

RESEARCH INTERESTS

Databases, Distributed Systems, Cloud Computing. I am a PhD graduate from UCSB, where I worked with professors Divyakant Agrawal and Amr El Abbadi. The focus of my PhD thesis was on data infrastructure problems related to the variety of big data. These included challenges arising out of the use of multiple database engines for storing application data, and distribution heterogeneity (read-heavy, hot-spots etc) within workloads. I have also worked on other distributed data management related challenges: geo-replicated transaction management, data reconfiguration, read-enhancements in consensus protocols and distributed transaction concurrency control. I now work on various large-scale database problems at Salesforce.

EDUCATION

University of California, Santa Barbara (UCSB) *Santa Barbara, California* *Sep 2011 – March 2018*
Masters and Ph.D. in Computer Science **Cumulative G.P.A: 3.97 / 4**

National Institute of Technology, Tiruchirappalli *Tiruchirappalli, India* *2005 – 2009*
Bachelors of Technology, Computer Science and Engineering **Cumulative G.P.A: 8.58 / 10**

PROFESSIONAL EXPERIENCE

Salesforce, Senior Member of Technical Staff *San Francisco, California* *Apr 2018 – Present*

- Member of the Database group at Salesforce. Working on various distributed database challenges and problems related to database access layer, LSM (Log-structured merge) and multi-tenancy.

Microsoft Research, Research Intern *Seattle, Washington* *June 2015 – Sep 2015*

- Member of the Database group at Microsoft Research, mentored by Justin Levandoski. Built a mechanism for efficient materialized view maintenance in an in-memory transactional engine.

HP Labs, Research Intern *Palo Alto, California* *June 2014 – Sep 2014*

- Member of the Analytics Lab. Designed a speculative execution based mechanism to support high transactional throughput on databases optimized for processing analytical workloads (like the column-oriented database Vertica).
- The different projects during the internship has been filed as 3 patent applications.

Amazon, Software Development Engineering Intern *Seattle, Washington* *June 2013 – Sep 2013*

- Member of the Lookout Team, which provides Anti-DDoS (Distributed Denial of Service) support for AWS services.
- Developed a service providing key-value mapping from numeric IDs to hostnames. The service ensures that there are no duplicates, is highly available, and provides low-latency access from different geographical regions.

Amazon, Software Development Engineering Intern *Seattle, Washington* *June 2012 – Sep 2012*

- Member of the Lookout Team. Developed a distributed monitoring system which measures the health of RDS database partitions and uses the data to improve the query results and the operational management of the system.

Yahoo! R & D, Senior Software Developer *Bangalore, India* *June 2009 – Aug 2011*

- Member of the team working on the problem of Malvertising (malicious advertising). As a part of Display Ad Flow, this team developed automated tools for ensuring that no malicious ads are displayed on Yahoo and partnered pages.

SELECTED ACADEMIC RESEARCH PROJECTS

Typhon: Consistency Semantics for Multi-Representation Data Processing *Jan 2016 – May 2017*

- Typhon defines consistency semantics for data spread across heterogeneous engines. Using a notion of entities, which

link the data in diverse representations, Typhon defines implicit causal guarantees, which capture the logical order intended by application clients. A protocol is designed to provide the implicit causal guarantees defined in Typhon.

Janus: A Hybrid Cloud Store for Transactional and Analytical Processing

Jan 2014 – Dec 2015

- Janus is a hybrid cloud datastore, which supports efficient execution of different workloads by storing data in different representations. A data movement pipeline based on in-memory redo-only transactional logs and a graph dependency management algorithm, is designed to continuously ensure up-to date data at the different representations.

Improving Read Scalability in Raft-like Consensus Protocols

Sep 2016 – Apr 2017

- Combined quorum reads with traditional leader-based reads in Raft, to offload the leader and better utilize the cluster in read-heavy scenarios. Implemented in CockroachDB, an open-source distributed SQL database, which uses Raft.

Squall: Live Reconfiguration of Partitioned Main Memory Databases

Feb 2013 – Dec 2013

- Squall is a pull-based scheme for live reconfiguration of data in partitioned main memory databases, while supporting distributed transactions, with no disruption of service. Implemented in H-Store, an open-source in-memory database.

Minimizing Commit Latency of Transactions in a Geo-Replicated DataStores

Feb 2013 – Dec 2013

- Lower-bound study and protocol design for minimizing commit latency of serializable transactions in geo-replicated settings.

