

WHITEPAPER

ReCast Modernizer

Reengineering BI migration to ThoughtSpot

Authors: **Sheba Qureshi**, Ramanath KP

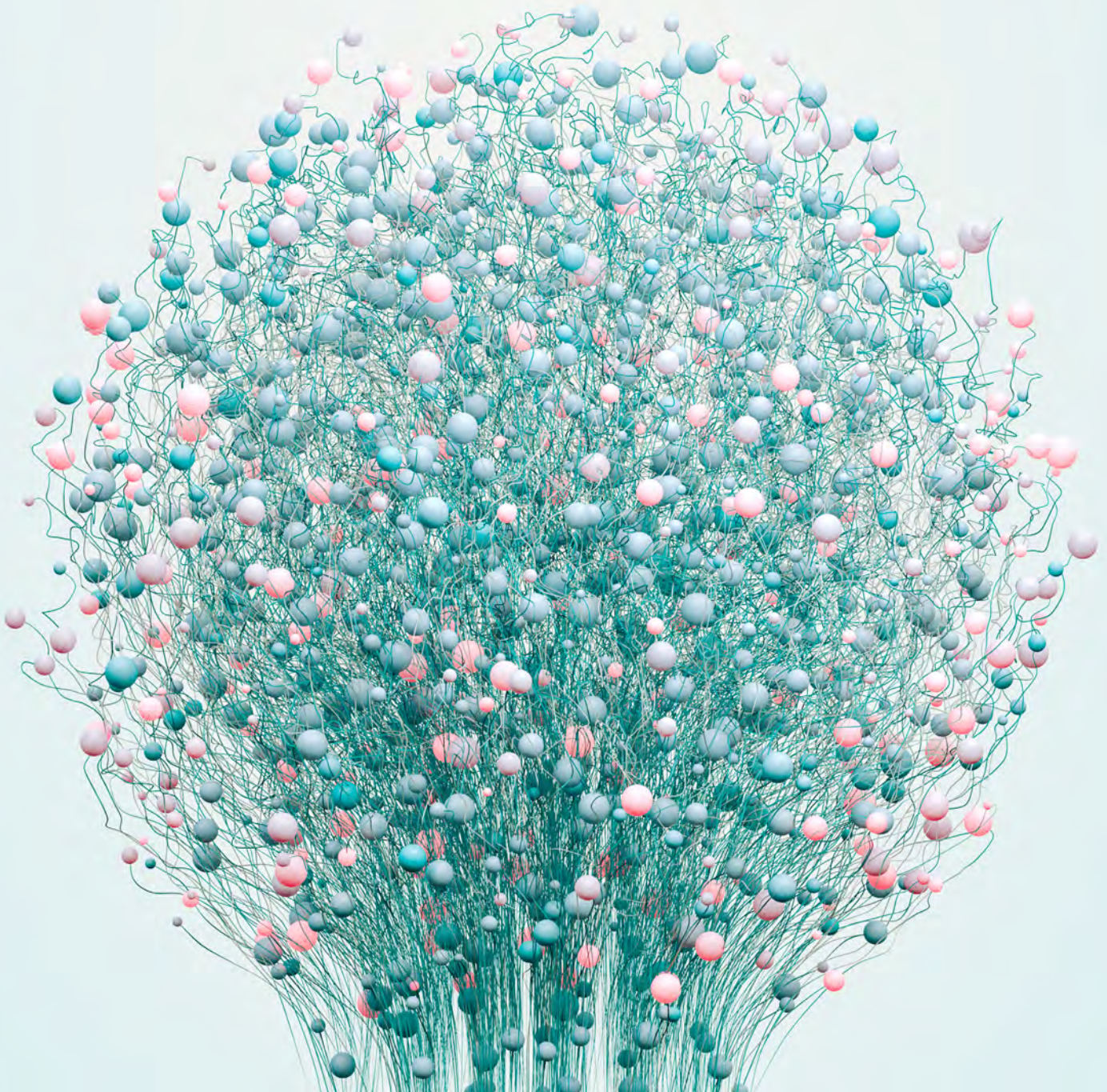


Table of Contents

1	BI architecture evolution	3
2	ReCast – BI migration	4
3	ReCast product offering	5
4	ReCast Workflow	8
5	ReCast data dependency workbench	9
6	ThoughtSpot Modernizer	11
7	ReCast Migrator	11
8	ReCast Reviewer	12
9	Feature conversion/comparison	13
10	About the Author(s)	14



