GST System Architecture and Rollout Strategy

1. **Vision**: To provide one online digital platform to taxpayers for GST compliance (Registration, Payment of Taxes, Filing of Returns, Refund, Appeals) and to tax officers for performing all statutory functions (Processing Registration & Refund applications, Appeal handling, Assessment).

2. Key Objectives :

- a. One platform for all statutory compliances.
- b. Easy integration with applications of other departments/agencies/banks etc.
- c. Paperless processes in GST and minimal human interface of taxpayers with tax officers.
- d. Assignment of multiple roles and tasks to a single tax officer based on role based workflow for approval.
- e. Allowing taxpayers to move across different jurisdictions, in case of change in location.
- f. Integrated assessment/audit-appeal process.
- g. Dashboard for taxpayer and tax officer to give them status of various items at one place
- h. Use data of GST for policy making as well as for other works like GDP estimation.

3. Architecture Principles adopted:

- a. **Platform approach**: To develop every function/module (GST Registration, Return Filing, Payment, Assessment etc.) of GST System as a service (single or multiple APIs), which can be used by GST portal as well as by third party applications. This is on lines of open scalable architecture followed by Aadhaar, Facebook, Google for providing taxpayer wider choice for GST compliance.
- b. **Openness**: Adoption of open API, open standards and Open source products (in line with Government of India's policy in this regard)
- c. No vendor lock-in and replace-ability: Design GST System using loose coupling principle, so that underlying service providing software (Database, Messaging service, File store etc.) can be replaced with minimal effort and without vendor lock-in. This was to ensure that non replicability of a component can be avoided.
- d. **Use of commodity hardware**: Use commodity grade hardware, which is commonly available with any hardware vendor, and avoiding specialized appliances & hardware as much as possible, to avoid vendor lock-in at Infrastructure layer also.
- e. **Security and Privacy**: The system ensures privacy and security of taxpayers' data both at rest (while stored in data store) and in motion (while sending data to portal or through API to third party systems)
- f. Scalability of hardware and software:
 - i. Modular and scalable design of system

- ii. Loose coupling through open stateless (i.e. request happens in complete isolation) API and messaging
- iii. Data Sharding and parallel processing
- iv. Horizontal scaling of compute, Network and storage using state of the art DevOps techniques
- g. **Manageability and Lights-out Operation**: Automated deployment of application using Continuous Integration (CI)/Continuous Development (CD) pipeline and completely automated monitoring of Application and Infrastructure for quick identification and resolution of issues thus reducing dependence on human monitoring.
- h. Reliability



GST Design Principles – Figure -1

- i. Zero loss of data (Data should not be secured once acknowledgement is given to taxpayer)
- ii. Non repudiation of information i.e. ability of system to prove legally that the information stored in system is provided by user, very similar to physically signed data.
- iii. Unauthorized access and alteration to the Data uploaded in the GST system shall be prevented.
- i. **Availability**: System should be available 99.9% for Stakeholders all the times. In case of any disaster (flood, earthquake...), it should be able to continue operation with minimum downtime from the other Data Center.
- j. **Data driven decision making**: System should aid data driven decision making by providing various dashboard and analytics to stakeholders.
- k. **Reconstruction of Truth**: In case of any corruption or alteration of data in main data store, system should be able to construct the original data.

4. Overview of Architecture

Presented here is a logical Architecture view showing various layers (database, Middleware, API, View, Processing Layer and External Integration Layer) of the GST System.



Logical GST Architecture - Figure-2

5. Design principle and Techniques used for handling some unique requirements of GST System:

During the initial days when the GST Law & Rules were getting finalised, GST Technical team focused on creating common framework like return signing framework, Authentication and Authorization framework, Encryption/Decryption framework, Common Processing Framework, Common Communication Framework, Automated Build and Deployment Platform. This helped in creating a robust backend platform/foundation, which was later leveraged by all modules of GST System like return/registration etc.. This has immensely helped GST system in rolling out functionality (Registration and Return Filing) after few days of finalization of form.

- GST System had a unique requirement of processing and storing large amount of invoice data like any large internet scale application and also keeping ledger, registration and payment data with the same consistency as bank. To address this GST System has decided to leverage two best data store to handle these specific requirement HBASE (Big Data Store) for keeping large amount of invoice data and MySQL (RDBMS) for keeping payment/ledger/registration meta data.
- To improve the turnaround time of transactions, GST System has divided operations into different independent **synchronous** (*immediate processing and acknowledgement like payment*) and **asynchronous** (*immediate acknowledgement for receiving the request and post processing like in invoice submission*) steps. For asynchronous steps, taxpayers were given feedback that their request will be processed within 20 minutes. This has helped in providing taxpayer a good experience and same time prevented system from system congestion.