TEACHING EXPERIENCE

- Teaching Assistant for six computer science courses at UCSB, including: ‘Advanced Distributed Systems’, ‘Advanced Databases’, ‘Automata & Formal Languages’, and ‘Data Structures and Algorithms’.

MENTORED STUDENTS

- Tanuj Mittal – MS (Project: Improving Read Scalability in Raft-like Consensus Protocols)
- Darshan Maiya – MS (Project: Resource Allocation for Distributed Machine Learning Executions)
- Ravi Kumar Suresh Babu – MS (Project: Reducing Aborts in High Contention Transactional Workloads)
- Angela Yung – Undergraduate (Project: Deterministic Sequencing of Transactions for Multi-Representation Data)

SERVICE AND HONORS

- Reviewer for **TKDE** and **DAPD** Journals (2017-2020). External Reviewer for **NDA 2017**, **Euro DW 2017**, **DAFSAA 2017**, **BIGCOMP 2017**, **EDBT 2017**, **SBBB 2016**, **CloudDM 2016**, **NETYS 2015**, **ICDE 2015**, **SIGMOD 2014**, **ETBT 2014**.
- Vice-Chair of **GSWC (Graduate Student Workshop in Computing)** 2014, held at UCSB.
- Recipient of several **travel grants** (UCSB Doctoral Senate grant, Sigmod 2015, ICDCS 2017, HotCloud/ATC 2017).
- Qualified for the **South Asia Onsite Regional of ACM-ICPC** (International Collegiate Programming Contest) 2008.
- Previously active member of **Topcoder** and **SPOJ**. Achieved **World Rank 316** and **India Rank 51** on **SPOJ**.
- All India Rank of 648 in **AIEEE 2005** (All India Engineering Entrance Examination) (Among ~.5 million test takers).
- Past member of **YEFI** (Yahoo Employee Foundation India), which supports community efforts to promote education.

SELECTED PUBLICATIONS

M-DB: A Continuous Data Processing and Monitoring Framework for IoT Applications - **IEEE Ithings 2019**
Vaibhav Arora, Mohammad Javad Amiri, Divyakant Agrawal, Amr El Abbadi

Dynamic Timestamp Allocation for Reducing Transaction Aborts– **IEEE Cloud 2018**

Vaibhav Arora, Ravi Kumar, Sujaya Maiyya, Divyakant Agrawal, Amr El Abbadi, Xun Xue, Zhiyanan, Zhu Jianfeng

Janus: A Hybrid Scalable Multi-Representation Cloud Datastore – **IEEE TKDE – Volume 30 Issue 4 2018 (Journal)**

Vaibhav Arora, Faisal Nawab, Divyakant Agrawal, Amr El Abbadi

Typhon: Consistency Semantics for Multi-Representation Data Processing – **IEEE Cloud 2017**
Vaibhav Arora, Faisal Nawab, Divyakant Agrawal, Amr El Abbadi

Leader or Majority: Why Have One When You Can Have Both? Improving Read Scalability in Raft-like Consensus Protocols – **HotCloud 2017**

Vaibhav Arora, Tanuj Mittal, Divyakant Agrawal, Amr El Abbadi, Xun Xue, Zhiyanan, Zhujianfeng

Multi-Representation based Data Processing Architecture for IoT Applications – **ICDCS 2017**

Vaibhav Arora, Faisal Nawab, Divyakant Agrawal, Amr El Abbadi

Minimizing Commit Latency of Transactions in Geo-Replicated Data Stores – **SIGMOD 2015**

Faisal Nawab, Vaibhav Arora, Divyakant Agrawal, Amr El Abbadi

Squall: Fine-Grained Live Reconfiguration for Partitioned Main Memory Databases – **SIGMOD 2015**

Aaron J Elmore, Vaibhav Arora, Rebecca Taft, Andrew Pavlo, Divyakant Agrawal, Amr El Abbadi

MaaT: Effective and Scalable Coordination of Distributed Transactions in the Cloud – **VLDB 2014**

Hatem Mahmoud, Vaibhav Arora, Faisal Nawab, Divyakant Agrawal, Amr El Abbadi

PATENTS

[Speculative Execution of a Stream of Changes](#)

Vaibhav Arora, Alkis Simitsis, William K. Wilkinson

[Read-Optimized Database Changes](#)

Vaibhav Arora, Alkis Simitsis, WK Wilkinson

[Database Transfer of Changes](#)

Vaibhav Arora, Alkis Simitsis, WK Wilkinson