESS IN DATA WAREHOUSES



STEP 1) EXTRACTION

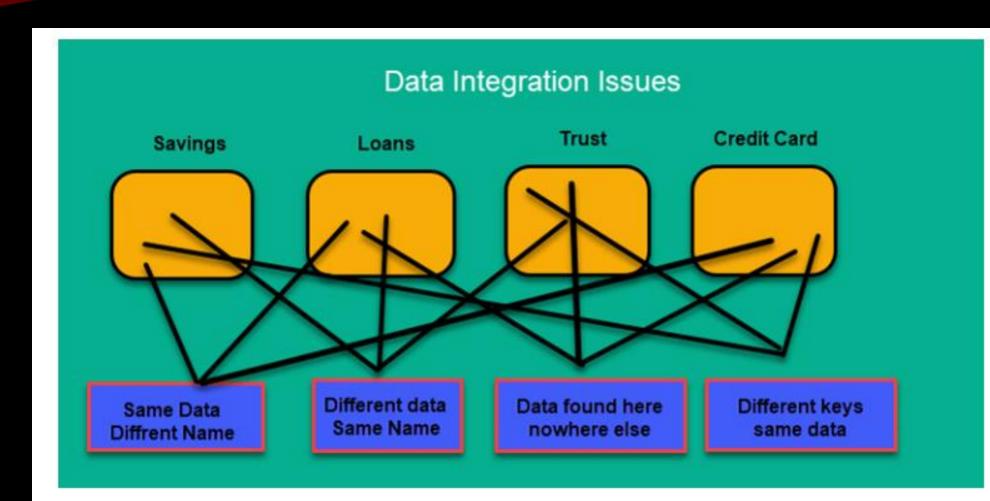
- In this step, data is extracted from the source system into the staging area.
 Transformations if any are done in staging area so that performance of
 source system in not degraded. Also, if corrupted data is copied directly
 from the source into Data warehouse database, rollback will be a
 challenge. Staging area gives an opportunity to validate extracted data
 before it moves into the Data warehouse.
- Data warehouse needs to integrate systems that have different
- DBMS, Hardware, Operating Systems and Communication Protocols. Sources could include legacy applications like Mainframes, customized applications, Point of contact devices like ATM, Call switches, text files, spreadsheets, ERP, data from vendors, partners amongst others.
- Hence one needs a logical data map before data is extracted and loaded physically. This data map describes the relationship between sources and target data.

Three Data Extraction methods:

- Full Extraction
- Partial Extraction- without update notification.
- Partial Extraction- with update notification
- Irrespective of the method used, extraction should not affect performance and response time of the source systems. These source systems are live production databases. Any slow down or locking could effect company's bottom line.
- Some validations are done during Extraction:
- Reconcile records with the source data
- Make sure that no spam/unwanted data loaded
- Data type check
- Remove all types of duplicate/fragmented data
- Check whether all the keys are in place or not

STEP 2) TRANSFORMATION

- Data extracted from source server is raw and not usable in its original form. Therefore it needs to be cleansed, mapped and transformed. In fact, this is the key step where ETL process adds value and changes data such that insightful BI reports can be generated.
- In this step, you apply a set of functions on extracted data. Data that does not require any transformation is called as **direct move** or **pass through data**.
- In transformation step, you can perform customized operations on data. For instance, if the user wants sum-of-sales revenue which is not in the database. Or if the first name and the last name in a table is in different columns. It is possible to concatenate them before loading.



Following are Data Integrity Problems:

- Different spelling of the same person like Jon, John, etc.
- There are multiple ways to denote company name like Google, Google Inc.
- Use of different names like Cleaveland, Cleveland.
- There may be a case that different account numbers are generated by various applications for the same customer.
- In some data required files remains blank
- Invalid product collected at POS as manual entry can lead to mistakes.

Validations are done during this stage

- Filtering Select only certain columns to load
- Using rules and lookup tables for Data standardization
- Character Set Conversion and encoding handling
- Conversion of Units of Measurements like Date Time Conversion, currency conversions, numerical conversions, etc.
- Data threshold validation check. For example, age cannot be more than two digits.
- Data flow validation from the staging area to the intermediate tables.
- Required fields should not be left blank.
- Cleaning (for example, mapping NULL to 0 or Gender Male to "M" and Female to "F" etc.)
- Split a column into multiples and merging multiple columns into a single column.
- Transposing rows and columns,
- Use lookups to merge data
- Using any complex data validation (e.g., if the first two columns in a row are empty then it automatically reject the row from processing)

STEP 3) LOADING

- Loading data into the target datawarehouse database is the last step of the ETL process. In a typical Data warehouse, huge volume of data needs to be loaded in a relatively short period (nights). Hence, load process should be optimized for performance.
- In case of load failure, recover mechanisms should be configured to restart from the point of failure without data integrity loss. Data Warehouse admins need to monitor, resume, cancel loads as per prevailing server performance.

