Microsoft

Microsoft Azure Platform for GST Suvidha Providers

A guide to developing and hosting GSP solutions on Azure Platform

Prepared for

GSP Technical and Business Decision Makers
Friday, 21 April 2017
Version 1.1

Prepared by

Mandar Samant

Contributors:
Anil Kumar Sharma

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.

Revision and Signoff Sheet

Change Record

Date	Author	Version	Change reference
29-03-2017	Anil Kumar Sharma	0.1	
8-04-2017	Mandar Samant	1	
21-04-2017	Mandar Samant	1.1	Added Section Monitoring

Reviewers

Name	Version approved	Position	Date
Mandar Samant	1	Area Architect	8-4-2017
Mandar Samant	1.1	Area Architect	22-4-2017

Table of Contents

Table of Figures.

1 Introduction

Onset of GST regime in India is a critical and a very strategic transition for India in the realm of indirect taxation with far reaching impact on ease of doing business, commerce, services & manufacturing, logistics and country's economic progress in general. As the country transitions to these new set of rules, GSPs are positioned as critical levers to enable this IT enabled transitions.

GSPs have limited time to understand the new set of rules, which are evolving even as this document is being drafted, develop a solution and host and run it on a secure and scalable platform.

The purpose of this document is to help you, GSPs, understand how Azure DC platform available in India can help address most of the key challenges in building and operating a successful GSP solution.

Do note that this document is not to be read as a prescriptive guidance but as a guiding note only. Based on your choice of technology mix the components will vary to fit the needs and may not be covered in this document.

2 GST, a short primer

Goods and Services Tax (GST) is essentially one indirect tax system for the whole nation, which will make India one unified common market, right from the Manufacturer to the consumer. it is a broad-based, comprehensive, single indirect tax which will be levied concurrently on goods and services across India.

The Central and State indirect taxes that may be subsumed by GST include Value Added Tax (VAT), Excise Duty, Service Tax, Central Sales Tax, Additional Customs Duty and Special Additional Duty of Customs. GST will be levied at every stage of the production and distribution chains by giving the benefit of Input Tax Credit (ITC) of the tax remitted at previous stages; thereby, treating the entire country as one market.

Due to the federal structure of India, there will be two components of GST - Central GST (CGST) and State GST (SGST). Both Centre and States will simultaneously levy GST across the value chain. For interstate transactions, an Integrated GST (IGST) will be applicable which will be settled back between the center and the states.

Goods and Services Tax Network, (GSTN) a, non-Government, private limited company has been formed to provide the IT infrastructure to State/Central government and taxpayers. It has setup a central solution platform to technically enable the GST system including registration of taxpayers, upload/download of invoices, filing returns, State/Central Government reporting, IGST settlement etc. This platform, the **GST Platform**, has been setup from day 1 as an open API (GST API) based platform allowing various authorized parties to exchange information at scale.
The core IT Strategy for GST had been identified way back and is accessible here http://www.gstn.org/images/pdf/The_IT_Strategy_for_GST.pdf.

GSTN has further identified GSPs who can wrap the GST Platform and offer various value added services to its customers (i.e. taxpayers) and further downstream sub-GSPs or registered application service providers (ASPs).

GSTN will be setting up the GSTN portal to service the tax payers, the GSPs are to implement a similar portal albeit with more value-added services for their customers. Such GSPs who offer value added services (like Application Service Providers) can offer plain vanilla GST API wrappers or provide much richer facilities.

The following section details a high-level solution capabilities the GSP portal would need.

3 GSP Solution: High level capabilities.

3.1 Business Requirements

Below are few examples of the Business requirements for GSPs as called out by GSTN (indicative and not exhaustive):

a) The GST Suvidha Providers are envisaged to provide innovative and convenient methods to taxpayers and other stakeholders in interacting with the GST Systems for activities e.g.
   - Registration of tax payers' business entity.
   - Uploading of invoice details
   - Filing of returns
   - Apply for refunds and so on and so forth, as permitted by the GST Act and Rules

b) Provide services from a portal like the GST System Portal, with additional value add and innovative user interface that the GSP will offer to the end user.

c) Provide Mobile Apps to users that shall offer functionality like the GST portal.

d) Provide Accounting software/ERP having complete accounting features as well as functionalities like upload invoices and filing returns, matching of invoices and providing buyer options to view mismatched/matched/missing invoices in auto populated GSTR2 and further upload only mismatched & missing invoices to GST System.

e) Provide plain vanilla access to GST APIs provided by GSTN.

f) Provide enriched API like Bulk uploading of invoices, conversion of invoice data formats (CSV to JSON, XML to JSON), latest HSN code tax rate, prepare GSTR1/GSTR2/GSTR3, validation of returns for latest tax rate and amount etc.

g) GSTN will provide the API only through MPLS to the GSPs to ensure the GST system is only open to trusted/secure and registered GSP systems thus reducing threat of attacks. It will also ensure that there is no single point of failure in case of DOS type of attacks as there will be many GSPs through which our services will be still be accessible.

h) Every GSP will be provided a unique license key and ability to generate more sub license keys. Through sub license keys GSP could provide controlled access of APIs to Application providers who shall have agreements / relationships with the GSPs as their partners.

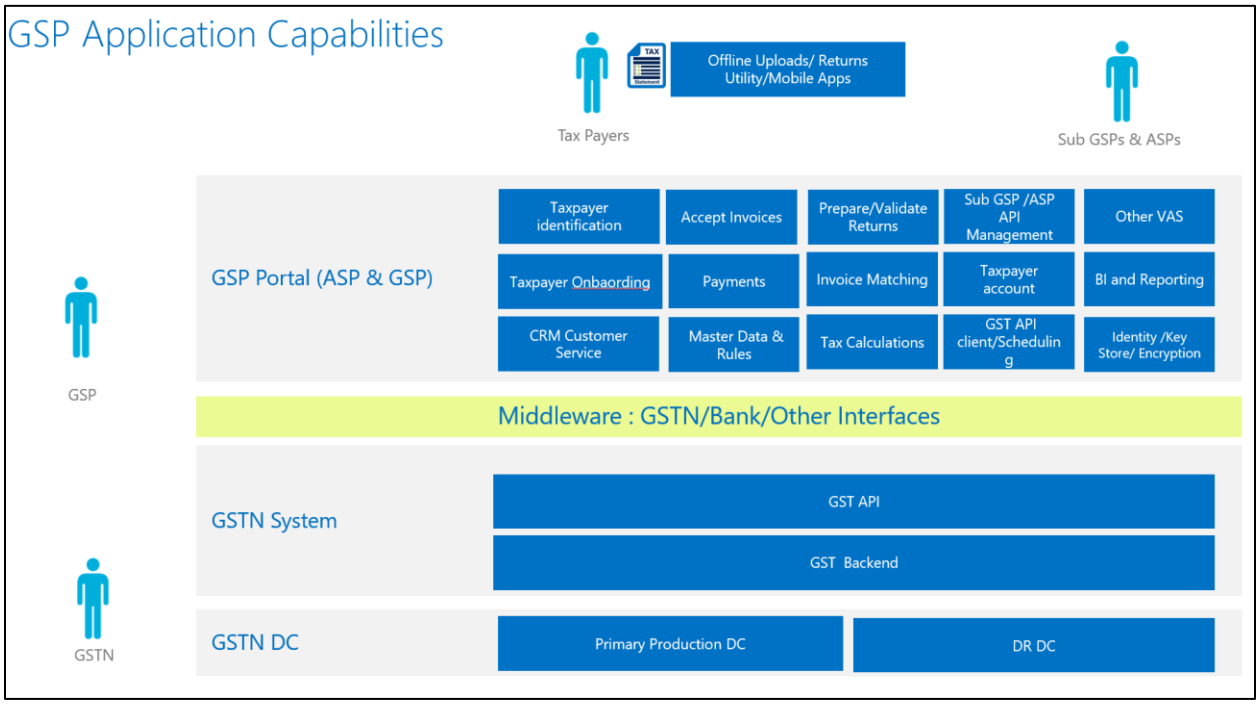

Figure 1 GSP Solution Capabilities

The figure above shows a *very high level set of capabilities* needed for the GSP application (the **GSP Portal**) and the various actors involved.

The capabilities desired are as follows:

Taxpayer identification and Onboarding:

The GSP should allow customers (taxpayer) to register and necessary workflows to authorize, validate the identity of the taxpayer and provision the tax payer in the system.

The provisioning process could range from simple setup of master data to a complex process of provisioning backend services (e.g. separate databases , encryption keys/certificate exchange, multi factor

authentication, API keys exchange, File Transfer Locations etc.) . Each taxpayer may have multiple logins and multiple roles e.g. preparer, auditor, consultant etc. who will need different kinds of access.

Once the taxpayer is onboarded they can log in to the system securely or exchange data with the GSP Portal via its APIs.

Offline Utility

The offline utility in the form of Excel or thick client executable will allow the taxpayers (esp. small entrepreneurs) to prepare their outward supplies, inward supplies, monthly and annual returns. This utility then in turn moves the data to GSP website which interacts with GST Platform.

A robust offline utility with adequate security mechanisms can allow the taxpayer to do what-if calculations and check-recheck the data with consultants/CAs before uploading to GSP portal.
This utility would need to at least maintain feature parity with GSTN Excel based utility which would be under development (http://gstn.org/images/pdf/tender/MSOU_RFP%20Volume-I%20%20Excel%20based%20offline%20utility_20170329.pdf)

Accept Invoices & Prepare Validate Returns

This Module will accept the purchase/sales ledgers from the taxpayers. This could be in form of online website modules, upload from *Offline utility*, file transfers and API calls. This module will also be responsible to convert the data uploaded by the taxpayer into the JSON format specified by the GSTN
The validation module will validate tax computations (rates, corrections etc.) as well as do other checks such as duplicates and so on.