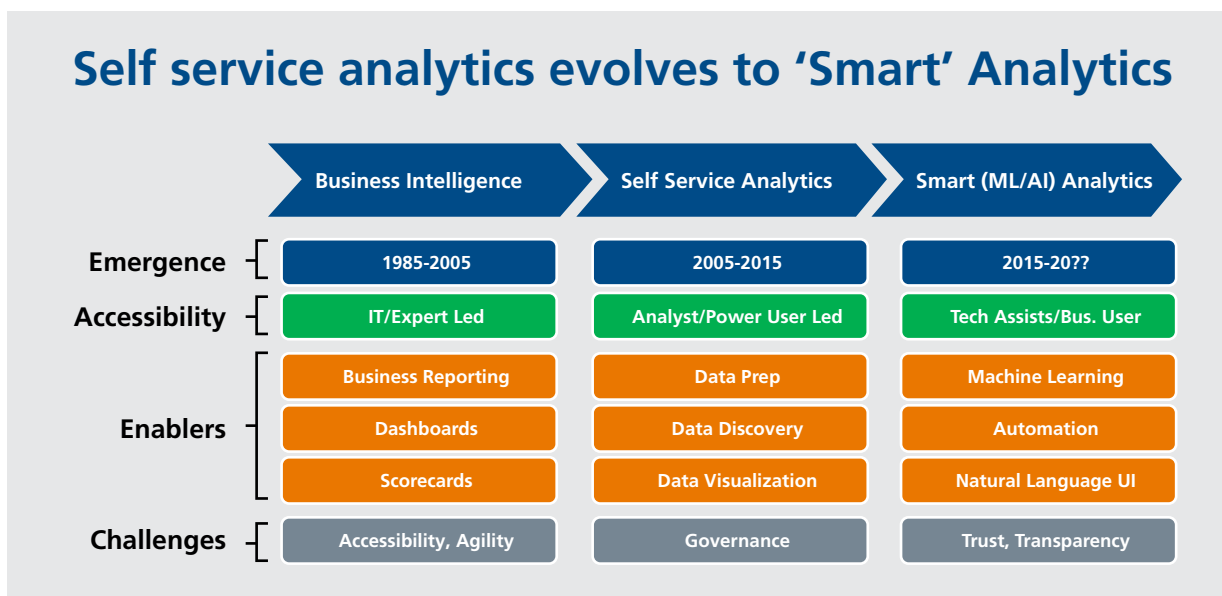
BI architecture evolution

Riding on data democratization, data literacy, and data cloud investment, BI modernization is the next strategic bet toward building a truly data-driven decision-making organization. The operational impetus to reduce cost, cloud adoption directive, and enhance user experience is further accelerating the migration from legacy BI to modern BI tools.

In recent times, as self-service BI has become more popular, it has enabled business users to explore the data and get insights without relying heavily on analysts and engineers. Ease of use and simplicity are the expectations from the BI tools of the current generation. With the integration of the latest technology like Artificial Intelligence (AI) and Natural Language Processing (NLP), BI self-service has risen to the next level.

NLP-based BI predominantly uses AI programming to understand inputs in the form of query or texts and converts them into business-understandable solutions. NLP and AI are affecting all enterprise software but will completely transform how we build, analyze, and consume data and analytics.

