LTIMindtree

Data Quality Management & Analytics techniques in Upstream

The recent advent of upstream data has compelled O&G firms to intensify its data quality management efforts. They need to better manage the hurdles like quality of the data and the complexity in analyzing that data, before they can interpret the basins and improve reservoirs. This paper discusses the importance of improving the quality of the wealth of data that O&G firms are sitting on and suggests simple steps to analytics driven future.

Introduction

The Oil & Gas (O&G) industry conducts extensive analyses in various phases such as geological, geophysical, geochemical and drilling in the exploration and production of oil and gas to improve the yield and to remain competitive. In the competitive and challenging O&G industry, rigorous data management and analytical tools are very essential and with their deployment storing and accessing data has become much easier as compared to the earlier printed format.

Data plays a vital role in any analysis at every point for making valuable decisions in every aspect of hydrocarbon exploration and production. However, the data available in the Exploration and Production Industry is usually found in complex formats, and most of it is contained in hard physical files that were collected over a period of time (may be a few decades). This makes it cumbersome to search for the correct files from a whole lot of data, especially when the data is not in any specific order. That's where Data Quality Management and Analytics play an important role in helping O&G companies deal with the collection, analysis, loading, quality of data, and ensures that the uploaded data meets corporate standards. Data Analytics is applied on the qualified data by generating the reports and charts for better decision-making in the upstream industry.

Data Quality Management is the process of providing solutions (in each of the above phases) that would include collection, processing and storing of data. The processing or analyzing of data involves various business rules to qualify and quantify the data according to the organization's requirements. The capturing, storing and accessing of valuable data is made easy by O&G data management tools. Data Analytics is employed in aggregating the data from various data sources by improving the visualization of data for industry needs.

Why Data Quality Suffers

Legacy systems – Improper organization of stored data

Data visualization – Lack of visual representation of data in the form of reports and charts

Data formats – Data stored in structured and unstructured format

Multiple versions of master data – Historical data sharing methods have created multiple copies of information. In a 1992 paper, Blaine Taylor (1992, p.193) states that 80% of a geoscientist or engineer's time is spent looking for information, verifying its accuracy, and formatting the data into analytical or modeling tools for analysis.

Lack of standard taxonomy/ data structure

Human error

Multiple users, locations and languages

Limited stewardship of data creation and maintenance

Following is the classification and example of poor quality data:

Duplicate data – Multiple UWI's for same UWI

Incomplete data – Missing the spud date for a well, missing a well incompletely

Inconsistent data – Different surface locations in drilling data as compared to geological data

Inaccurate data - Incorrect geodetic datum

Out of Date – Not using the latest directional well plan

Not fit for Use – Incorrect naming standards of curve mnemonics

Such data cannot become useful as it is and must be identified and corrected for further use.

Enhancing the Quality of Data

The activities involved in the identification and correction of defects are:

Measuring the defects in data, then processing the identified defects and grading the quality of the data.