Types of Loading:

- Initial Load populating all the Data Warehouse tables
- Incremental Load applying ongoing changes as when needed periodically.
- Full Refresh —erasing the contents of one or more tables and reloading with fresh data.

Load verification

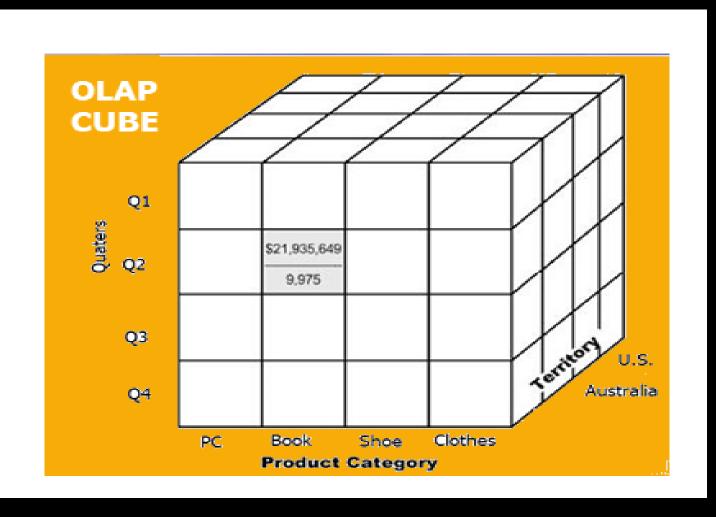
- Ensure that the key field data is neither missing nor null.
- Test modeling views based on the target tables.
- Check that combined values and calculated measures.
- Data checks in dimension table as well as history table.
- Check the BI reports on the loaded fact and dimension table.

OLAP (ONLINE ANALYTICAL PROCESSING): CUBE, OPERATIONS & TYPES

WHAT IS ONLINE ANALYTICAL PROCESSING?

- OLAP is a category of software that allows users to analyze information from multiple database systems at the same time. It is a technology that enables analysts to extract and view business data from different points of view.
 OLAP stands for Online Analytical Processing.
- Analysts frequently need to group, aggregate and join data. These
 operations in relational databases are resource intensive. With OLAP data
 can be pre-calculated and pre-aggregated, making analysis faster.
- OLAP databases are divided into one or more cubes. The cubes are designed in such a way that creating and viewing reports become easy.

OLAP CUBE



- At the core of the OLAP, concept is an OLAP Cube. The OLAP cube is a data structure optimized for very quick data analysis.
- The OLAP Cube consists of numeric facts called measures which are categorized by dimensions. OLAP Cube is also called the **hypercube**.
- Usually, data operations and analysis are performed using the simple spreadsheet, where data values are arranged in row and column format. This is ideal for two-dimensional data. However, OLAP contains multidimensional data, with data usually obtained from a different and unrelated source. Using a spreadsheet is not an optimal option. The cube can store and analyze multidimensional data in a logical and orderly manner.

HOW DOES IT WORK?

- A Data warehouse would extract information from multiple data sources and formats like text files, excel sheet, multimedia files, etc.
- The extracted data is cleaned and transformed. Data is loaded into an OLAP server (or OLAP cube) where information is pre-calculated in advance for further analysis.

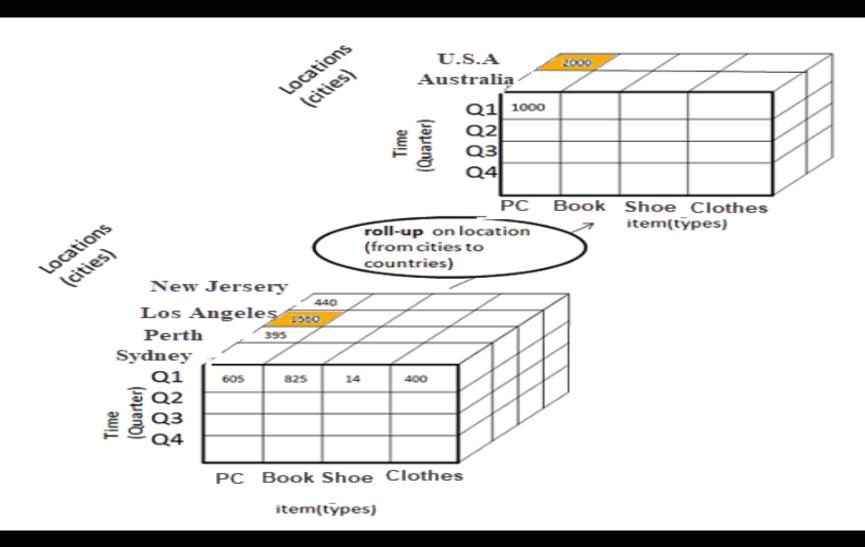
BASIC ANALYTICAL OPERATIONS OF OLAP

- Four types of analytical operations in OLAP are:
- Roll-up
- Drill-down
- Slice and dice
- Pivot (rotate)

1) ROLL-UP

- Roll-up is also known as "consolidation" or "aggregation." The Roll-up operation can be performed in 2 ways
- Reducing dimensions
- Climbing up concept hierarchy. Concept hierarchy is a system of grouping things based on their order or level.