Considering the latest trend, a lot of BI tools are increasingly adding NLP to their products to ease dataset accessibility by supporting NLP-based search. Even as a lot of powerful tools in the market have added NLP functionality, it still is an additional feature. Over the past ten years or more, visual-based data discovery tools (e.g., Tableau, Qlik, Tibco Spotfire) have disrupted the traditional BI market (e.g., IBM Cognos, SAP BusinessObjects), which is now slowly moving towards NLP-based data discovery tools (e.g., ThoughtSpot).



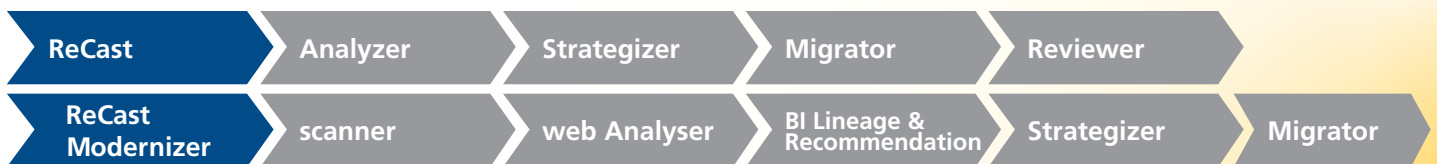
'Smart' analytics evolution



ReCast – BI migration

LTIMindtree’s ReCast is one such tool that consistently works towards improving the automation process for BI tool migration. We have various scanner tools for extracting metadata from the client’s BI landscape. With our experience of migration, we always found that the client’s BI landscape needed a revamp, redesign, or re-architecting. Some common development fallacies are duplicate reports, unused reports, too many similar kinds of reports with filters, etc. Our state-of-the-art tools help detect the various fallacies and improve the BI architecture. Our tool chains guide the migration process and create various strategies for various target BI tools. Using this approach, we have successfully migrated traditional BI tools like Power BI, Tableau, AWS Quicksite, etc.

We are now in the age of the AI revolution. ThoughtSpot is an AI-powered BI analytics, a new generation BI representative. Our traditional BI migration process needs an AI shift to cater to these new BI tools. ReCast Modernizer is our brand-new approach to migration. We redesign our tools using the latest technologies like graph data model, semantic search, domain knowledge graph, synonym generator, and the latest UI for lineage and decision making.



ReCast and Modernizer workflow

With the ReCast modernizer, we have introduced more components for better analysis and recommendation generation. Also, knowledge graph of the domain gives more context keywords or synonyms, which helps improve search.

Challenges for BI migration:

BI migration initiatives are complex and marred with challenges associated with a lack of documentation of legacy BI implementation, database scheme and data drift, the magnitude of migration, automation complexity, and shortage of talent.

Here are a few ThoughtSpot adoption challenges

- Extended learning curve
- No tool to depict the existing BI landscape
- What to improve or rationalize on the current BI landscape
- Huge rework required to migrate old reports to ThoughtSpot
- Lack of trained resources and longer time for adoption

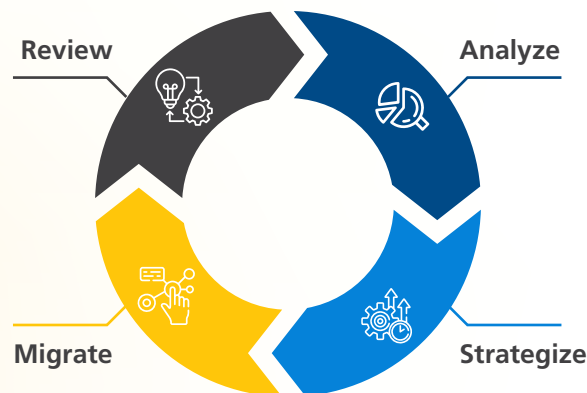


ReCast product offering

Standard BI to BI migration

ReCast automates the report migration lifecycle starting from report inventory gathering and analysis, rationalization, and migration, to post-migration validation. LTIMindtree has leveraged Robotic Process Automation (RPA) and AI technologies to develop this BI migration automation tool.