GST Portal



Asynchronous Processing of Invoice Upload – Figure-3

- GST system has followed API based development to ensure flexibility and integration with other applications. This allows an external application to call the applications of the GST system with specific inputs and receive specific output, without knowing the further details or internal logic of the GST system application. For instance, the "Search Taxpayer" API expect GSTIN of taxpayer as input and provide details of the taxpayer as output.
- GST system has to share a large amount of data with various agencies, including States and Central Board of Indirect Taxes and Customs (CBIC) which have their own back-end systems for their tax officers. To address data sharing and reconciliation issues, GST System has designed a new Application Programming Interface (API) based data sharing framework from scratch, this has feature of on demand pulling, robust reconciliation like information of total records with hash of records for integrity of data etc. with every file. This framework is now getting followed by many government system requiring large amount of data transfer. For online validation and work flow based integration also GST System has used API framework with great success. Currently more than 130 entities are integrated with GST System and it is handling on an average more than 5 million API hits per day.

• One of the major theme of GST System was Taxpayer convenience and experience. To address this GST System had come up with online portal (which is accessible from both desktop and mobile) having all GST services related to GST compliance and various offline tools for preparing returns. But as there were various type and group of taxpayers like large corporates having more than 10 Lakhs invoice/month to MSME having less than 10 invoices. The companies doing manufacturing to providing service having fully automatic accounting system to MSME having completely manual operation, it was getting difficult for GST System to address varying requirement. To address this GST system has open all the functionalities of return and ledger to various software companies through APIs. This has enabled many GST Suvidha Providers (GSP)/Application Service Providers (ASPs) to provide services to various type of taxpayers and support them in GST compliance.



GST System API Integration View for

Taxpayers – Figure 4



GST System API Integration View for Tax Officer – Figure 5

- To handle increasing taxpayer base, the GST System software have been designed for horizontally scalability i.e. capacity of the GST system can be increased by simply adding more commodity hardware without making any change in Application and affecting taxpayer experience. A completely automated Development Operations (DevOps) techniques have been used to instantly deploy more resources at the time of peak load and reduce resources without any downtime. This has helped GST System to utilize its resources efficiently and provide superior user experience.
- To achieve high availability and zero data loss requirement GST System has deployed every component be it web servers, application servers, storage layer and application in high availability mode i.e. every component is deployed in pair, so that in case of failure of one , another mirror component can take up without affecting GST system. To ensure "Zero Data Loss" and continuity of business in case of any disaster, the concept of Data Center/Near Data Center (DC/NDC) and Data Recovery/Near Data Recovery (DR/NDR) across two distinct geographies have been implemented. The introduction of Near DC (NDC) & Near DR (NDR) will help in recovering data in case of sudden disaster as replication of data across two distinct geographies takes time.
- To ensure security & safety of GST Portal and the taxpayer's data, state of the art security software and infrastructure components have been deployed in all the data center. Almost all the integration with external systems have been

through a secure Multiprotocol Label Switching (MPLS) connectivity on top of it Hypertext Transfer **Protocol** Secure (HTPPS) protocol is used to securely transport data from GST System to any other integrating system. To ensure complete anonymity of taxpayer data, APIs have been designed to take encrypted data from ASP/GSP system, which can be decrypted by only GST System. A unique authentication and authorization framework for API has also been designed, which is now emulated by various GSP and Government systems. A unique custom framework has also been developed to protect the stored data from tampering.

6. **Size, nature and methods of sourcing architecture and design teams**: The architecture of GST System was conceptualized by then Chief Architect of UIDAI in Request for Proposal (RFP) of GST System. Subsequently, the architects of Goods and Services Tax Network (GSTN) have architected and designed the GST System following the Architectural principles envisioned in RFP which have been followed by the Managed Service Provider (MSP) in implementation. GST System has extensively adopted proven framework like Spring Boot, Hibernate, Spring Batch Framework, Spring API framework and various design patterns to make GST System design flexible and scalable.

7. Architecture Governance

This is ensured by presenting all the technical changes for review by the respective technical leaders to the Architecture Governance Board.

