CHAPTER I

INTRODUCTION

1.1 INTRODUCTION

1.1.1 Data Warehouse

It provides data storage architectures and gear for industry people to regulate systematic, comprehend and use the data to make planned decisions. In today’s rapidly changing world a huge number of companies initiate that Data Warehouse systems are precious gear in the contest. In the past few times many companies have spent millions in structuring Data Warehouses at the enterprise level. Many people think that with the increase of contest in every trade, data storage is the newest and must have advertising mace in such a way to keep consumers through culture more of their needs (Ruilaian HUI et al., 2011; Stella Gaitzu et al., 1999).

It has been identified Data Warehouses in a lot of ways make it tricky to place a precise description. Freely speaking, Warehouses refer to data that is handled independently when compared with the foundation Operational Databases. It allows the amalgamation of a mixture of application systems and supports information dispensation by providing a concrete stand of consolidated past data for analysis (Paulraj punniah et al., 2001; Sam Anohry et al., 1997).

WH Inmon specified that, the main originator in building a Data Warehouse is “subject-oriented, integrated, time-variant, and nonvolatile compilation of data in support of management's judgment making process”(Inmon et al., 1991). The present description is short, but comprehensive Data Warehouse Omani features large. The keywords, Object-oriented, integrated, alternative time, and is not worried discriminates Warehouses from other systems like Relational Database systems, transaction processing systems and file systems (Praveen Sharma et al., 2009).
- **Object-oriented**: A Warehouse is organized around key subject such as customers, suppliers, products, sales. Instead of focusing on day after day operations and transaction processing of the business it focuses on historical collection of data and provides analysis to decision-makers. Thus, Data Warehouses naturally present uncomplicated and terse overview about meticulous issue by not including data which is unhelpful in supporting the decision-making process.

- **Integrated**: Usually created by integrating varied sources like data from relational, flat files, rules and records of online transactions clean-up and data combination techniques of data are applied to make sure steadiness in naming conventions, programming structures, attribute measures, and so on.

- **Time variant**: The data stored is attached a time stamp to present information from a chronological point of view as information in the data repository containing element of time may be completely or clearly.

- **Non volatile**: Warehouse is always a separate repository of the converted data from functional data in an operating atmosphere. As a result of this it does not require Transactional Data Base coping, revival and concurrency control mechanisms and generally feels necessity for only fewer operations to access the data mainly for the initial download of data and admittance of data. In amount, Warehouse is reliable and linguistically which sets out as one which is actually implementing the decision support model and serving the company to make tactical decisions by providing valiant information.

Data Warehouse seen often as structural design fabricated by integrating data from diverse sources to sustain a variety of ad hoc queries, structured, critical reporting and decision-making (Jiawei Han et al., 2012).
Basing on the foregoing, Warehousing is seen as a process of organizing and making use of it. Building a Warehouse requires data amalgamation, data clear up and data consolidation. To be benefitted Warehouse always requires set of decision support techniques which paves way for the "knowledge personnel" such as managers, analysts and executives, for exercising the Data Warehouse swiftly in order to easily get an impression of the data for making resonance decisions basing on the information in the repository. Various writers used the word "Data Warehousing" as to pass on purely as the progression of building a Data Warehouse, while DBMS term Warehouse generally points to managing and using the Warehouses.

Many organizations are using this information to sustain big business decision making actions which includes

- Purely focusing on the analysis patterns of customer purchasing (like trade inclination, trade point in time, financial plan phases, and appetite to spend).
- Merchandise positioning, merchandise management schemes by comparing sales of part performance and geographic regions to tweak strategies of production.

- Looking for sources in return after analyzing the operations.

- Consumer relationship management, conducting ecological corrections and running the outlay of the company's resources.