The core aspect of migration is to transfer accurate information from source to target systems. To ensure that the migrated data is accurately represented, it is necessary to first assess the source system thoroughly and create a report migration strategy to achieve the objective. The approaches might be different, either manual or semi-automated, but the steps remain the same. Migration life cycle:



This migration methodology covers the entire data life cycle – right from the semantic layer, reports, users and groups, schedules, distributions, reports, or cube refresh.

■ ReCast Analyzer

Analyzes legacy BI implementation, inventory and rationalize reports, and dashboard to be migrated.

■ ReCast Strategizer

Prepare a knowledge base for migration by mapping features between the legacy and new BI tool.

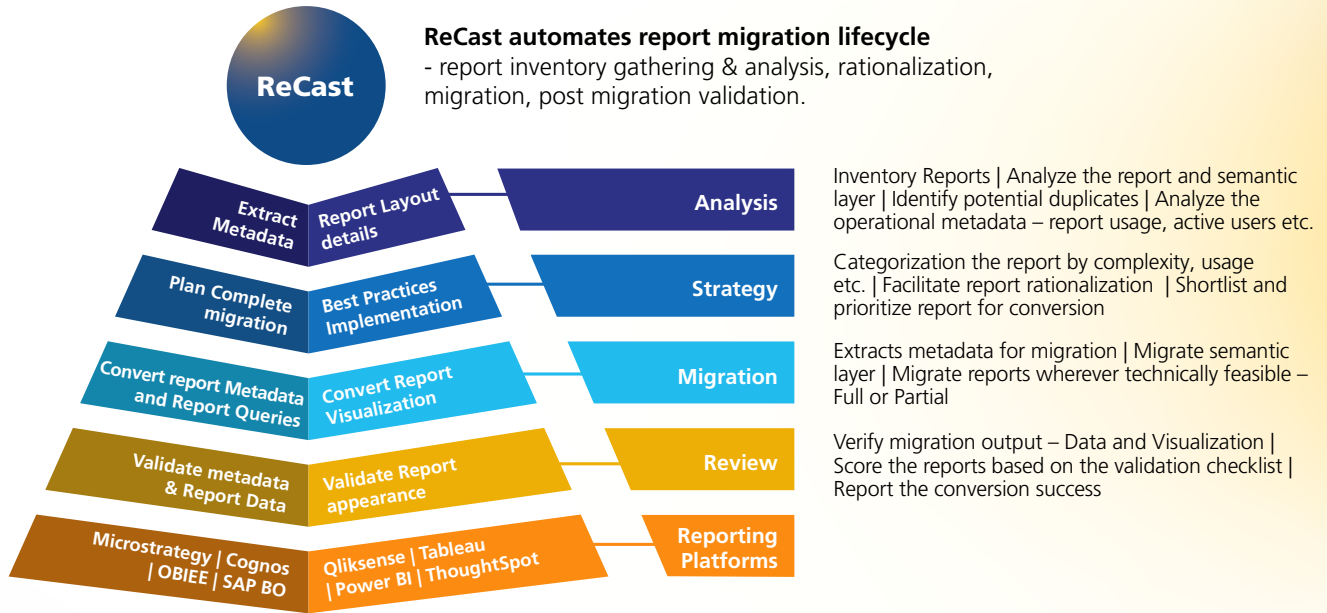
■ ReCast Migrator

Automate migration of reports, dashboard, and associated objects to new BI tool.

■ ReCast Reviewer

Validate and assure the quality of the migrated reports and dashboards





ReCast lifecycle

Standard BI to AI-powered BI migration

Recently we have observed a lot of traction for ThoughtSpot as the target BI system for migration. With all the emerging trend, a problem arises with the speedy migration of traditional BI tools to NLP-based BI platforms. To address this, we have worked on automating the migration process for Tableau to ThoughtSpot.