CONSIDER THE FOLLOWING DIAGRAM23

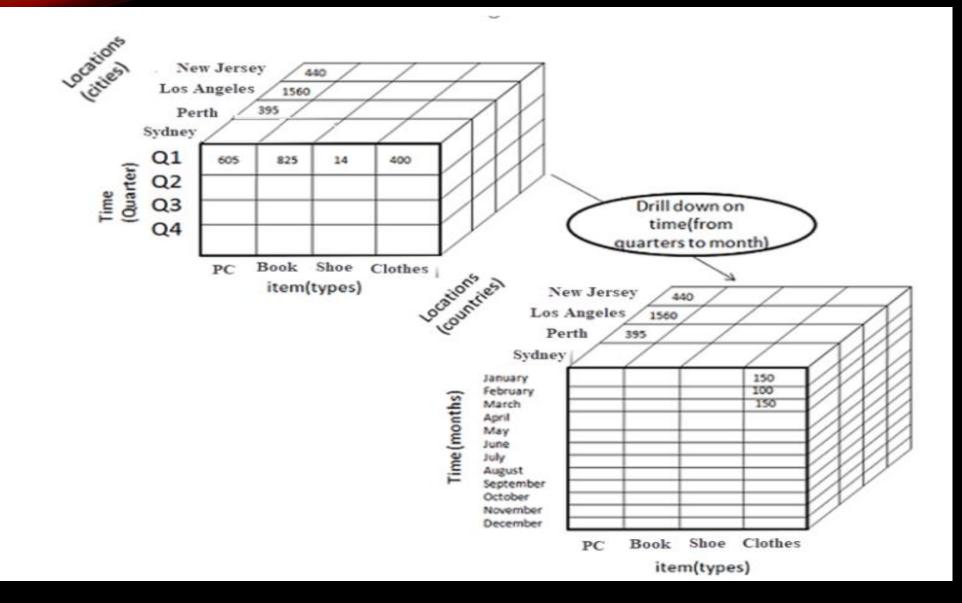


CONTINUE...

- In this example, cities New jersey and Lost Angles and rolled up into country USA
- The sales figure of New Jersey and Los Angeles are 440 and 1560 respectively. They become 2000 after roll-up
- In this aggregation process, data is location hierarchy moves up from city to the country.
- In the roll-up process at least one or more dimensions need to be removed.
 In this example, Quater dimension is removed.

2) DRILL-DOWN

- In drill-down data is fragmented into smaller parts. It is the opposite of the rollup process. It can be done via
- Moving down the concept hierarchy
- Increasing a dimension



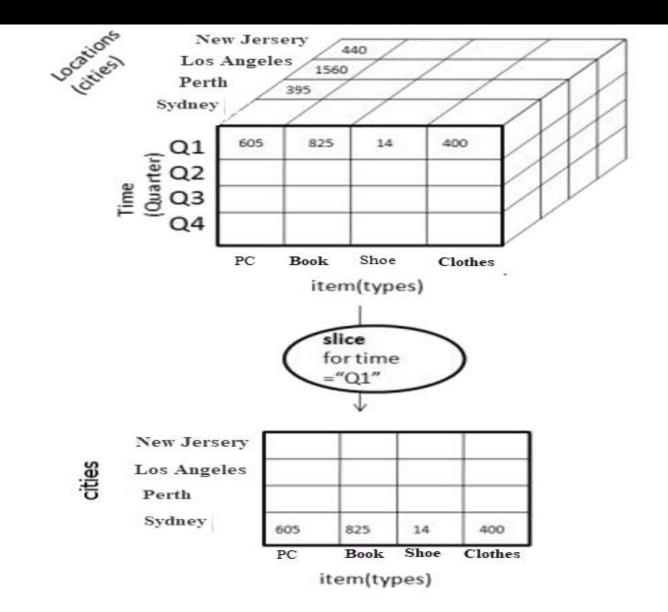
CONTINUE..

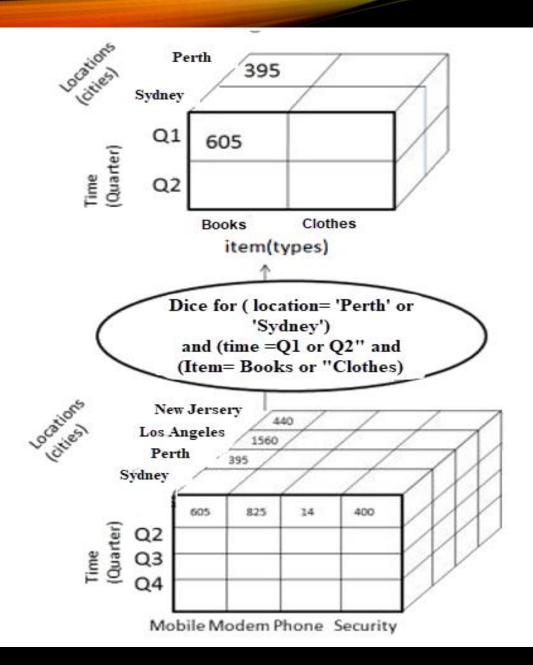
- Quater Q1 is drilled down to months January, February, and March.
 Corresponding sales are also registers.
- In this example, dimension months are added.