Data storage is also very functional from the standpoint of integration of the varied data base. Many Organizations classically accumulate different types of statistics and preserve the Databases from complications, assorted, independent, dispersed sources of information, so as that they can integrate this data and provide effortlessly effective admittance which is highly advantageous, but a challenge and has spent considerable effort in data factory and research society to achieve this purpose.

The customary approach of Database to diverse Data Warehouse integration is to fabricate covers and integrators (or intermediaries) a head of multiple heterogeneous data. A diversity of data Carpenter (DB2) and blade facts (Informix) products fit in to this grouping. When you pose a query at site of the client, a data dictionary about the data is used in translation of query to the appropriate queries in the heterogeneous individual sites where the queries are placed and send to confined query processors and outcome is integrated into a global answer group coming from different locations. This query-based loom requires multifarious information filtering and integration of processes; compete for assets with treatment in the confined sources. It is incompetent and likely to be costly for the recurrent queries, especially for queries that require groups.

Data storage provides a motivating option to the customary approach of diverse Database amalgamation described earlier. Instead of using the query-based loom it uses update driven approach where information from a variety of heterogeneous origins is incorporated early and deposited in the ware house for directly querying and
Data Warehouses would not enclose for latest information unlike transaction databases and assures that the Data Warehouse brings a immense-performance incorporated as heterogeneous system where the information is hackneyed, processed in advance, incorporated, explicated, summed up and reformed into one semantic data storage. Moreover the query dispensation in Data Warehouses do not hinder with dispensation at confined sources.

Data Warehouse systems provide knowledge recruits a data study which helps in making decisions. The method which systematizes the at hand data in different forms so as to have room for the varied requirements of dissimilar users which is known as On-line Analytical Processing (OLAP) (Paulraj punniah et al., 2002; T. P. Nadeau et al., 2002).

OLAP system is considered as the data analysis oriented towards the market and is used by managers, executives and analysts. OLAP system governs huge amounts of chronological data and governs information at diverse levels of detail to facilitate the summary and assembly to increase the ease in making an informed decision. OLAP system uses a database called the planned model. Various schemes star scheme, snowflake scheme and the fact constellation are used to represent data. OLAP system integrates information collected and stored from various organizations. Access to OLAP systems and process is only reading since data is historical data. To access the OLAP system, it requires complex queries (Gupta et al., 1997).

1.1.2 OLAP Operations

OLAP systems provide drill-down drilling levels down to lesser levels of aspect drill through to summarize OLAP using a diverse situate of hierarchies of other dimensions and are saved in the resource Data Warehouse. Roll-up rolling up to hierarchical levels of aggregation. Roll-up or drill up, drill down or drills through are enormously functional methods of OLAP systems which sustain multi-dimensional analysis (Helen Hasan et al., 2001).
Once Data Warehousing developed in the form of a multidimensional data cube different analytical tools are used to perform the direct querying and complex analysis of data. OLAP is well suited for data analysis and is mainly used to access the data online and to analyze the data. OLAP provides the user responsive atmosphere for influencing data analysis and are specially designed for analyzing on very large database. Slicing, Dicing and Drilling are the operations which permit the user to vision the data at different degrees of cipher.

### 1.1.2.1 Slicing
This operation selects one dimension from the specified cube ensuing in a sub cube. For example the client posed the following query “skeptical the record of all substance that were sold in the third quarter in Hyderabad”. For this kind of queries slicing operation is used.

\[
\text{Slice}_{\text{name}} = \text{Q C}\ [\text{Quarter, City, Product}] = \text{C}\ [\text{City, Product}]
\]

### 1.1.2.2 Dicing
This operation selects more than two dimensions from the cube given and defines a sub cube. For example the client posed the following query “show me the listing of all objects that were hoarded in quarter two and three in Hyderabad and Bangalore. For this kind if queries dicing operation are used

\[
\text{Dice}_{\text{time}} = Q2\ \text{and} \ Q3\ \text{and location} = \text{“Hyderabad” or “Bangalore” C}\ [\text{Quarter, City, Product}] = \text{C}\ [\text{Quarter, city, product}] \text{where quarter and city are truncated domains such as} \ \{Q2, Q3\} \text{ and} [\text{Hyderabad, Bangalore}] \text{respectively}
\]

