

Role: Sr. Data Engineer.
cum Data Architect)
Experience: 4+ years
Updated: 21st/Jan/22

Burhanuddin Bhopalwala
Mobile: +91-7000577856
burhanuddinbhopalwala.cse@gmail.com
<https://github.com/burhanuddinbhopalwala>



SUMMARY

Tech-Charger || Data-Enthusiast || GSoC 2017-18 & Mentor || 5x AWS Certified || Data Engineering & MLOps || AWS Big Data Certified || AWS Database Specialty Certified || AWS Solutions Architect Certified
5+ Publications on [Medium Towards Data Science](#) || 3+ Publications on [DEV Community](#) || Over 700+ Software Developers & Data Engineers followers WW on Medium & DEV Community || [upGrad](#) Data Engineering & Data Science/Streak News Letter Content writer - [here](#) || Certification and Supporting docs [here](#)

EDUCATION

- Graduate in Software Development Focused Data Engineering & Machine Learning (MLOps) from Electronics & ICT Academy, Indian Institute of Technology (IIT) Roorkee, India in association with CloudX Lab in 2017-18
- B. Tech in Computer Science Engineering (CSE) from Jaypee Institute of Information Technology (JIIT), Noida, Uttar Pradesh, India. Aggregate GPA: 3.7 / 4.0 (Graduated in 2014-17)

ACHIEVEMENTS & CERTIFICATIONS

- In GSoC 2017 - Proposed & Implemented Integrated Rule-Oriented Data System (iRODS) Client or plugin for ImageJ for International Neuroinformatic Coordinating Facility – Belgium in 2016-17
- CODECHEF Certified (A Directi Global Programming Platform for Competitive Programming WW) DATA STRUCTURES & ALGORITHM PROGRAMME (CCDSAP) 2018 – Successfully passed the certification exam in 2018
- AWS Certified SOLUTIONS ARCHITECT (95%) certified in 2020
- AWS Certified BIG DATA & ANALYTICS – Specialty (85%) certified in 2020
- AWS Certified DATABASE & SECURITY – Specialty (90%) [DevSecOps] certified in 2021
- Graduate Aptitude Test in Engineering Exam (GATE) 2017 Qualified with percentile above 99.9%
Certification and Supporting docs [here](#)

SKILLS

- Comprehensive knowledge of Data Structures, Algorithms, Computer Networks, Databases, Operating Systems as Qualified GATE 2017 with 99.9% percentile
- Proficient Languages: Python 3, Ruby, JavaScript (ES6), Golang, Scala
- Data Engineering (DE) Skills:
 - Basic/Supporting DE Skills
 - Software Development - System Design, Solutions Architect, SDLC, Agile (JIRA), Linux + Git
 - Backend Development - SAML/OAuth2, FaaS serverless frameworks (Ruby/Jets), Client/Server-side Caching, HTTPs req/res architecture, Distributed tracing & Incident response/OnCall systems (PagerDuty)
 - Core Data Engineering & Cloud Skills
 - Open-Source Big Data Engineering: ETL Data pipelines using Debezium Change Data Capture (CDC), Elasticsearch, Kibana (ELK) Stack, Real-time & Batch Streaming using Apache Kafka, Apache Superset, & other Open Source MPP Database PostgreSQL. Big Data processing using Apache Spark [RDD, DataFrame, DataSets], Spark structured streaming, Adv Spark operations like narrow transformations, caching, broadcast & bucketed joins, Spark MLLib
 - Cloud Data Engineering: AWS Kinesis, AWS MKS (Managed Kafka Service), AWS Glue, AWS DMS (Managed replication). Data Lake building using AWS S3, AWS Athena, AWS DynamoDB, MPP DB Redshift, AWS EMR
 - Database Administration (DBA) – Defining user level roles & permissions for dev/prod, SQL Data Modelling (indexes, partitioning, bulk inserts, windowing), Fine-tuning SQL leveraging query stats, PostgreSQL Databases for heavy production workloads by minimizing the Replication time (WAL segments), Query Planner, Connectivity, Locking & Logging