Invoice Matching

As various taxpayers upload their outward supplies, on set dates in the subsequent month, GSTN will share the information back with the matched taxpayers so that they can take the relevant input tax credit. This module will analyze the data received from calls to GST API and then allow the taxpayer to match inward supplies invoices uploaded by their suppliers with their own data and apply corrections identify missing invoices and so on and so forth. If this module is not executed properly, it might lead to increasing

calls to GSP support desks and customer dissatisfaction. *User friendly, extremely fast and feature rich invoice matching service would be key to providing customer delight.*

GSTN API Jobs and Scheduling

This module will be the interface to GST APIs exposed by the server and make on-demand and scheduled calls to the GST APIs and exchange information.

This module will also keep a log of all data sent and received from GST API as well as securely store the Keys issued by GSTN and the keys it issues to sub-GSPs.

Core Accounting / Taxpayer Account/Payments

This module will cater to maintaining the sales purchase and purchase ledgers of each taxpayer as well as taxes paid. The taxpayer account will keep a record service charges paid by the taxpayer.

Master Data & Rules

This module will manage all the master data related to HSN codes, tax rates, state specific tax rules, critical dates etc. Strict access control, maker/checker facility and auditing would be needed to ensure that no unintentional changes are done to this data.

BI and Reporting

Each taxpayer would need to report and visualize their data, be it invoice data or returns and tax computations. ASPs can provide rich reporting and visualization layers to allow taxpayers to analyze their supplies as well as tax liabilities including basic forecasting, historical analysis, allowing export to Excel for further slicing and dicing etc.

Sub GSP /ASP Management and GSP API

The GST ecosystem allows for potential sub-GSPs and ASPs to be identified. These modules will enable registration of sub-GSPs and further allow exposing the GST API wrapped as-is or enhanced to the sub-GSPs. (**GSP API**)

The GSP API can also be exposed to high value taxpayers with mature Accounting ERP systems so that they can upload the data via GSP API to GST API as a pass-through transaction

The core GST API will undergo changes over time and the GSP system should also be able to enable API versioning of its APIs.

4 Azure Platform for GSP Solution

Microsoft Azure is Microsoft's public cloud offering, in India Microsoft has invested in 3 large datacenter locations, highest for any public cloud provider in India. The datacenters offer a large variety of IAAS, PAAS and SAAS services in secure and trusted environment. This section shares high level detail of Azure offering and why it is the best platform for secure and scalable hosting of GSP solutions.

4.1 Trust and Scale

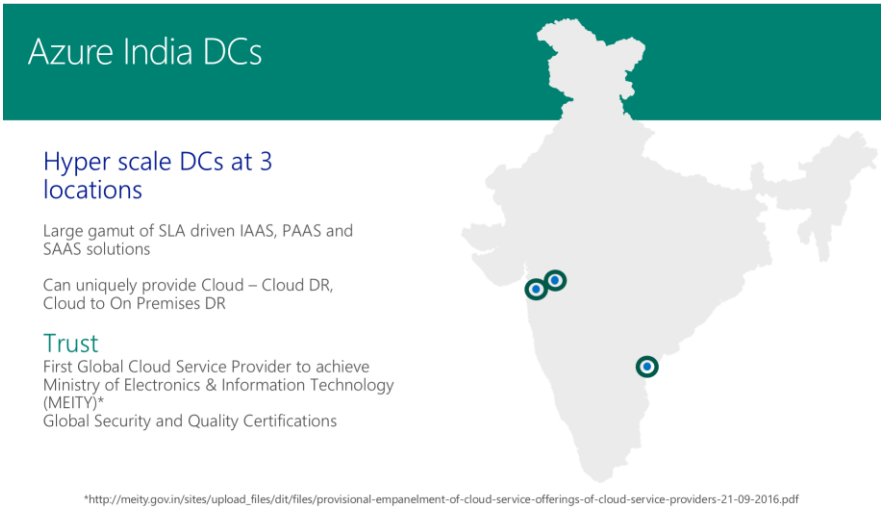

Figure 2 Azure India Data Centers

Microsoft operates **3 large DC facilities** in India in the Mumbai, Pune and Chennai region. These are designed to offer hyperscale could services across IAAS, PAAS and SAAS solutions. With 3 locations in India Azure can provide pure cloud DR capabilities, which other public cloud vendors are yet to match.

Microsoft has been tirelessly pursuing process and security certifications across various standards as well as government bodies. Azure is designed with an aggressive compliance strategy that helps customers address business objectives and industry standards & regulations. Our security compliance framework includes test and audit phases, security analytics, risk management best practices, and security benchmark analysis to achieve certificates and attestations.

Figure 3 Trusted Cloud

Azure India Services has been empaneled by MEITY for hosting govt. solutions (G2G and G2B) and is the only large public could provider to be listed till date.

Microsoft Azure Trust Center available here (https://www.microsoft.com/en-us/trustcenter/CloudServices/Azure) is single source of all compliance and certification related information when it comes to Azure Services.

ISO and other certificates for Azure India services can be located on the trust center here (https://servicetrust.microsoft.com/Documents/ComplianceReports), this includes the IS 27001 certification mandated by GSTN.

About 90% of Fortune 500 companies as well as multiple government bodies trust Azure with about 120,000 new subscriptions coming in every month! This gives Azure credibility as key, scalable and secure platform to host GSP solutions.

4.2 Open Source solutions on Azure

Not only does Azure support Open Source solutions, as well as non-Microsoft commercial solutions, many of its own services are powered by Open Source software. The support spans almost the entire life cycle of solution development i.e. from dev-test tooling, scripting and management to hosting OSS and CSS solutions.

The following diagram shows some of the OSS solutions supported on Azure.

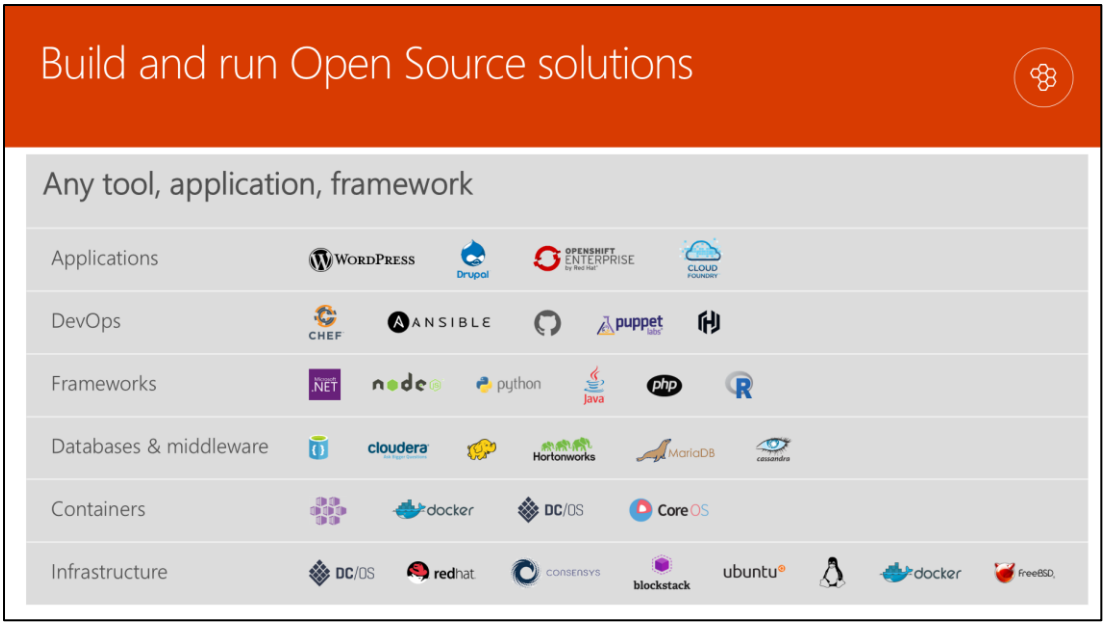

Figure 4 Oper Source Solutions Support on Azure [Not Exhaustive]

In addition to OSS, Azure is a great platform to host CSS solutions such as Oracle DB, SAP and many x86 based platforms which run on Linux or the Windows OS. The following URL, https://azure.microsoft.com/en-us/overview/choose-azure-opensource/ , is a good source to start reading on Open Source on Azure.

More general information on Azure is available at https://azure.microsoft.com/en-us/ .

The following sections detail features of Azure and how they can map to GSP solution capabilities.

5 Mapping GSP Solution platform capability to Azure

This section maps various platform requirements of GSP solution to Azure capabilities and features.

5.1 PAAS and IAAS

Before we begin with mapping exercise, it is important to mention that Azure cloud platform offers many services under the PAAS offering. For many of such PAAS services, similar service/features can be achieved by setting up the VMs and installing and maintaining the product on Azure IAAS. To give an example, you

can setup a Redis Cache farm on Azure IAAS VMs or get a fully managed Azure Redis Cache as a PAAS solution.

If your team has the relevant operations expertise to run and scale the solution as well as take care of DR needs etc. you can setup the Redis Cache farm on Azure VMs. If you don't want to be concerned with Redis Cache operations and need a managed Redis Cache then go for the managed PAAS service. For the consuming application, the same Redis Cache SDK would be used and the applications won't notice the difference in the underlying platform.