### 1.1.2.3 Drilling
This operation is meant for moving up and down along the classification hierarchies. Drilling operation is of two types:

Drill-up operation is moreover called roll-up. This operation switches from a detailed to an aggregate level within the same classification hierarchy. For example the client posed the following query “show me the list of all items sold” this type of queries can be answered using drill-up operations. From the given example this operation
implements data cube aggregation, also hiking the dimension hierarchy or by dimension diminution, such as aggregation of data level of the city to the level of the province or state.

Roll-up time \( C \{ \text{Quarter, city, product} \} = C \{ \text{Quarter, state, product} \} \)

The measures stored in the data cells, \( C \{ Q1, \text{Hyderabad, CBZ} \} \) and \( C \{ Q1, \text{Secunderabad, CBZ} \} \) are added to determine the measurement to be stored at aggregate level \( c \{ Q1, \text{Andhra Pradesh, CBZ} \} \).

Drill-down is the repeal of roll-up. This operation switches by an aggregate level leading into a further comprehensive level within the same classification hierarchy. In the drill-up operation the dimensions are reduced while as in the drill-down operation additional dimensions are introduced.

For Example if a client posed the following query “Show me the list of items sold in Jan to Dec” this type of queries can be quirked using this drill-down operation from the given example the Quarters are switched to more detailed level

Drill-down time \( C \{ \text{quarter, city, Product} \} = C \{ \text{Jan, City, Product} \}, C \{ \text{Feb, City, Product} \}, \) to \( C \{ \text{Dec, City, Product} \} \).

Data Warehouses and OLAP tools bases up on multi-dimensional data model. Views data from the outline with respect to cube data which allow conjunct mining at compound levels of notion.

With a huge and growing quantity of data hoard in files, databases and other archives there is an urgent need to build up a strong media to analyze, interpret and extract knowledge from data which will help in the decision-making process.

1.1.3 Data Mining
Data Mining also predominantly can be stated as knowledge discovery in databases (KDD) citing the elicitation of inherent information which is self-evident, not known previously and can be useful data in databases. Many treated DM and KDD as two
things having a same meaning Data Mining is in reality an element of the knowledge discovery process.

The KDD is a procedure of summarization having smaller number of steps primarily coming from an underdone data collection to some structure of new knowledge (Jiawei Han et al., 2006; El-Hajj et al., 2003; Liu Xmying et al., 2008) which is a continual process consisting of the subsequent steps.

- **Data cleaning:** which also known as refinement of the data in which the clattered and immaterial statistics are detached from the group.

- **Data Integration:** It is a juncture where numerous data source varied in many cases, may be pooled into a general resource.

- **Selection of data:** It is a stair where data pertinent to study determined and salvaged from data group.

- **Data transformation:** The amalgamation of data where specific data is converted in the appropriate forms for conducting mining.

- **Extract the data:** It is the decisive step that applies smart techniques to take out patterns that can be useful.

- **Assessment of Style:** A juncture where firmly attractive patterns of knowledge representation are recognized basing upon given measures.

- **Knowledge representation:** The final juncture in which the knowledge discovered is prominently illustrated to the user. These imaging proficiencies are used as basic step to help out users to comprehend and infer the Data Mining outcomes. It is familiar to unite some of these strides collectively. For example, cleaning of data and integration of data can be performed collectively as pre-generation stage to generate Data
Warehouse. Selection of data and data conversion can also be pooled where the coalition of the data is the outcome of the Selection or as in the case of Data Warehouses the Selection is done on altered data. KDD is a continual process, once the revealed knowledge is accessible to the user, the costing measures can be improved, mining can be further polished, new data can be preferred or further altered, or new data resources can be incorporated in order to get diverse but more suitable results.