Systematically eliminating defects in data, using specialized data quality software, comparing the data with regulatory repositories and using specialized software to make the corrections as required. Whenever the corrections need to be done manually, the technical team goes back to the drawing board and makes the corrections manually. Once corrections are made, the quality of data is re-assessed.
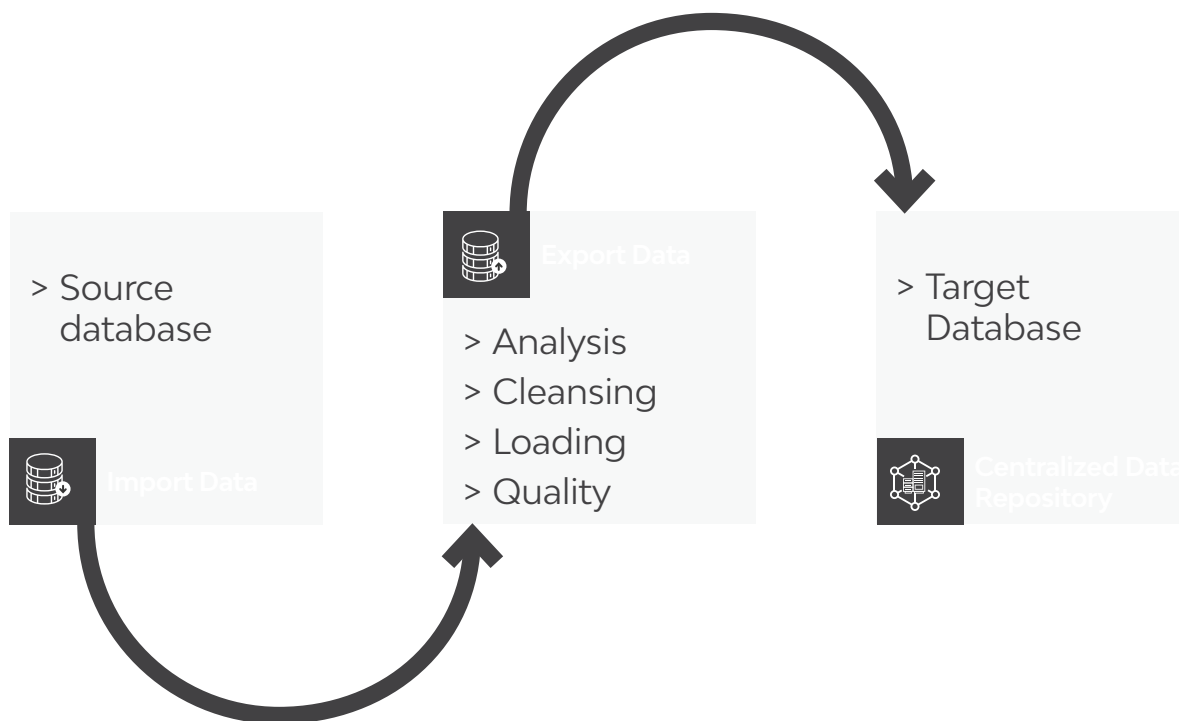
> Source database

