

# Anirban Bhattacharjee

✉ anirban.bhattacharjee@vanderbilt.edu

📞 Anirban2404

in anirban-bhattacharjee-76285a19

📠 615-601-6772

## SUMMARY

---

Proactive, and curious-minded Ph.D. student working in areas of Cloud Computing, Distributed Systems and BigData Analytics, with 3 years of industrial experience in full-stack software design and development.

## EDUCATION

---

### Vanderbilt University

PhD in Computer Science

Nashville, TN

Expected August 2019

### West Bengal University of Technology

Bachelor of Technology in Computer Science and Engineering

Kolkata, West India

June 2009

## KEY SKILLS

---

- **Programming:** Python, Java, NodeJS, PHP, Shell Scripting, C, C++, HTML/CSS, Ansible
- **Database and Data Warehouse Technologies:** Oracle, MySQL, Postgres, Microsoft SQL Server, Cassandra, MongoDB, Amazon DynamoDB, Azure CosmosDB, Redshift, SQL Data Warehouse, Informatica
- **Machine Learning/AI Platforms/Libraries:** Scikit-learn, TensorFlow, Keras, Weka, Spark MLlib, PyTorch
- **Cloud and Big Data:** OpenStack, Azure, AWS, Docker, Kubernetes, Hadoop, Spark, Storm, Kafka, Pig, Hive, ZooKeeper, Raft
- **Development and Design Tools:** Eclipse, Visual Studio, Android Studio, Matlab, Jupyter Notebook, PyCharm, Vim, AngularJS, Spring, Informatica, Git/SVN, Maven, Latex

## INDUSTRIAL EXPERIENCE

---

### ARM Inc.

Cloud Research Summer Intern

Austin, TX

May 2018 – Aug 2018

- Developed Scalable and Agile Engineering Analytics Platform(EAP) for large datasets on Amazon AWS Stack and Microsoft Azure stack using AWS and Azure SDK for Python.
- Used Stream Processing Engines(Kinesis, Event Hubs), EMR, DataBricks, DynamoDB, CosmosDB, ElasticSearch, Redshift, SQL DW, Visualization tools(Kibana, PowerBI, QuickSight) to build end-to-end EAP.
- Evaluated the read/write performance of different datasets and data formats reside on several distributed file systems ( HDFS, Ceph, and S3) to design DataLake infrastructure.

### TATA Consultancy Servies Limited

Systems Engineer

Kolkata, India

Dec 2009 – Aug 2012

- Developed workflows using PL/SQL, SQL code, and Informatica DW tool to extract, load, transform, and analyze data efficiently as per the business requirements.
- Developed automated task scheduler to run the workflows using UNIX Shell Scripting and PL/SQL as per business and workflows dependencies, and to generate visual reports on IBM Cognos GUI.
  - Reduced 30% of manual processing time.
- Build software for Multi-Room Hotel using JAVA MVC, and Oracle and implemented multiple user-level views with security.

## RESEARCH EXPERIENCE

---

### Vanderbilt University

Graduate Researcher/Developer

Nashville, TN

Aug 2012 – Present

### StreamlinedML: End-to-end Machine Learning and Data Analysis Framework

Sep 2017 - Present

- Designed NodeJS and MongoDB based interactive drag-and-drop UI for building predictive analysis pipeline including data preparation, training, visualization, and validation.
- Designed a Python wrapper to integrate a diverse set of Machine Learning(ML) and statistical analysis algorithms and libraries from various (CPU/GPU intensive) frameworks such as TensorFlow, Scikit-learn, Apache Spark MLlib (Python-based).
- Distributed and Scalable AI/ML platform for training with multi-source data aggregation.
- Integrated interactive programming and API interfaces for advance users.
- Proactive optimization of resources to handle dynamic workloads for Deep-Learning prediction jobs (TensorFlow-based) across heterogeneous distributed platforms.
  - Guaranteed the SLOs 99% of the time.
  - Performed 50-95% better than the naive approach.