Similar to above example Azure offers Azure DocumentDB (NoSQL database with wire protocol similar to MongoDB), Azure Search, Azure SQL etc. Which can be included in your solution and immediately benefit from managed services.

Based on the solution components you choose you can include PAAS offerings in the solution architecture and get to offer better SLAs, availability, managed DR capability as well as reduction in costs. PAAS services have documented SLAs for various performance tiers, high availability, DR (with defined RTO and RPO as applicable) and pay as you go billing. Being a managed service the relevant license costs are included in the service billing.

Over a period Microsoft will continue to offer more services under the PAAS model. **In general, a PAAS service will offer a better TCO over a period for equivalent feature set**, allowing GSPs to lower their solution costs without compromising on features and in-fact offering a richer feature set and peace of mind.

5.2 Lift-and-Shift to Azure

In many cases there would be solutions or modules which are already functioning and tested for GSP feature set. Azure is a great platform to lift and shift these applications and immediately benefit from benefits Azure offers.

These would fit into the Azure IAAS model and it is very likely that equivalent VM (compute), storage (disks & SSD) as well as network topologies can be quickly setup and applications, or even VMs running on on-prem private clouds can be shifted to Azure.

The following links provide more information these services:

a> Azure Windows and Linux VMs : https://azure.microsoft.com/en-us/services/virtual-machines/

b> Azure networking : https://docs.microsoft.com/en-us/azure/virtual-network/

c> Azure Storage : https://azure.microsoft.com/en-in/services/storage/

Over the life of the solution, and as mentioned above, some of the platform components can be replaced by equivalent PAAS components thus benefiting from best of both worlds.

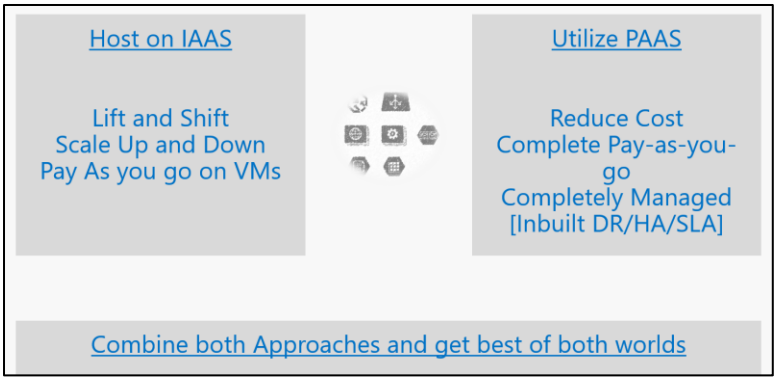

Figure 5 Combine PAAS and IAAS

5.3 Architecture: Drivers and Principles

This section discusses some of the solution architecture drivers and the architecture principles and solution imperatives.

Drivers	Principles and imperatives
GST is new tax regime, there will be tremendous flux in the ecosystem as taxpayers, GSPs and govt. bodies settle down to a rhythm. Tax payer anxiety must be actively managed.	Solution UX and GSP experience should make the transition to GST process easier for the taxpayer. Consider adopting modern UX practices, profuse help and guidance on the portal, pro-active customer support engine and other best practices to offer the customer a great worry-free experience.
High volumes of invoice processing, spiky and, sometimes, unpredictable.	
Pricing GSP services to taxpayers must be competitive.	Ensure core tax computation modules are flexible and robust.
GSP Solution will need to be extremely secure as these will be targeted by hackers and mischief mongers alike.	Target very high availability for the solution, at least 99.95+ uptime.
Govt. would need oversight over the operations and process at regular intervals.	Design the processes to allow and recover from failures, consider idempotency while processing files and data. Core platform should allow restoring state from previous well known states in case of irrecoverable issues in tax computations etc.
	Allow processing modules to scale in and scale out to handle workload with the right capacity.
	Adopt platform which gives *pay as you go* and transaction level costing where possible to get

	cost predictability and flexibility and prevent upfront cost investments. Aggressively adopt scalability best practices: Loose coupling, queuing, distributed processing, caching/in-memory data processing, state less processing, high IOPS storage, avoid SPOF etc. Adopt aggressive Security practices from the datacenter layer up to application layer, consider adopting secure lifecycle development practices. Build solution telemetry for real time monitoring of both technical, security and functional errors from day one. Will be needed for govt. audits too. Ensure all components esp. stored state, are designed for recovery in case disaster strikes.

The GSPs need to come across as **trust worthy partners** to the govt. agencies and more importantly to the anxious taxpayers. This implies focusing on the core tax operations and adopt a trustworthy platform partner who brings in a scalable, robust and a secure platform. This is where we feel Microsoft Azure India DCs offer the right platform to host and run their solutions.

The following sections map most of the options available on Azure to build a complete GSP solution.

Note: Not all Azure Services are available at all locations / DC , do check the link in Appendix A below for up-to-date information on service availability.

5.4 Azure Platform mapping

The following diagram and sections below shows a typical map of platform technology components needed to build a GSP solution.

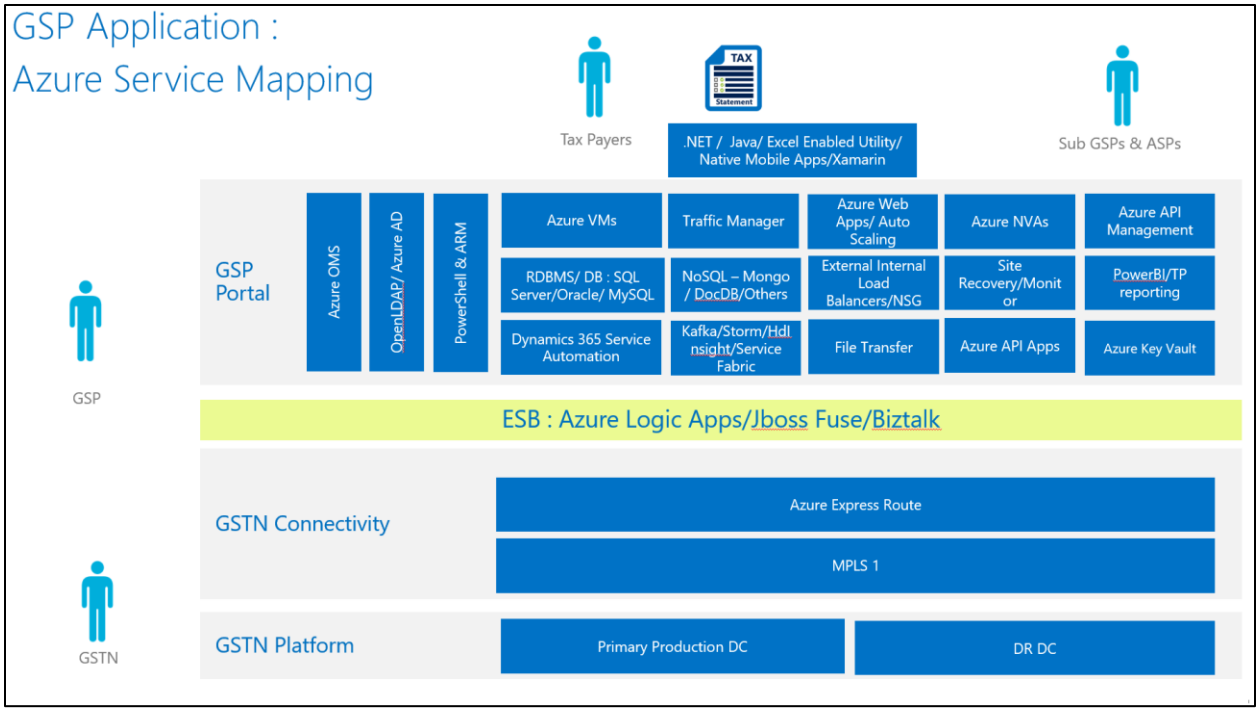

Figure 6 Azure Platform mapping

5.4.1 Azure IAAS platform

Azure VMs, Azure Storage and **Azure networking** provide a great platform to host GSP cloud solutions. It is not possible to cover all the relevant features in this document.

Azure VMs

Azure provides VMs of various sizes and capabilities targeted for various workloads (compute intensive, dev test, high memory etc.). For typical GSP use the D (and DS) series are recommended as they come with high end Intel Xeon based processor hosts and memory combination along with SSD backed storage (DS).

For all Virtual Machines that have two or more instances deployed in the same Availability Set, Azure guarantees Virtual Machine Connectivity to at least one instance at least **99.95% of the time**. [https://azure.microsoft.com/en-us/support/legal/sla/virtual-machines/v1_5/]

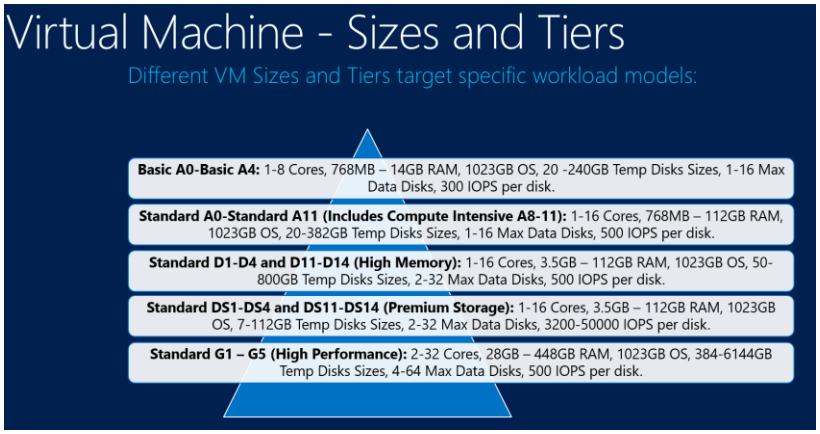

Figure 7 Indicative VM Offerings

In addition to predefined images with base OS, Microsoft also provides images pre-loaded with software which includes the licensing costs. E.g. SQL Server Image. As shown below.

