Siddhartha Chandra

Berlin, Germany

sc3653@columbia.edu | linkedin.com/chandrasiddhartha | siddharthachandra.com | github.com/siddhartha-chandra

PROFESSIONAL SUMMARY

Data platform and backend engineer with 15+ years building large-scale systems across Berlin, New York, Mumbai, and Chandigarh. Deep expertise in Spark, Airflow, Kafka, and production ML pipelines. Built a data deduplication pipeline processing 10M+ entities daily at 10X speed, architected cross-team logistics hubs at Zalando, and led a tech incubator managing 1.3M EUR fund across 60+ startups. Currently at Zalando SE architecting fulfillment systems that serve millions of users across Europe.

CORE COMPETENCIES

Distributed Systems | Data Pipeline Architecture | Apache Spark | Apache Airflow | Kafka | Kotlin | Machine Learning Ops | AWS, GCP, Azure | Agentic Development | Monitoring & Observability | Operational Excellence | Python & Scala | SQL & NoSQL

WORK EXPERIENCE

Zalando SE

May 2025 - Present

Senior IC, Backend & Data - Berlin

- Co-authored tool for **weekly operational reviews** used by 6 teams in the department
- Architected cross-team system enabling **multi-carrier injection hub** in Logistics
- Evolved system for improved fulfilment offering, increasing avg. daily shipping by ~120 units for Netherlands & Sweden
- Architected bridge application consolidating context for fulfilment planning

Six Ladders

Jun 2024 - Present

Co-founder - Mumbai

- Established architecture and mentored developers in building a suite of educational offerings for students
- Projected 2026 yearly revenue: 0.43M EUR; MAU: 300; DAU: 100

Sardar Patel Technology Business Incubator

May 2024 - May 2025

CEO - Mumbai

- Supported ~60 tech startups with a team of 6; managed fund of 1.3M EUR
- Established 5 strategic partnerships; built technical arm from scratch

Sardar Patel Institute of Technology

Aug 2023 - May 2025

Professor of Practice - Mumbai

- Taught Data Structures (undergrad), Advanced Algorithms (MTech), Business Analytics (undergrad); advised 14 technical project groups
- Designed revised Data Structures syllabus as course convenor, along with 2 professors
- Published 5+ papers in international conferences

Collective[i]

Oct 2016 - Apr 2022

Senior Data Science Engineer - New York / Remote

- Created daily data deduplication pipeline of **Spark jobs over Airflow** for ~10M entities, unlocking **10X speed gain**
- Co-authored system for predicting win probability, close time, and customer revenue based on deal pipeline for sales representatives of B2B clients
- Built strong engineering foundations for the Data Science team: mentored developers, added dashboard monitoring, and played a critical role in legacy code transitions and technical-debt payoffs

x.ai

Oct 2014 - Sep 2016

Data Scientist - New York

- Built CRF-based email parser that outperformed Mailgun's production system by ~15% F1 on signature detection
- Creator and keeper of dashboards for measuring annotation quality, model performance, and KPIs

Infosys

Sep 2008 - May 2012

Senior Systems Engineer - Chandigarh

- Recognised for execution excellence at organizational level for implementing i18n for a client

EDUCATION

MS in Computer Science - Columbia University — 2013 - 2014

New York, United States

BE in Information Technology - Mumbai University — 2004 - 2008

Mumbai, India

SELECTED PUBLICATIONS

Advanced AI-driven Solutions for Virtual Fitting	IJIRT
Enhancing Accessibility for People with Disabilities	IJIRT
Hybrid Optimisation & Explainability-Driven Framework for Creditworthiness Assessment	IEEE
Wavelet-Enhanced Transformer Architecture for Dermatological Image Analysis	IEEE
Comparative Analysis of Optimized ML/DL Models for Credit Risk Prediction	IEEE
Revify: An Agentic Opinion Mining System for Product Reviews	IEEE

SKILLS

Languages	Python, Scala, Kotlin, Java, SQL
Frameworks	Apache Spark, Apache Airflow, Kafka, Spring Boot, Flask, Hadoop
Tools & Cloud	AWS, GCP, Azure, Docker, CI/CD, Git
Data & ML	Applied ML, NLP, CRF, Scikit-learn, PostgreSQL, MongoDB, Neo4j
Engineering	Distributed systems, Data pipeline architecture, System design, Legacy migration, Monitoring & observability, Operational excellence