CASE STUDY

UNLOCKING THE VALUE OF DATA: GSTN ADOPTS ADVANCED ANALYTICS TO IMPROVE TAX COMPLIANCE

Infosys®
Navigate your next

Goods and Services Tax Network (GSTN) was incorporated on 28th March 2013 with a mission to implement a new Indirect Taxation regime, Goods and Service Tax (GST), across India. GST taxation was implemented to provide standard indirect taxation across states within India. Since the new GST regime came into effect from 1st July 2017, GSTN has brought its 1.3 crore taxpayers on the new taxation model. To support massive transformation involving taxpayers, GSTN partnered with Infosys to build a highly scalable and responsive application platform that serves the needs of Tax Payers, Tax Officers, and other stakeholders.

Phase one and two of the GST system implementation was focussed on building the core functionalities needed by taxpayers and officers. With the core functionalities in place, GSTN's next focus was to leverage the data available to generate actionable insights using a combination of BI tools and AI/ML-based models to help improve compliance, detect/prevent fraud, and support policymakers. Phase-3 of GST system implementation was aimed to design and develop an advanced analytics platform for Central and State Tax Authorities and policymakers.

GSTN formed a Business Intelligence and Fraud Analytics (BIFA) unit in March 2019, responsible for the execution of GST Phase-3. Infosys partnered with the BIFA unit of GSTN to deliver the goals set out for the Phase-3 of the programme.

The Objective

The key objectives of this program, as listed below, were aimed to empower policy decision-makers, operations & the compliance team with real-time actionable insights:

- Detection of tax evasion
- Revenue assurance
- Creation of statistical insights for policymakers
- Provision of value-added services for Taxpayers

> **"Before we started on the BIFA journey, we attempted to study the analytics capabilities available to tax officers in different countries across the world, to understand the best practices from each. We wanted BIFA to be the best analytics and fraud detection platform in the world from a tax compliance perspective. However, very little is available in public domain and hence we had to rely on knowledge of our investigators and use modern AI/ML Models on the vast amount of data we have."**
>
> **- Prakash Kumar, CEO, GSTN**

Leveraging the data mine: Adoption of advanced analytics to gain actionable insights

With over three years of operations, GSTN had accumulated a large amount of data related to taxpayer demographics (~13 mn), invoices (~9 bn), eWay bills (~1.42 bn), tax payments (~150 mn), and the tax returns filed (~530 mn). Data from external sources such as Customs, Direct Taxation system were also available for deriving insights.

Infosys formed a core team to strategize, architect and implement a Data and Analytics Platform to meet the objective set out by GSTN for this initiative.

The Solution: Build a Data and Analytics Platform with a foundational enterprise data lake and advanced analytics capabilities

A scalable **Enterprise Data Lake** was setup leveraging open source components that sourced data and made it available in easy-to-access formats on the data lake platform. The Infosys Genome Solution was leveraged to optimize the underlying data layer and created a unified user access portal for the users to consume, based on their entitlements. A robust access control process was implemented to ensure compliance with data security and privacy guidelines.