3) SLICE:

- Here, one dimension is selected, and a new sub-cube is created.
- Following diagram explain how slice operation performed:

- •Dimension Time is Sliced with Q1 as the filter.
- •A new cube is created altogether.





DICE:

 This operation is similar to a slice. The difference in dice is you select 2 or more dimensions that result in the creation of a sub-cube

New Jersey (cities) Los Angeles Perth Sydney 605 825 400 14 Book Shoe Clothes PC item(types) Pivot 605 PC 825 Book (types) 14 Shoe Clothes 400 New Perth Sydney Angeles Jersey Location (Cities)

Locations

4) PIVOT

• In Pivot, you rotate the data axes to provide a substitute presentation of data.

 In the following example, the pivot is based on item types.

Benefits of using OLAP services

- OLAP creates a single platform for all type of business analytical needs which includes planning, budgeting, forecasting, and analysis.
- The main benefit of OLAP is the consistency of information and calculations.
- Easily apply security restrictions on users and objects to comply with regulations and protect sensitive data.

Drawbacks of OLAP service

- Implementation and maintenance are dependent on IT professional because the traditional OLAP tools require a complicated modeling procedure.
- OLAP tools need cooperation between people of various departments to be effective which might always be not possible.

WHAT IS OLTP?

• Online transaction processing shortly known as OLTP supports transaction-oriented applications in a 3-tier architecture. OLTP administers day to day transaction of an organization.

The primary objective is data processing and not data analysis

EXAMPLE OF OLTP SYSTEM

- An example of OLTP system is ATM center. Assume that a couple has a joint account with a bank. One day both simultaneously reach different ATM centers at precisely the same time and want to withdraw total amount present in their bank account.
- However, the person that completes authentication process first will be able to get money. In this case, OLTP system makes sure that withdrawn amount will be never more than the amount present in the bank. The key to note here is that OLTP systems are optimized for transactional superiority instead data analysis.

Other examples of OLTP system are:

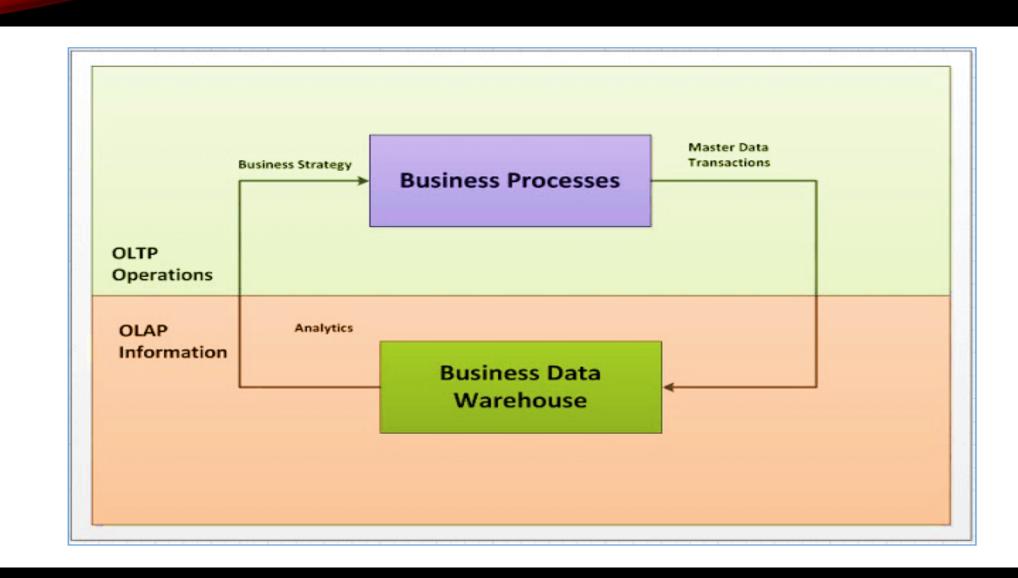
- Online banking
- Online airline ticket booking
- Sending a text message
- Order entry
- Add a book to shopping cart

Benefits of OLTP method

- It administers daily transactions of an organization.
- OLTP widens the customer base of an organization by simplifying individual processes

Drawbacks of OLTP method

- If OLTP system faces hardware failures, then online transactions get severely affected.
- OLTP systems allow multiple users to access and change the same data at the same time which many times created unprecedented situation.



