Atul Gupta

Mobile#: *9646000509*

Email-id: [*guptaat033@gmail.com*](mailto:guptaat033@gmail.com)

### Professional Profile

Experience of around 16 + yrs. in IT and an expert in Big data and in cloud solutions architecture designs, Cloud Security, Devops (troubleshooting expert) and Network planning, automating cloud deployments and implementing IaaS, PaaS, SaaS cloud service solutions. Led many projects from the Proof of Value stage to full maturity where insight is used to automate business decisions or is integrated into business intelligence platforms at the enterprise level.

## Education

M.S in Software Systems from BITS, PILANI with CGPA 7.96

B.E in computer science with 71% from GCET under NIIT/REC

***Technical Skills:***

* ***AWS*** *– Redshift, RDS, SQS, SNS, AWS IOT, API Gateway, Polly, Lex, Rekognition, Comprehend, S3, EMR, Lambda, Kinesis, DynamoDB, CloudFront, CloudWatch, Cloud Formation, Cloud Pipeline, CodeBuild, CodeDeploy, CodeStar, AKS, ECS*
* ***GCP****- Dataflow, Pub/Sub, GCS, Spanner, Big Query & Apache Beam, Deep learning development with TensorFlow, Cognitive Linguistics, AI chatbots, NLP, Cloud Build, Artifact Registry, Spinnaker, Tekton, GKS*
* ***Azur****e – App service, Azure Functions, AKS, Azure container Instances, Azure AD, Time Series Insights*
* **Big data tools & technologies** – *StreamSets, Airflow, Druid, SparkML, Apache Spark, Apache HBase, Apache Flume, SQOOP, Apache Kafka & Avro, MongoDB, Hive, Neo4j, NIFI, Cassandra, Elastic Search, Kubernetes, Docker, Sckikit-learn, NumPy, Python & Scala*
* *JIRA, Chef, Puppet, TeamCity, Jenkins, GIT, NEXUS*

***Profile summary include but not limited to:***

* Expertise in leading design of big data systems, SOT analysis, design alternatives, production tech-stack selection & Implementation
* Senior Data Lake Architect & strategic advisor for Big data tools & Technologies
* Working closely with customers on their pressing issues and helping them in resolving complex technical and document sols
* Worked across technical & business teams to harmonize data asset
* Expert in the platform build & multi-tenant architecture using cutting-edge technologies, ROI analysis & re-engineer business processes
* Expert in Big Data capability building, COE management & managed services for Key Big Data Accounts
* Expertise in cloud and has extensive work-exp there and extensively & advocated for the best practices & proper design ensuring successful delivery
* Implement large scale infra using CI/CD pipeline and industry best practices
* Design and generate code following software coding and design standards, application solution documentation and established practices that are consistent with the overall sys arch
* Design IaaS, PaaS and SaaS Solutions

# ***Deloitte: Certified Senior Architect /Specialist Master***

# Environment: AWS, GCP, Azure, Apache Beam, Kafka, Airflow, Replicator, Dremio, Alluxio, Minio, Codepipeline, GitHub, Jenkins & Zenko

# Multi-cloud (AWS, GCP, Azure) data acquisition & Ingestion

# Multi-cloud config driven data curation and conformance

# Multi-cloud data storage and cross-cloud data synchronization using Kafka Replicator

# Data Serving and Distribution using Dremio

# Design & implementation of Next-Gen Pricing model

# Implementing real time ops & analytics datastore on GCP

# Cost assessment & extrapolation across the clouds

# plugging model into the application & executing predictions

# Led to the cross-functional teamwork atmosphere using agile delivery lifecycle

# Continuous delivery using codepipeline and integrating with GitHub

# Environment: Pyspark, AWS S3, EMR, Lambda, RDS Postgres, Redshift, StreamSets, ABC framework, CodeDeploy , Nexus Airflow

# 

# Pipelines development using StreamSets SDC and control hub

# Data lake development using AWS s3, RDS Postgres (ERP) and Redshift (Derived Store)

# Config driven Pyspark framework End-to-End set-up & ETL scripts development

# Customized for the basic ABC framework inclusion

# Data validation checks on sample datasets

# Drive data quality across the product vertical and related business areas

# Automating deployment using CodeDeploy and automated upgrades

# End-to-End testing using ABC framework

# 

# ***Ernst & Young: Senior Big data lake Architect***

Environment: Gcp, Big query, Gcs Bucket, G-Cloud Function, Apache Beam, K8S, Docker, Cloud Dataflow, Cloud Shell, Terraform, Cloud Build, Gsutil, Dataproc, Vm Instances, Cloud Sql, MySQL, Posgres, Artifact registry, Spark, Hive, Spark-Sql

# Built and architected multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation in GCP

# Developed apache beam transformations scripts and executed it in cloud Dataflow to run Data validation between raw source file and Big query tables

# Scala and spark based configurable framework to connect common Data sources like MYSQL, Postgres, and load it in Big query

# Process and load bound and unbound Data from Google pub/sub topic to Big query using cloud Dataflow with Python

# Setup Kubernetes Cluster, Docker Image for running jar Files to load data to Data Lake

# Developed Terraform plan using cloud build runs

# Worked on custom classification & product reviews for automatic categorize customer documents

# Push and Pull images using Artifact Registry

# 

# ***Standard Chartered Bank: Asst Vice President for data delivery***

Environment: Azure Data Factory, Blob Storage, Apache NIFI, HBase, Spark, SparkML, Azure Cloud Services, SQL Azure, Active Directory, Data Factory, Data Lake,K8S,Terraform,ACS

