THE DATA-POWERED ECONOMY: MAKING TAXATION SIMPLE

Tax fraud, legacy tax policies and operational issues reduce India's GDP by as much as 1%. Combining real-time insights with artificial intelligence systems can uncover nefarious tax activity, drive new tax policy and help SMBs – the engine of the economy – grow richer in these trying times.

Being able to track information has always been a taxman's dream. Ensuring that money owed is money paid can save governments across the world billions of dollars. Worldwide, the International Monetary Fund (IMF) estimates that $650 billion annually is lost due to tax evasion, with about a third of that figure related to developing countries.[1]

In this complex tax environment, timely use of data is the new oil. If governments understand what is being sold and where, and at what price, they can recoup much of the tax owed. They can also ensure that struggling businesses are given the right amount of support through new tax policies. Further, a system that collates information on commodities across geographies can be monetized, with data about consumption and generation of goods at a zip-code level helping budding entrepreneurs build new businesses, further benefiting the state.

The solution is to make taxation simpler and data driven at every level

But this isn't easy to do. The sheer amount of data being produced by businesses is increasing exponentially. Firms are now part of a wide ecosystem of partners, suppliers and buyers. Supply chains, and the goods and services they create, are subject to fierce headwinds from COVID-19. Collating, analyzing and democratizing nuggets of information from such disparate systems is a headache for even the most talented data science and policy governance gurus.

The solution then is to make taxation simpler and data driven at every level. Reduce the amount of bureaucracy, and collect data together in one repository, developing insights on goods

and services imported, exported, manufactured and consumed outside and within jurisdictions. With this sort of intelligence, governments and industries can crack down on and deter fraud, increase revenue assurance, and benefit the wide-ranging small and medium-sized business entities across the world that struggle to understand how best to buy, sell and grow in these trying times.

Though such a solution can be used in any economy, the focus of this paper relies on one particular case study taking place within India. More particularly, it homes in on the new goods and services tax (GST) reform, introduced in July 2017, and the development of the goods and services tax network (GSTN) that was created to implement the reform.

GSTN – Making data work for good

Before the GST reform, if you wanted to buy a certain laptop in India, the tax paid in Calcutta would be very different to that paid in Bangalore. Different taxes were paid on manufactured goods, services, sales, imports and exports. Further, some types of taxes were collected by the central government, and some by the state.[2]
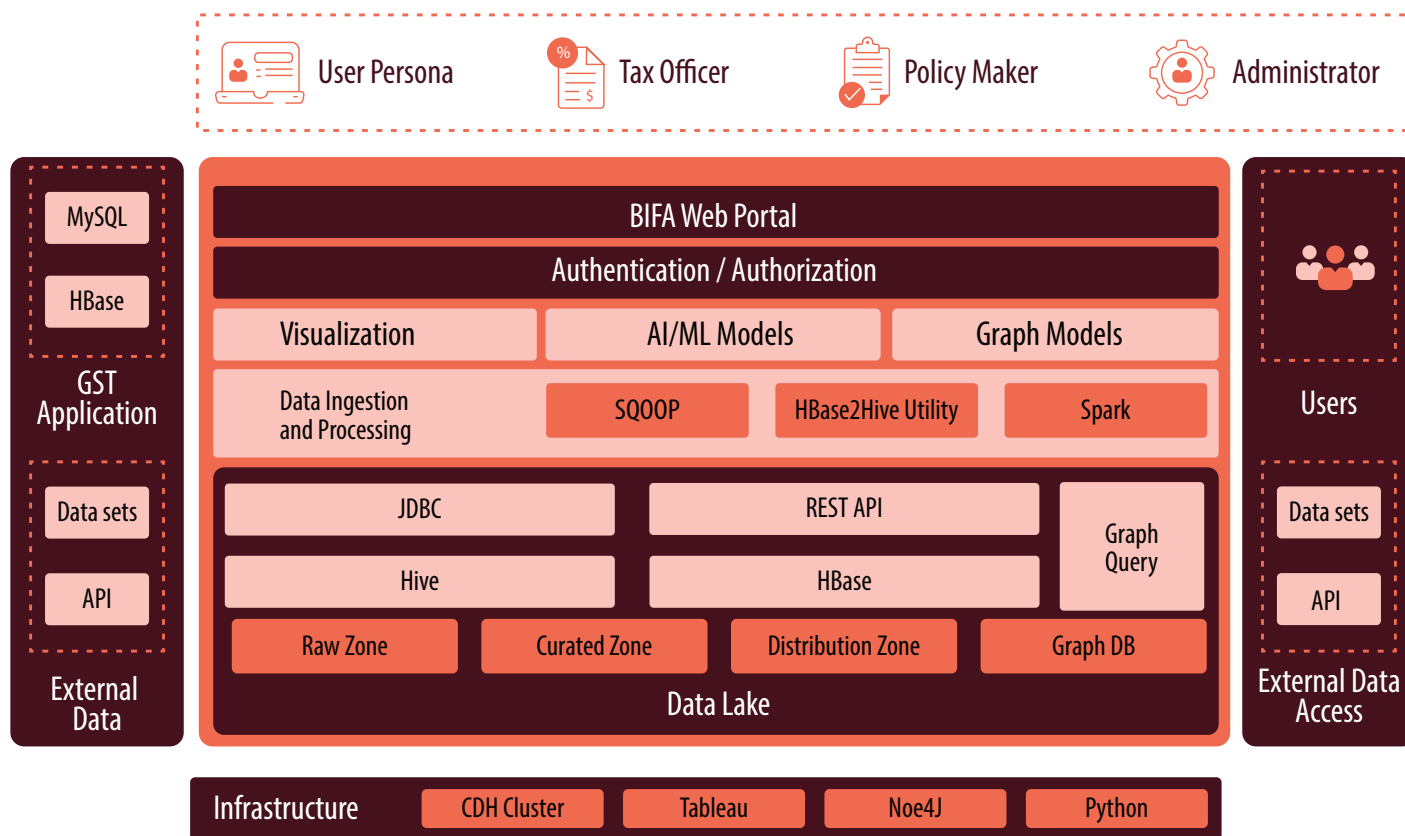
This complexity cascaded on the consumer, who lost out due to a lack of purchasing power and investment potential. For their part, state governments would lose money, since they didn't have the right to impose service taxes within their jurisdictions. Lack of robust information gathering led to lots of discrepancies, which allowed fraudulent activity to flourish. Further, tax data was housed in many different systems, in silos that made real-time information transfer between states and central government next to impossible.

