Data Analytics and Reporting Using Business Intelligence

ArunPrakash. R[1], Shanmugapriya.T[2]
[1, 2] Department of Information Technology
[1, 2] SNS College Of Technology, Coimbatore, India

Abstract- *Data analytics is the process of managing and analysing the data from data warehouse and convert them into useful information.The business intelligence is the technology that deals with handling the big data and data warehouse to convert those data into useful information. The information from the business intelligence technique helps in the best decision making for organizations. The collection of object units that are distinct from one another and an co-ordination of those information is called as data. Database management system is a collection of organized information in a regular structure and the inter-related relationship between those data. Then the wide range of database are collected together and combines to form data warehouse. Handling this relationship and converting those data into some useful information is called big data analytics. Business intelligence provides some of the services to handle those data and analyse the data and to convert them into reports. The paper consists of some discussed techniques and those services which is used for data mining and data analytics.*

Keywords- data analytics; business intelligence; data mining; data warehouse; big data

I. INTRODUCTION

Today huge amount of data is being gathered together in many important areas like e-commerce, social network, finance, health care, education, banking and ticket reservation. Data are any facts, information, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different databases and different formats. Data mining is the process of analysing data from different outlook and summarizing it into useful information. The information that can be used to increase revenue, cuts costs, or both. It has become increasingly popular for mining such big data in order to generate reports that helps to make business decisions and to provide better personalized, higher quality services. Big data analytics refers to the process of collecting, organizing and analysing large sets of data to discover patterns and other useful information for decision making. Big data analytics help's organizations to better understand the information contained within the data and also help identify the data that is most important to the business and future business decisions.

There are more solved techniques to mine big data and business intelligence is one of the technology that provide effective optimal and analysed reports. Business intelligence (BI) is the tool and set of techniques used for the transformation of raw data into meaningful and useful information for business analysis purposes.

II. BUSINESS INTELLIGENCE

Business intelligence is the set of techniques and tools for the transformation of raw data into meaningful and useful information for business analysis and decision making purposes. BI technologies are capable of handling large amounts of unstructured data and to convert as structured data, helps to identify, develop and create new strategic business opportunities. As in [3] the goal of BI is to allow user for the easy interpretation of those large volumes of data. Identifying new opportunities and implementing an effective strategy based on insights can provide businesses with a competitive market advantage and long-term stability.Most companies collect a large amount of data from their business operations. To keep track in that information, a business would need to use a wide range of software programs, like Excel, Access, text documents and different database applications for various departments throughout their organization. Using multiple software programs makes it difficult to retrieve information from the database in a timely manner and to perform analysis of the data.

BI can be used to support a wide range of business decisions ranging from operational to strategic. Basic operating decisions in business include product positioning or pricing. Strategic business decisions include priorities, goals and directions at the large regional level. In all cases, BI is most effective while combining the data derived from the market in which a company operates the external data with data from company sources internal to the business such as financial and operations of internal data. When external and internal data are combined, it can provide a more complete picture in effect, creates an "intelligence" that cannot be derived by any set of singular data.

Generally the above systems will illustrate business intelligence in the areas of customer profiling, customer support, market research, report making, avoiding unwanted data, market segmentation, product profitability, statistical analysis, and inventory and distribution analysis.

Business intelligence combines a large set of data analysis applications, including ad hoc analysis, querying, enterprise reporting, online analytical processing (OLAP), online transaction processing (OLTP), mobile BI, real-time BI, operational BI, cloud and software as a service BI, open source BI, collaborative BI and location intelligence. BI technology also consists data visualization like designing charts, other graphics to mention information, tools for building BI dashboards and performance scorecards that display visualized data on business metrics and key performance indicators in an easy-to-grasp way. BI applications can be bought separately from different vendors also part of a unified BI platform from a single vendor.

The information available from BI solutions should allow end users to dynamically slice and dice measures (or facts) across different dimensions and further at different levels of the dimension hierarchy.

The DW stores the dimensions and fact data, which allows us to fetch static reports directly from the DW itself. However, the DW may not be able to handle the flexibility of dynamically slicing and dicing by different dimensions or various levels of the dimension hierarchy. Thus we have cubes, which virtually store the aggregated data for each measure, level of dimension hierarchy, and dimension. The cubes are data models which virtually appear to store pre-aggregated measures data across different levels of the dimension hierarchy, thereby giving end users the flexibility to dynamically slice and dice the measures at different levels.