Migrating Tableau to ThoughtSpot Workflow

Between the source and target system, there are different layers, and each layer accomplishes unique operations using various technologies. The source adapter is used by Tableau Analyzer to connect to Tableau, which helps in extracting the metadata. This is used by the Modernizer Engine to convert into equivalent ThoughtSpot objects. The Migrator Engine, with the help of the Target Adapter, consumes the converted objects and imports them into ThoughtSpot.



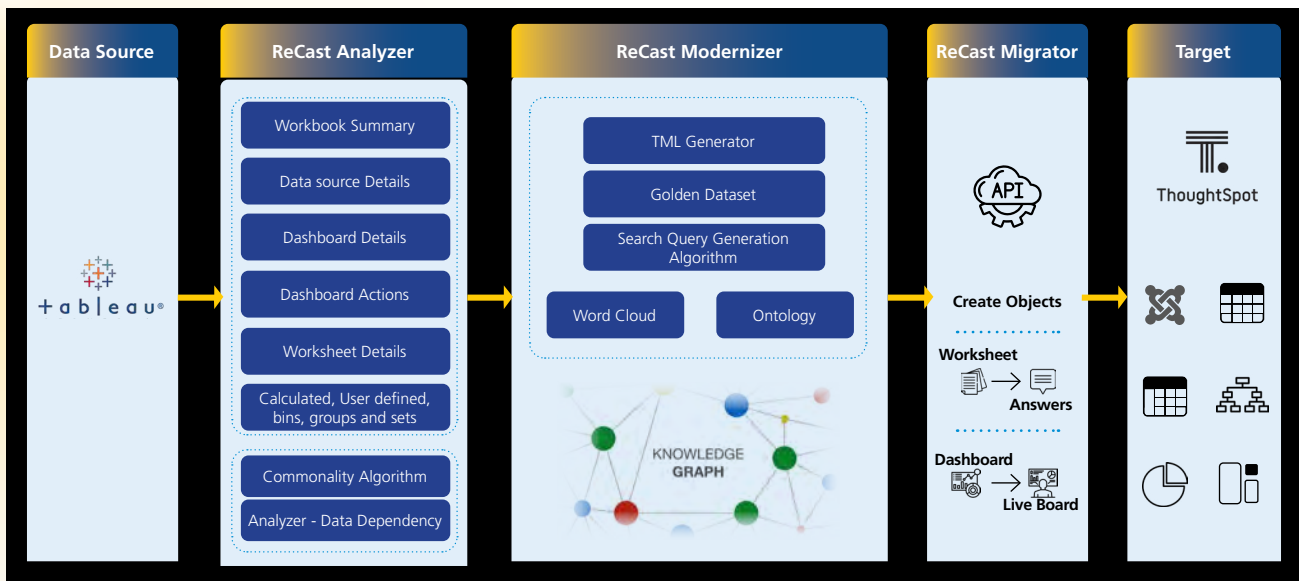
Source Adapter: Connects to Tableau.

Analyzer Engine: Extracts metadata information for analysis and runs the algorithm against objects present in the model. The extracted information is stored in the in-built ReCast metadata tables for further analysis and migration. This information is used by graph DB to generate an object data dependency chart.

Modernizer Engine: The extracted metadata and data dependency chart help us to migrate and rationalize the workbooks. The Modernizer Engine generates function mapping documents using a lexical algorithm, visual mapping documents using a visual conversion algorithm, and position conversion for answers. It uses ontology, word cloud, and data dependency chart to create a golden dataset to give the best search experience.

Migrator Engine: The inputs from the Modernizer Engine are consumed to build the objects for ThoughtSpot. With the help of the Target Adapter, the objects are then imported into ThoughtSpot.

Target Adapter: Connects to ThoughtSpot by API.