But when the new tax regime was introduced in 2017 – known as "One Nation, One Tax" – tax laws were standardized for almost 100% of goods and services. Any given commodity was taxed the same right across the country, and the number of different tax rates was reduced significantly (from over 15 to just five). By optimizing efficiency and equity, while bringing tax GDP growth under a common market, India's stock in the world crept up, even as Moody's Investors Service said that the new regime would increase government revenues through improved tax compliance.[3]

Within this new paradigm, an outfit called GSTN was created. Infosys was brought in to help GSTN create the core infrastructure for the new tax network. Such infrastructure involved building a new digital platform to support such a massive tax reform. About six million taxpayers and their opening balance information was migrated to this platform as part of the transition to the new GST tax regime. Over the next two years (2017-2019), the taxpayer base increased to 13 million, the number of invoices reached 9 billion, and 150 million tax payments along with 530 million tax returns were recorded in the system.[4]

This migration made up phase one and phase two of the project, with the digital platform used by both taxpayers and tax compliance officers. Then, with phase three, started in 2019, the focus shifted to creating actionable insights from this data. A Business Intelligence and Fraud Analytics (BIFA) unit was formed to make this vision a reality across a number of data analytics use cases. Infosys created an analytics and AI/ML platform with a foundational data lake which housed the entirety of the data processed in phases one and two (around 200 TB), along with data from other external systems like Indian customs (Figure 1).

Figure 1. The analytics and AI/ML platform used by GSTN

User Persona	Tax Officer	Policy Maker	Administrator

GST Application
- MySQL
- HBase

External Data
- Data sets
- API

BIFA Web Portal

Authentication / Authorization

Visualization	AI/ML Models	Graph Models

Data Ingestion and Processing | SQOOP | HBase2Hive Utility | Spark

JDBC	REST API	Graph Query
Hive	HBase	

Data Lake

| Raw Zone | Curated Zone | Distribution Zone | Graph DB |

Users

External Data Access
- Data sets
- API

Infrastructure | CDH Cluster | Tableau | Noe4J | Python

Source: Infosys

These analytics use cases were three-fold: (1) to stop tax fraud occurring, thereby improving revenue assurance at the government level; (2) to help government policymakers design new policies that benefit the Indian economy; and 3) to use this trove of information to help SMBs increase revenues through greater visibility into their supply chains and working capital.