It acquires its name from the correlations between searching for expensive information in a big database and mining rocks for a layer of precious ore. Both imply either purifying from outsized quantity of material or cleverly prying the material to exactly pinpoint where the values exist in. However, a Hoosier, since mining for gold in rocks is habitually called “Gold Mining” and not “Rock Mining”, thus by equivalence Data Mining should have been called “knowledge Mining” instead. Nonetheless, Data Mining became the customary term and very swiftly a drift that even overclouded more common terms such as knowledge discovery in databases that portray a more complete process (Hangchan et al., 2006; Lizengwer et al., 2004; Mishra et al., 2011). The type of patterns that can be detected on data depends on the Data Mining tasks engaged. In general, Data Mining errands are divided into two types: descriptive Data Mining that portray the common characteristics of the presented data and predictive functions to extract data which is trying to do forecasting based on inferred data available.

• Characterization: Characterization of data is to summarize the common features of the substance and produces so-called distinct policy. Usually it retrieves related data to the specified class of user database by running a query through the unit to take out the core and summarize the data at diverse levels of abstraction. For instance, one might desire to distinguish consumers who hire more than 30 films per year regularly from video store. With the concept hierarchies on the attributes unfolding the target group, the attribute-oriented orientation method can be worn.
• Discrimination: It produces so-called rules of differentiation and mainly the comparison between the broad-spectrum features of substance among the categories cited to as the objective layer and the contradictory layer. For example, one might want to contrast the common distinctiveness of clients who hired more than 40 films in the past time with those who rent expense of less than 5. The techniques used to distinguish the data are very comparable to the proficiencies used to characterize the data except that the results of discrimination statements include relative measures.

• Analysis: Association study is to discover what is usually called association rules. The completion of the study is to discover regularity of items that occur collectively in transaction databases basing on the threshold of support and classifies as groups of recurring item. An added threshold confidence is used as a conditional probability which seems to be an element in the deal when any more items recur to determine the rules of association. Correlation analysis is used to analyze the market basket study.

• Classification: It is to organize data in certain categories also known as categorization under supervision which utilizes given labels of the class rating to regulate items in a group of data. Categorization approach usually uses the working out set where all the items by now allied with class labels known. Categorization algorithm reads from the working out set and constructs a model which is used to categorize new items. Example illustrates, after the start of the credit policy managers of video parlor to analyze the behavior of customers and face their own credit-to-face and the designation of the customer accordingly who received credits with possible signs of "secure", "dicey" and "too dicey." Analyzing the rating might produce sculpt that can be used to agree to or decline credit applications in the prospect.

• Prediction: which fascinated great consideration given to predict the probable’s successfully in the context of the business repercussions. One can attempt to guess some of the data which is not available or guess class label for some of the data. The concluding is related to categorization. Once the categorization model fabricated based on the training set, the class label of an object can be expected stationed on the attribute value of the object and classes. But often it is alluded to predict the
expectations of the arithmetical value which are missing or trends in the raise or reduce of the data at relevant point. The major proposal is to utilize outsized number of previous values to regard possible potential values.

• Clustering: Like category, assembly is to organize data in layers. On the other hand, unlike categorization in clustering class poster is unidentified and assembly algorithm discovers the admissible classes also called as unverified categorization, as the categorization is not imposed by a specific category labels in many ways the whole rally is stationed on the principle of widening the resemblance by involving items in the similar category (resemblance within category) and to reduce the resemblance among items of diverse classes (similarities between class).

• Outliers: Outliers analyses are data essentials which are not classified in a particular category or group also known as repudiation or stupefaction is often very vital to recognize them. While in outliers noise can be measured and disposed of in some utilization which can expose vital knowledge in other areas which can be very large and value the study.