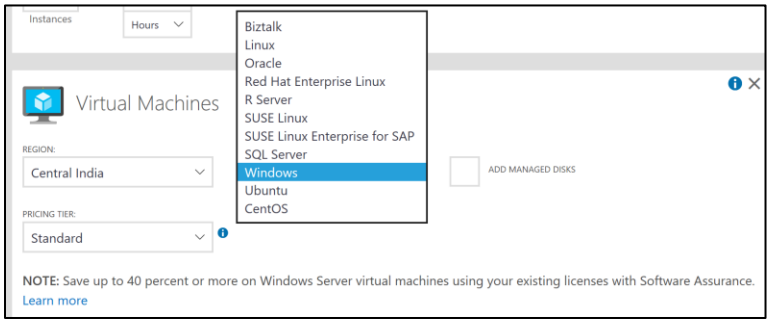

Figure 8 Sample of pre-configured VM templates available

[Do note to check with the software OEM for licensing clarifications for non-Microsoft commercial software images] . **Azure Marketplace** has a range of pre-built templates ready for deployment e.g. VMs with File transfer management solutions.
Note: Only 64-bit OS' are supported.

5.4.2 Azure Networking

Azure Virtual Networking provides networking backbone and offers critical facilities such as firewalls, HTTP/S endpoints, multiple NIC VMs, IP4/IPV6 support, network isolation, vnet Peering, user defined routing of traffic and VPN connectivity to model a secure networking layer.

Azure also offers both External (for public /internet facing endpoints) and Internal (within the VNET) **load balancing** to route traffic and performs other tasks such as SSL offloading.

A range of commercial **network security virtual appliances** (NVA) from leading security vendors such as Barracuda, CISCO etc. are available too. These can be combined with Azure **NSG (Network Security**

Groups) to direct and route the network traffic with flexible rules. User defined Routes **(UDR)** can further help route traffic going out of subnets.

Once we have the solution deployed on Azure DCs, **Azure Traffic Manager** can provide a single endpoint to the customer and then reroute traffic to DR endpoints in case of a disaster solution.

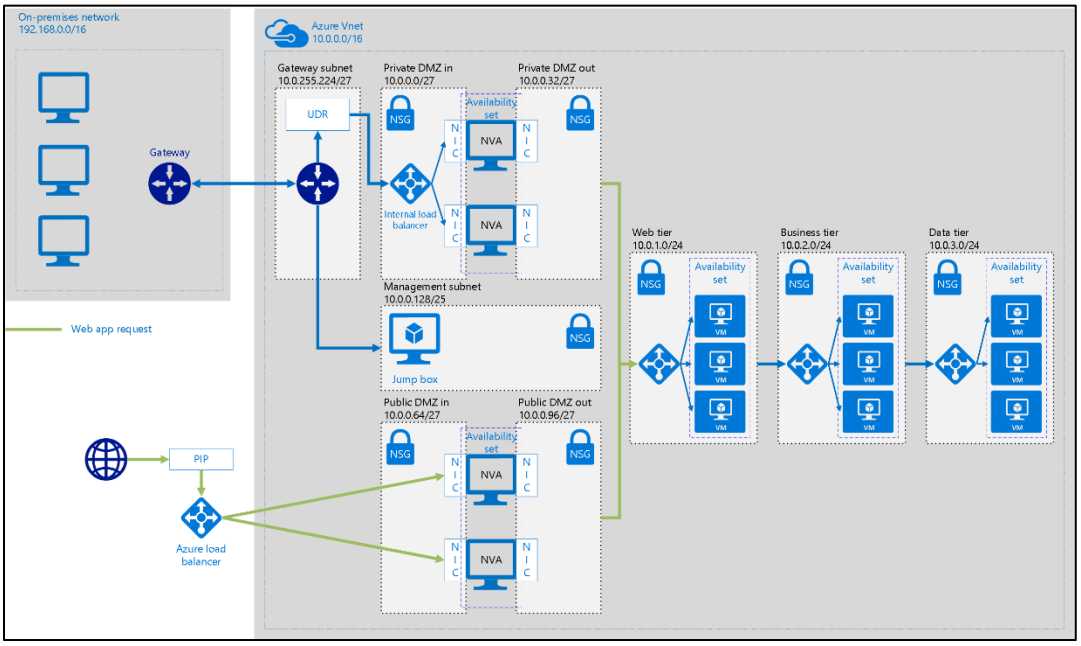

Figure 9 Sample Network layout on Azure

GSTN will be exposing their APIs to GSPs only over an MPLS connection. The **Azure ExpressRoute (ER)** offering allows connecting the Azure cloud network to on-premises networks of the GSTN Datacenters over MPLS. There are, currently, five networking providers who offer such connectivity in India.

Airtel :
http://www.airtel.in/about-bharti/media-centre/bharti-airtel-news/enterprise/airtel+expands+cloud+services+portfolio+with+launch+of+connexion

Aryaka :
http://www.aryaka.com/blog/planning-to-deploy-microsoft-azure-is-your-enterprise-network-up-for-the-challenge/

SIFY:
http://telecom.sify.com/azure-expressroute.html

Tata Communications:
http://www.tatacommunications.com/lp/izo/azure/azure_index.html

Global Cloud Exchange
http://globalcloudxchange.com/cloud-platform/cloud-x-fusion/cloud-x-fusion-for-azure/

5.4.3 Azure Storage

Azure Storage provides robust storage engine to power the VMs and rest of the PAAS offerings. Azure maintains 3 copies of data [synchronously replicated] to provide local redundancy and customers can choose to have Geo redundancy to get additional redundancy in case of disasters [geo-redundancy] .

Azure offer standard storage [hard disks] and premium storage backed by SSD drives. Offering a range from 500 IOPS to 50000 IOPs per disk. Which, combined with local and geo redundancy, provides a robust storage solution.

On this base, other services such as Azure Files [Managed SMB 2.1 File storage], Azure Disks, Blob storage, Table [name/value] and Queues are offered.

While planning the storage to consider the boundaries documented at https://docs.microsoft.com/en-us/azure/storage/storage-scalability-targets to plan the storage solution.

More information on Storage is available at https://azure.microsoft.com/en-us/services/storage/files/

5.4.4 Web Servers – Portal and APIs

The GSP portal would be the primary interface to the taxpayer and would need to run on a scalable web server platform (Web Site + APIs) supporting and designed using modern web application paradigms.

Azure IAAS, as mentioned earlier, supports many flavors of Windows and Linux which can run most popular Web Servers such as IIS (Internet Information Server), Apache Tomcat and others of your choice.

These Web servers can be load balanced using the **Azure Load External Load balancer**, the applications can further be tiered by creating separate VM farms to expose business logic via web APIs and using an **internal load balancer**.

To achieve high availability the web server VMs need to be configured in an **Availability Group (AG)** which ensures that Azure will not update the hosts or local power outage will not affect your high availability.

In addition to the above set Azure offers a managed **Azure Web Apps** service where you can benefit from auto-scaling high availability and other enterprise grade features. Azure Web Apps supports both Windows and Linux environments allowing apps to be built using .net, PHP, Java and Node.js . More information is available here (https://azure.microsoft.com/en-us/services/app-service/web/)

Core APIs can be built using **Azure VMs running web servers** (as mentioned earlier) or using the managed **Azure API Apps platform**. The API apps platform, which is part of the **Azure App Service,** allows you to build scalable APIs using the language of your choice. This environment can be integrated with Azure VNETs allowing out to expose traditional services running in isolated VNETs via modern REST based APIs.

It also offers automatic metadata generation, CORS, SSO support and so on. More information is available here. [https://azure.microsoft.com/en-us/services/app-service/api/]

5.4.5 RDBMS

Azure supports almost all popular RDBMS products like **SQL Server, Oracle DB, MySQL, Maria DB etc.** on Windows and Linux VMs as relevant.

Provisioning the databases is mostly automated using scripts and can be achieved by few clicks on the Azure Portal or creating PowerShell , CLI scripts and ARM templates. A simple walkthrough of creating MySQL database is provided here https://docs.microsoft.com/en-us/azure/store-php-create-mysql-database

Azure also offer some of these as managed services i.e. in a PAAS model e.g. **Azure SQL Database**.

Azure SQL Database offers four different service tiers and can use up to 4 TB storage. After initially picking a service tier and performance level, customer can scale a single database up or down dynamically based on actual requirement. This PAAS is built in the same engine as SQL Server 2016 (v12) including in-memory OLTP. Good comparison of SQL Server and Azure SQL v12 is available here. https://docs.microsoft.com/en-us/azure/sql-database/sql-database-features.

To make building multi-tenant applications or separating customer databases, Azure offers a management and resource sharing optimization in form of **Azure SQL Elastic pool.** This service allows you to pool a bunch of Azure SQL Databases which then share a common performance resource and can offer the best price/performance to GSPs who manages large number of databases dedicated to clients e.g. For large customers GSPs can offer dedicated RDBMS database as a premium performance service.

