



# 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



December 2016



October 2014



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%**.

## Achievements:

- **Hackathon winner** at Salesforce, and best CoP team awarded, in Oct-2022.
- Awarded for **significant contributions on 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.

## Profile Summary:

- **Leadership and Interpersonal Skills:**
  - ❖ Led a team of 12 members in a Supply Chain project at Apple, 6 members at IBM.
  - ❖ Excellent interpersonal skills and ability to work independently with multiple clients simultaneously.
- **Big Data and Analytics:**
  - ❖ Developed ETL data pipeline applications (including streaming data) in Big Data ecosystems using Apache PySpark and Kafka. Exposure in Impala, and Flink.
  - ❖ Created schema-oriented Parquet files and loaded them into HIVE (external/internal tables) and MapR DB using Python. Built multi-threaded scripts for PySpark parallel processing.
  - ❖ Expertise in SQL and OLAP databases such as Delta Lake, Snowflake, and Redshift.
  - ❖ Having exposure with Analytics / data-visualisation tools such as Tableau, Power-BI.
- **Machine Learning and AI:**
  - ❖ Delivered products based on Linear/Logistic Regressions, Classifications, Lasso, Ridge, Decision Tree, XGBoost, ARIMA Time Series, Neural Network, GAN, Text Mining, and K-Means Clustering.
  - ❖ Built ML models on image data using TensorFlow with Nvidia GPUs and performed text mining for value insights. Experience in AI-driven data analytics.
- **Software and Backend Engineering:**
  - ❖ Decomposed monolith applications into microservice/API based architecture using Docker/Kubernetes.
  - ❖ Handled unstructured and semi-structured data in XML and JSON using Kafka-Spark streaming.
- **Data Architecture and Integration:**
  - ❖ Expertise in huge volume data migration/integration and implementing Data-Warehouse repositories using ETL tool [SSIS] and SQL Server.
  - ❖ Provided end-to-end ETL solutions with deployment to Azure Cloud, data analysis, and SQL DBA.
  - ❖ Managed Azure API Management and Azure Web Apps, converting business requirements into technical solutions. Data visualization in web apps using JavaScript and HTML.
- **Technical Proficiency:**
  - ❖ Proficient in programming languages such as Python, Scala, Rust and similar.
  - ❖ Strong experience in developing data products, including data marts, data semantic layers, analytics, and third-party system integrations.
- **Strategy and Communication:**
  - ❖ Worked closely with higher management, business, and product stakeholders to understand BI & AI-driven analytics needs. Created data strategy, architecture, and roadmap.
  - ❖ Assisted in the preparation of presentations and formal analyses in support of executive decision-making and strategy development. Promoted a culture of experimentation and data-driven innovation.
- **Additional Attributes:**
  - ❖ Passionate, positive, can-do attitude, and willing to take ownership of problems for full resolutions.
  - ❖ Strong interpersonal, communication, and presentation skills.
- **Availability and Budget Management:**
  - ❖ Took on responsibility for platform's 24x7 support availability.
  - ❖ Oversaw budgetary concerns, including cloud expenditure.



## Work Experiences:



July 2022 - Present



June 2020 - July 2022



November 2017 - May 2020



December 2016 - November 2017

<b>Project:</b> VDP (Security Domain), Sales AI	<b>Role:</b> Module Lead
<b>Tools/Technology:</b> AWS, Big-Data, Spark-Kafka, Redis, Python, Airflow, MySQL/NoSQL, ETL, AI, ML, Ops	
<b>Details:</b> <ul style="list-style-type: none"><li>i. Gather business requirements in Agile/Scrum, collaboration with cross-team &amp; stakeholder.</li><li>ii. Building <b>vulnerability management applications</b> over high volume of data, and deploy in AWS.</li><li>iii. Owning a module to rebuild the ticketing application with distributed systems with AWS through Kerberos security. <b>Translating customer needs into technical solution.</b></li><li>iv. Proposed <b>redesigned</b> app, added <b>data streaming</b>, did <b>PoC</b>, &amp; showed a <b>demo to leadership</b>.</li><li>v. <b>Architecting &amp; re-designing</b> data processing engines/pipeline to <b>reduce</b> overall latency/runtime.</li><li>vi. Developing <b>distributed data platforms &amp; pipelines</b> using <b>Spark DAG, Airflow, Kafka, AWS-S3, NoSQL databases [document DB, key-value pair, columnar database, Graph-DB] &amp; Grafana.</b></li><li>vii. Using CI-CD pipeline, deployment through <b>EKS, Kubectl/K8s, Terraform, Jenkins</b> etc.</li><li>viii. Building customs LLM &amp; forecasting models, focusing on increasing Sales based on requirements.</li></ul>	