• Evolution and deviation analysis: It relates to examine data associated to a point which changes in a timely manner and builds evolutionary trends in data which assent to distinction, comparison, collection of time relevant facts. Deviation analysis on the further hand sees the differences involving the exact values and probable values exerts to find the cause of the deviation from the expected values (Harinarayan et al; 1996).

1.2 MINING ASSOCIATION RULE
Association rule mining supports for exhilarating exciting relationships between items in a specified data situate. Confidence and support are the measures of interestingness as they respectively give the benefit and sureness of the convention which is discovered (Agarawal et al., 1994; Han et al., 2000; Savasere et al., 1995). A Support 2% of the base assembly means that 2 percent of all transactions in study show that laptops and antivirus are redeemed collectively. A confidence of 50% says that 50%
of consumers who have redeemed a laptop also bought the antivirus. Usually, the rules of association contemplated fascinating if both satisfy the minimum support and confidence inception. These inceptions can be rested by users or dominion experts. Mining association rule in large Databases is a two-step process:

Step 1: Discover frequent item sets. As definition says, all of this will occur not completely as often as a predetermined least count support.

Step 2: Create a sturdy association convention from regular item sets.

As Definition says these conventions must meet the least support and the least confidence. Supplementary interestingness measures if they so wish. The later is the easier of the two and on the whole recital of the mining association rules is determined by the step first (Sarawagi et al., 1998; Kalnis et al., 2002; Cheung et al., 1996; Zaki et al; 2000; Park et al., 1995; Shintani et al., 1996; Agarawal et al., 2001; Webb et al., 2000).

1.2.1 Mining Recurrent Item Sets Using the Horizontal Data Format

APRIORI is efficient algorithm for mining frequent element bent for rules of logical association. The algorithm name is stationed on the reality that algorithm uses past knowledge of the frequent element bent equities. APRIORI operates a repeated approach known as level of wise search, where K element bents are used to discover (k+1) item sets. First set of frequent 1-itemsets is discovered which is denoted as \( L_1 \). \( L_1 \) is used to find \( L_2 \); in turn \( L_2 \) is used to find the \( L_3 \), and so on, until no more frequent item sets as can be originated. It requires full check of the database for finding of each \( L_k \). For bettering the competence of level wise invention of frequent item sets Apriori property is used which condenses the exploration space. As it should be all non-empty subsets of recurrent item set repeated recurrently.

By definition, if an item set \( I \) do not gratify the least support verge \( r \), then \( I \) is not recurrent, i.e., \( P (I) < r \). If an item F is added to the item set \( I \), then the resultant item set (i.e., \( I [F] \)) cannot transpire more recurrently than \( I \). Thus, \( I [F] \) is not recurrent either i.e., \( P (I [F]) < r \). This chattels correlates to a particular grouping of properties which is capped as anti-monotone in the wisdom that if a set does not through a check all of its supersets will not pass the same check as well.
Here for an example $P_{k-1}$ is used to explore $P_k$ subsisting of join and prune measures.

1. Join step: To uncover $k$, a rest of candidate $k$-item sets is accomplished by combining $P_{K-1}$ with itself. This rest of candidates are termed as $C_k$. Let $P_1$ and $P_2$ be item sets in $P_{K-1}$. The notation $P_{i[j]}$ refers to the $j^{th}$ item in $P_i$ (e.g., $P_{1[k-2]}$ refers to the second to the past item in $P_1$). By convention, Apriori speculates that elements within an operation are classified in escalating lexicographic arrangement. The join is observed, where components of $P_{K-1}$ are blend able if their first ($k-2$) elements are in general i.e., components $P_1$ and $P_2$ of $P_{K-1}$ are blended if $(P_{1[1]} = P_{2[1]}) \land (P_{1[2]} = P_{2[2]}) \land \ldots \land (P_{1[k-2]} = P_{2[k-2]}) \land (P_{1[k-1]} < P_{2[k-1]})$. The stipulation $P_{1[k-1]} < P_{2[k-1]}$ purely assures that no equivalents are accomplished.