Export Data

> Analysis
> Cleansing
> Loading
> Quality

> Target Database

Import Data

Centralized Data Repository

Fig 1: Processing of data

Source Database
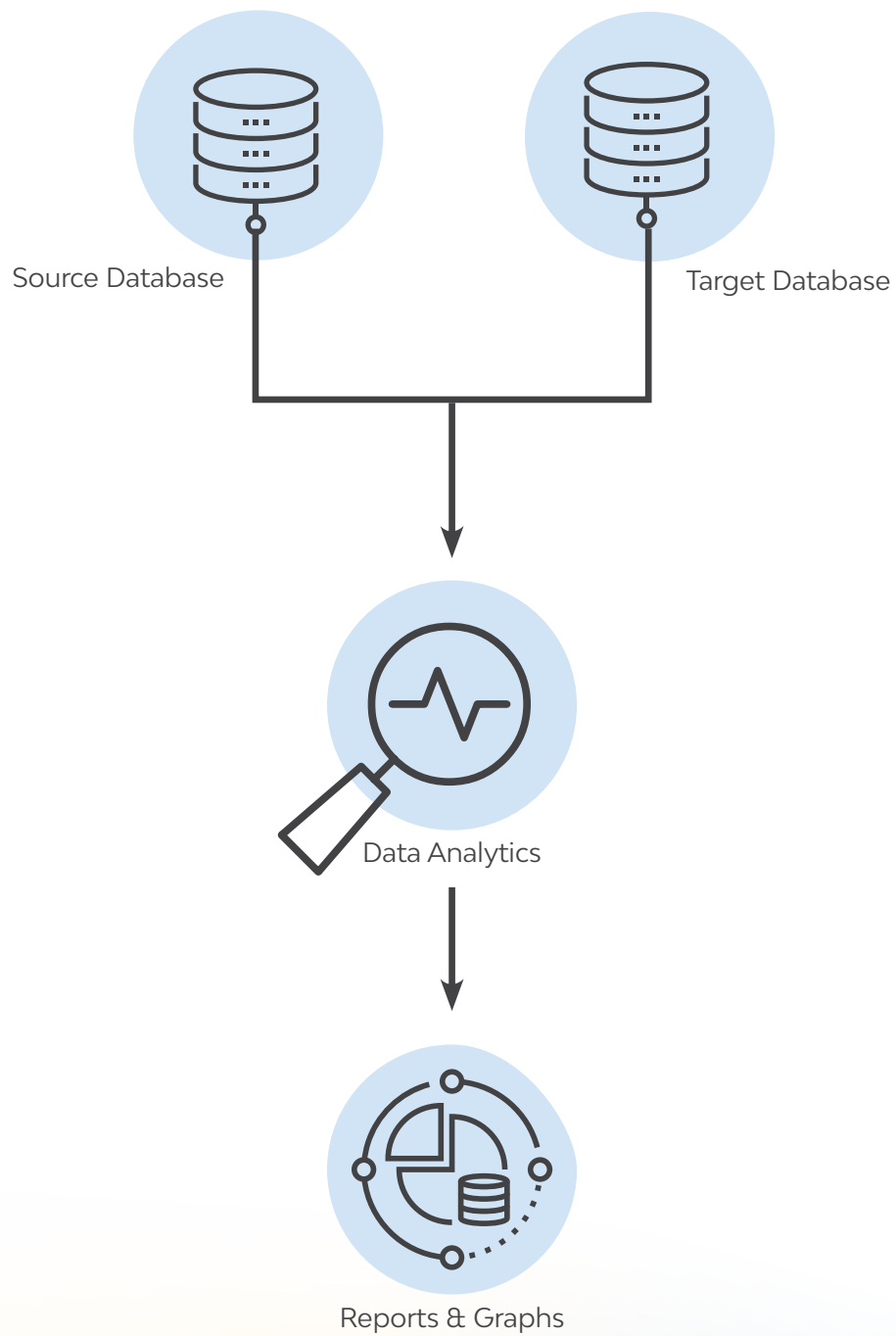
Target Database

Data Analytics

Reports & Graphs

Fig 2: Comparing data

Data Quality Methodology

In order to improve poor quality of data we must implement a process for measuring data defects and systematically eliminate them. This is where a predefined process of data management plays a crucial role in qualifying where the data stands on a scale of quality. This can be better explained by using data from an oil well as an example.
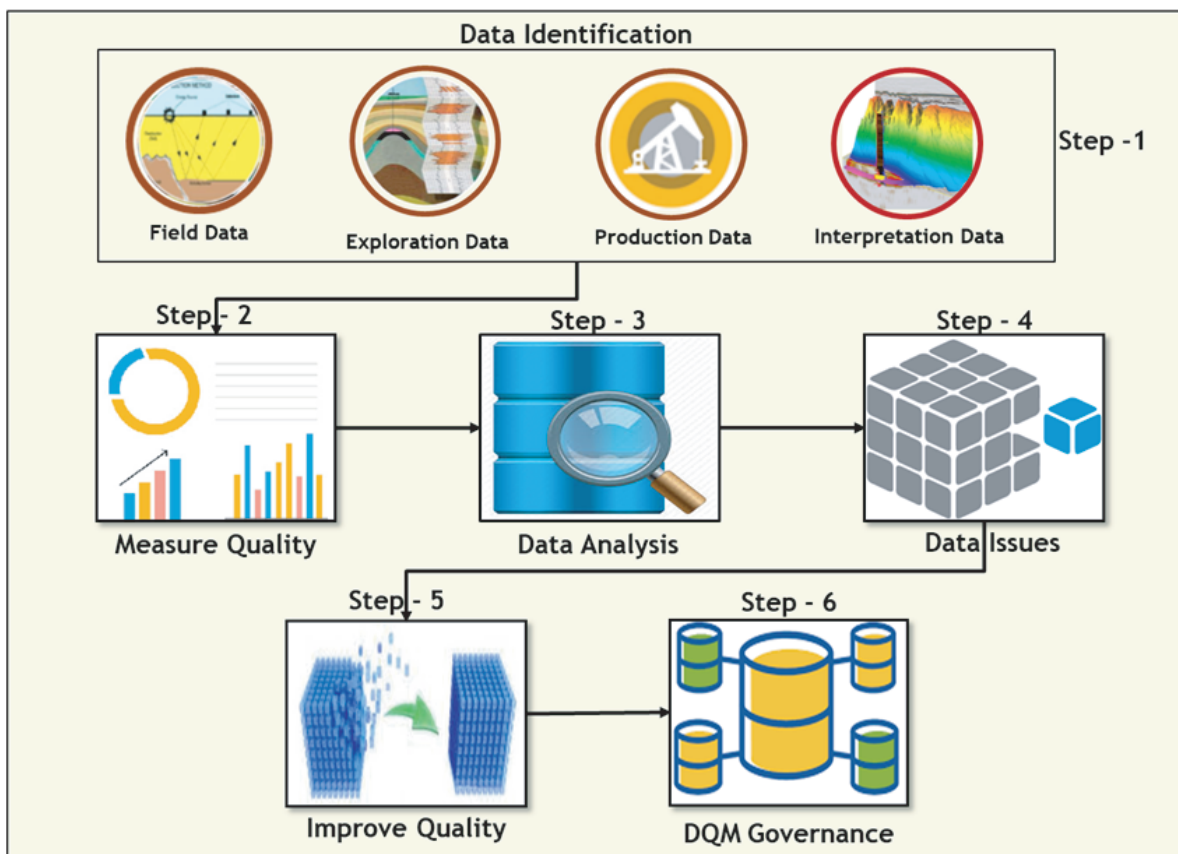

