



Shubham Mallick



+91-9734389354



shubham.kolkata@gmail.com



<https://www.i-shubham.in/>



<https://www.linkedin.com/in/i-shubham/>



Hyderabad, India



USA Business Visa (B1/B2)

Qualcomm Hackathon, Innovation Maestro, & Super-Qualstar Winner | Apple/PwC Awardee | Microsoft/IBM Certified Professional



Experiences:

Present

salesforce

July 2022



June 2020

Qualcomm

November 2017

IBM

December 2016



October 2014

Cognizant

April 2012



Educations:

- ❖ Master of Technology in Data Science & Engineering from BITS, Pilani.
- ❖ Bachelor of Technology in Computer Science & Engineering from W.B.U.T.
- ❖ Higher Secondary (12th) from W.B.C.H.S.E. with 84.10%.
- ❖ Secondary (10th) from W.B.B.S.E. with 85.37%.

Profile Summary:

- 13+ years of extensive experiences in architecting scalable & distributed systems, and skilled in Software/Backend Engineering with Big-Data Analytics, Cloud & Machine Learning / Gen AI.
- An experienced team leader (Led a Tech team of 12 members of Supply Chain CPT project at Apple), with excellent interpersonal skills, ability to work independently with multiple clients' simultaneously.
- Decomposing monolith application to microservices [API] based architecture using Docker/Kubernetes.
- Building ETL/data pipeline application (including streaming data) in Big-Data ecosystem using Apache-PySpark & Kafka. Creating schema-oriented Parquet files and load them into HIVE (external/internal table) and MapR DB using Python. Created multithreaded script for PySpark parallel processing.
- Experienced in delivering products based on Linear/Logistic Regressions, Classifications, Lasso, Ridge, Decision Tree, XGBoost, ARIMA Time Series, Neural Network, GAN, Text Mining, K-Means Cluster.
- Built ML models on image data using Tensorflow with Nvidia GPUs, and Text mining for value insight.
- Expertise in building application / data-pipeline which can handle huge Volume of data, converting business requirements into technical solutions. Data visualization in web apps using JavaScripts, HTML.
- Handling unstructured & semi-structured data in XML, JSON using Kafka-Spark streaming.
- Passionate, positive, can-do attitude, and willing to take ownership of problems for full resolutions.

Achievements:

- **Hackathon winner** at Salesforce, and best CoP team awarded.
- Awarded for **significant contributions to successful go-live of Apple's Supply-Chain project**, in 2021.
- Stood in **Top 3** in **Qualcomm Worldwide Online Machine Learning Contest** in 2019.
- Won the **WINNER** title of **Innovation Maestro 2018-19** in Qualcomm OneIT, based on **ML**.
- **Qualcomm OneIT HaQkathon 2018** winner, got chance to present my prototype to **Qualcomm CxO**.
- Went to final round in **Microsoft AI Challenge India 2018**, which was based on Bing search engine.
- **Super-Qualstar** awarded for rapid prototypes & executions on innovative ideas in Qualcomm.
- **Qualstar** awarded for rapid prototyping on an innovative idea using **Python & NLP** in Qualcomm.
- **Qualstar** awarded for optimized **QCT Forecast Bias Model** development using **Python & Sklearn**.
- **QBuzz 2018 finalist**. QBuzz is a platform organized by Qualcomm for new Innovative Paper-Presentation in cutting-edge technologies. I presented an innovative idea in AI/ML domain.
- **Certificate** for my contribution to **OneIT Innovation Program** using Machine Learning in Qualcomm.
- **Microsoft Certified Data Warehouse professional**, *Data Warehouse in Microsoft SQL Server* - (70-463).
- **Client Appreciation Certificate** awarded for outstanding performance in Data Migration in PwC.
- **Appreciation Certificate "KUDOS"** awarded for quick successful DM deliverables in PwC.
- **"STAR Performer"** awarded for outstanding performance in **Data Migration** project at PwC.
- Received **Client Appreciation Certificate** for outstanding performance in Data Analytics in PwC.