Source to target end to end flow



ReCast Workflow

Tableau Analyzer

The Tableau Analyzer module, which is a component of the ReCast Migration tool, performs a comprehensive analysis of Tableau reports available on the server. It extracts report metadata, such as report objects, dimensions, measures, column instances, and worksheet details, as well as data source information, including tables, filters, dimensions, measures, and folders. By leveraging a similarity engine algorithm, the module identifies potential duplicates, allowing businesses to reduce redundant reports.

Moreover, the Tableau Analyzer module provides additional value to businesses by categorizing reports based on complexity as simple, medium, and complex, streamlining the reporting process. The insights generated by the module can be used to optimize report creation and management, ultimately leading to enhanced data-driven decision-making. Overall, this is a powerful tool that provides significant value to businesses that utilize Tableau for their reporting needs.

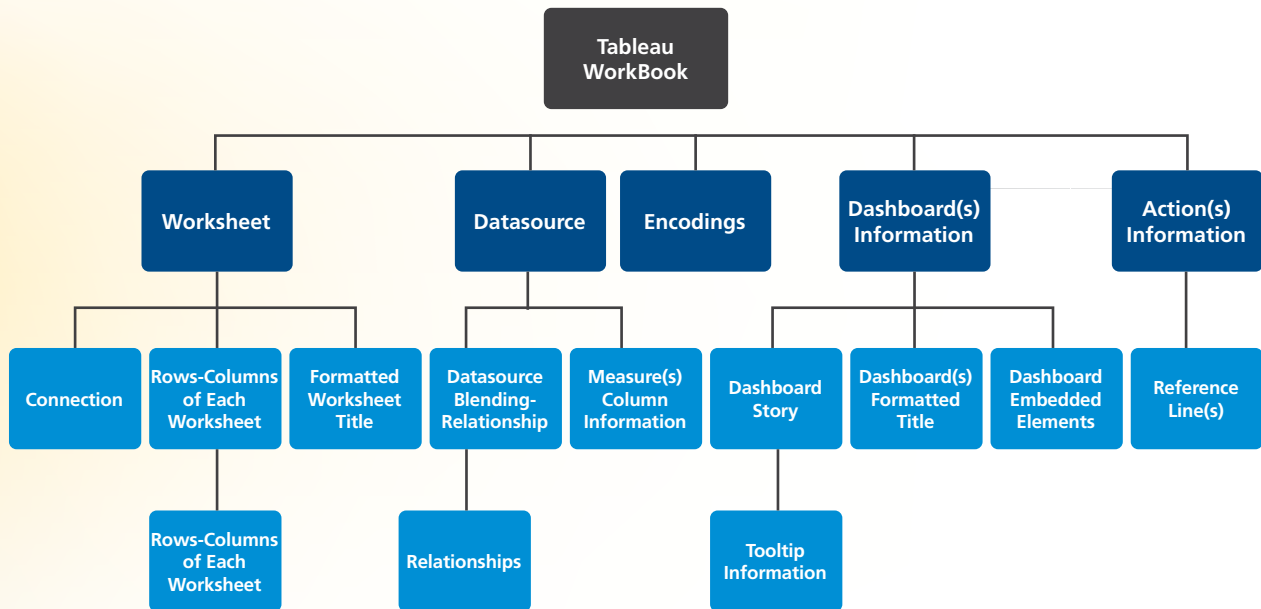


Tableau Analyzer output



Commonality detection algorithm: Finds the set of all similar reports in various file paths across the project. We developed our state-of-the-art algorithms based on data source tables, columns, filters, visualization, and GUI filter to rank report commonalities with clusters.

Complexity analysis algorithm: Complexity is attributed to a report on various factors like multiple data sources, multiple SQL, complex matrix calculation, numbers of visualizations, and many other factors. A customizable template allows tuning the complexity calculation as per the project requirements.

Variable dependency analysis: An introspection of data model variables dependency shows the hierarchy and pinpoints the excess variables and unused variables.