The cube may not be required in a BI solution if the solution requires only static canned reports that are consumed by end users directly without any need for dynamic ad hoc reporting.

Business intelligence data typically stored in a data warehouse or indata marts that hold subsets of a company's information. Before it is used in BI applications, raw data from different source systems must be integrated, consolidated and cleansed using data integration and data quality tools to ensure that users are analysing accurate and consistent information.

In addition to BI managers, business intelligence teams generally include a mix of BI architects, BI developers, business analysts and data management professionals; business users often are also included to represent the business side and make sure its needs are met in the BI development process. To help with that, a growing number of organizations are replacing traditional waterfall development with Agile BI and data warehousing approaches that use Agile software development techniques to break up BI projects into small chunks and deliver new functionality to end users on an incremental and iterative basis. Doing so can enable companies to put BI features into use more quickly and to refine or modify development plans as business needs change or new requirements emerge and take priority over earlier ones.Microsoft Corporation also provides business intelligence techniques through SQL server 2012 with all three services together as MSBI. It is the only product that provides all three services of business intelligence as a single product.

A. ETL

ETL is short for extract,transform, load, three database functions that are combined into one tool to pull data out of one database and place it into another database.

1. **EXTRACTING:** Extracting the data from source systems, data from different source systems is converted into one consolidated data warehouse format which is ready for transformation processing.
2. **TRANSFORMING :** Transforming the data may involve the following tasks: 1.Applying business rules, 2.Cleaning, 3.Filtering, 4.Splitting a column into multiple columns and vice versa, 5.Joining together data from multiple sources, 6.Transposing rows and columns, 7.Applying any kind of simple or complex data validation.
3. **LOAD :** Loading the data into a data warehouse or data repository other reporting applications.

IV. INTEGRATION SERVICE

Integration service is a ETL process service that transforms all kinds of data into centralised database. It uses ETL process for transforming the raw data into centralised database. Integration service is a platform for data integration and workflow applications. It features a fast and flexible data warehousing tool used for data extraction, transformation, and loading (ETL). The tool may also be used to automate maintenance of SQL Server databases and updates to multidimensional cube data. It can convert all kinds of data like excel, text, xml data, raw data and other kinds of data into centralised or cube. The multidimensional data is that combination of x, y and z axis.

SQL Server Integration Services (SSIS) is the anchor in a trilogy of products that make up the Microsoft SQL Server Business Intelligence (BI) platform. SSIS along with Analysis Services and Reporting Services round out a platform that clearly puts Microsoft on the map in the enterprise Business Intelligence arena. In its simplest form,

SSIS is an enterprise - level extract, transform, and load (ETL) development tool. However, SSIS is not just a fancy wrapper around an import wizard. In a drag - and - drop development environment, ETL developers can snap together intricate workflows and out - of - the - box data - cleansing flows that rival custom coding and expensive third - party tools. For edge cases, the model is easily extensible and custom components can be developed in .NET languages to simply snap into the framework. However, custom coding most likely will not even be necessary. With the latest version of SSIS, novice developers can use the embedded Visual Studio Tools for Applications (VSTA) development environment to custom code workflow tasks and data pipeline transformations in VB or C# .NET languages.

The SSIS runtime engine and its complementary programs actually run the SSIS packages. The engine saves the layout of packages and manages the logging, debugging, configuration, connections, and transactions. Additionally, it manages handling the events when they are raised within the package. The core strength of SSIS is its capability to extract data, transform it, and write it out to an alternative destination. Data sources are the conduit for these data pipelines and are represented by connections that can be used by sources or destinations once they have been defined.