# Worked on COLT & lifecycle transformation model

# Worked on moving data from apps from 1500 TP systems to Data Lake

# Worked on Data Factory Editor, to create linked services, tables, data sets, and pipelines by specifying JSON definitions for these artifacts

# Built a solution using Azure managed services StorSimple and Blob storage to archive on-premises data to cloud

# Led migration of Virtual Machines to Azure Virtual Machines for multiple global business units

# Prepared capacity and architecture plan to create the Azure Cloud environment to host migrated IaaS VMs and PaaS role instances for refactored applications and databases

# Worked with Terraform Templates to automate the Azure Iaas virtual machines using terraform modules and deployed virtual machine scale sets in production environment

# Deploying (K8s) cluster with Azure Container Service (ACS) from Azure CLI and utilized Kubernetes and docker for the runtime environment of the CI/CD system to build, test and deploy

# Lead a team of 18-20 software professionals

# Analytics based on customer transactions & implementing event processing logic

# Build NRT sys handling data up to few mins & the related reporting

# 

# ***United Health Group(OPTUM): Senior DataLake Architect***

# Environment: Apache KAFKA 0.9, Sqoop2, Apache HBase1.0 , Hive2.0 , IBM CDC, ECG and Spark 1.6, DF2.0 framework , Talend MDM,GitHub,Apache Atlas and Ranger

# Data fabric 2.0 is used for building a tenancy model which would get data from across several UHG data sources and ingesting them in a BDPaaS environment for consumption by various users and applications.

# As a part of core team involved in standard functions to Framework Mapping

# Involved in Data repository organization Framework

# Execution Framework provisioning

# Assess and interpret customer needs and requirements as a SA

# Involved in Designing and Development of Data Fabric2.0 Framework

# Providing functional and technical expertise in migration from DF1.0 to DF2.0 project

# Involved in ETL, designing, provisioning, security module and Data lake creation and managing junior team-mates around 10

# Maintain Technology assets, and the applicability and capability of each

# 

# ***General Electric: Senior Big DataLake Architect***

Environment: Druid0.9.1, Apache Hive 2.0, Apache HBase 1.0, Sqoop 1.4.6,Kafka, HVR, Sparkline

& Apache Spark 1.6

# Designed and Implemented Consumption/Serving Layer using Alitbase HDB & Druid+Spark.Lead Corporate Data Lake

# platform architecture and maintenance for the Finance IT Data COE, partnering with IT Program managers, client

# leaders, end-users and business stakeholders, to drive simplification and IT operations SLAs

# Lead end to end solution architecture for entire Finance Data lake, which comprise of 40+ ERP systems integrated into the data lake

# Finance Data Lake 2.0 design and implementation based on Lambda architecture with Hortonworks Data Platform 2

# Managed a team of 3 Big data Engineers

# Designed and Implemented Data Ingestion pipeline using Kafka and HVR

# Designed and Implemented Batch Processing Layer on Talend + Spark

# 

# ***Infosys Tech Ltd: Solution Architect/Lead***

# Environment: AWS S3, EC2, ELB, Cloud watch & SimpleDB

# Deploying, managing, and operating scalable, highly available, and fault tolerant systems on AWS

# Migrating an existing on-premises application to AWS

# Implementing and controlling the flow of data to and from AWS

# Lead a team of 15 SE & SSE

# ***MIET(EMC) : Consultant***

Environment: Core Java, Web Services, UML, Enterprise Architecture, WebLogic,

Hibernate, XMLBean, MQ Series, Sql Server.

# Document management, XML integration and development of multi-tier applications

# Project management, technical coordination and coaching

# Functional analysis, design, implementation in Microsoft, Java, and BPM

# Integration, deployment in production and support to the end user

# ***Certifications:***

# Certified Big data developer and Specialist

# Trained broadly on Big data tools & technologies

# IBM Certified SOA Associate [2008]: Experience of Articulating the value of a SOA solution in the context of the business strategy

# IBM Certified Solution Advisor - Cloud Computing Architecture V3

# ITIL Certified in IT Service Management:

# Knowledge of complete IT Service Incident, Change and Problem management

# Certified Advanced Database Administrator for Linux, Unix and Windows-734

# Certified Database Administrator for Linux, Unix and Windows ~DB2 V9 :731

# Certified Database Associate for Linux, Unix and Windows ~ ~DB2 V9 :730

# ***White Papers***

# Streaming in open world: Presented to the BU and was circulated to various project groups in the organization

# CDC tuning approach paper to potentially reduce latency using Kafka

# Paper on extending approach to File-based systems with similar objectives of reducing latency and keeping costs minimal (open source)

# ***Rewards and Recognitions:***

# Awarded with several BRAVO Awards with Quarterly Best Performance

# Many Appreciations from client regarding several automations performed

# Won various accolades for Best Resolve Rate, Highest Revenue Generator and Customer Satisfaction.

# 

# ***Declaration***

I hereby declare that the information provided above is true to my knowledge and belief.

Atul Gupta