DIFFERENCE BETWEEN OLTP AND OLAP

| Parameters | OLTP | OLAP |
|----------------|---|---|
| Process | It is an online transactional system. It manages database modification. | OLAP is an online analysis and data retrieving process. |
| Characteristic | It is characterized by large numbers of short online transactions. | It is characterized by a large volume of data. |
| Functionality | OLTP is an online database modifying system. | OLAP is an online database query management system. |
| Method | OLTP uses traditional DBMS. | OLAP uses the data warehouse. |

| Query | Insert, Update, and Delete information from the database. | Mostly select operations |
|----------------|---|--|
| Table | Tables in OLTP database are normalized. | Tables in OLAP database are not normalized. |
| Source | OLTP and its transactions are the sources of data. | Different OLTP databases become the source of data for OLAP. |
| Data Integrity | OLTP database must maintain data integrity constraint. | OLAP database does not get frequently modified. Hence, data integrity is not an issue. |

| Response time | It's response time is in millisecond. | Response time in seconds to minutes. |
|------------------|---|--|
| Data quality | The data in the OLTP database is always detailed and organized. | The data in OLAP process might not be organized. |
| Usefulness | It helps to control and run fundamental business tasks. | It helps with planning, problem-solving, and decision support. |
| Operation | Allow read/write operations. | Only read and rarely write. |
| Audience | It is a market orientated process. | It is a customer orientated process. |

| Query Type | Queries in this process are standardized and simple. | Complex queries involving aggregations. |
|------------|--|---|
| Back-up | Complete backup of the data combined with incremental backups. | OLAP only need a backup from time to time. Backup is not important compared to OLTP |
| Design | DB design is application oriented. Example: Database design changes with industry like Retail, Airline, Banking, etc. | DB design is subject oriented. Example: Database design changes with subjects like sales, marketing, purchasing, etc. |
| User type | It is used by Data critical users like clerk, DBA & Data Base professionals. | Used by Data knowledge users like workers, managers, and CEO. |
| Purpose | Designed for real time business operations. | Designed for analysis of business measures by category and attributes. |

| Performance metric | Transaction throughput is the performance metric | Query throughput is the performance metric. |
|-----------------------|---|--|
| Number of users | This kind of Database users allows thousands of users. | This kind of Database allows only hundreds of users. |
| Productivity | It helps to Increase user's self- service and productivity | Help to Increase productivity of the business analysts. |
| Challenge | Data Warehouses historically have been a development project which may prove costly to build. | An OLAP cube is not an open SQL server data warehouse. Therefore, technical knowledge and experience is essential to manage the OLAP server. |

| Process | It provides fast result for daily used data. | It ensures that response to the query is quicker consistently. |
|----------------|---|---|
| Characteristic | It is easy to create and maintain. | It lets the user create a view with the help of a spreadsheet. |
| Style | OLTP is designed to have fast response time, low data redundancy and is normalized. | A data warehouse is created uniquely so that it can integrate different data sources for building a consolidated database |

WHAT IS DIMENSIONAL MODEL IN DATA WAREHOUSE?

WHAT IS DIMENSIONAL MODEL?

- A dimensional model is a data structure technique optimized for Data warehousing tools. The concept of Dimensional Modelling was developed by Ralph Kimball and is comprised of "fact" and "dimension" tables.
- A Dimensional model is designed to read, summarize, analyze numeric information like values, balances, counts, weights, etc. in a data warehouse. In contrast, relation models are optimized for addition, updating and deletion of data in a real-time Online Transaction System.

- These dimensional and relational models have their unique way of data storage that has specific advantages.
- For instance, in the relational mode, normalization and ER models reduce redundancy in data. On the contrary, dimensional model arranges data in such a way that it is easier to retrieve information and generate reports.
- Hence, Dimensional models are used in data warehouse systems and not a good fit for relational systems.

ELEMENTS OF DIMENSIONAL DATA MODEL

<u>Fact</u>

• Facts are the measurements/metrics or facts from your business process. For a Sales business process, a measurement would be quarterly sales number

Dimension

- Dimension provides the context surrounding a business process event. In simple terms, they give who, what, where of a fact. In the Sales business process, for the fact quarterly sales number, dimensions would be
- Who Customer Names
- Where Location
- What Product Name

In other words, a dimension is a window to view information in the facts

Attributes

- The Attributes are the various characteristics of the dimension.
- In the Location dimension, the attributes can be
- State
- Country
- Zipcode etc.

Attributes are used to search, filter, or classify facts. Dimension Tables contain Attributes

Fact Table

- A fact table is a primary table in a dimensional model.
- A Fact Table contains
- Measurements/facts
- Foreign key to dimension table

