

# DHRUV KUMAR

---

CONTACT INFORMATION	A-506, R&D Building, IIIT Delhi New Delhi-110020 India	<i>Mobile:</i> +917742710554 <i>E-mail:</i> dhruv.kumar@iiitd.ac.in <i>URL:</i> kudhru.github.io
RESEARCH INTERESTS	Big Data Analytics, Machine Learning, Distributed Systems, Human-Computer Interaction	
CURRENT AFFILIATION	<b>IIIT Delhi</b> , New Delhi, India <i>Assistant Professor</i>	<b>Aug 2022 - Present</b>
EDUCATION	<b>University of Minnesota</b> , Twin Cities, United States <i>PhD in Computer Science</i>	<b>Sep 2017 - Mar 2022</b>
	<ul style="list-style-type: none"><li>• CGPA: <b>4.0 / 4.0</b></li><li>• 3M Science and Technology Fellowship.</li><li>• <b>Relevant Courses:</b> Distributed Systems, Machine Learning, Databases, Non-Linear Optimization, Dispersed Data Driven Computing, Matrix Theory, Probability and Statistics.</li></ul>	
	<b>Birla Institute of Technology and Science (BITS)</b> , Pilani, India <i>Bachelor of Engineering (Hons.) Computer Science</i>	<b>Aug 2010 - May 2014</b>
	<ul style="list-style-type: none"><li>• CGPA: <b>9.92 / 10.0</b></li><li>• <b>Rank 1</b> in class of 2014 of Computer Science, comprising 120 students.</li><li>• <b>Rank 4</b> in entire class of 2014, comprising 800 students.</li></ul>	
RESEARCH EXPERIENCE	<b>Microsoft Research, Bengaluru, India</b> <i>Post-Doctoral Researcher</i> <b>Mentors:</b> <i>Ranjita Bhagwan, Venkat Padmanabhan</i>	<b>Mar 2022 - Jul 2022</b>
	<ul style="list-style-type: none"><li>• Optimizing network and compute utilization for Microsoft Azure.</li></ul>	
	<b>Distributed Computing Systems Group, UMN, Twin Cities</b> <b>Project:</b> <i>Geo-distributed Analytics</i> <b>Advisor:</b> <i>Abhishek Chandra</i>	<b>Sep 2017 - Mar 2022</b>
	<ul style="list-style-type: none"><li>• Systems for optimizing latency, cost and WAN traffic in geo-distributed analytics.</li><li>• Implementation done on top of Apache Flink and Apache Spark.</li></ul>	
	<b>Google Cloud, Sunnyvale</b> <b>Project:</b> <i>Learning to prefetch data for disk servers</i> <b>Mentor:</b> <i>Mustafa Uysal</i>	<b>May 2019 - Aug 2019</b>
	<ul style="list-style-type: none"><li>• Identified an efficient strategy for constructing ground truth and proposed a deep neural network based architecture for predicting the prefetch data.</li><li>• Initial evaluation showed an improvement of upto 20% in cache hit rates using the proposed model over the existing approach.</li></ul>	
	<b>ADAPT Lab, BITS-Pilani</b> <b>Project:</b> <i>A New Distributed Computing Framework for Data Mining</i> <b>Mentors:</b> <i>Navneet Goyal, Poonam Goyal, and Sundar Balasubramaniam</i>	<b>Apr 2013 - Oct 2014</b>

- Designed and implemented data mining algorithms such as OPTICS, SLINK, DBSCAN for shared memory and distributed memory models.
- Implemented using OpenMP and OpenMPI libraries in C.

TEACHING  
EXPERIENCE

Cloud Computing (Monsoon 2022), Distributed Systems (Winter 2023) at IIT Delhi  
Graduate Teaching Assistant for Operating Systems (Fall 2020) at UMN, Twin Cities

SELECTED  
PUBLICATIONS

**Dhruv Kumar**, S Ahmad, A Chandra, R Sitaraman. *AggFirstJoin: Optimizing Geo-Distributed Joins using Aggregation-Based Transformations*, accepted in **IEEE/ACM CCGrid 2023**. [\[Link\]](#)

**Dhruv Kumar**, J Wolfrath, A Chandra, R Sitaraman. *Towards WAN-Aware Join Sampling over Geo-Distributed Data*, in **ACM EdgeSys 2022**. [\[Link\]](#)

J Wolfrath, N Sreekumar, **Dhruv Kumar**, Y Wang, A Chandra. *HACCS: Heterogeneity-Aware Clustered Client Selection for Accelerated Federated Learning*, in **IEEE IPDPS 2022**. [\[Link\]](#)

**Dhruv Kumar**, S Ahmad, A Chandra, R Sitaraman. *AggNet: Cost-Aware Aggregation Networks for Geo-distributed Streaming Analytics*, in **ACM/IEEE SEC 2021**. [\[Link\]](#)

Y Wang, J Wolfrath, N Sreekumar, **Dhruv Kumar**, A Chandra. *Accelerated Training via Device Similarity in Federated Learning*, in **ACM EdgeSys 2021**. [\[Link\]](#)