Connection Managers are used to centralize connection strings to data sources and abstract them from the SSIS packages themselves. In fact, the connections created by a Connection Manager are typically created in the registry of a machine and not stored in the package itself — although it can encrypt this information and store it. This allows user to deploy the SSIS package with a configuration file that can set the full value of the connection properties at runtime. One nice thing is that we can even configure the connection offline and completely design an SSIS package without connecting to the server until it is ready to test. SSIS will not use the connection until it begin to instantiate it in the package. This provides the ultimate in lightweight development portability for SSIS. Data Flow Task is simply another executable task in the package. The Data Flow replaces the simple black arrow data pump that we may be familiar with from legacy DTS packages. If this is not familiar, this arrow describes what the Data Flow does, wonderfully. The Data Flow Task is the pumping mechanism that moves data from source to destination. However, in the case of SSIS, we have much more control of what happens from start to finish. In fact, we have a set of out of the box transformation components that snap together to clean and manipulate the data while it is in the data pipeline. Transformations are key components within the Data Flow that allow changes to the data in the data pipe. Transformations is used to split, divert, and remerge data in the data pipe. Data can also be validated, cleansed, and rejected using specific rules. Transform Components in the SSIS Data Flow affect data in the data pipe in - memory. This is not always the panacea for ETL processing, especially under high - volume data processing. A core component of SSIS is the notion of a package best parallels an executable program that maintains workflow and business logic. Essentially, a package is a collection of tasks snapped together to execute in an orderly fashion. Precedence constraints are used to connect the tasks together and manage the order in which tasks will execute, based on what happens in each task or specialized rules. The file - based code is marked up using the development environment and can then be saved and compiled for deployment to a SQL Server as a filein the file system or can be saved into the msdb database metadata.

The integration service in all the data analytics techniques is integrated by ETL processing, and the transformation includes converting an unstructured data into structured data. In business intelligence techniques the transformed data into centralised database and two dimensional data into three dimensional data or cube.

V. ANALYSIS SERVICE

Analysis Services is an online analytical processing (OLAP) and data mining process in business intelligence. The analysis service is used as a tool by organizations to analyze and make sense of information possibly spread out across multiple databases, or in disparate tables or files. These services include Integration Services, Reporting Services and Analysis Services.

Analysis Services includes a group of OLAP and data mining capabilities and comes in two flavours - Multidimensional and Tabular. Analysis Services is the technology to develop Online Analytical Processing (OLAP) solutions. In simple terms, analysis service creates cubes using data from data marts / data warehouse for deeper and faster data analysis.

Cubes are multi-dimensional data sources which have dimensions and facts as its basic constituents. From a relational perspective dimensions can be thought of as master tables and facts can be thought of as measureable details. These details are generally stored in a pre-aggregated proprietary format and users can analyse huge amounts of data and slice this data by dimensions very easily. MDX- Multi-dimensional expression is the query language used to query a cube, similar to the way T-SQL is used to query a table in SQL Server.

A core component of SSIS and DTS is the notion of a package . A package best parallels an executable program that maintains workflow and business logic. Essentially, a package is a collection of tasks snapped together to execute in an orderly fashion. Precedence constraints are used to connect the tasks together and manage the order in which tasks will execute, based on what happens in each task or specialized rules.

The package is compiled into a .DTSX file that is actually an XML - structured file with collections of properties. Just like other .NET solutions, the file - based code is marked up using the development environment and can then be saved and compiled for deployment to a SQL Server as a file in the file system or can be saved into the msdb database metadata.

The tabular approach uses relational modeling constructs such as tables and relationships for modeling data, and the xVelocity in-memory analytics engine for storing and calculating data. Unlike the multidimensional approach, the tabular approach doesn't require data to be organized in a star or snowflake schema, as it relies on compressed columnar storage of data. This makes data modeling pretty much easier with the tabular approach. The tabular model uses DAX for querying and defining calculations, which is relatively easier to learn and master compared to MDX. The tabular model supports the in-memory xVelocity mode and Direct Query mode equivalent to ROLAP in the multidimensional approach. However, Direct Query mode only supports data marts hosted on SQL Server. It currently does not support any other data sources.

A core component of SSIS and DTS is the notion of a package. A package best parallels an executable program that maintains workflow and business logic. Essentially, a package is a collection of tasks snapped together to execute in an orderly fashion. Precedence constraints are used to connect the tasks together and manage the order in which tasks will execute, based on what happens in each task or specialized rules. The package is compiled into a .DTSX file that is actually an XML - structured file with collections of properties. Just like other .NET solutions, the file - based code is marked up using the development environment and can then be saved and compiled for deployment to a SQL Server as a file in the file system or can be saved into the msdb database metadata.