Dimension table

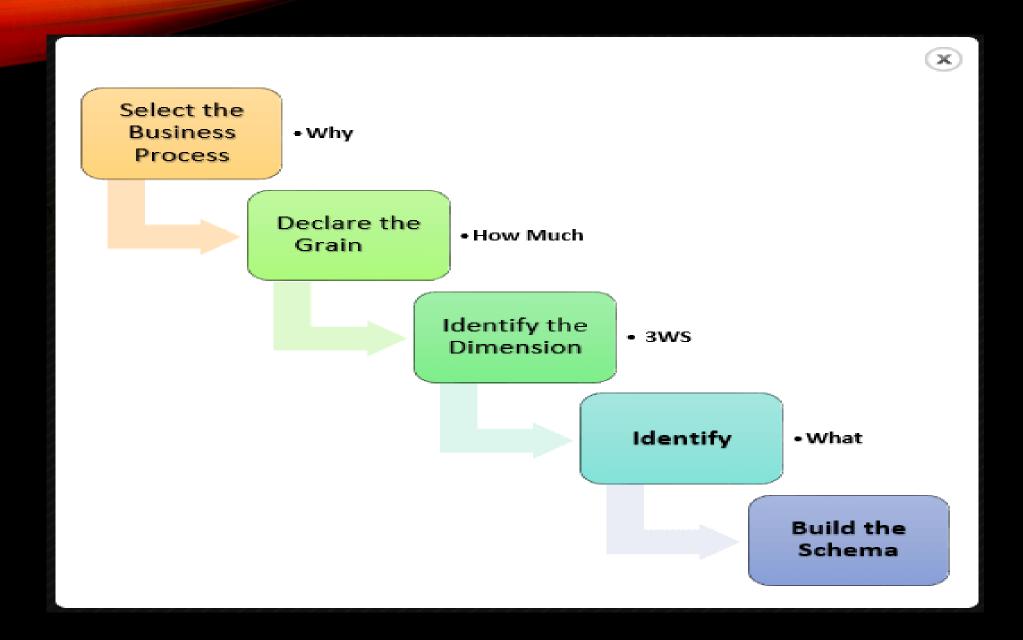
- A dimension table contains dimensions of a fact.
- They are joined to fact table via a foreign key.
- Dimension tables are de-normalized tables.
- The Dimension Attributes are the various columns in a dimension table
- Dimensions offers descriptive characteristics of the facts with the help of their attributes
- No set limit set for given for number of dimensions
- The dimension can also contain one or more hierarchical relationships

STEPS OF DIMENSIONAL MODELLING

The accuracy in creating your Dimensional modeling determines the success of your data warehouse implementation. Here are the steps to create Dimension Model

- Identify Business Process
- Identify Grain (level of detail)
- Identify Dimensions
- Identify Facts
- Build Star

The model should describe the Why, How much, When/Where/Who and What of your business process



Step 1) Identify the business process

 Identifying the actual business process a datarehouse should cover. This could be Marketing, Sales, HR, etc. as per the data analysis needs of the organization. The selection of the Business process also depends on the quality of data available for that process. It is the most important step of the Data Modelling process, and a failure here would have cascading and irreparable defects.

Step 2) Identify the grain

- The Grain describes the level of detail for the business problem/solution. It is the
 process of identifying the lowest level of information for any table in your data
 warehouse. If a table contains sales data for every day, then it should be daily
 granularity. If a table contains total sales data for each month, then it has monthly
 granularity.
- During this stage, you answer questions like
- Do we need to store all the available products or just a few types of products? This decision is based on the business processes selected for Datawarehouse
- Do we store the product sale information on a monthly, weekly, daily or hourly basis?
 This decision depends on the nature of reports requested by executives
- How do the above two choices affect the database size?

Example of Grain:

- The CEO at an MNC wants to find the sales for specific products in different locations on a daily basis.
- So, the grain is "product sale information by location by the day."

Step 3) Identify the dimensions

 Dimensions are nouns like date, store, inventory, etc. These dimensions are where all the data should be stored. For example, the date dimension may contain data like a year, month and weekday.

Example of Dimensions:

 The CEO at an MNC wants to find the sales for specific products in different locations on a daily basis.

Dimensions: Product, Location and Time

Attributes: For Product: Product key (Foreign Key), Name, Type, Specifications

Hierarchies: For Location: Country, State, City, Street Address, Name

Step 4) Identify the Fact

• This step is co-associated with the business users of the system because this is where they get access to data stored in the data warehouse. Most of the fact table rows are numerical values like price or cost per unit, etc.

Example of Facts:

- The CEO at an MNC wants to find the sales for specific products in different locations on a daily basis.
- The fact here is Sum of Sales by product by location by time.

Step 5) Build Schema

 In this step, you implement the Dimension Model. A schema is nothing but the database structure (arrangement of tables). There are two popular schemas

Star Schema

• The star schema architecture is easy to design. It is called a star schema because diagram resembles a star, with points radiating from a center. The center of the star consists of the fact table, and the points of the star is dimension tables. The fact tables in a star schema which is third normal form whereas dimensional tables are de-normalized.

Snowflake Schema

• The snowflake schema is an extension of the star schema. In a star schema, each dimension are normalized and connected to more dimension tables.