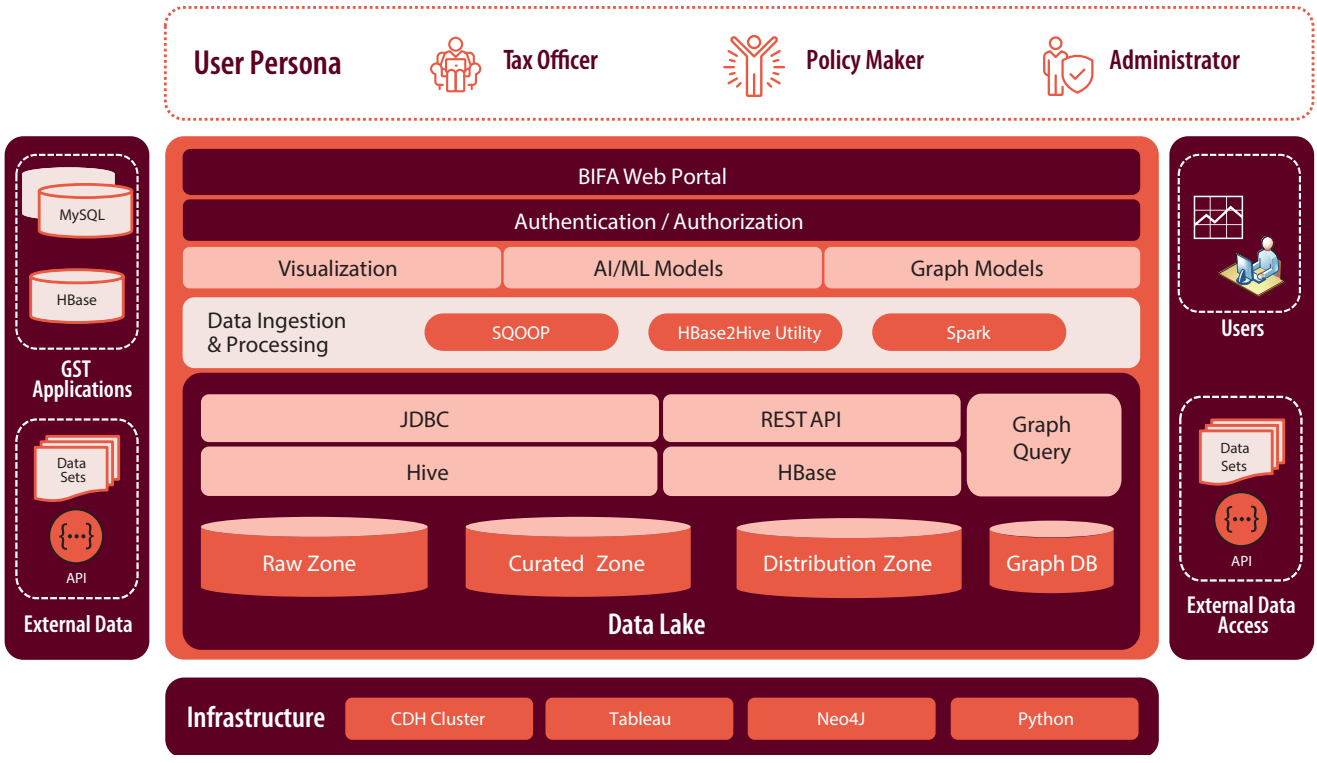


Fig 1: Data and Analytics Platform

Implementation scope involved:

Development and construction of genome data models to support analytics use cases

Migration of approximately 100TB of historical data from over three years from GST's application data stores to a data lake

Setting up of graph database with 200 million nodes and two billion edges that represented the taxpayer ecosystem data

Implementation and setup of daily data refresh from GST application's sources to the data lake and automation of the daily batch process

Development and deployment of analytics use cases

Uncovering data patterns using Advanced Analytics

Once the data across internal and external sources were made available on the data lake, the focus was to identify analytics use cases that could utilize the data and deliver insights to GSTN for any action.