It is critical to get data replicated to DR site for any RDBMS solution that is chosen and Azure SQL Database, being a PAAS service, provides managed DR. The RPO/RTO metrics are documented here (https://azure.microsoft.com/en-us/blog/azure-sql-database-point-in-time-restore/)

Topologies and templates for various databases deployed on IAAS VMs are available and documented on the OEM websites as well as Microsoft Azure document site.
e.g.

SQL Server Always-on Availability Groups on Azure : https://docs.microsoft.com/en-us/azure/virtual-machines/windows/sql/virtual-machines-windows-sql-high-availability-dr

Oracle Data Guard : https://docs.microsoft.com/en-us/azure/virtual-machines/windows/classic/configure-oracle-data-guard

MySQL Galera Cluster : http://download.microsoft.com/download/6/1/C/61C0E37C-F252-4B33-9557-42B90BA3E472/MySQL_HADR_solution_in_Azure.pdf

Clear DB Managed Service on Azure : http://w2.cleardb.net/azure/

Maria DB on Azure : https://azure.microsoft.com/en-us/blog/mariadb-enterprise-cluster-on-azure/

5.4.6 NoSQL Processing

GSP providing enriched API like Bulk uploading of invoices, conversion of invoice data formats (CSV to JSON, XML to JSON), latest HSN code tax rate, prepare GSTR1/GSTR2/GSTR3, validation of returns for latest tax rate and amount etc. It would also be necessary to log calls to GST APIs which need JSON formatted data for audit purposes and made searchable in case of issues.

Following options can be considered.

JSON support in SQL Server and Azure SQL
Microsoft SQL Server and Azure SQL Server offers integrated JSON processing support. This makes ingesting and extracting data in JSON format very easy. The support is documented here [https://azure.microsoft.com/en-us/blog/json-support-is-generally-available-in-azure-sql-database/]

MongoDB is an open-source NoSQL database and is supported s first class citizen on Azure . [https://docs.mongodb.com/ecosystem/platforms/windows-azure/]

Document DB, is completely managed NoSQL database offered by and globally scalable NoSQL platform . It supports Mongo dB wire protocol and existing code which relies on Mongo DB can be seamlessly upgraded to Document DB. A **third-party** blog post comparing DocumentDB and Mongo DB is available here [https://medium.com/@th0maswe1ss/azure-documentdb-vs-mongodb-6d5806c16239]

GSPs can also configure their NoSQL DB of choice on Azure VMs e.g. CouchBase [https://www.couchbase.com/partners/microsoft-azure], RavenDB etc.

5.4.7 Reporting and BI

Numerous reporting and analytic solutions are available. In case of GSP applications it would be necessary to provide the tax payer view of their ledgers with various filters. There will be no or limited need to combine data from multiple taxpayers at a time.

Azure supports multiple reporting tools which can be hosted on Azure VMs and available on the Azure Marketplace e.g. **JasperSoft BI** (https://www.jaspersoft.com/press/jaspersoft-launches-business-intelligence-offering-on-windows-azure) , **Tableau** (https://azure.microsoft.com/en-us/blog/tableau-server-in-the-azure-marketplace/) , SAP Crystal Reports etc.

Microsoft SQL Server offer **SSRS** (SQL Server Reporting Service) a proven reporting tool which provide scalable report generation service. Combined with ODBC driver for Document DB, JSON support in SQL Server, you can combine both tabular and JSON data in one reporting tool. SSRS strength lies in scalability and tabular/pivot reporting with ability to export reports to various formats such as Excel and PDF.

Note: Excel and CSV export are going to be key asks from tax payers.

Microsoft also offers **PowerBI** and **PowerBI Embedded** as powerful dashboarding tool to enable rich visualizations for your customer. (https://docs.microsoft.com/en-us/azure/power-bi-embedded/power-bi-embedded-what-is-power-bi-embedded)

5.4.8 Data Processing Options

GSP solution would need a Platform for building high performance data integration solutions, transformation and business rules checking solution. Validating tax computations in each invoice/line items, matching outgoing supplies uploaded by the suppliers with incoming supplies invoices, calculating returns for filings would need a scalable processing engine.

Based on how the solution in architected Azure offers multiple options.

Azure HDInsight provides managed **Apache Kafka and Apache Storm** services to handle such modelling. Where invoices are ingested as stream of incoming data points and then processed on distributed network of processors. Azure Marketplace also offers other vendors who support single click deployment of their Kafka, Storm, Hadoop flavours e.g. **Bitnami Kafka** : https://azure.bitnami.com/ , **HortonWorks Hadoop** : https://channel9.msdn.com/Shows/Data-Exposed/Deploying-Hortonworks-HDP-on-Microsoft-Azure

Azure Service Fabric, is a managed fabric to model a **micro service architecture** when it comes to processing customer invoices and tax data, this platform allows building redundant /distributed and coordinated services which can be scaled out to process invoices and then scaled down during non peak times. This would require some involved solution design and custom .net code to process data.

Microsoft SQL Server Integration Services (SSIS) is a platform for building high performance data integration solutions, including extraction, transformation, and load (ETL) packages for data warehousing. Newly announced **ODBC adapter for DocumentDB** allows mixing of tabular and noSQL /JSON processing.

Most of the customers prefer ‘ease of use’ GUI based AND ‘ETL/ELT’ at scale and SSIS has all the GUI interface that can fulfill the ‘ease of use’ requirement.
For a pure cloud managed service to process data from multiple disparate sources and formats, **Azure Data Factory,** ADF can be considered, ADF is great as an ELT tool – you move data from one place to another and then execute some set of transformations on it through USQL, Hive, Stored Proc, etc. depending on where your data lands. [Note: Since SSIS is not available as a PaaS service, consider using Azure (IaaS) SQL VM’s, installing and configure SSIS to load the data warehouse and use SSIS as an ETL tool on Azure].

Azure Storage Queues, **Azure Service Bus Queues** or plain old RDBMS tables can be used as a Queueing mechanism to control job sequencing.

Along with processing engine a job scheduling engine would be needed to trigger various processing jobs. This must work with a queuing mechanism and can be modelled using databases, queuing and service bus products hosted on Azure IAAS. Below are some of the options.

SQL Server integration services and SQL Server Agent: For pure data processing jobs, SSIS and SQL Server Agent **provide** a robust and proven technology to move and process data.

Custom .NET Windows Services: For very granular processing control.

Quartz Scheduler : Quartz is a distributed scheduling engine to manage Java based jobs.

5.4.9 Caching Layer

Most master data e.g. HSN codes, merchant metadata [GST TINs etc.] will be repeatedly accessed in data processing and **Azure Redis Cache** [a completely managed Redis Cache service] will allow the application processing to remain responsive even as processing and user load increases.
This service leverages the low-latency, high-throughput capabilities of the Redis engine to enable building fast, scalable applications.
More information available here https://azure.microsoft.com/en-us/services/cache/

5.4.10 API Management Layer

GSTN has adopted an Open API based architecture while designing the GST Solution and similar approach can be adopted by the GSPs.

The GSP solution will need to expose Web API (REST based APIs) to :

a> Interact with offline utilities
b> Expose GST APIs to large customers who are interested in purely accessing the GST API [pass through]
c> Expose enriched APIs to Sub GSPs and ASPs

Whichever way you decide to build your APIs, Azure offers and **Azure API Management**, a completely managed, turnkey solution for publishing APIs to external and internal consumers. One can quickly create consistent and modern API gateways for existing back-end services hosted anywhere, secure and protect them from abuse and overuse, and get insights into usage and health. In addition, an associated developer portal, automate and scale developer onboarding to help get your API program up and running.

API policies allow the GSPs to change the behavior of the API through configuration. Policies are a collection of statements that are executed sequentially on the request or response of an API. Popular statements include format conversion from XML to JSON and call rate limiting to restrict the amount of

incoming calls from a developer, JWT token validation, restricting connections from identified IP addresses and many other policies are available.

Inbuilt analytics dashboard provides real-time view of API traffic, issues encountered, bandwidth consumed and more.