Fig 3: Data identification

Following is the process involved in Data Quality Management of typical data from a well:

Identify the data that needs to be improved

During this phase the team works to define the data types to be improved.. Examples of such data types are logs, well header, well paths, etc. After assessing the data types to be improved by consulting the business stakeholders, the data team can dive deeper to focus on the specific fields for improving the data. Examples of specific data fields in well header data are the surface location of a well, well UWI, operators, etc. The next step is to work on quality requirements for the data fields, i.e. decide if an 80% accuracy level acceptable for the data field or does it need to be at least 99% accurate. Finally, the decision should be made on the percentage of measuring the data accuracy of each data type.

Measuring the quality of the well data

During this phase of the project, the team will translate the measurements defined in step one into specific quality rules and will then load the set of rules into the software tool designed to measure data quality. This tool helps in measuring the quality metrics of data before and after the data cleanup process and we can gauge how much data has been improved.

Analyze the existing well data

During this phase, the rules are run against the dataset. Based on the rules defined, the erroneous data is identified. At this point, a root cause analysis is conducted to understand the underlying reason why so many wells failed a certain rule. If there is an underlying systematic problem, then changes are recommended to the work process to avoid future problems with the data.

Group problem wells

Once a set of problem wells are identified, they are grouped based on similar problems. If a group of wells are failing a certain rule, then the team creates a correction rule that is always applied to the data. For example, if surface location is missing for a group of wells, a rule is created that pulls the surface location from the G&G database and loads it into the drilling database automatically every end of the day. Grouping wells also helps in dividing the cleanup work in the next step.

Improve the quality of the well data

This is the most important step. Once the lists of wells with data issues have been identified, the team then systematically figures out how to eliminate these defects. This is done by using software tools to compare the data with other data sources and perform corrections using the tool or manually. After this, the team reverts to Step 3 and checks if the existing well data meets the quality goals defined in Step 1.

Control future well data problems

Now that data has been cleaned and deemed fit for use, how will an O&G company ensure that it is properly maintained and kept error free, and that any new data added is also good quality? Software tools can be set to automatically reassess data and send reports of data quality when it falls short of established standards. Henceforth, it is very important for data quality governance strategy to be applied on all the future data migrating to the corporate database.

Rules Employed in Data Quality

Based on the problems in data, data standards should be defined for assessment of rules to be used for measuring data quality. There are several aspects to data quality, listed below, that determine the typical assessment rules:

Completeness
– Does the data source have all the required attributes?
For example does each source of data contains Operator name under well operator field.

Consistency
- Do the attributes of each item agree between data sources?
For example does the well have the same surface location in each data source.

Uniqueness
– Are there multiple items that really refer to the same item?
For example, is the same well listed several times with different UWI(s) and or name(s).