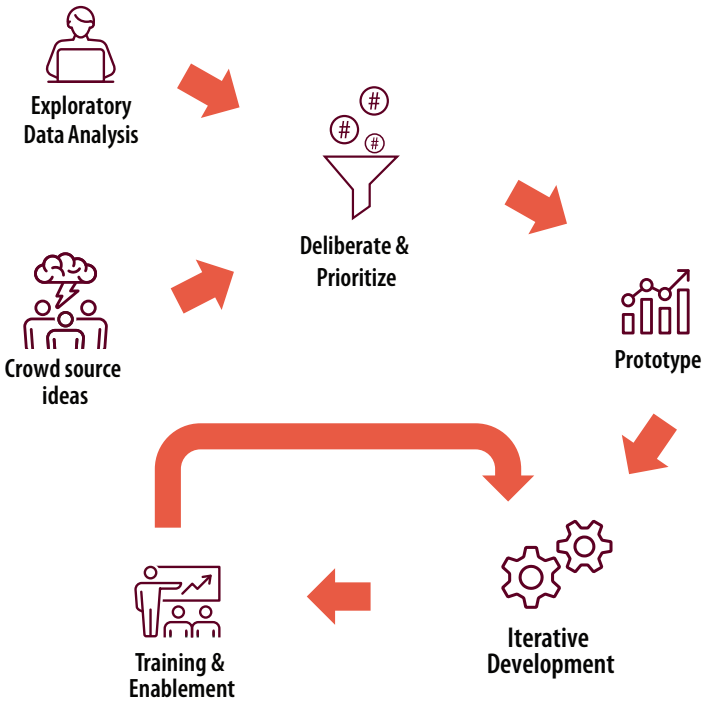
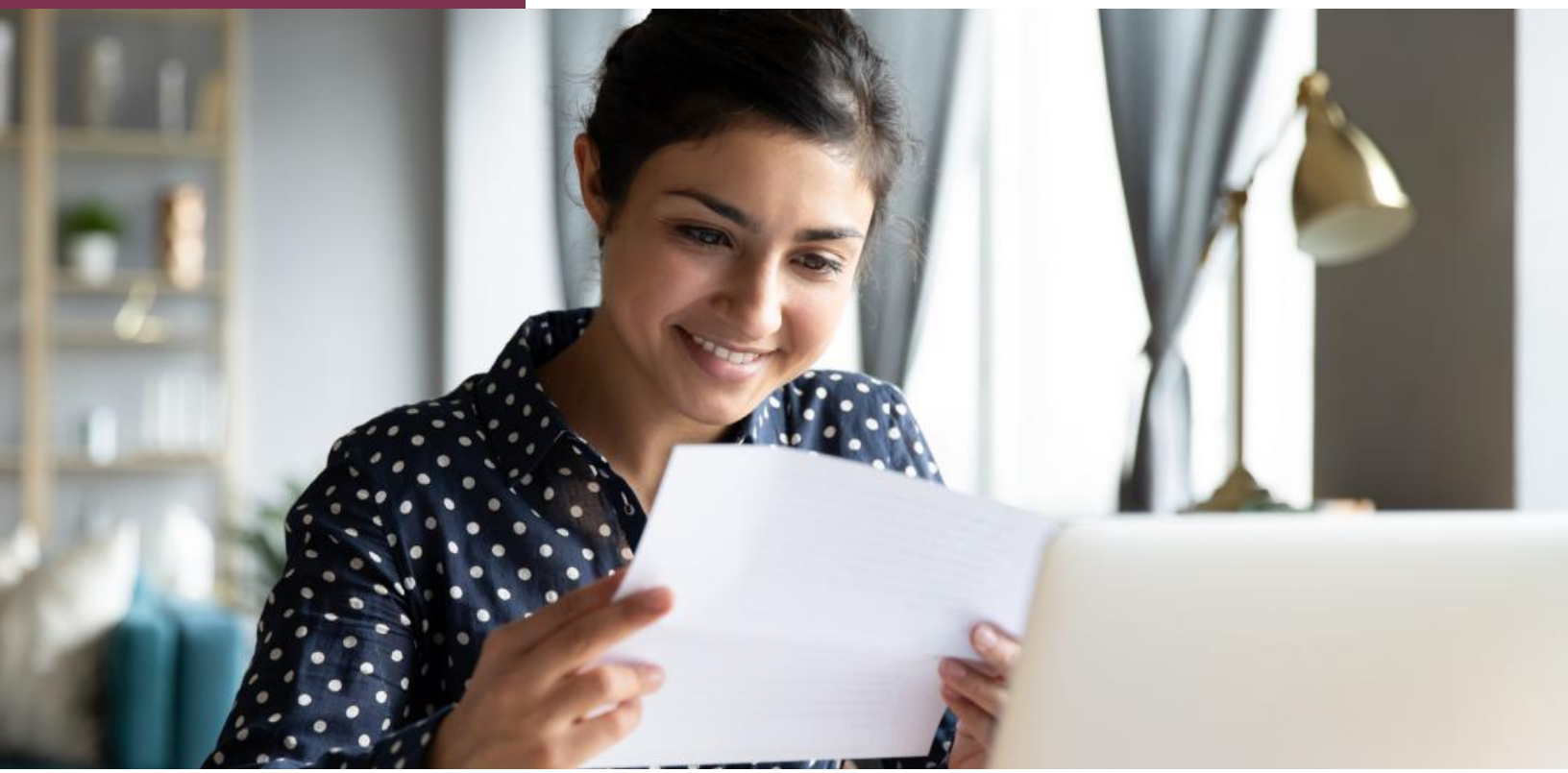


Fig 2: Analytics Use-Case Development Cycle



Crowdsource of Ideas

Ideas/use-cases were gathered from Tax/Enforcement Officers, Tax Practitioners, policymakers, which were taken up at subsequent stages of refinement and prioritization. Another method used was to interrogate the experienced investigators to find out heuristic approach used by them to identify fraudsters or malpractices.

Exploratory Data Analysis

EDA was conducted on the raw data from upstream GST application to identify data patterns indicative of fraudulent behavior.

Deliberated & Prioritized

All ideas were deliberated within GSTN's BIFA team, and with due consultation from other stakeholders like policymakers and enforcement officers, wherever required.