2. Prune step: $C_k$ is a superset of $P_k$ i.e., its components may or may not be recurrent, but all of the recurrent $k$-item bents are incorporated in $C_k$. A search of the database to decide the tally of each entrant in $C_k$ would outcome in the fortitude of $P_k$ i.e., all claimants having a count no less than the minimum support count are frequent by definition, and therefore belong to $P_k$. $C_k$, nevertheless, can be gigantic, and so this could perk up serious calculation. To lessen the extent of $C_k$, any $(k-1)$-item set that is not recurrent will not be a subset of a recurrent $k$-item set. Hence, if any $(k-1)$-subset of a candidate $k$-item set is not in $P_{k-1}$, then the candidate cannot be recurrent also and so can be detached from $C_k$.

The subset measuring could be completed swiftly by sustaining a hash tree of all recurrent item sets and accomplishing association rules from recurrent item sets once they are found from transactions in a database $M$, it is uncomplicated to produce sturdy association regulations from them (where strong association rules satisfy both least support and least confidence). This can be done using equation for confidence, where the conditional probability is uttered in terms of item set support count: $\text{Confidence (A|B)} = \frac{\text{support count (A|B)}}{\text{support count (A)}} = \frac{\text{support count (A\cap B)}}{\text{support count (A)}}$; where support count $(A \cup B)$ is the amount of transactions accommodating the item sets $A$ and $B$, and support count $(A)$ is the number of transactions accommodating the item set $A$. Basing on this equation the rules of association can be accomplished as supervened.
For each recurrent item set \( h \), generate all non-empty subsets of \( h \). For every non-empty subset \( r \) of \( h \), output the rule \( r \) if support count \( (h) \) support count\( (r) \), where \( \min\text{conf} \) is the minimum confidence threshold. Since the rules are accomplished from recurrent item sets then everyone routinely gratifies minimum support. Recurrent item sets can be subservience in advance of time in hash tables along with their counts so that they can be achieved quickly.

### 1.2.2 Mining Recurrent Item Sets Using the Vertical Data Format

The Apriori and FP-growth (Han et al., 2000) methods mine recurrent patterns of a group of transactions in Tid item set arrange where TIS is the identity of the deal and item collection is a set of items acquired in transaction TID known as the horizontal data arrangement. Instead it can also display the data in the form of item TID_set, where the item is the item name and TID_set is a set of transaction identifiers that contain the ingredient known as vertical form of data as shown in Figure 1.2 to Figure 1.4. First turn horizontally coordinate data to the vertical form by searching the data situate once. Support count of item group is basically the extent of a group TID set item set (GUO Yi-Ming et al.; 2010). Starting with \( K = 1 \) the recurrent item groups can be used to build a candidate \((k+1)\) -item sets. This process reruns every time by increasing \( K \) by 1 every time until no frequent item groups or groups of any item candidate can be initiated.

Advantages of this algorithm:
- Better than some constructive in generating a candidate \((k + 1)\) -item range of groups as a recurring item
- There is no necessitate to examine the Database to discover support groups \((k +1)\) item (for \( k \geq 1 \)) since TID_set all your item support.

<table>
<thead>
<tr>
<th>Item Set</th>
<th>TID_SET</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>T1,T4,T5,T7,T8,T9</td>
</tr>
<tr>
<td>I2</td>
<td>T1,T2,T3,T4,T6,T8,T9</td>
</tr>
<tr>
<td>I3</td>
<td>T3,T5,T6,T7,T8,T9</td>
</tr>
<tr>
<td>I4</td>
<td>T2,T4</td>
</tr>
<tr>
<td>I5</td>
<td>T1,T8</td>
</tr>
</tbody>
</table>

**Figure 1.2: 1-Item Vertical Data Format**
1.3 MOTIVATION
After a cautious study on the above disadvantages in existing approaches, the present study found that Data Mining approaches can be effectively used in Materialized View Selection and Maintenance process and also can be remodeled to get better results. Based on the above, the present thesis worked in the direction to develop novel schemes to resolve above problematic and vital issues in Materialized View Selection and Maintenance, by using advanced Data Mining techniques and improving the efficiency.