BI element lineage: A lineage model of semantic data layer - reports, tables, columns, matrix, SQL, visualization, etc., on a graph database gives clear visualization of the re-engineering strategy of the BI landscape.

ReCast data dependency workbench

A tool for data model review, rationalization, and corrective action. It is a web-based work area for data landscape view.

Key features

Centralized view: Organizations use multiple BI tools for various historical reasons. Though modern BI tools give the lineage of their own BI components, but the consolidated view is not possible for multiple BI tools. In a data dependency tool, we can visualize your single or multiple BI data models in unified location (SAP BO, Tableau, Cognos, MSTR, etc.). The entire BI landscape is visible using this tool.



Segmented view: A segment-wise view of BI objects is also possible from this tool. Users can see the report to table to column relationships, among others.

- Report-> table-> column
- Report -> query-> measure-> calculated measure
- Report->visualization-> filter (Dynamic)
- Commonality cluster: Similar reports based on tables

Gap analysis

- A gap analysis is a method of assessing the loopholes of a BI architecture or design limitation of existing BI tools. It helps you find unused and duplicate reports and variables based on usage.
- A gap analysis may also be referred to as a needs analysis, needs assessment, or need-gap analysis. Users can utilize the lineage topology to find out the new areas reports density need to improve.

BI recommendations

- Optimization suggestions based on an existing data model.
- Golden dataset: A golden dataset is a validated and integrated data set that is properly annotated without bias. It's simply a centered dataset used to develop various reports to ensure they contain the same data and are processed the same way. Though it is commonly referred to as hand-labeled, making it very high-quality data. But our tools' inbuilt algorithms will find out the optimum columns used for all reports, hence a golden dataset will be generated by the recommendation engine.
- A needs analysis, needs assessment, or need-gap analysis always shows the BI topology and BI report density as per domain. It helps to recognize the custom change request on the upcoming BI model and integrate it with the existing data model.



ThoughtSpot Modernizer

ThoughtSpot Modernizer consumes the data output from Tableau Analyzer and data dependency chart and converts the objects from Tableau to equivalent ThoughtSpot objects.

Modernizer uses ThoughtSpot Modeling Language (TML) approach to create objects like connections, tables, worksheets, answers, and live boards. TML is a Yet Another Markup Language (YAML) based language that ThoughtSpot uses to build and represent its objects.

Data source details extracted from Tableau Analyzer are used to create connection, tables and worksheets in ThoughtSpot. Workbook detail extracts are used to identify the sheets and visualizations and are migrated into the best possible equivalent visualization in ThoughtSpot Answers. The metadata of sheets in respective dashboards in Tableau is used to pin equivalent answers to live boards in ThoughtSpot.

Output from data dependency chart will be helpful in rationalizing the reports. This data, combined with ontology data, creates a golden worksheet that will boost the search experience of the end users.

ReCast Migrator

The objects created based on the TML structure are imported into ThoughtSpot for individual objects. The migrator uses ThoughtSpot's REST API SDK to import the objects into ThoughtSpot with the help of the Target Adapter.



ReCast Reviewer

Validate BI migration through this review and validation framework that measures the success of automation. The discrepancy, limitation, and exception post migration will be gauged, and corrective measures will be prescribed. A report-wise analysis will help to determine the success of the automated migration.

PDF diff: Compare PDF-like source BI to target BI report.