Prototype

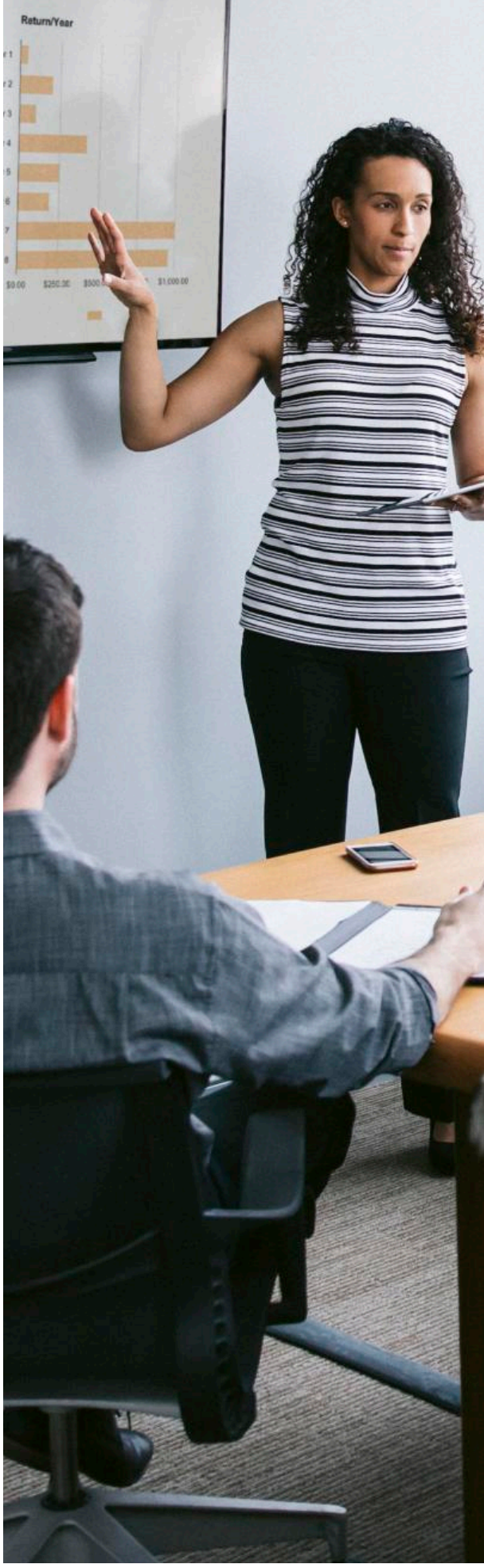
Prioritized use-cases were then prototyped with data from the Data Lake to assess the feasibility and extent of the impact.

Iterative Development

The use-cases were built iteratively to include features based on feedback and results from the prototype.

Training & Enablement

Users across states and the center were enabled through training programs for each release cycle.



The outcomes

Detection of Tax Evasion

- **Reconciliation:** Identified mismatches between returns filed by taxpayers and identified potential frauds
- **Anomaly detection:** Used AI/ML techniques to identify potential fraudulent transactions
- **Outlier Analysis:** Used un-supervised learning techniques to identify outliers within similar peer-groups
- **Community Detection:** Used graph algorithms, identified communities of taxpayers involved in fraudulent activities
- **Fraud Propensity Assessment:** Leveraged AI/ML techniques to perform Fraud propensity analysis of individual taxpayers & their transactions

Revenue Assurance

- Enabled early warning signal using AI/ML techniques to identify potential tax payment default
- Developed models to help identify entities that were not covered under GST
- Enabled Tax filing compliance analysis

Enabled Policy-makers with statistical insights

- **Sectoral Analysis:** Provided comparative and trend analysis of turnover, tax and other elements across different Geography/ Industry Sectors/ Products/Services
- **Import/Export Analysis:** Provided comparative and trend analysis of import and export across various Geography/ Industry Sectors/Products/ Services
- **Supply Chain Analysis:** Provided view of the flow of different commodities from manufacturing/import till consumption/export.
- **What-If Analysis:** In the process of developing tools to help policymakers in evaluating the impact of tax rate changes on tax collections

Value-added services for Tax Payers

- Know your business: We are currently in the process of enabling taxpayers with a 360-degree view of their business, insights on relative performance with their peer group, trend analysis of financial parameters etc.

> Within three months of launching the BIFA tools, tax officers across the country have been able to uncover 50 million USD of fraud and have initiated the due process for recovery for the same. The partnership with Infosys was a critical factor in the success of the BIFA where it brought its seasoned team of data scientists and its platform. Infosys brought in people with required tech acumen and expertise. What impressed me most about the Infosys partnership is the collaborative working where they not only did the development but also brought ideas on new models and use cases.

- Prakash Kumar, CEO, GSTN

For more information, contact askus@infosys.com

Navigate your next