➤ **Advanced Cloud Data Engineering Skills**

- **Data Security & DevSecOps:** Proactive, Reactive and Detective controls with regards to Data Encryption in Transit, in Rest, **Versioning** & Intelligent **storage tiering** for Data Lakes on Cloud
- **MLOps, Machine Learning** (Optimization approach less Probabilistic), **Deep Learning** - Understanding of MLP, CNN, GAN, RNN, Autoencoders), AWS MLOps (model building, training & deployment using **AWS SageMaker & AWS Developer Suite (CI_CD)** & **AWS Lambda & API G/W**), **Docker & Kubernetes**, **AWS SAM & CDK frameworks (IaaS)**, ***Quick Learner**

PATENTS AND PUBLICATIONS (Data Engineering & Machine Learning, MLOps)

- **5+ publications in Towards Data Science** (Problem-solving + Big Data) and **2+ publications in DEV Community** (Software Engineering) - [Medium - Towards Data Science](#), [DEV Community](#)
- **At Fareye Logistics (2019):** PATENT is in process for Dynamic ETA processing engine – ML & MLOps Processed TRUCKS-Long Haul Journeys data using DBSCAN Clustering for Point of Interest (POI's) and final consolidation using Dynamic Programming to predict the ETA (long haul journeys/trips) from any route POI in O(1) time. MLOps: Containerized the model using docker, scaled using Kubernetes pods to expose the endpoints
- In 2017 (4th Year), Published: Real-time Image Processing using Multilayer Perceptron (MLP-CNN) using Denoising & Deblurring techniques (Super Resolution using GAN's) as feature engineering under the guidance of Dr. Satish Chandra (satish.chandra@jiit.ac.in), Associate Professor @ **JIIT**. Also PATENT is in process at National Research Development Corporation (NRDC) Delhi
- In 2016 (3rd Year), Communicated for **Institute of Engineering and Technology (IET)** journal about Model View Controller (MVC) ORM fast query optimization & extraction under the guidance of Dr. Suma Dawn (suma.dawn@jiit.ac.in), Assistant Professor (Senior Grade) @ **JIIT**
Certification and Supporting docs [here](#)

WORK EXPERIENCE

- **From [2020 - Present] Lead Cloud Data Engineer @Aramex Logistics.**

- **Big Data/Hiring & Building DE Team from scratch:** Building the team of Data & MLOps Engineers for Aramex. Being the first in the team includes setting up end-to-end rails for candidate screening, shortlisting and **interviewing with multiple tech rounds**, and managing the candidate pipeline. Closely worked with DS team to enabled framing the interview questions/coding/leadership rounds to ensure hire and develop the best
- **Big Data/Seasonal Forecasting ETL & MLOps pipeline:** Setting up the pipeline for generating forecast on demand. The underlying forecasting was leveraged by taking a mean of AWS state of the art CNN-QR Algorithm, FB Prophet, and other exponential moving average algorithms. The entire infrastructure and its deployment consist of services using **MPP database AWS RedShift along with other AWS tech such as AWS S3, Forecast, SageMaker, Lambda set up AWS CloudFormation**

As phase 2, built the end-to-end automation to save the cost, efforts. This was accomplished by building the dashboards using **AWS Quick sight** and embedded forecasting algorithm to serve the business faster and in a cost-efficient way.

- **Big Data/MLOps for Data Science Modules:** The part of the job consists of assisting the Data Science (DS) team for making the DS module production-ready to achieve ease of maintenance, scalability, and best deployment practices by doing code optimization utilizing memory and disk profiling, removing code smells and finally implementing the **test-driven (TDD) approach as per Google test pyramid** for the use case, etc
- **Big Data/ETL Data Pipeline using SQL Data Modelling - for Rolling Business Metrics:** Implementing Data Pipelines for calculating both static and rolling business metrics or KPIs. These pipelines assist the Business to generate these Metrics on-demand based on seasonality. The underlying architecture consists of a data warehouse **AWS RedShift** views and analytics pipeline using **AWS Redshift, QuickSight, DynamoDB, Lambda, SES** (for final results shipping)

- **DevOps - CI/CD:** Supporting CI_CD pipelines & migrations to **AWS Developer Suite** across all the **ML use cases at Aramex**. This is accomplished using **AWS SAM** to easily replicate the migration with ease and minimal dev efforts across all the ML microservices and use-cases

These CI / CD pipelines (s) developed were less with all the Coding standards for Python via pre-commit hooks. To name a few **Flake8 linting, AutoPep8 formatting, Bandit, Safety dependency checks, etc.**

- **AdHoc & Support Task:** Supporting the DS team for Adhoc analytics task, attending the failure events for the deployed use cases to provide RCA, Infrastructure related task

- **From [2018 SDE 2 to 2020] promoted to SDE 3 (Data Engineering) @Fareye Logistics Pvt. Ltd. Delhi NCR.** Fare ye is Fareye a major logistic firm focused on last-mile and long-haul deliveries. I took charge of backend authentication OAuths and optimization at DB and application tier. Ensuring **high production backend service & data pipeline outputs**. Set up a process to collect the failure data and choose to buy and migrate to the managed AWS cloud and implemented services/control to efficiently run and scale.

