

Soubhik Chakraborty

Waltham, MA 02452 • 617-906-1795 • shoubhik.c@gmail.com • LinkedIn: [soubhik-chakraborty](#) • Github: [soubhik-c](#) • Visa: F1

Summary

Presently doing Master's in Business Statistics, I have 20+ years of programming expertise developing [OLTP](#) applications and [HTAP](#) databases, architecting & implementing web scale data platforms.

Recent Education

Bentley University, Waltham, MA

Master's in Business Statistics

GPA: **3.52** May 2023

Recent Professional Experience

Big Data Consulting, Pune, MH (India)

April 2018 – August 2021

Architect

- Architected 15+ petabytes datalake constituting dense storage hardware design & object storage, fast nosql datastore, near realtime analytics pipeline and programmer in autonomous networking AI engine, onboarding 1000s of network applications on a 3-way disaster recovery data platform for Rakuten Mobile Inc.
- Scaled EFK datalake to ingest 500-1000k msgs/sec. with zero message loss.
- Feasibility study of on-prem S3 datacenter at 0.10 paise with ~1.5 PUE.
- Scaling of Nvidia Xavier NX GPU to run various video analytics.
- Kubernetes using sriov/ daemonsets & statefulsets helm charts/ cloud automation using ansible.

VMware, Pune, MH (India)

June 2017 – February 2018

Staff Engineer

- Implementation team of [Distributed Network Encryption \(NSX security\)](#).

Research Assistant Projects

- Behavioral Economics
September 2021 – present
A boardgame [BuyWord](#) developed in python
- Intragroup study of Racial discrimination
November 2021 – present
Cross validation using SPSS multivariate analysis
- Postpartum depression clinical trial of [Punjiri](#)
December 2021 – present
Design of the statistical experiment

Recognitions

- VMware Individual Excellence Award
for outstanding contributions in the India Center
- VMware Hackday Competition
Runners up – graphical navigation of distributed query plan using d3js.
- Opus Appreciation for outstanding contribution
on behalf of Janney Montgomery Scott, Wall Street, New York.

Patent(s) & Publications

Patent Number: [10162841](#)

[HTAP on Spark](#)

SIGMOD 2016, ACM

[STA - Unified Engine](#)

SIGMOD 2016, ACM

[Interactive Analytics](#)

CIDR 2017

Technical Skills C/C++, core java, RDBMS, nosql, python, scala, spark, R, kubernetes, distributed systems, linux

Previous Professional Experience

SnappyData Technologies (Tibco ComputeDB), Pune, MH (India)

March 2016 – June 2017

Staff Engineer

- First few lead engineers developing a [HTAP](#) database.
- Thought leadership in [Approximate Query Processing](#)
- Implemented several [Spark optimizations](#) like adding [CBO](#), extending whole stage code generation, min-max column statistics etcetera in Spark 1.6/2.0.
- Spark streaming & parallel ingestion over [microsoft CDC](#), docs [here](#).

VMware, Pune, MH (India)

Feb 2007 – February 2016

Staff Engineer (Mar'2016 – Feb'2016)

Sr. Member of Technical Staff [May'2010 – Mar'2013]

Sr. Technical Lead, GemStone Systems (Feb'2007 – May'2010)

- Developed [distributed caching](#) product (now VMware Tanzu).
- Developed Pivotal nosql [MPP](#) database later known as Tibco [ComputeDB](#).
- Key Features: *lockfree distributed transaction/ total ordering & distributed Joins/ distributed querying, and query plan optimization (local and global), visualization, resource monitoring & cancellation/ gc pause reduction/ global hash index/ WAN replication/ dynamic dependency injection for UDFs/ ODBC driver using thrift/ realtime stream processing/.*
- Key Technical Innovations: *concurrent hashmap/ concurrent skiplist/ compact concurrent hashmap & hashset/ disk optimized iterator/ early conflict + eventual consistency [mvcc](#) based distributed transactional operations/ light weight highly optimized query parser (LL), byte-to-byte comparators for different datatypes esp. unicode string & bigdecimal/ multistage + multinode code generation/ heuristics-based query compilation (300+ lines)/ vectorization of scans/ atomic distributed DDL operations/ predicate based data eviction from in-memory cache/ data colocation and automatic hash-indexing to avoid shuffles during reads/ memory management of nodes per shard/.*
- Key Language Features used: *java concurrent apis, unsafe apis, off heap memory allocation, nio, directbytebuffers, selectors, custom gc finalizers.*

OSS Contribution

Optimized syslog-ng Elasticsearch java-plugin

Pull request [2728](#)

[BuyGame](#) – Academic project

Developer

Before 2007 Professional Experience

Opus Software Solutions, Chennai, TN (India)

October 2000 – January 2007

Sr. Technical Consultant

- m27 -- a C++ framework that helped focusing on “design by contract” methodologies.
- mTalk -- a messaging framework in C++ like today’s kafka that supported data-dependent routing for load balancing, guaranteed delivery by message persistence, publish-subscribe real time at scale using async notification.
- Trading System – deployed in Janney Montgomery Scott, Wall St. NY supporting NYSE/AMEX/PENN exchanges, internationally NSE/BSE (India), HKSE (Hong Kong) stock exchanges supported.
- Key Features: *Classes as business objects, [BPML](#) workflow modelling as [FSA](#)/ Order management system supporting program trading/ Realtime Risk management system with sub-microsecond latency/ Market data feed with sub-millisecond latency/ post trade management & back office settlement*
- Key Technical Innovations: *C++ compiler for C++ classes as business objects supporting Object-Relational mapping/ auto-remoting of class methods/ automatic object caching and lifetime management/ class modifiers for publish-subscribe/ optimistic locking in microsoft SQL server store procedures/ lineage-based order history tracking/ deployment automation in microsoft windows.*
- Technology: Trading protocols -- NYSE’s CMS, FIX, Compiler – Lex/Yacc

Netlink Technologies, Chennai, TN (India)

March 2000 – October 2000

Programmer Analyst

- IZone – ISP Satyam’s countrywide internet browsing terminal
- Key Features: *session management of internet users, presenting their internet usage time, providing re-charge options, billing & MIS reporting, user rights management.*

Risha Software Solutions, Jabalpur, MP (India)

August 1997 – September 1999

Programmer Analyst

- Inventory Management and Resource Scheduling for [DGQA](#), Ministry of Defence.
- Key Features: *optimal schedule for thousands of ammunitions’ proofing (a quality check procedure by firing them) based on 42 factors/ inventory adjustment module implementing rollback feature in FoxPro.*
- Key Technical Innovations: *Operational Research problem with 42 mutually dependent parameters, where each parameter derived was based on multiple other factors and input parameters like stock availability, heating chamber optimal transition & availability, armour plate reservations, components used.*

Prior Education

[CDAC](#), Center for Development of Advanced Computing, India

Diploma in Advanced Computing, *Computer Science*

January 2000

Govt. Autonomous [Science College](#), India

Bachelor of Science, *Mathematics, Physics & Chemistry*

April 1997

Technical Skills *C/C++, core java, RDBMS, nosql, python, scala, spark, R, kubernetes, distributed systems, linux*