Skills

Project	Big Data	MapR-Hadoop, Spark, HIVE, Parquet, HDFS, NoSQL, MongoDB, Databricks, Delta-Lake, Kafka, Flink, Snowflake, YARN
	Backend Engineering	Microservices, API, Django, FastAPI
	Machine Learning	Supervised/Unsupervised Learning, Regressions, Classifications, Neural Network, TimeSeries, TextMining, NLP, LLM, Tensorflow/Keras, Scikit-Learn
Tools & Technology	Programming Language	Primary: Python Secondary: Javascript, Rust, C#, Java, GoLang, R, Shell, Perl
	Cloud Technology	Apple Cloud: Kafka, Redis AWS: EMR, EC2, Spark, S3, RDS, EKS, Docker, Athena, Airflow Azure Cloud: Blob Storage, Cloud SQL, Service Bus, C#
	Framework	Django, Flask, FastAPI, Rust-Tokio, Bootstrap
	Business Intelligence	SSIS / SQL Server Data Tool, SSRS, MSBI/ETL/BIDS, Data Quality Services (DQS), Data Cleansing
	Other Tech	R-Studio, Tableau/QlikView, ServiceNow, JIRA, SSAS-Cube, MDX
	Tools	GitHub, Visual Studio, Pycharm



Work Experiences:



July 2022 - Present

Project: VDP (Security), Sales AI Analytics	Role: Module Lead/Owner, Performance Engineering
Tools/Technology: AWS, Big-Data, Spark, Kafka, Redis, Python, Go, Airflow, Gen AI, MySQL/NoSQL, ML	
Details: <ol style="list-style-type: none">Agile Collaboration: Worked with stakeholders and teams to gather requirements and deliver solutions.Performance Engineering: Enhanced app performance, Generative AI & Prompt Engineering, using JMeter.Vulnerability Management & Ticketing Systems: Built high-volume AWS applications and architecting a distributed ticketing system with Kerberos security.Application Redesign: Proposed and implemented a redesign with data streaming, conducted a PoC, and presented the solution to leadership.Data Pipelines & Optimization: Built and optimized pipelines with Spark, Airflow, Kafka, AWS-S3, and NoSQL databases to reduce latency.CI/CD & AI Models: Automated deployments with EKS, Terraform, Jenkins, and built LLMs and forecasting models to improve sales.	



June 2020 - July 2022

Project: Global Allocation (Supply Chain), CPT (Capacity Planning Tool), MDAT
Role: Tech/Team Lead, Module Owner
Tools/Technology: AppleCloud, AWS, Python-Django, Kafka, Redis, Spark, Parquet, NoSQL, ETL, Snowflake
Details: <ol style="list-style-type: none">Gather business requirements in AGILE/SCRUM methodology & implementation.Led a team of 12 members on technology stack in Supply-Chain domain [DC-CPT] and mentoring juniors/peers/cross-team members. Building custom ETL solutions using Spark.Architected backend data engineering pipeline using Spark/EMR/AWS from RDBMS/Oracle.Migrated current ETL process to Apache Airflow and AWS as per business requirements.Developed scalable Kafka queues in Apple Cloud for data streaming throughout applications.Built/Developed Analytics layer of application for data manipulation, used Redis cache.Migrated data-lake from Oracle to AWS, using Snowflakes, Spark, HIVE, AWS etc.Experienced in transformation of a monolith into microservice [API] based architecture using Docker-Kubernetes-Python-Django, deployed in AWS-EKS.Led code reviews & documentation, collaborating with business stack-holders, cross-teams, and make the product delivered.Hired people, built team for custom solutions.



November 2017 - May 2020

Project: ECAP (Engineering Compute Analytics Platform)	Role: Team Lead, Module Owner
Tools/Technology: Machine Learning, NLP, Hadoop, Spark, Python, HIVE, Kafka, Snowflake, Netezza	
Details: <ol style="list-style-type: none">Demonstrated strong development processes in Agile. Also playing role as a Scrum Master.Built ML models on image data using Keras, applied GAN. Training Neural models & tuning hyper parameters like Learning Rate, Epochs, Batch Size, Activation Function etc.Built ML models for QCT Chipset forecasting using regressions (Lasso, Ridge, GradientBoosting, XGBoost, DecisionTree) from scikit-learn, & plot using matplotlib, plotly.Parameter tuned of ML models using GridSearchCV & RandomSearchCV.Data Masking by Data Dictionary, Data Wrangling collected from Heterogenous sources.Process and resources (CPU, RAM etc.) optimizations for Chipset in RTS using ML algorithms.Experienced in building solutions using Columnar databases, scraping data from QlikView dashboards, and loading into HIVE using Py-Spark. Worked on Log-Stash from 'ELK' stack.Built Big-Data-Pipelines in Hadoop Eco-System using Spark & MapR-DB for ML models. Loading streaming data into MapR-DB in HDFS using Spark. Creating schema-oriented Parquet files, HIVE (external/internal table) using Beeline/Spark. Set-up message-brokers using Kafka.Replaced ETL tool (like Informatica) using Apache-Spark data pipeline, scheduled on YARN.Experienced in NLP models using Word2Vec, Glove, BERT models.Collaborated with cross functional teams, mentoring interns, hiring people, building team.	