Thus, ensuring high productivity, low operation costs together with minimal output delay and resolution time. Responsible for enlightening the team of **11+ members** with best and state of the art Software practices and setting up the processes for the team and **hiring support for the team in setting up Coding tests and interviews**. Extended my role to add ML, MLOps capabilities to the product for predicting high accuracy and low latency ETAs. Key projects are as follows:

- **Big Data/Near Real-time Streaming/ETL Pipeline - using DEBEZIUM (bin log Change Data Capture (CDC)) & Kafka:** Setting up and maintaining the ETL pipeline using DEBEZIUM MySQL bin log connector and Apache KAFKA for Staging DB and ELK, later moved to managed migration service called AWS Data Migration Service (DMS) for Real-time DB Migration and CDC or On-going Replication
- **Big Data/Real-time Streaming - AWS S3 Data Lake using AWS Kinesis & Data Querying pipeline using AWS Glue Data Catalog & AWS Athena:** Architect & build S3 Data Lake from scratch using Streaming GPS Device pings data. Transforming data using AWS Lambda in near real-time and storing it into Amazon S3 and using AWS Glue Data Catalog, setting up querying data pipeline with AWS Athena for querying S3 data
- **Big Data/DBA Optimization - Bulk DB Inserts by partition query optimization for Vehicle data & Consolidations:** For problems like converging Displacement to Distance for the vehicle/trucks data when odometer km (cumulative distance reading of vehicle) plus, Algorithm for deriving the distance from previous pings (handling out-of-order data). In-memory file queue (like Redis) implementation for processing pings file in Bulk, and saving as it partitions wise
- **Big Data/DBA Optimization - Automation, Monitoring & AWS RDS/Database Optimizations:** Using a Serverless framework to shift all the CRON Jobs to AWS Lambda and Monitoring using AWS CloudWatch & Grafana & PagerDuty. RDS Maintenance includes VACUUMING, Controlling Replica Lag, optimizing long-running queries
- **Big Data/DBA Replication - SQL Fine-tuning and Partition Maintenance using AWS DMS and Lambda services:** Set up the RDS Staging for QA, Development for UAT purposes using AWS DMS and automated the monthly Partition by setting up serverless services to clean partition, performing vacuum and managing the WAL size for RDS checkpointing, WAL senders for replication & WAL compression for RDS storage optimization. Which overall leads to tuning RDS for high availability workloads at the @month end. The serverless service was set up using Ruby jets serverless framework as IaasC
- **Big Data/Dashboarding & Visualization – ELK pipeline & Anomaly Detection using Kibana Machine Learning for Anomaly Detection:** Implementation of Elasticsearch (using Debezium ETL) for KPI Dashboard for the end-users, Aggregation on unstructured schema-less JSON data. And Anomaly Detection (Point Anomaly – CPU, Memory & Network) using Kibana Machine Learning Implemented K-Nearest Neighbor's with dynamic Cluster # generation using Knead for Non-Globular Convex sets. Benchmark set as DUNN Index & # of correctly identified Anomaly for a minor Cluster
- **Big Data/Dashboarding & Visualization - Using Apache Superset, communicated and for setting up an on-premise consolidated data warehouse platform using Apache Superset:** Setting up the Apache Superset to assists the analytics team for consolidate the data collection and help them to build and visualize the data and cost-effectively. Underlying subtasks include setting up Apache Superset on auto-scaling, data seed using SQL Alchemy ORM, and building the SQL queries for smart searches
- **ML & MLOps - ML Dynamic ETA (Transit time-long haul journeys/trips) Engine using DBSCAN Clustering & Dynamic Programming (DP):** Collected data from Relational database and Amazon S3 Data Lake for each journey/trip POI and clustered them using DBSCAN Clustering and derived the ETA for each POI-checkpoint using its subsequent poi by consolidating the Clustering Result using Dynamic Programming paradigm formula in O(1) time. Model iteration (for finding minPts & eps for DBSCAN), development & deployed using **AWS SageMaker & AWS Lambda, API G/W**. Benchmarking has been