Every Analysis Services object can and must have an identifier. This includes cubes, dimensions, hierarchies, levels, members, and so on. By using the identifier of an object to reference the object in Multidimensional Expressions (MDX)

statements. Data Analysis Expression (DAX) queries are used to analyse the tabular data model.

VI. REPORTING SERVICE

Reporting Services is a server-based report generation software system to generate reports from data warehouse. It is part of suite of Microsoft SQL Server services, including integration services and analysis services. While analysis services enables users to construct special databases for fast analysis of very large amounts of data, and while integration services enables users to integrate data from many sources outside SQL Server, reporting services enables users to quickly and easily generate reports from SQL Server databases. The service can be customized in reports so the user can design their reports and kinds of reports they receive.

The final layer of the BI solution is the reporting layer where users derive meaningful information from reported data. The reports might be in the form of dashboard reports that display highly summarized data for executive users along with key performance indicators(KPIs) and visual indicators, or detailed reports that display each transaction that occurred as required by information workers. The reports can also be classified as static reports, which are designed and developed by developers from the data warehouse, and are consumed by end users as is. Dynamic ad hoc reports, on the other hand, are exposed to end users via cubes, allowing them to dynamically slice and dice the data. The reports might have to be delivered to end users via an email (push), or users might browse the report on demand (pull). The reporting solution should be able to cater to all types of reports needed.

A dataset is typically a query or database object reference used to retrieve a set of records for reporting. SSRS reports can consume data from a variety of data sources. The query language used for a dataset is specific to the data provider or processing extension specified in the data source. A dataset must have only one data source, but a data source, whether embedded or shared, can serve any number of datasets. Because the query command text is processed by the data provider, queries must be in the native query language of the data source. For example, a dataset for SQL Server uses the T-SQL query language. Chart reports provide a varied range of visualization options. Aggregated data is presented graphically and plotted on a linear two-dimensional or three-dimensional grid, in circular pie slices, or a radial plot space. Combinations of chart types may be combined to make comparisons and to correlate graphical data displays. In Figure 5.1, a Pareto chart employs a column chart type to range categories in descending order while a line chart type shows cumulative values.

Nearly all relational database products are queried using a form of Structured Query Language (SQL),which means that a query created for one database product may be somewhat portable to a different data source. Most database products implement a form of SQL conforming to the ANSI SQL standard. Microsoft SQL Server, for example, conforms to the ANSI 92 SQL standard, and other products may conform to other revisions like ANSI 89 SQL or ANSI 99 SQL. Beyond the most fundamental SQL statements, most dialects of SQL are not completely interchangeable and will require some understanding of their individual idiosyncrasies. SSRS is a great tool for BI developers to build canned static reports for end users. SSRS is the most flexible reporting platform with vast set of visualizations such as gauges, indicators, and maps. SSRS allows reports to be exported to various formats including PDF, Excel, Word, and HTML. Further, the reports can be delivered by a subscription to file share, SharePoint list, or email.

Regardless of the language compatibility, technical users of many database platforms typically apply a convention that is common within their communities. For example, in Oracle circles, it's considered proper form to use non-ANSI joins by matching key columns in the WHERE clause, rather than in the FROM clause. Using the JOIN and ON operators is more common in the Microsoft realm. The query syntax is interpreted by the data provider and target database engine, so it is free to use the syntax appropriate for that particular data platform. For SQL Server Analysis Services, queries are written using the MDX query language. For the casual user, designing reports for SSAS cubes requires little orno knowledge of MDX. Thus the reports from the BI are well analysed and compared which helps in the better decision making in the organisations.

VII. APPLICATIONS

The business intelligence technology is effective in handling big data. The technology can be used wherever the database exists. One of the real time application of BI is online shopping. To analyse the sales rate of each product, season sales (winter products, summer products) regular customers, sales rete based on locations, feedbacks, area needs to be focused to increase the sales, online marketing, etc., The government process like death rate calculations year by year, reasons of diseased deaths, natural disasters, weather precautions, agricultural results for all seasons and profit/loss from products, export and import analysis, transportation, rural and urban development schemes and so on. In educational institutions to analyse the students results, to categorise the student, to increase the results output, to decide the better lesson plan, mark analysis, attendance maintenance

of both students and staffs, etc., in the corporate companies analysing the projects, work output, employee details, salary packages, work output, company profits analysis etc., some of the applications like ticket reservations here BI helps to focus on customer details, time interval of each customers journey helps to provide special offers and it attracts more customers and increases the profit. BI on online shopping plays vital role on attracting the customers to purchase the products by analysing their previous activities, it helps to compare the activities through social networking also.