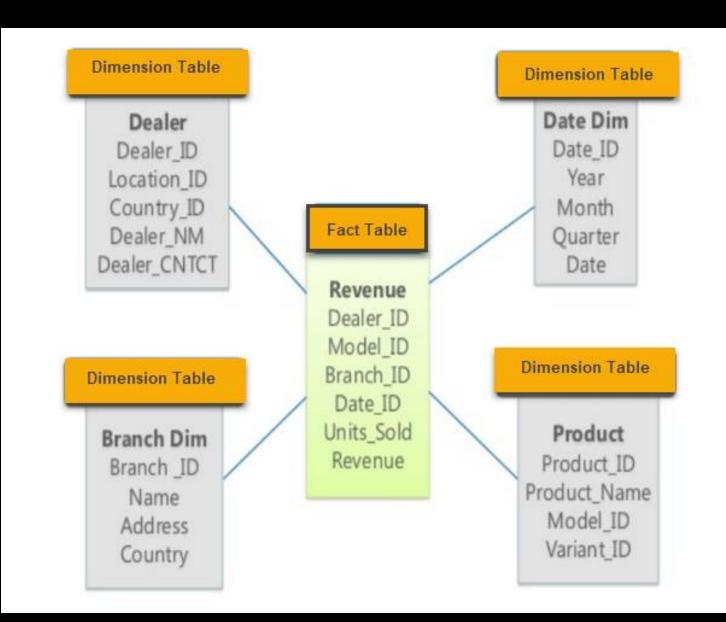
STAR AND SNOWFLAKE SCHEMA IN DATA WAREHOUSING

- What is Multidimensional schemas?
- Multidimensional schema is especially designed to model data warehouse systems. The schemas are designed to address the unique needs of very large databases designed for the analytical purpose (OLAP).
- Types of Data Warehouse Schema:
- Following are 3 chief types of multidimensional schemas each having its unique advantages.
- Star Schema
- Snowflake Schema
- Galaxy Schema

WHAT IS A STAR SCHEMA?

• The star schema is the simplest type of Data Warehouse schema. It is known as star schema as its structure resembles a star. In the Star schema, the center of the star can have one fact tables and numbers of associated dimension tables. It is also known as Star Join Schema and is optimized for querying large data sets.

For example, as you can see in the above-given image that fact table is at the center which contains keys to every dimension table like Deal_ID, Model ID, Date_ID, Product_ID, Branch_ID & other attributes like Units sold and revenue.

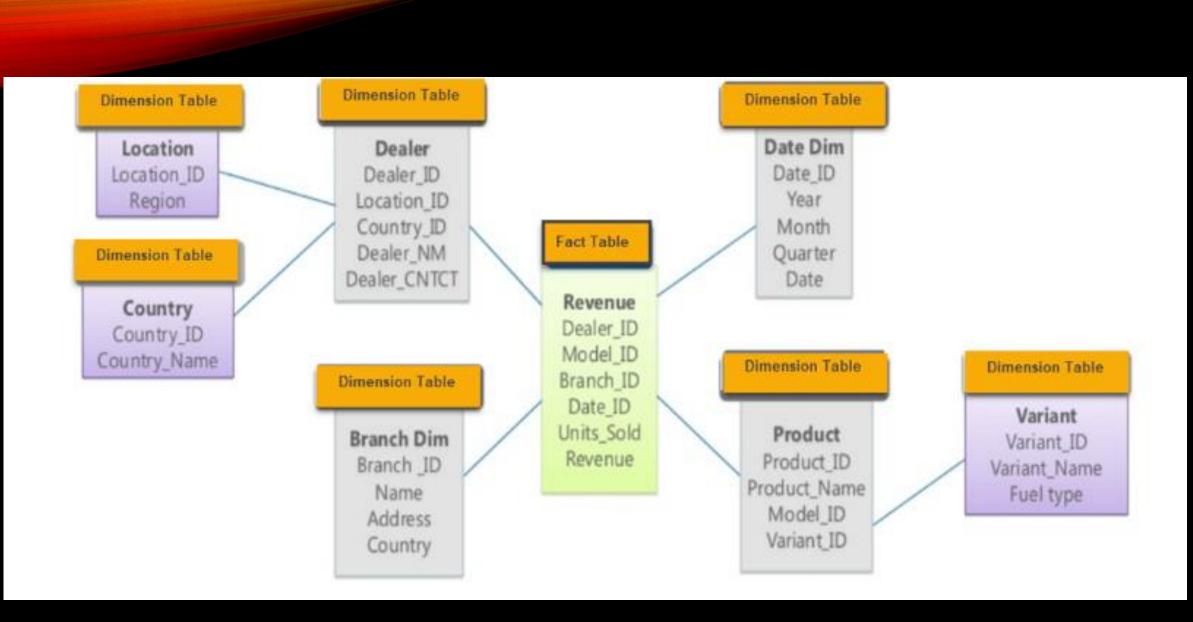


Characteristics of Star Schema:

- Every dimension in a star schema is represented with the only one-dimension table.
- The dimension table should contain the set of attributes.
- The dimension table is joined to the fact table using a foreign key
- The dimension table are not joined to each other
- Fact table would contain key and measure
- The Star schema is easy to understand and provides optimal disk usage.
- The dimension tables are not normalized. For instance, in the above figure, Country_ID does not have Country lookup table as an OLTP design would have.
- The schema is widely supported by BI Tools

WHAT IS A SNOWFLAKE SCHEMA?