> The data platform was used to stop fraud, design new tax policies and help SMBs grow in trying times

Wrestling with tax evasion

The tax evasion use case was the first to be tackled by the new BIFA unit. Housed with data scientists and AI experts, new AI and machine learning tools were developed to unearth nefarious activity occurring throughout the country. The AI/ML system was able to uncover:

- **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 unsupervised learning techniques to identify outliers within similar peer groups
- **Community detection:** Used graph algorithms to identify communities of taxpayers involved in fraudulent activities
- **Fraud propensity assessment:** Leveraged AI/ML techniques to perform fraud propensity analysis of individual taxpayers and their transactions

GST-related arrests are happening every week now in India, a testament to how good the new system actually is. In August 2020, in Odisha, the system was used to bust a GST invoice racket allegedly involving Rs 712 crore ($9 billion).[5] One man was arrested in October by GST officers in Gurugram on charges of creating and operating fictitious firms on forged documents and passing fake tax credit to the tune of over Rs 392 crore ($5 billion).[6]

Giving policymakers statistical insights

For most movement of goods, the information is captured by the GSTN system. Likewise, any business-to-business transaction is also captured, along with quantity, price and other

anonymized demographics. For policymakers, this truly is a gold mine. A government official – the Delhi Finance Minister, say – can use insights captured to propose new legislation that keeps revenue generation in his or her state. Because the official knows what goods are being manufactured, where and by whom, along with details of consumption, the official understands what sectors are doing well, and which need more help.

These insights can be offered as a trend across multiple financial quarters, say, with deeper dives into how well each jurisdiction is doing across certain commodities. Because commodities and services are grouped to form a hierarchy of commodities, products and sectors, a policymaker can carry out analysis at any level of this hierarchy, potentially using this data to propose initiatives that would otherwise be missed.

> The BI platform supports initiatives like "Self-Reliant India", making sure products and services work to lift the Indian economy

For example, the data might show that 90% of electronic components consumed in Delhi are imported from China. The policy would then encourage as much electronics manufacturing as possible to occur within India. In this paradigm, insights from the digital BI platform can potentially be used to support the execution of initiatives like "Aatma Nirbhar Bharat," otherwise known as "Self-Reliant India."

Along with this sort of sectoral/import and export/and supply chain analysis, insights can be found that also help policymakers evaluate the impact of tax-rate changes on tax collections. If

ministers in Delhi, Gujrat and Kerala have different tax rates for the same commodity based on unit price, the What-If analysis would arrive at a revenue neutral tax rate (taking into account shipment volumes, declared liability, realized cash, etc.) that increases operational efficiencies without compromising on tax revenue in these regions – a precious insight.

Growing business through a data marketplace

There are about 13 million registered business taxpayers in India. Out of these, 70% are micro business and SMBs with turnover of less than 10 million rupees. Such businesses need help. They don't have the money to spend on sophisticated ERP systems from the likes of SAP or Oracle. From the new GSTN system, the BIFA team can act as a source of valuable insights for these smaller players, giving them live data on not only GST details but overall sale and purchase trends with industry-specific trend analysis. In short, the system is acting as a data marketplace for trend analytics, helping businesses to invest in the right way, ensuring their supply chains are more resilient and grow in troubled times. Further, with government analyzing this sort of data in detail, fraudsters are less likely to act.

A government economy powered by data

Of course, other use cases for the system are possible. While tax evasion is a compelling use case, the GSTN tools and techniques used here are applicable to other domains, including reducing fraudulent insurance claims, money laundering and the use of shell companies in circular trading (to inflate revenues and thereby get better credit).

Revenue assurance can also be amplified. AI and ML models can

be used across the data lake and processing funnel to identify potential default on tax payments. The system can also be used to develop models that help identify entities that are not even covered under GST.

Further, the system and AI/ML models can be used to model the impact changes in one sector will have on another. For instance, if we know that in October 2020, tax collection in manufacturing increased significantly, authorities can use this to predict that CPG firms are likely to have increased collections in January 2021. This increases GDP resilience and ensures markets are more robust while reducing compliance costs.

This solution is of course applicable to governments the world over, increasing tax compliance and using data for the social good to help SMBs when money runs dry. With COVID-19 forcing businesses to shut and as federal loans put an exacting demand on government coffers, using a data-driven tax system with central authority can ensure money is kept within national borders, consumers get the full benefit of goods and services tax, and businesses are given a lifeline just when they need it.

References

1. Countries versus corporations: the great global tax race, John Plender, Jan 14, 2020, FT

2. A Comprehensive Analysis of Goods and Services Tax (GST) in India, Anand Nayyar & Inderpal Singh, Feb 2018, Research Gate

3. GST will boost revenue, positive for Indian credit profile: Moody's, July 2, 2017, Business Standard

4. Goods and Services Tax Network

5. GST fraud Odisha tax officials bust Rs 712 crore GST fraud, arrest one person, Debabrata Mohanty, Aug 19, 2020, Hindustan Times

6. Tax fraud: Delhi man held on charges of forging 392cr input tax credit, Oct 30, 2020, Mint

Authors

Prince Thomas

AVP – Infosys

prince.thomas@infosys.com

Harry Keir Hughes

Senior Consultant – Infosys Knowledge Institute

harrykeir.hughes@infosys.com

Infosys® | Knowledge Institute

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision making on critical business and technology issues.
To view our research, visit Infosys Knowledge Institute at infosys.com/IKI

Infosys®
Navigate your next

For more information, contact askus@infosys.com