



# 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

pwc

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:

- 12+ 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	<b>Primary:</b> Python <b>Secondary:</b> Javascript, Rust, C#, Java, GoLang, R, Shell, Perl
	Cloud Technology	<b>Apple Cloud:</b> Kafka, Redis <b>AWS:</b> EMR, EC2, Spark, S3, RDS, EKS, Docker, Athena, Airflow <b>Azure Cloud:</b> 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

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



June 2020 - July 2022

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



November 2017 - May 2020

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



December 2016 - November 2017

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



October 2014 - December 2016

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



April 2012 - October 2014

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

## 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 <sup>th</sup>	W.B.C.H.S.E.	MATH	PHYS	CHEM	BIOS	Overall
		195/200	152/200	167/200	165/200	84.10%
10 <sup>th</sup>	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