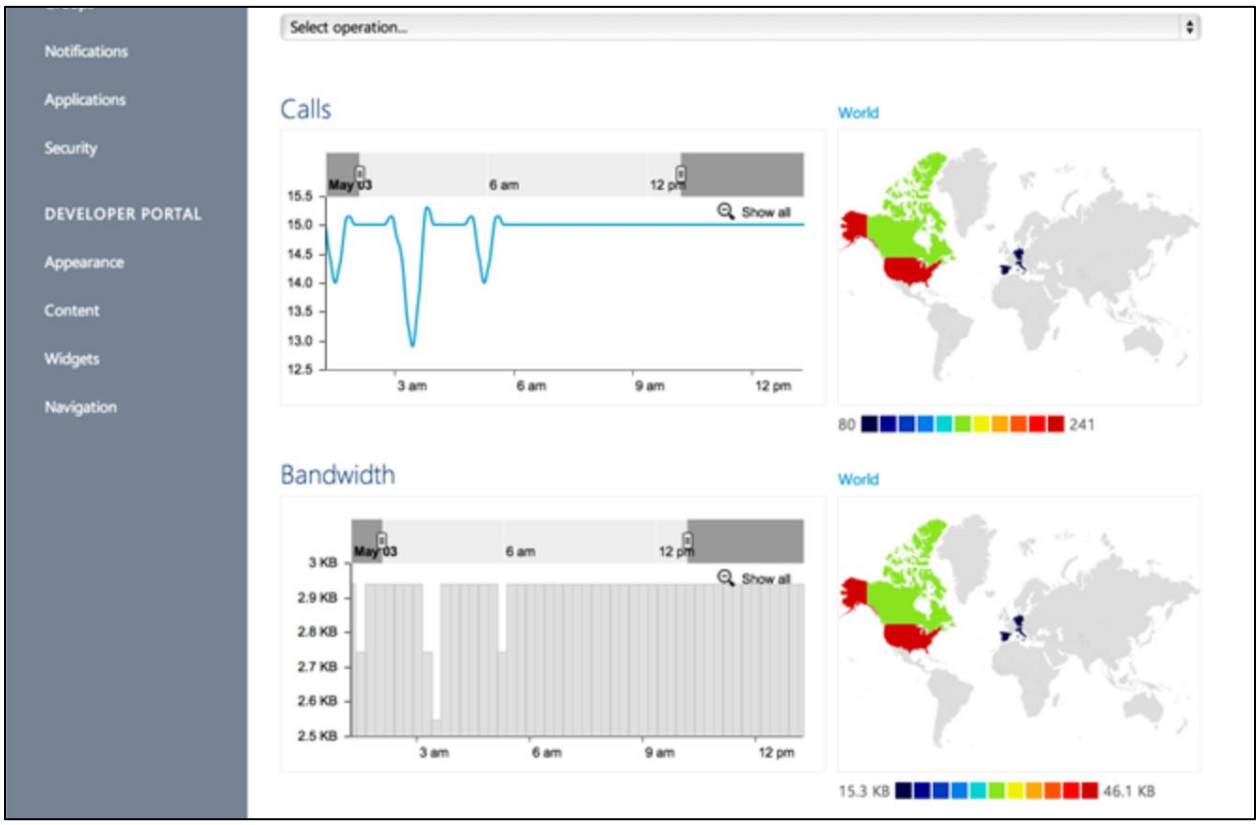

Figure 10 Sample API Management dashboard

In addition to Azure API Management, there multiple OSS options available on Azure on Azure VMs including..

- **NGINX** Plus : A well-established HTTPS Gateway management solution [https://www.nginx.com/products/nginx-plus-microsoft-azure/]
- WSO 2 : WSO2 API Manager (APIM) is a fully open source, complete solution for creating, publishing and managing all aspects of an API and its lifecycle

5.4.11 Integration layer and Enterprise Service Bus

GSPs will need to consume and keep transferring data to and from GST APIs and on the other side allow flexibility on the taxpayer side to upload/provide data through various options.

Large tax payer might want to share data directly via their ERP system or upload CSV files, there would be a need store and forward responses back to taxpayer systems, retry mechanisms, throttling of calls and so on and so forth. An Enterprise Service Bus provides

Traditionally EAI (Ent. Application servers) such as **Biztalk Server**, **JBoss Fuse, IBM Integration Bus** etc. provide such EAI feature set and are available on Azure IAAS (https://access.redhat.com/ecosystem/ccsp/microsoft-azure for JBOSS Fuse)/ (https://developer.ibm.com/integration/blog/2016/04/04/installing-ibm-integration-bus-microsoft-azure/).

Azure provides one more option of completely managed integration platform, **Azure Logic Apps**.

Azure **Logic Apps** provide a way to simplify and implement scalable integrations and workflows in the cloud. It provides a visual designer to model and automate your process as a series of steps known as a workflow. There are many connectors across the cloud and on-premises to quickly integrate across services and protocols. A logic app begins with a trigger (like 'When an account is added to Dynamics CRM') and after firing can begin many combinations actions, conversions, and condition logic.

The advantages of using Logic Apps include the following:

- Saving time by designing complex processes using easy to understand design tools
- Implementing patterns and workflows seamlessly, that would otherwise be difficult to implement in code
- Getting started quickly from templates
- Customizing your logic app with your own custom APIs, code, and actions
- Connect and synchronise disparate systems across on-premises and the cloud
- Build off of **BizTalk server**, **API Management**, **Azure Functions**, and **Azure Service Bus** with first-class integration support

Logic Apps is a fully managed iPaaS (integration Platform as a Service) allowing developers not to have to worry about building hosting, scalability, availability and management. Logic Apps will scale up automatically to meet demand.

Logic Apps brings speed and scalability into the enterprise integration space. The ease of use of the designer, variety of available triggers and actions, and powerful management tools make centralizing your APIs simpler.

Easy to use design tools - Logic Apps can be designed end-to-end in the browser or with Visual Studio tools.
Connect APIs easily - Even composition tasks that are easy to describe are difficult to implement in code. Logic Apps makes it easy to connect disparate systems.

Extensibility baked-in - Logic Apps is designed to work with your own APIs and code; you can easily create your own API app to use as a custom connector, or call into an **Azure Function** to execute snippets of code on-demand.

Real integration horsepower - Logic Apps can easily leverage the power of BizTalk, Microsoft's industry leading integration solution to enable integration professionals to build the solutions they need.

Here are a couple of examples:

- Move invoice files uploaded to an FTP server into Azure Storage for processing
- Keep retrying calls to customer or GST API in case there is a error
- Connect to customer SAP environment using Enterprise Connectors

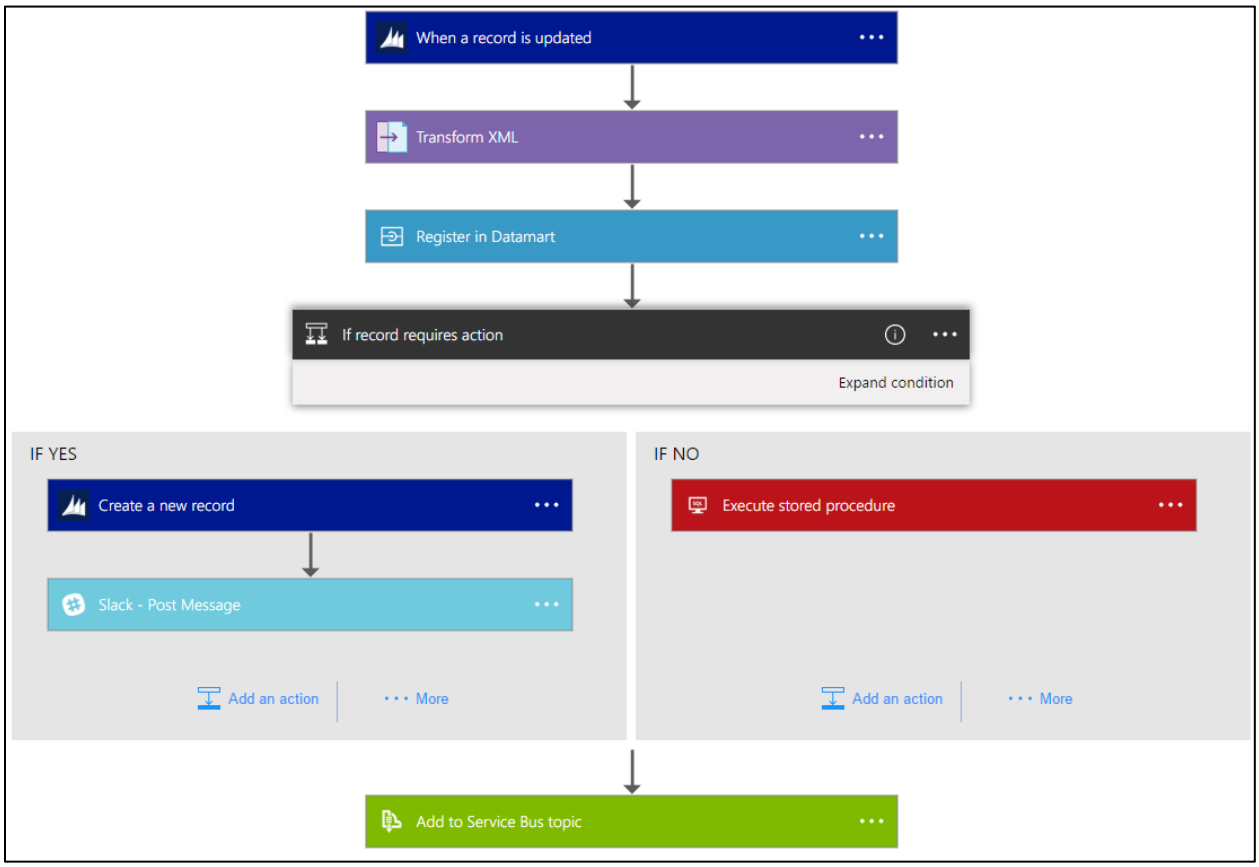

Figure 11 Sample Logic App Workflow

5.4.12 File Transfer System

Many taxpaying organization esp. with legacy accounting systems would need a mechanism to upload invoice data files (preferably compressed and encrypted) to secure file upload locations. The following options are available:

a> **Windows (inbuilt IIS FTPS Service)** or Linux based FTPS servers integrated with Azure AD domain services for user credential integration and with SSL enabled for secure FTP uploads. The uploaded file shares can be stored in customer specific folders on Azure File Shares and monitored by data processing jobs for processing.

b> Dedicated File transfer management solutions such as **EFT** - Globalscape's Enhanced File Transfer (EFT) service that is available on Azure Marketplace.

c> **Go Anywhere** Managed File transfer for Azure (https://www.goanywhere.com/platforms/microsoft-azure)

5.4.13 Dynamics 365 Service Automation

Finally, GSPs need to offer world class customer support to guide the customers through this transition and overall to differentiate services from other GSPs.