8. Role of Capacity Building

GST System brought a huge change in the way users (taxpayers, tax consultants and tax officers) used to interact with the tax systems prior to introduction of GST. Thus, capacity building exercise was undertaken much before the GST application was rolled out. For 60,000 odd tax officials, train the trainer approach was adopted under which 3000+ tax officers were trained as master trainers. They in turn trained remaining officials as well as the taxpayers and tax consultants. For tax consultants and taxpayers, mixed method of workshop/seminar, webinars and CBTs have been used.

The capacity building was facilitated by the following modes:

Face to face: For master trainers in the initial period by getting them at one location and conducting hands on training in batches.

Workshops and seminars: For tax consultants and taxpayers, large number of workshops and seminars were conducted in various parts of the country in partnership with local tax administration and industry/trade associations.

Webinars: After a year of rollout, additional training was imparted through webinars. During COVID period, webinar has been the only medium of capacity building.

Short duration videos: Keeping the taxpayers in view today, short videos of 3 to 5 minutes duration were developed on functionalities of various modules in multiple languages. GSTN encouraged others to use the videos to have voice

over in multiple languages to increase the reach. Today, all new changes are accompanied with short videos.

Learning Management System (LMS): Since beginning of the project, LMS has been provided to 60,000 tax officers.

Social Media: GSTN has used effectively social media (YouTube, Twitter and Facebook) and has dedicated channel on them. All new launches (new module or functionality or change) are disseminated through these mediums as well.

9. Challenges in implementation, and how addressed

Challenges	How they were addressed
Designing a system of an unprecedented tax regime at the scale of merging tax systems of 36 States/UTs and CBIC into a single system without any reference to start with where the architecture was supposed to be flexible, yet robust enough to handle traffic of ~1.3 crores taxpayers who file monthly returns and details of business to business invoices.	The use of open source technologies and platform design approach, enabled GST System to operate without tight integration within or external systems. The choice of technology principles, tools and architecture also provided for highly available fault tolerant (HAFT) system ensuring no failure.
Designing a system on which various external entities/systems could be connected.	This was achieved by adopting platform approach which enabled GST system to integrate with banks, Reserve Bank of India, GST Suvidha Providers, Central Board of Direct Taxes (CBDT), Ministry of Corporate Affairs (MCA) and other systems requiring interaction with the GST system.
Designing a system where requirements were changing fast on account of changes made in the law and rules	Using Agile and Minimum Viable Product (MVP) methodology to create services with standard interfaces, so that changes in one module does not affect other modules. Hence we were able to minimize the impact of change and contain rework.
Information and data security in GST system	Multi-tier design ensured that the sensitive and valuable information stays deep into the technology System. GSTN has heavily invested into state-of-an-art technologies, processes & governance framework to ensure maximum possible security to stakeholder data. "Need-to-know-need to do' principle to establish
	proper role based data access, which provides access of data only to the required stakeholder.

Training more than 60,000 tax officers in a short period of time on the new system.	All data in flight/movement is always encrypted. Sensitive data sets within GST Data system is also encrypted at rest as well as on flight. This was done by training 3000+ master trainers who in turn not only trained remaining tax officers but also the taxpayers and consultants through industry and trade associations. More than 40 short duration videos apart from detailed user manuals formed part of capacity building. Large number of webinars in various languages have been conducted. After every major release webinars are conducted in multiple languages which are also live-cast on YouTube.
Enabling large number of return filing on the last date of filing	GSTN has designed application to segregate the processing into two parts:a) Uploading of Invoicesb) Processing of the Uploaded Invoices.This will enable the taxpayer to upload data
	regularly without waiting for processing. This approach have allowed to handle high concurrency. Strategy has been devised to upload data through multiple channels like Offline Tools, API and from the GST Web Portal.
Disaster recovery capabilities	In order to handle zero data loss in case of disaster, new concept of DC/NDC and DR/NDR across two distinct geographies ensures continuity in business have been implemented.
Development of tools to detect tax frauds.	Use of <i>Artificial intelligence</i> (AI) and <i>Machine learning</i> (ML) has been made to develop more than 50 tools under project Business Intelligence and Fraud Analytics (BIFA).
	Availability of granular data (at invoice level with Harmonized System of Nomenclature) for the whole country has been a great enabler for not only developing tools for investigators but also tools for policy makers which help do what-if- analysis, forecasting and advanced analysis at commodity level comparing states/sub-states and exports/imports. Tie-up with experts from IIT Delhi helped in this designing the various BI Models.