## **CloudCAMP: Self-Service Cloud Deployment and Management Platform**

*Sep 2014 - Aug 2017*

- Build NodeJS and MongoDB-based fault-tolerant, highly-available and extensible self-service framework by abstracting cloud and application infrastructure specifications using Model Driven Engineering.
- Build NodeJS and MySQL-based DSML to transform abstract business model to DevOps-specific (Ansible) infrastructure code and orchestration code for different cloud providers and hardware.

## **STRATUM: Server selection framework for Deployment and Migration**

*Jan 2017 - Present*

- Developed and compared multiple Python-based (Scikit-learn) ML models for training and predicting application performance by collecting system metrics using collectd and storing data in InfluxDB.
- Build time series model on large real-world dataset for workload prediction using ARIMA and Facebook Prophet's Python SDK.
- Automated migration (including distributed co-ordination and state-machine replication) of latency-sensitive applications (Docker containerized) to minimize cost while satisfying performance and latency constraints across the edge/IoT and cloud.

## **Evaluation of Security Algorithms on Time-Triggered Networked Control Systems**

*Dec 2012 - Aug 2014*

- Implemented and evaluated Message Authentication (HMAC) protocol in TTech device's Linux real-time kernel.

## **PUBLICATIONS**

---

- **BARISTA: Efficient and Scalable Deep Learning Prediction Serving using Serverless Computing | 2019 IEEE International Conference on Cloud Engineering(IC2E)**
- **FECBench: A Holistic Interference-aware Approach for Application Performance Modeling | 2019 IEEE International Conference on Cloud Engineering(IC2E)**
- **STRATUM: A Serverless Framework for Lifecycle Management of Machine Learning based Data Analytics Tasks | 2019 USENIX Conference on Operational Machine Learning**
- **A Model-Driven Approach to Automate the Deployment and Management of Cloud Services | 2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)**
- **UPSARA: A Model-Driven Approach for Performance Analysis of Cloud-Hosted Applications | 2018 IEEE/ACM 11th International Conference on Utility and Cloud Computing (UCC)**
- **(WIP) CloudCAMP: Automating the Deployment and Management of Cloud Services | 2018 IEEE International Conference on Services Computing (SCC)**
- **Performance Interference-Aware Vertical Elasticity for Cloud-Hosted Latency-Sensitive Applications | 2018 IEEE 11th International Conference on Cloud Computing (CLOUD)**
- **PADS: Design and implementation of a cloud-based, immersive learning environment for distributed systems algorithms | 2018 IEEE Transactions on Emerging Topics in Computing Journal**
- **INDICES: exploiting edge resources for performance-aware cloud-hosted services | 2017 IEEE 1st International Conference on Fog and Edge Computing (ICFEC)**
- **Computation and communication evaluation of an authentication mechanism for time-triggered networked control systems | 2016 Sensors Journal**
- **Performance evaluation of an authentication mechanism in time-triggered networked control systems | 2014 7th International Symposium on Resilient Control Systems (ISRCS)**

**Google Scholar Link** : <https://scholar.google.com/citations?user=nt-jeioAAAAJ&hl=en>

## **ACADEMIC PROJECTS**

---

### **A framework for Hadoop Performance Prediction**

*Apr 2013*

- Benchmarked different Hadoop applications with various configurations for building the performance model.
- Predicted the cost and performance (using WEKA) of Hadoop based MapReduce applications in the cloud environment.

### **Android Image Synchronization Application**

*Dec 2012*

- Downloaded images from webpage and stored in content provider.
- Synchronized the state of the Content Providers among multiple devices.

### **Protein-protein interaction prediction using domain-domain interaction with Multilayer Perceptron**

*Apr 2011*

- Developed a JAVA based prediction model to predict interactions between two protein sequences by exploiting all possible combinations of constituent domains.
- The overall MLP prediction accuracy achieved in the range of 66.70%, using twofold cross-validation.

## **AWARDS AND ACHIEVEMENTS**

---

- Awarded for Best Paper in IEEE International Conference on Cloud Engineering (IC2E), 2019 and International Symposium on Resilient Control Systems (ISRCS), 2014
- Participated and Awarded for quizzes in corporate and college competitions.
- Volunteered in Social Welfare Activities.