To manage customer complaints and service requests in a predictable and SLA bound manner, Microsoft offers **Dynamics 365 for Customer Service** platform from India DCs. It is a completely managed SAAS solution which enables GSPs to quickly setup their customer support engine. It is not possible to cover all the relevant features but you can get started with the features here [https://www.microsoft.com/en-us/dynamics365/customer-service]

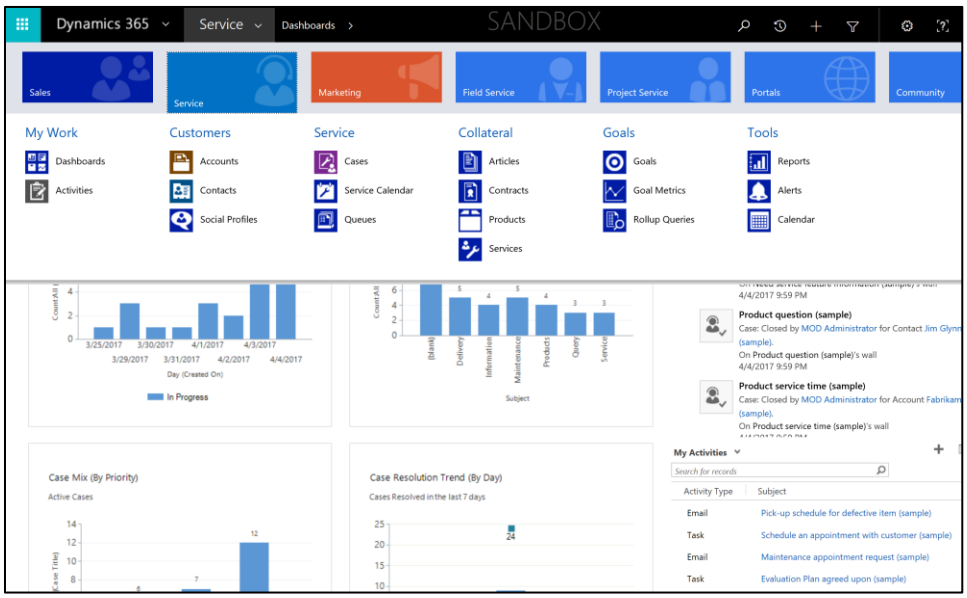

Figure 12 Dynamics 365 for Customer Service sample Dashboard

5.4.14 Security: User Identity and credentials

Information security is paramount for GSPs. Apart from identifying customer's, they also need to segregate which GSP employee accesses which part of the solution at module level as well as operations level.

A well laid out GSP solution would have multiple modules e.g. tax payer Portal, API Gateway, File transfer (SFTP/FTPS), monitoring modules etc. Setting up independent authentication principles for operations can help isolate monitoring.

If Sub GSPs are involved, then their credentials would need to be managed separately.

In general :

- **Reduced credential store attack surface**: - Each module may have internal /GSP users with elevated rights and if one account gets hacked than it can be used to attack the whole directory tenant and compromised the whole user database.

- **Multi-Layered Security -**Tax payer needs to use of latest Multi-Layered Security Measures, they will also authenticate themselves by providing user ID, OTP

- **Managed Identity Solution** - GSP will need to onboard the users and provide them with appropriate access to the Application.

- **User Authentication / Authorize -** Tax payer will need to Authenticate themselves and have access on the resources they are authorized for.

PAAS mapping: Azure AD is managed directory solution like Windows AD, but made for the cloud world and is a good fit for the solution**.** Azure Active Directory is a comprehensive identity and access management cloud solution that provides a robust set of capabilities to manage users and groups. It helps secure access to on-premises and cloud applications

AAD Multi-Tenant - Azure Active Directory (Azure AD) - Meets the Application design requirement to have multi-tenant cloud based directory and identity management service. One can create multiple AAD tenants under a Azure subscription to isolate credential stores e.g. one tenant for taxpayers and other for internal users.

AAD High Availability- The Multi-Tenant, Geo-Distributed, High Availability Design of Azure AD allows makes it a good choice for credential store

AAD Authentication is performed through protocols such as SAML, WS-Federation, and OAuth.

Query Azure AD- It's possible to query Azure AD, but instead of using LDAP, REST APIs called AD Graph API are used. These all work over HTTP and HTTPS. **LDAPS and Windows Domain Controller emulation -** More relevant for lift and shift scenarios, **Azure Active Directory Domain** sitting on top of the AAD tenant

enables a Windows DC (domain controller) layer on top of Azure AD which enable traditional applications which need a Windows DC or an LDAP/S (read) endpoint for authentication.

AAD Multi-Factor Authentication (MFA)- Azure Active Directory provide enterprise class capabilities spanning self-service, Multi-Factor Authentication (MFA).

OAuth -Azure Active Directory (Azure AD) uses OAuth 2.0 to enable customer to authorize access to web applications and web APIs in the Azure AD tenant. It is used to perform authentication and authorization in most application types, including web apps and natively installed apps

For managing the Azure environment, Azure offers Role Based Access control (RBAC) mechanism to control which operations employee can manage which of Azure Services [https://docs.microsoft.com/en-us/azure/active-directory/role-based-access-control-configure]

Other Credentials Stores:

Numerous other credential stores e.g. Classic Windows Servers ,Linux and OpenLDAP or other stores can also be used on Azure IAAS platform

Security : Protecting data

While we are on the security topic, data protection can be enabled at all layers of the solution platform.

Offline utility: Standards based encryption using digital certificates for symmetric and asymmetric encryption

User Authentication: Secure credential principle and MFA, as discussed in this document. Azure AD also monitors and raises alerts using machine learning to identify inconsistent access patterns.

Encrypting data at rest : SQL Server and Azure SQL Service offers TDE (Transparent data Encryption)

Encrypted Disks : Azure Storage supports Disk level encryption to host the Windows and Linux VMs

Azure Key Vault [key storage]: If needed, Azure provides a hardware (HSM) backed key storage service called Azure Key Vault, it is integrated in many Azure services and provides an SDK to access the keys.

Firewalls and network security appliances: Azure provides inbuilt network security features and supports several industry leading security appliances through Azure Market place e.g.

- Azure Web Application firewall : https://azure.microsoft.com/en-us/blog/azure-web-application-firewall-waf-generally-available/
- Azure NSGs : Network security groups, sets up traffic rules
- Checkpoint - Next Generation Firewall - (Firewall, IPS, Anti-Malware, Email-Gateway and VPN) [https://www.checkpoint.com/products/vsec-microsoft-azure] ,
- Barracuda Application and Network Security Appliances [https://docs.microsoft.com/en-us/azure/app-service-web/app-service-app-service-environment-web-application-firewall]

5.4.15 Monitoring and Compliance

The GSPs need to put in monitoring processes to ensure:

a> Solution Availability
b> Compliance
c> Cloud spend analysis

While there are many commercial and open source Enterprise Monitoring tools to monitor VMs, Azure Portal offers many OOB options to monitor Azure environment. Apart from inbuilt monitoring in Azure Portal dash boards, Microsoft offers products such as:

- **Azure Monitor**:

  Offers visualization, query, routing, alerting, auto scale, and automation on data both from the Azure infrastructure (Activity Log) and each individual Azure resource, Diagnostic Logs

- **Azure Application Insight:**

  Deep inspect application performance and availability

- **Log Analytics**:

  Provides a holistic IT management solution for both on-premises and third-party cloud-based infrastructure (such as AWS) in addition to Azure resources. Data from Azure Monitor can be routed directly to Log Analytics so you can see metrics and logs for your entire environment in one place.

The sections below how these tools can be used to achieve monitoring and compliance goals.

Monitoring Solution Availability
Solution availability is key for any business-critical system, while we take all the steps to ensure that the solution is designed for high availability and stability it is still prudent to ensure that the solution availability is monitored at all times. It not only helps in ensuring high customer satisfaction but also helps in identifying opportunities for improvement.

The following aspects of monitoring needs to be considered:

- **Monitor 'availability' as experienced by the end user**

External monitoring of the Portal solution ensures web pages are available and are rendered in acceptable time limit and you see availability as perceived by the taxpayer. This is the outside-in view of the solution and is the bare minimum monitoring which should put in place. You can monitor your landing (anonymous access) URL or create a customer web page which performs basic checks (e.g. fires few queries on the database, logs to NoSQL, generates a minimal set of reports to ensure all components are functional). **Azure App Insight** offers ability to 'ping' your application URLs from various locations and setup alerts if application endpoints monitored go inactive. It also offers multi-step tests and can be used to perform function specific monitoring e.g. test invoice upload pages, test GSTR 1 upload pages, test GST API wrapper

pages etc. More information on web availability monitoring is available here, https://docs.microsoft.com/en-us/azure/application-insights/app-insights-monitor-web-app-availability.

- **Monitor individual solution components**
    - *Monitor Application Performance*: Azure App Insight offers Application Performance Monitoring (APM) for multiple web platforms. Once your application developers embed the necessary App Insight libraries in your code, you start getting deep insight into how users are using your application, which web pages are not performing, user and session counts, page views and so on. This is not limited to web portal, you can embed the SDK in business logic code and get custom insight monitoring. The figure below shows how AppInsight works.