1.4 PROBLEM IDENTIFICATION
After a careful, thorough and decisive exploration in literature survey the present study found that due to the increase in data of the Warehouse repository there seems
to be a huge requirement of advance techniques in the area of Partial Materialization. The existing approaches yield better results using the Data Mining techniques in Materialized View Selection and Maintenance, but these methods are restricted to table level \ dimension level. There is a great scope to propose new Selection and Maintenance method which can ably create a solution by developing attribute level as the data cells in MOLAP are arranged in attribute levels.

Thus it has become a very important area for researchers experimenting with new methodologies for proposing advanced View Selection (MVS) and View Maintenance (MVM) methods to use the existing data in an efficient way. The present work, thereby introduces certain novel methods which can overcome the shortcomings of these existing techniques so as to improve the Selection and Maintenance process and then investigate the suitability of them in terms of performance in applications.

1.5 OBJECTIVES OF THE PRESENT STUDY
This section will briefly outline the objectives of the proposed work. The main intention of this thesis is to put forward a novel advent for Materialized View Selection (MVS) and Materialized View Maintenance (MVM) and to improve the performance of Materialized View Selection by proposing Improved Materialized View Selection (IMVS) approach. Based on the modifications and performance evaluation, some modified and novel methods were proposed which retain some of the best characteristics of the available MVS methods and produce even better results.

The objectives of the survey are encapsulated as follows:

1. To study the existing Materialized View Selection and Maintenance algorithms.
2. To propose novel framework for Materialized View Selection.
3. To implement Materialized View Selection for getting better results.
4. To compare the existing algorithms with the derived results of newly proposed algorithms.
5. To compare the gain measure of MVS approach with the existing algorithms.
6. To propose novel framework for Materialized View Maintenance.
7. To compare the efficiency of MVM with the existing methods.
8. To propose novel framework for Improved Materialized View Maintenance.
9. To evaluate the evaluation outcome of the existing algorithms
   With the derived outcomes of newly proposed algorithms

1.6 SCOPE OF THE PRESENT THESIS

To meet the above problem statement and derived objectives the present thesis is divided into six chapters.

Chapter 1 provides the basics of Data Warehousing, Data Mining, OLAP operations, Association Rule Mining and outlines the objective of the study for possible solution of the problems identified. It also provides the complete scope of the thesis.

Chapter 2 provides a details study of literature survey conducted for the identification of area and problems relating to objectives of the thesis.

Chapter 3 presents one of the main contributions of this thesis, a novel Materialized View Selection approach using Association rule Mining. Section 3.2 presents the MVS frame work with different individual components. Section 3.3 presents the experiments conducted and comparative study between the new proposed approach and existing approaches. Section 3.4 presents the Cost model and Gain Measure of the proposed MVS approach.

Chapter 4 presents second main contributions of this thesis, Materialized View Maintenance (MVM). Section 4.2 presents the MVM frame work with different individual components. Section 4.3 the experimental comparisons with traditional and recent algorithms is presented and complete analysis of the results is done.

Chapter 5 presents third contribution of this thesis, Improved Materialized View Selection (IMVS) approach. Section 5.1 gives the introduction of this chapter and
presents details regarding Improved Materialized View Selection approach and the base algorithm Improved Apriori. Section 5.2 presents the framework of the IMVS approach. In Section 5.3 the experimental comparisons with traditional and recent algorithms are presented and complete analysis of the results is done.

Chapter 6 Deals with overall summary and future scope of the present thesis. Further a brief summary is given at the end of each chapter.