<b>Project:</b> Global Allocation (Supply Chain Integration), CPT (Capacity Planning Tool), MDAT
<b>Role:</b> Tech/Team Lead, Module Owner
<b>Tools/Technology:</b> Apple Cloud, AWS, Python-Django, Databricks, Kafka, Redis, Spark, NoSQL, ETL, Snowflake
<b>Details:</b> <ul style="list-style-type: none"><li>i. Gather business requirements in <b>AGILE/SCRUM</b> methodology &amp; implementation.</li><li>ii. <b>Leading 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>iii. <b>Architecting backend data engineering pipeline</b> using <b>Spark/EMR/AWS</b> from RDBMS/Oracle.</li><li>iv. Migrating current ETL process to <b>Apache Airflow</b> and AWS as per business requirements.</li><li>v. Developing scalable <b>Kafka queues</b> in <b>Apple Cloud</b> for <b>data streaming</b> throughout applications.</li><li>vi. Building/Developing Analytics layer of application for data manipulation, used <b>Redis cache.</b></li><li>vii. Migrating data-lake from Oracle to AWS, using Snowflakes, Spark, HIVE, AWS etc.</li><li>viii. Experience in <b>transformation of a monolith into micro-service [API]</b> based architecture using <b>Docker-Kubernetes-Python-Django</b>, deployed in <b>AWS-EKS.</b></li><li>ix. Leading code reviews &amp; documentation, collaborating with business stack-holders, cross-teams, and make the product delivered.</li><li>x. Hiring people, building team for custom solutions.</li></ul>

<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> <ul style="list-style-type: none"><li>i. Demonstrating strong development processes in <b>Agile</b>. Also playing role as a <b>Scrum Master</b>.</li><li>ii. <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 using Kafka</b>.</li><li>iii. Building <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>iv. Building 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>v. <b>Parameter tuning</b> of ML models using <b>GridSearchCV &amp; RandomSearchCV</b>.</li><li>vi. <b>Data Masking</b> by <b>Data Dictionary, Data Wrangling</b> collected from <b>Heterogenous</b> sources.</li><li>vii. Process and resources (<b>CPU, RAM</b> etc.) optimizations for <b>Chipset</b> in <b>RTS</b> using <b>ML</b> algorithms.</li><li>viii. Experiencing 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>ix. Building <b>Big-Data-Pipelines</b> in <b>Hadoop Eco-System</b> using <b>Spark &amp; MapR-DB</b> for <b>ML models</b>.</li><li>x. Replacing <b>ETL</b> tool (like Informatica) using <b>Apache-Spark data pipeline</b>, scheduled on <b>YARN</b>.</li><li>xi. Experienced in NLP models using <b>Word2Vec, Glove, BERT</b> models.</li><li>xii. Collaborating with cross functional teams, <b>mentoring</b> interns, hiring people, building team.</li></ul>	

<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> <ul style="list-style-type: none"><li>i. Gather business requirements in AGILE/SCRUM method &amp; <b>design Data Model</b>.</li><li>ii. 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>iii. Create <b>MongoDB</b> database, <b>migrate DataMart</b> from SQL Server to MongoDB using SSIS custom scripts, and process for Data Analytics.</li><li>iv. Deploy <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 &amp; deploy in Azure</b>.</li><li>v. Leading a team of 6 members and grooming juniors to make them industry ready.</li></ul>	



October 2014 - December 2016



Cognizant

April 2012 - October 2014



## Additional Responsibility:



## Academic Project:



## Academic Details:



## Interests:



## Personal Details:

<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>	
i.	Data Warehouse migration in AGILE/SCRUM methodologies, understanding business requirements & <b>design Data Model</b> for SQL Server Data Warehouse.
ii.	<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 & jQuery.
iii.	<b>Create &amp; deploy 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.
iv.	<b>Design Data-Mart</b> and load data from SQL Server DW through SSIS packages, developing <b>SSRS</b> and <b>PowerBI</b> report for Data Quality check, after & before Migration.
v.	DBA tasks ( <b>Back up strategy and SQL jobs, Recovery mode, Performance tuning, Index rebuilding</b> , Availability groups), Integrate <b>DQS (Data Quality Service)</b> with SSIS.
vi.	<b>Developing C# WCF web</b> services to expose data to external clients.

<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>	
i.	Production Support project of 24X7 hours for MetLife, handling <b>bulk production data &amp; process to SQL server through SSIS</b> packages.
ii.	<b>Monitoring of running live Production jobs through IBM Tivoli/Maestro</b> , solving issues on <b>Ticketing System</b> on priority basis.
iii.	<b>Automation</b> of existing process to minimize day-to-day manual efforts & errors.
iv.	<b>Upgrade of existing SSIS packages, tuning existing Stored Procedures</b> , solving Production issues with priority as per <b>Business tickets</b> .

- Appointed as Mentor for newly joined folks for internships, and lateral joiner to Qualcomm.
- 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 from India's premier institutes on technical competencies (Analytics, SQL, Data-Warehouse, DB Migration, ETL), & received appreciation from Executive Director for quality training delivered at PwC.

- **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.

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%

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

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