VIII. CONCLUSION

Business Intelligence is one of the emerging fields which provide a data analysis service for accessing the data through data warehouse. Recently it was used for the field of e-commerce. Research done in Business Intelligence to find the particular data among big data based on the data mining approach. For this we use the approach called SSIS and SSRS services. The framework which concentrates on creating and maintaining a centralized database to parse the requested data from that database instead of handling the database directly. It reduces the direct human involvement and makes the data mining computational. The Business Intelligence on educational institute makes the report making computational and reduces the man power in data analytics. The great value of BI consists in making every little business process measurable and thus create, aggregating all data and combining all measures among each other, a sustainable basis of decision making for operational units such as Marketing or Procurement. Business Intelligence enables the business to make intelligent, fact-based decisions. BI provides holistic analyses for the development of sustainable strategies. With the various BI Solutions offered by the market, we can cope with this difficult task technically as well as substantially. BI tomorrow will focus more on real-time events and predicting tomorrow's headlines. The MSBI provides the automated scheduled report generation for those database which is updated day by day. It generates report for those data periodically as per schedule by triggering the process.

REFERENCES

[1] IleladewaAdeoyeAbiodun, and Cheng WaiKhuen. "A Preference-Based Multi-Agent Data Mining Framework for Social Network Service Users Decision Making", World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:6, No:11, 2012.

[2] Mariam Adedoyin-Olowe, Mohamed MedhatGaber and Frederic Stahl, "A Survey of Data Mining Techniques for Social Network Analysis" School of Computing Science and Digital Media, Robert Gordon University Aberdeen, AB10 7QB, 2014.

[3] Yafeng Lu, Feng Wang, and Ross Maciejewski "Business Intelligence from Social Media :A Study from the VAST Box Office Challenge", Arizona State University, IEEE Computer Society, September/October 2014.

[4] Ville H. Tuulos and Henry Tirri, "Combining Topic Models and Social Networks for Chat Data Mining", Complex Systems Computation Group, Helsinki Institute for Information Technology, FIN-02015 HUT, Finland.

[5] Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding, "Data Mining with Big Data", Ieee Transactions On Knowledge And Data Engineering, Vol. 26, No. 1, January 2014.

[6] ltugAkay, AltugAkay and Bj¨orn-ErikErlandsson, "Network-Based Modelling and Intelligent Data Mining of Social Media for Improving Care", Ieee Journal Of Biomedical And Health Informatics, Vol. 19, No. 1, January 2015.

[7] Rina Fitriana, Eriyatno, Taufik Djatna, "Progress in Business Intelligence System   research : A literature Review", International Journal of Basic & Applied Sciences IJBAS-IJENS Vol: 11 No: 03.

[8] Jun Huan, Wei Wang, Jan Prins, " Efficient Mining of frequent subgraph in the Presence of Isomorphism," Department of Computer Science University of North Carolina, Chapel Hill.

[9] AltugAkay, AltugAkay and Bj¨orn-ErikErlandsson, "Network-Based Modelling and Intelligent Data Mining of Social Media for Improving Care", Ieee Journal Of Biomedical And Health Informatics, Vol. 19, No. 1, January 2015.

[10] Rina Fitriana, Eriyatno, Taufik Djatna, "Progress in Business Intelligence System   research : A literature Review", International Journal of Basic & Applied Sciences IJBAS-IJENS Vol: 11 No: 03.

[11] Pasquale De Meo, Emilio Ferrara, Domenico Rosaci, And Giuseppe M. L. Sarné, "Trust And Compactness In Social Network Groups", Ieee Transactions On Cybernetics, Vol. 45, No. 2, February 2015.

[12] Tom E. Yoon, Biswadip Ghosh, Bong-KeunJeong, "User Acceptance of Business Intelligence (BI) Application: Technology, Individual Difference, Social Influence, and Situational Constraints", 47th Hawaii International Conference on System Science, 2014.