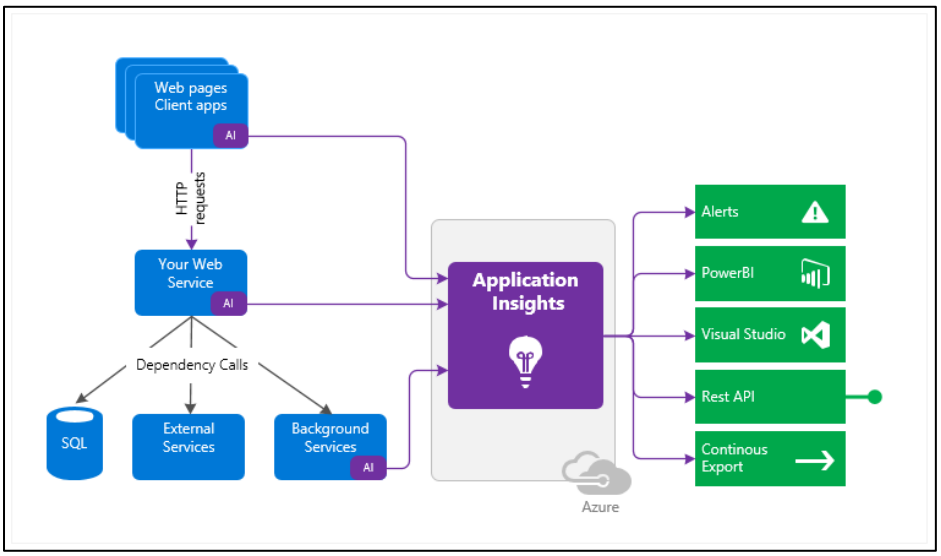

Figure 13 How Application Insights works

This enables live monitoring of your solution where you can also see the live stream of metrics on a customizable dashboard.

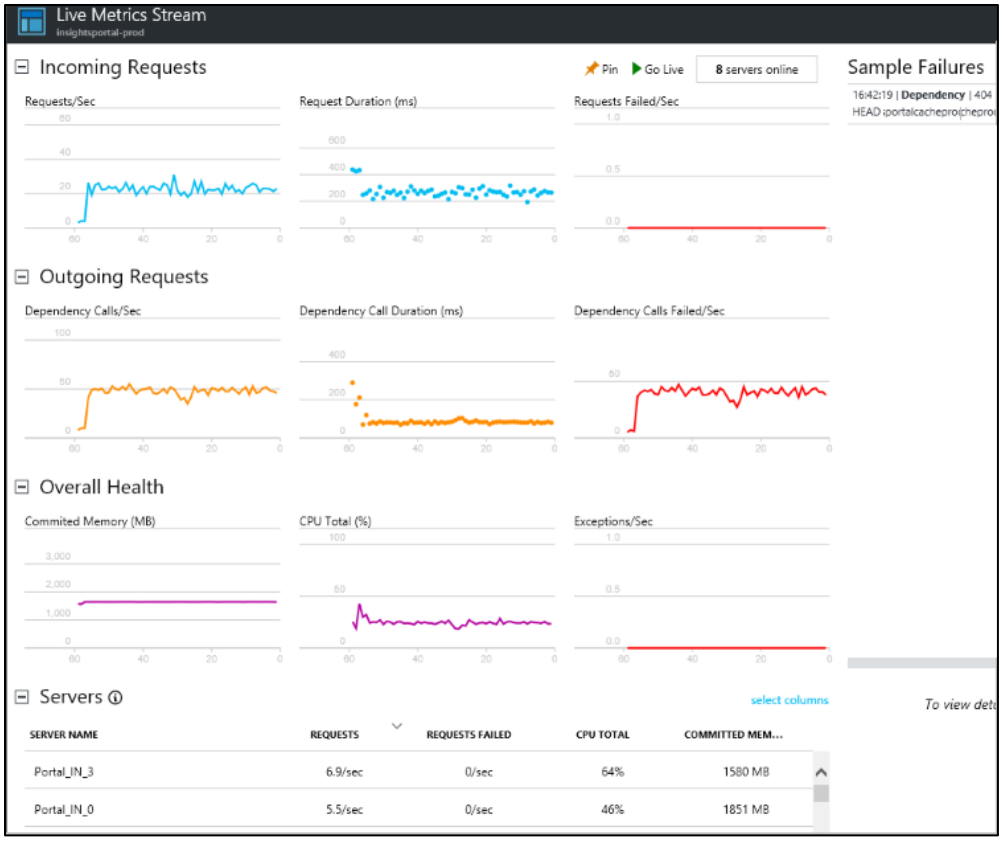

Figure 14 Live Metrics Stream Dashboard

- *Monitor VMs*: Azure Portal itself (out of the box) offers multiple ways of monitoring VM metrics in terms of Dashboards. You can create multiple dashboards which club individual metrics across various layers of the solution. As shared earlier, **Azure Monitor** provides inbuilt monitoring of IAAS resources in the resource group. Apart from visualization, you can setup Alerts which can fire custom scripts to remediate the situation. You can also use popular monitoring solutions to install agents inside VMs to monitor guest operating system parameters and collate it. Azure **Operations Management Suite** is one such service which gives you end to end monitoring, more on OMS later.
- *Monitor PAAS components*: Most PAAS components need not be **directly** monitored, these are monitored by Microsoft Azure teams to ensure performance, availability and SLA, still if you need to trouble shoot you can directly monitor some PAAS components e.g. Azure Monitor allows monitoring of Azure SQL Server operations.

Monitor Compliance

Compliance needs to be built into the system in terms of security rules (authentication and authorizations), it cannot be bolt-on later. For this one needs to plan 'operations' for the solution, as one would do for any other on-premises secured solution i.e. identify operations staff who will operate the

solution e.g. DBAs, backup operators etc. and then lock permissions to these users to access only those components which they have access to, document the processes and so on

Once proper security operations planning is incorporated, Azure can ensure that security logs are collected analyzed for threats and vulnerabilities and violations are isolated. In addition, live threat monitoring (e.g. connection from active botnets) and alerts are offered via Azure **Operations Management Suite** (OMS). OMS Security & Compliance offers monitoring of security events such failed logins, user logons, malware protection status, brute force attacks and offers baseline security analysis for Windows VMs, SQL Server etc.

Figure 15 Sample of OMS solutions security analytics

Azure OMS combines covers 4 areas of:

a> Insight and Analytics
b> Automation and Control
c> Security and Compliance &
d> Protection and Recovery

It combines many of the offerings mentioned earlier e.g. App Insight, Log analysis and they need not be procured separately.

> **OMS is highly recommended for GSP solution monitoring.** More on OMS her https://www.microsoft.com/en-us/cloud-platform/operations-management-suite

Cloud Spend Analysis

While not of direct functional importance, monitoring cloud spend can give you indirect insight into usage of various services involved and then you can make changes in the service tiers to optimize the spend. Azure Portal itself gives adequate information of cloud spend and your Azure CSP (Azure Cloud Service Provider) can provide detailed reports of cloud consumption.

Most Azure resources allow you to **'tag'** them with markers which are then available in usage reports e.g. you can tag all resources needed in invoice matching with 'invmatch' tag and then pivot the usage data to see how much is being spend on invoice matching.

Azure offers billing APIs [https://docs.microsoft.com/en-us/azure/billing/billing-usage-rate-card-overview] which allow you to pull in billing data and programmatically analyze it.

Various third party ISVs have built solutions on top of these billing APIs to give end to end spend monitoring e.g. CloudCruiser (https://www.cloudcruiser.com/partners/microsoft/) and Cloudyn

6 Conclusion

While we have tried to cover most of the important features of GSP solution and Azure platform, it is not possible to cover the host of services offered in a single document. Dev/Test/ ALM support and tooling, Azure mobile Services, support for containers e.g. Docker, Windows Containers etc. are some of the ever-growing palette of services available on the Azure platform which you can decide to leverage to build a robust GSP solution and not covered in this document.

To summarize, Microsoft Azure Platform offers a wide range of services to build secure and scalable GSP app with the development and platform tools of our choice. Documented platform SLAs along with the advantage of combining IAAS and PAAS services allow you to offer the best GSP platform at predictable costs.

For more information, do connect with Mr. Manish Lodha at Manish.Lodha@Microsoft.com .

7 Glossary and Definitions

- API: Application Programming Interface
- APIM: API Management
- ASP : Application Service Provider
- BI: Business Intelligence
- CRM: Customer Relationship Management
- DC : Data Center
- DR : Disaster Recovery / Data Center
- GSP: GST Suvidha Provider
- GST : Goods and Services Tax
- IAAS: Infrastructure as a Service
- IOPS: Input Output Operations per Second
- JSON: Open standard human readable format for data transfer
- MEITY: Ministry of electronics and information technology
- NIC : Network Interfaces
- NVA: Network Virtual Appliance
- OSS : Open Source Software/Solution
- OOB: Out of the Box
- PAAS: Platform as a Service
- REST: Representational state transfer
- SAAS: Solution as a Service
- SLA: Service Level Agreement
- SPOF: Single Point of Failure
- SSD: Solid Stage Drives / Flash Storage
- UX/UI : User Experience / User Interface

8 Appendix A – Web Resources

Azure Infographics Posters which cover the range of services available on Azure: https://azure.microsoft.com/en-us/resources/infographics/

MEITY: Provisional Empanelment http://meity.gov.in/sites/upload_files/dit/files/provisional-empanelment-of-cloud-service-offerings-of-cloud-service-providers-21-09-2016.pdf

Open Source On Azure: https://azure.microsoft.com/en-us/overview/choose-azure-opensource/

Azure Pricing Calculator: https://azure.microsoft.com/en-us/pricing/calculator/

Azure SLAs documented at https://azure.microsoft.com/en-us/support/legal/sla/

Services Available in Azure India DC : https://azure.microsoft.com/en-us/regions/services/

<End of Document>