Y Wang, **Dhruv Kumar**, A Chandra. *Exploiting Data Heterogeneity for Performance and Reliability in Federated Learning*, Poster in **ACM/IEEE SEC 2020**. [\[Link\]](#)

**Dhruv Kumar**, A A Ramkumar, R Sindhu, A Chandra. *DeCaf: Iterative Collaborative Processing over the Edge*, in **USENIX HotEdge 2019**. [\[Link\]](#)

**Dhruv Kumar**, J Li, A Chandra, R Sitaraman. *A TTL-based Approach for Data Aggregation in Geo-Distributed Streaming Analytics*, in **ACM SIGMETRICS 2019**. [\[Link\]](#)

**Dhruv Kumar**, J Li, A Chandra, R Sitaraman. *TTL-based Approach for Data Aggregation in Geo-Distributed Streaming Analytics*, Poster in **OSDI 2018**. [\[Link\]](#)

OTHER  
PUBLICATIONS

S Rallapalli, A Jain, **Dhruv Kumar**. *Cloud-based neuro-fuzzy hydro-climatic model for water quality assessment under uncertainty and sensitivity*, in **Environmental Science and Pollution Research, Springer, 2022**. [\[Link\]](#)

R Srinivas, M Drewitz, J Magner, A P Singh, **Dhruv Kumar**, Y B Katpatal. *Simulating Landscape Hydrologic Connectivity in a Precise Manner Using Hydro-Conditioning*, in **Advances in Computational Modeling and Simulation, Springer, 2022**. [\[Link\]](#)

P Goyal, JS Challa, **D Kumar**, A Bhat, N Goyal, Sundar B. *Grid-R-tree: A data structure for efficient neighborhood and nearest neighbor queries in data mining*, in **JDSA Springer 2020**. [\[Link\]](#)

**Dhruv Kumar**, P Goyal, N Goyal. *An Efficient method for Batch Updates in OPTICS Cluster Ordering*, in **IJDATS 2018**. [\[Link\]](#)

P Goyal, S Kumari, A Sood, **Dhruv Kumar**, Sundar B, and N Goyal. *Exact, Fast and Scalable Parallel DBSCAN for Commodity Platforms*, in **ICDCN 2017**. [\[Link\]](#)

P Goyal, S Kumari, S Sharma, **Dhruv Kumar**, V Kishore, Sundar B, and N Goyal. *A fast, Scalable SLINK Algorithm for Commodity Cluster Computing Exploiting Spatial Locality*, in **HPCC 2016**.[\[Link\]](#)

P Goyal, S Kumari, **Dhruv Kumar**, Sundar B, N Goyal, S Islam, and JS Challa. *Parallelizing OPTICS for Commodity Clusters* in **ICDCN 2015**.[\[Link\]](#)

P Goyal, S Kumari, **Dhruv Kumar**, Sundar B, and N Goyal. *Parallelizing OPTICS for multicore systems* in **ACM COMPUTE 2014**.[\[Link\]](#)

PROFESSIONAL  
EXPERIENCE

**Several Startups**  
*Technology and Strategy*

**Apr 2016 - Aug 2017**

- Designed and implemented the entire back-end for three startups from scratch.
- The entire back-end functionality was exposed using RESTful APIs implemented using Django web framework and hosted using Amazon web services.
- Gained valuable experience in building scalable and secure back-ends for web and mobile applications.

**Goldman Sachs, Bengaluru, India**  
*Software Developer, Investment Management Division*

**Nov 2014 - Apr 2016**

- Improved the efficiency of risk-management system by suggesting improvements to the SQL queries going to Sybase IQ database.
- Assisted in migrating from Sybase IQ database to MemSQL database for faster access.
- Wrote APIs for accessing MemSQL database.
- Implemented a H2-database based server for allowing real-time updates to the tables residing in the servers.
- Learnt about the real life use-cases of databases.

**CSIR-CEERI, Pilani, India**  
*Machine Learning Intern*

**May 2012 - July 2012**

- Studied, compared and implemented various unsupervised machine learning algorithms.
- Learnt about the use of these algorithms in real world applications

HONORS AND  
AWARDS

3M Science and Technology Fellowship, UMN Graduate School

**Aug 2017 - Jun 2021**

Student Travel Grant, ATC 2019

**Jul 2019**

Student Travel Grant, SIGMETRICS 2019

**Jun 2019**

Student Travel Grant, OSDI 2018

**Oct 2018**

Merit Scholarship (Top 10 among 800 students), BITS Pilani

**Aug 2010 - May 2014**

Research Fellowship (Outstanding undergraduate thesis), BITS Pilani

**Jan 2014 - May 2014**

SERVICE

Reviewed Papers for Elsevier JPDC, IPDPS (2020, 2019), Big Data (2020, 2019), IISWC 2020, ICDCS 2020

TECHNICAL SKILLS

- **Programming:** C, Java, Python, OpenMPI, OpenMP, MySQL, Verilog, Matlab
- **Mobile and Web Technologies:** HTML, CSS, JavaScript, AngularJS, Django, Android
- **Cloud platforms:** Amazon web services