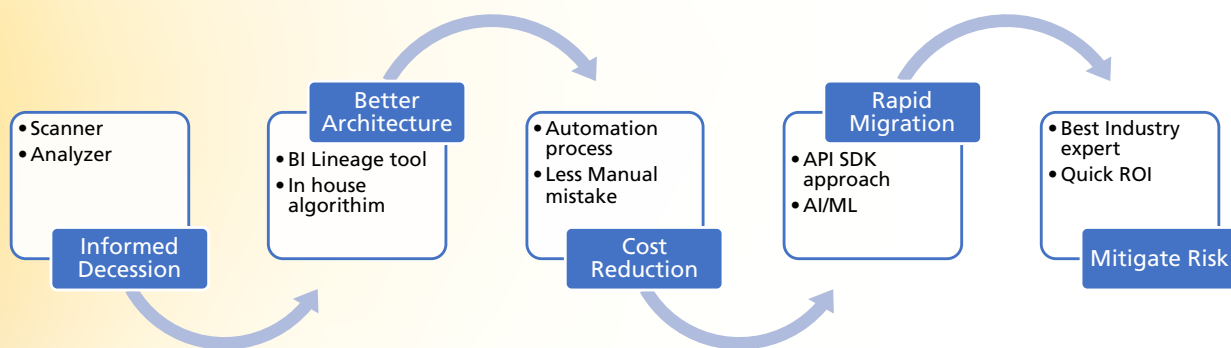
XML diff: Compare XML data, especially the metadata or GUI data extracted by BI tools.

Excel diff: Excel output for BI metadata will compare across multiple BI domains.

LTIMindtree clients who leveraged ReCast for BI migration were able to reduce the project risks associated with effort and schedule overrun. They also benefited from a 30% acceleration in time-to-market compared to the manual approach.

Key features

- Informed decisions: ReCast tools guide the right decision for the migration approach.
- Better architecture: Remove previous mistakes, and ensure better BI architecture.
- Cost reduction: Reduce the manual efforts and costs in BI migration through an automation approach.
- Rapid migration: Faster time-to-market with improved efficiency and automation.
- Mitigate risk: De-risk migration with key guiding principles, proven best practices, and automation.



ReCast Reviewer workflow



Feature conversion/comparison

The table below provides the feature conversion mapping while migrating from Tableau to ThoughtSpot.

Object	Tableau	ThoughtSpot	ReCast Migration
Connection	Supports live and extract mode	Supports live mode as queries are directly pushed to the data source	Extract mode gets migrated as live mode
Connection	Supports flat files, cloud data sources, and on-premise data sources	Supports popular cloud data sources and few on-premise data sources	Supports cloud data source migration
Table	Supports data blending, table joins, and unions across different data sources	Supports table joins across the same connection	Supports migration of table joins from the same connection
Table	Supports modifying column properties at the workbook level	Supports table and worksheets with semantic layer to modify column properties	Supports migration of column properties to worksheets for individual workbooks in Tableau
Visualization	Graphs are represented in sheets	Graphs are represented in answers	Supports migration of all the basic visualization, along with advanced visuals supported by ThoughtSpot
Reporting	Dashboards contain multiple sheets. Supports dashboard actions and custom-developed extensions	Live boards contain multiple answers	Supports migration of dashboards with sheets. ThoughtSpot supports dashboard actions and extensions.
Reporting	Story	Multi-tab live boards	Supports migration of story into live boards with tabs

Keeping up with the current market requirements, LTIMindtree's ReCast has added support for migrating from traditional BI tools to AI-NLP-based tools. ReCast has emerged as a tool by following the status quo. It is continuously evolving by incorporating and supporting additional features for a smooth migration process and has a lot of potential for more growth.



About the Author(s)



Sheba Qureshi, Specialist Data Scientist, has more than eight years of experience in various domains like retail, CPG, BFSI, and travel transport & hospitality. Worked majorly in building ML-based solutions, including data analysis, predictive analysis, statistical analysis, and data visualization.



Ramanath KP is a software engineer with 2+ years of experience in Java, JavaScript, Angular, Spring Boot, and REST API development. His expertise in these technologies allows him to create efficient and reliable software solutions that meet the needs of clients and users alike.





*Let's get to the
future, faster. Together.*

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700 clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by nearly 82,000+ talented and entrepreneurial professionals across more than 30 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen and Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit www.ltimindtree.com.