Validity
– Are the attributes of each item "incorrect"?
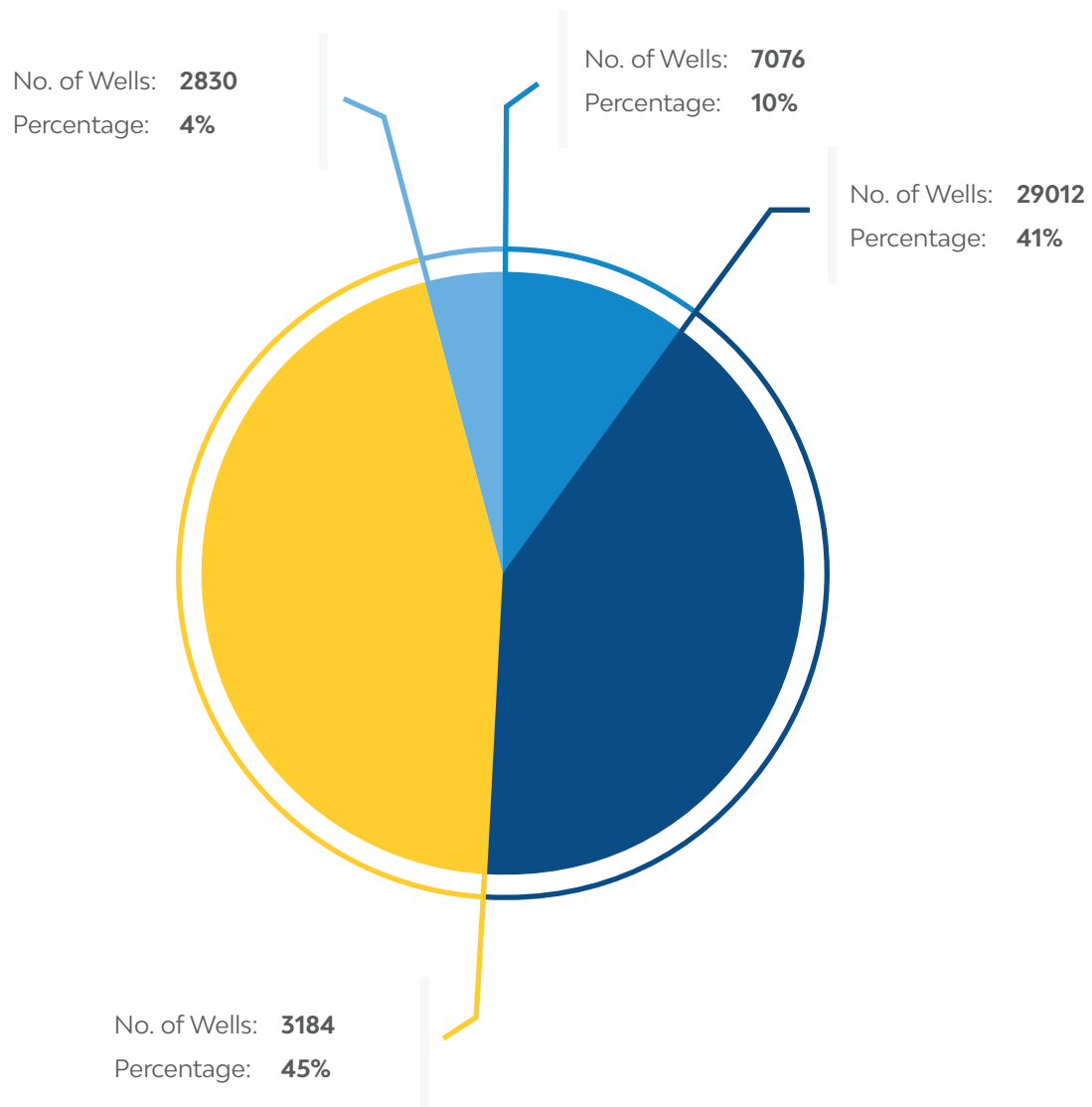For example, is the operator a valid operator for the area.

Content
- Do all data sources have all the required data?
For example, if the wellbore is deviated, then the wellbore is required to have a well path.

Data Quality Analysis

The following example gives an idea about, the number of wells failed under a project after executing data quality rules.

No. of Wells: **2830**
Percentage: **4%**

No. of Wells: **7076**
Percentage: **10%**

No. of Wells: **29012**
Percentage: **41%**

No. of Wells: **3184**
Percentage: **45%**

Rules: ■ Completeness ■ Consistency ■ Uniqueness ■ Validity

Data Analytics
– Modeling and Visualization

Data Analytics and Modeling is deployed by segregating the data from various data sources by creating a single data repository at corporate level involving master data management concepts by generating the various predictive models using multiple combinations of data. Data obtained by aggregating the data from various sources like LWD, MWD, SCADA, fracking, etc. can be used to generate predictive models. Better data visualization techniques will help in better decision-making based on the data. Production data can be analyzed on various assets for developing better production optimizing mechanism by building a flexible and cost-efficient data and analytics platform.