done using our Experimental formulation. PATENT in process

- **BigData/ML & MLOps - Order Delivery, Surface Delivery (for Last Mile Deliveries) prediction in near real-time using many-to-many RNN (LSTM):** A model that predicts multiple delivery predictions trained using many-to-many RNN. Feature Engineering includes Trigonometric Features as Order Deliveries and Surface Deliveries mostly behaves as sin/cos waves. Model beat the Facebook Prophet Library Model and hence the Benchmark. Trained, developed, and deployed using AWS SageMaker, AWS Lambda, and API G/W
- **BigData/ML & MLOps - Intelligent Stoppage Identification & Classification using Logistic Regression:** Forgiven a combination of routes or for particular routes and stoppages, identifying & classifying the stoppages for new routes or future journeys/trips for a particular route. Used Logistic Regression as features are Non-Collinear (Perturbation Test), and for a Low Latency solution. Later trained and deployed a Deep Learning based model using MLP and Sigmoid layer (For multi-class stoppage identification & classification)
- **Backend Architect - NodeJS Workflow Algorithm Development for Vehicle Data Consolidation & Monitoring ORCHESTRATION Engine (taken reference from UBER Cadence Orchestration Engine):** Workflow development includes consolidations in near real-time from pings-data to create duration-based-state-histories for the vehicles, POI (point-of-interest) consolidation & ALERTS (in the plant, over-speeding, hub wise, wrong routing) processing
- **Backend Architect - Leverage OAuth2 Standards for Delegating Authorization across different Microservices:** Using Authorization Code Flow & Client Credential Code Flow Implementation
- **DevOps - CI/CD (Continuous Integration/Continuous Delivery) & Containerization Backend NODE Apps:** Implemented Travis CI & Dockerize (Containerization-for auto-scaling with CI/CD for multiple team members) of Backend app using AWS ECS and currently also working to move it to Kubernetes using AWS ECR, AWS Managed Grafana and Prometheus (for monitoring)
- **From [2017 - 2018] as SDE 1 (Backend Engineering) at Fashion Jewelry E-commerce Giant Voylla Fashions (Amazon Subsidiary):** Responsible for achieving backend sprint tasks, documenting code changes, peer reviews. SDE1 is also accountable for repairing Adhoc failures, reducing defect rates, learning new techs, and always producing bug-free and quality code with actual work estimates provided. Key projects are as follows:
 - **Big Data/ETL-ELK pipeline** – Build an end-to-end first ETL pipeline to sink in order & payments data to Elasticsearch & Kibana (ELK). To help visualize the payments data, failure codes, and promotion trends. Thus, achieving higher-order success rates and reducing the promotional costs
 - **Backend Dev/Payment Gateway Implementation and Maintenance:** Implemented PAYTM, PAYU Payment Gateways, Transaction Request API, Transaction Status Check (aka Double Signature API), Refund API, Refund Status Check API for Out of Stock or Checksum failure
 - **Backend Dev/ETL Pipeline Key Order Functionalities Implementation and Maintenance:** Order Tracking, Return, and Replacement Functionality, Order Pipeline Development & Maintenance for smooth Customer experience and delight
 - **Backend Dev/True-Asynchronous Write API Implementation:** Pushing Order-ORM post to DB commit to ERP via AWS SQS to Celery as a Task Queue, which stores the Order ORM into ERPRDBMS
 - **Backend Dev/Caching Related Functionalities Implementation:** Appy hours & Admin Functionality + Building Automated Script to Purge the Cache at all the 3 Levels Memcached at Application Layer, Varnish at Web Layer and Amazon CloudFront CDN at Client Level
- **In 2015-16 – 6 months Internship at HCL as Core JAVA SE8 Backend Developer:** Implemented Desktop Application Plugin using Dropbox Core API which helps the user to upload and download the Data to and fro.

PERSONAL

- **Single, Age 26, Avid Chess and Cricket Player, Willing to travel and work**
 - **Always active in Community projects** like helping needy families like offering my services at hospitals when needed. **Assisting community older & needy members in setting up local vaccination drives**
 - **Multi-Linguistic:** Urdu (mother tongue), Hindi, Arabic (tutoring), Gujarati, English - (All working proficiency level)
-