December 2016 - November 2017

Project: Project Oaktree (Azure Cloud Migration)	Role: Team Lead & SME
Tools/Technology: Azure Cloud, Data Factory, SQL, Web ServiceBus, WCF, JavaScript, SSRS, MongoDB	
Details: <ol style="list-style-type: none">Gather business requirements in AGILE/SCRUM method & design Data Model.Data migration to Azure Cloud DB from heterogeneous legacy source systems, performing DBA tasks & end-to-end ETL solutions at Azure Cloud, along with DBA activities (Back up strategy and SQL jobs, Recovery mode, Performance tuning, Index rebuilding, Availability Groups).Created NoSQL (MongoDB) database, migrate DataMart from SQL Server to MongoDB using SSIS custom scripts, and process for Data Analytics.Deployed Azure Service Bus using C#.NET to interact with Broker message from client app, create C# WCF web services interact with Cloud Database & deploy in Azure.Led a team of 6 members and mentoring them to make industry ready.	



October 2014 - December 2016

Project: GCP-GDW & IPSR	Role: Senior BI Developer
Tools/Technology: SQL, SSIS, ETL, DBA, C#, JavaScript, TFS, Free Text search, MongoDB, Python	
Details: <ol style="list-style-type: none"> 1. Data Warehouse migration in AGILE/SCRUM methodologies, understanding business requirements & design Data Model for SQL Server Data Warehouse. 2. Predictive search/Free-Text Search (like Google search) from billions of rows in just 2-3 seconds from .NET application using C# web services & jQuery. 3. Created & deployed SSIS packages for Data Migration from heterogeneous source to SQL Server and provide end-to-end ETL solution. Performance tuning of SSIS (with SQL tuning) to load 50 million rows within limited time from remote data source to SQL server. 4. Designed Data-Mart and loaded data from SQL Server DW through SSIS packages, developing SSRS and PowerBI report for Data Quality check, after & before Migration. 5. DBA tasks (Back up strategy and SQL jobs, Recovery mode, Performance tuning, Index rebuilding, Availability groups), Integrate DQS (Data Quality Service) with SSIS. 6. Developed C# WCF web services to expose data to external clients. 	



April 2012 - October 2014

Project: MetLife USBPM	Role: Application developer
Tools/Technology: Visual Studio, SQL Server, T-SQL, SSIS, C#, ADO.NET, JavaScript, jQuery	
Details: <ol style="list-style-type: none"> 1. Production Support project of 24X7 hours for MetLife, handling bulk production data & process to SQL server through SSIS packages. 2. Monitoring of running live Production jobs through IBM Tivoli/Maestro, solving issues on Ticketing System on priority basis. 3. Automation of existing process to minimize day-to-day manual efforts & errors. 4. Upgrade of existing SSIS packages, tuning existing Stored Procedures, solving Production issues with priority as per Business tickets. 	

Additional Responsibility:

- Appointed as Mentor for newly joined folks for internships, and lateral joiner to Qualcomm, Apple, PwC.
- Active member of Core technical interviewers' panel for lateral hiring, and building team at Salesforce, Apple, Qualcomm, IBM & PwC.
- Appointed as a trainer for fresh campus hired graduates/associates on technical competencies (Analytics, SQL, Data-Warehouse, DB Migration, ETL), & received appreciation for quality training delivered at PwC.

Academic Project:

- **Online Job Portal System – University Final Year Project**
 - This is a web-based application developed in JAVA and Oracle 9i in 2011.
- **Online Air Line Reservation System – IETE Training Institute (Summer Intern)**
 - This is a web-based application developed in JAVA and Oracle 9i in 2011.
- **Computer Retail Business Management – IBM Certified Academy**
 - This is a SQL/PL-SQL application developed in Oracle 9i in 2010.

Academic Details:

Qualification	University/Board	Marks Obtained				
M. Tech	BITS, Pilani	7.67 CGPA				
B. Tech	W.B.U.T.	7.86 CGPA				
12 th	W.B.C.H.S.E.	MATH	PHYS	CHEM	BIOS	Overall
		195/200	152/200	167/200	165/200	84.10%
10 th	W.B.B.S.E.	Math	Ph. Sc.	Life Sc.		Overall
		91/100	95/100	88/100		85.37%

Interests:

- Hadoop, Big-Data, Machine Learning, Block-Chain, Cryptocurrency, Data Warehouse, Data Analytics, Business Intelligence, Python, R, Data Mining, Blog writing

Personal Details:

- DOB : September 11
- Gender : Male
- Language Known : English, Hindi, Bengali