Data Management in the Cloud

Availability of a corporate data store in cloud environment bring IT costs down but could also improve the way the business solutions managed organically. Cloud technology will provide O&G companies with the powerful processing capabilities they need, while facilitating interaction between multiple offices across the world. Cloud will also ensure enhanced security and protection of sensitive information by storing data in a virtual hub rather than on local servers, and at much low cost.

Cloud adoption will be one of the main IT trends for the O&G sector, along with the Internet of Things, intelligent rigs, volumetric calculations of the surface formations, and leak detection software. The benefits of the cloud include realizing value from projects much more quickly, and developing new solutions

that are flexible and easier to deploy. Most of the upstream data is maintained in structured and unstructured format. For example, most of the well data and seismic data are in flat file format, wherein a lot of massaging needs to be carried out for converting unstructured data into structured data.

In the O&G industry, decision-making involves multiple groups and various domains. To be able to take the right decisions, O&G companies need to communicate data quickly in a variety of ways on various platforms, including visual interpretation of data. Data visualization can help maximize assets and optimize production without compromising safety. For example, data visualization's ability to tap critical upstream data and convert it into colorful images can highlight priorities, offer meaningful information to decision makers and build momentum for all the domain projects.

Below is an example illustrating a dashboard can be prepared from multiple sources of data using analytical tools for better visualization techniques.
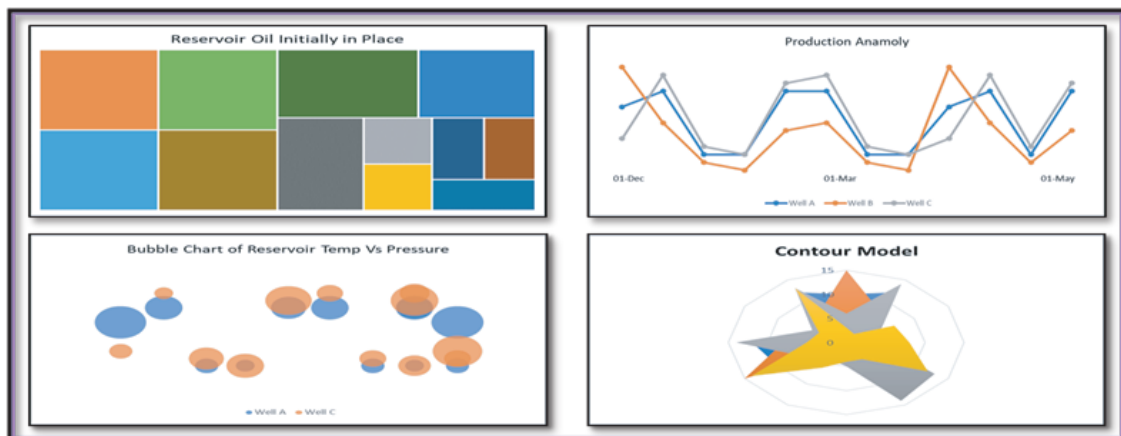
Conclusion

Effective O&G data management facilitates desktop access to numerous up-to-date databases, including those related to surface and subsurface land, wells, pressure, temperature, production, pipelines, core, reserves, seismic and logs. Combining data quality with analytics, improving the data quality standards and utilizing the analytical techniques for analyzing the data ensures that business users get high quality data, and allows them to better visualize data and analyze it for better decision-making.

But maintaining data quality is not limited to current data, and also extends to the historical and onboard data. This is done through retrieval, analysis, loading, verifying, validating, cleansing and analytical techniques for better-qualified data and analysis. There are several benefits of Data Quality Management and Analysis, such as:

- ✓ Cost savings resulting from the removal of redundant data and reduced manual efforts

- ✓ Better customer service with more accurate and relevant customer records

- ✓ Reduced operational costs

- ✓ Lower overhead costs through higher employee productivity

- ✓ Enhanced value of existing assets through analytical techniques

- ✓ Support for effective tactical and strategic decision making through more accurate analytics

- ✓ Increased revenues through knowledge and insights

LTIMindtree