- A Snowflake Schema is an extension of a Star Schema, and it adds additional dimensions. It is called snowflake because its diagram resembles a Snowflake.
- The dimension tables are normalized which splits data into additional tables.
 In the following example, Country is further normalized into an individual table.

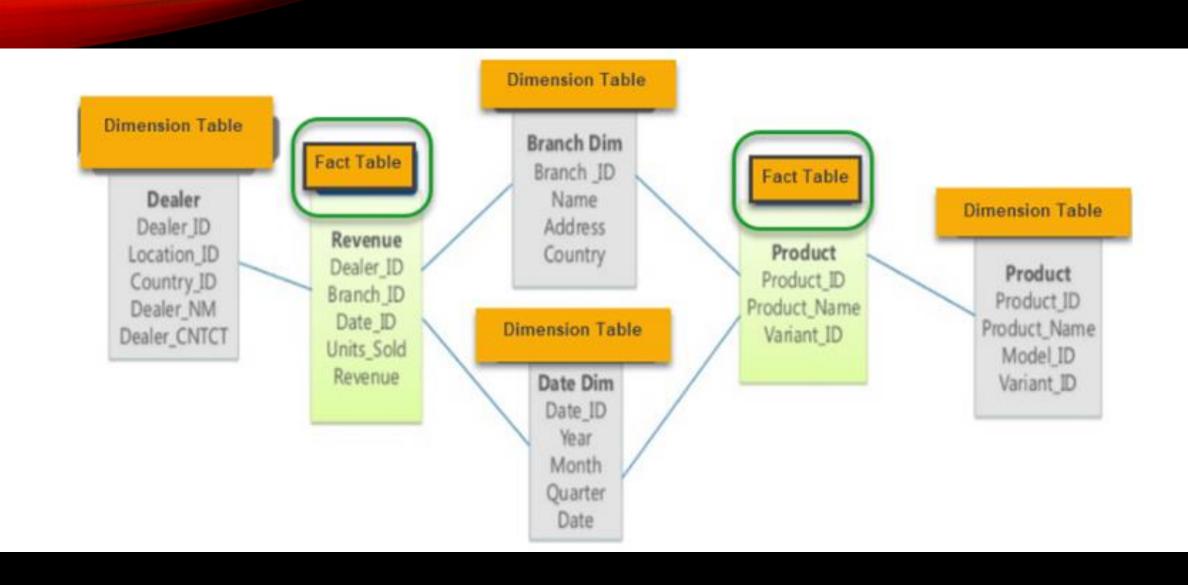


Characteristics of Snowflake Schema:

- The main benefit of the snowflake schema it uses smaller disk space.
- Easier to implement a dimension is added to the Schema
- Due to multiple tables query performance is reduced
- The primary challenge that you will face while using the snowflake Schema is that you need to perform more maintenance efforts because of the more lookup tables.

WHAT IS A GALAXY SCHEMA(FACT CONSTELLATION)?

 A Galaxy Schema contains two fact table that shares dimension tables. It is also called Fact Constellation Schema. The schema is viewed as a collection of stars hence the name Galaxy Schema.



- As you can see in above figure, there are two facts table
- Revenue
- Product.

In Galaxy schema shares dimensions are called Conformed Dimensions.

Characteristics of Galaxy Schema:

- The dimensions in this schema are separated into separate dimensions based on the various levels of hierarchy.
- For example, if geography has four levels of hierarchy like region, country, state, and city then Galaxy schema should have four dimensions.
- Moreover, it is possible to build this type of schema by splitting the one-star schema into more Star schemes.
- The dimensions are large in this schema which is needed to build based on the levels of hierarchy.
- This schema is helpful for aggregating fact tables for better understanding.

WHAT IS DATA MART?

- A data mart is focused on a single functional area of an organization and contains a subset of data stored in a Data Warehouse.
- A data mart is a condensed version of Data Warehouse and is designed for use by a specific department, unit or set of users in an organization. E.g., Marketing, Sales, HR or finance. It is often controlled by a single department in an organization.
- Data Mart usually draws data from only a few sources compared to a Data warehouse. Data marts are small in size and are more flexible compared to a Datawarehouse.

WHY DO WE NEED DATA MART?

- Data Mart helps to enhance user's response time due to reduction in volume of data
- It provides easy access to frequently requested data.
- Data mart are simpler to implement when compared to corporate Datawarehouse. At the same time, the cost of implementing Data Mart is certainly lower compared with implementing a full data warehouse.
- Compared to Data Warehouse, a datamart is agile. In case of change in model, datamart can be built quicker due to a smaller size.
- A Datamart is defined by a single Subject Matter Expert. On the contrary data warehouse is defined by interdisciplinary SME from a variety of domains. Hence, Data mart is more open to change compared to Datawarehouse.
- Data is partitioned and allows very granular access control privileges.
- Data can be segmented and stored on different hardware/software platforms.