

# Amitayush Thakur

✉ [amitayusht@gmail.com](mailto:amitayusht@gmail.com) | in [in/amitayush-thakur](https://in.linkedin.com/in/amitayush-thakur) |  [amit9oct](https://github.com/amit9oct) |  [amit9oct.github.io](https://amit9oct.github.io)

## EDUCATION

---

### Birla Institute Of Technology and Science, Pilani

*Master of Science (Hons.) Mathematics*

*Bachelor of Engineering (Hons.) Computer Science*

Pilani, India

*Aug. 2012 – July 2017*

*Aug. 2012 – July 2017*

*Cumm. GPA - 9.02/10*

## EXPERIENCE

---

### Software Engineer II

*Microsoft India Development Center*

*Microsoft Azure Backup (MAB) team*

September 2019 – Present

*Hyderabad, India*

- Developed technique for unsupervised pattern mining of service trace logs to predict root cause for the customer issues. The method was based on clustering similar logs and then training a decision tree on top of it to infer the reasoning behind the clustering. This inference was then used by service engineers to figure out the root cause for the customer issues. Mentored interns to implement the same.
- Developed & Designed a standalone micro-service to efficiently back up hierarchies of objects, such as File Shares, on Cloud. The micro-service split the entire backup into multiple smaller disjoint sub-tasks and then efficiently solved these sub-tasks in parallel—the idea was inspired from map reduce. Designed algorithms, along with proof of correctness, to figure out differences between two hierarchies of objects facilitating incremental backups i.e. only backing up the changed objects

### Software Engineer

*Microsoft India Development Center*

*Microsoft Azure Backup (MAB) team*

July 2017 - August 2019

*Hyderabad, India*

- Focused on creating micro-services, distributed systems, using formal methods (TLA+), unit testing, scaling, Azure Cloud Storage, File Systems
- Developed and Designed Scalable & Distributed Billing orchestrator for Azure Backup. The overall time to compute billing related information improved **6x** because of the distributed design.

### Research Intern

*Microsoft Research*

Jan 2017 - June 2017

*Bangalore, India*

- Worked on Program Synthesis with Microsoft's *PROSE (Program Synthesis using Examples) SDK* team. Work was related to use of ML in PL
- It involved synthesizing programs which were generated on a restricted grammar using input-output examples, and then using ML to rank generated programs to get the most useful program.

### Software Engineering Intern

*Microsoft India Development Center*

*Microsoft Azure Backup (MAB) team*

May 2016 - July 2016

*Hyderabad, India*

- Improved performance for recovering parts of data from large ( $\geq 1$  TB) backed-up disks, the performance went up from **10 hrs to 3 mins** using iSCSI and mocking certain layers of NTFS file system.

### Summer Research Fellow

*National Institute of Science Education & Research*

May 2015 - July 2015

*Bhubaneswar, India*

- Worked under Professor Brundaban Sahu in Computational Number Theory
- The main focus was studying the time complexity of Lenstra's Elliptic Curve Factorization Algorithm, and its relation to  $y$ -smooth numbers

### Project Trainee

*Bhabha Atomic Research Centre (BARC)*

May 2014 - July 2014

*Mumbai, India*

- Worked in development of Network Management System and automation modules.

## PUBLICATIONS

---

A. Gautam, A. Thakur, G. Dhanania and S. Mohan, “**A distributed algorithm for balanced multi-robot task allocation,**” 2016 11th International Conference on Industrial and Information Systems (ICIIS), Roorkee, 2016, pp. 622-627, doi: 10.1109/ICIINFS.2016.8263014. [[IEEE Link](#)]

## RESEARCH INTERESTS

---

**Primary Interests:** Theoretical Machine Learning, NLP, Robotics, AI, Theory Of Computation, Compilers, Distributed Systems

**Other Interests:** Cryptography, Number Theory, Quantum Physics

## PROJECTS

---

**TUTOR: Text to qUery generaTOR** | *SQL query synthesis from Natural Language*

[Github Link](#)

- TUTOR is a library to generate simple SQL query from natural language (English). Written In Python. It uses NLP tools (spaCy) to find out the entities present in the query, understand the PoS and dependency tree of the text query.
- It applies heuristics to summarize the Natural Language Query as an Abstract Meaning Representation (AMR) Tree which captures the semantic meaning of the query. It then uses this AMR and tries to create a valid Abstract Syntax Tree (AST) from it. After this a code generator runs over the AST and generates the SQL query using CFG (context free grammar) for SQL. This idea works good for simple queries.

**Say No To SQL** | *SQL query synthesis from Input-Output examples*

[Github Link](#)

- In C#. Takes csv based input and output tables as examples and synthesizes SQL query.
- It uses Microsoft's [PROSE](#) SDK

**Toy Compiler** | *Compiler & Runtime Env built for pedagogical purpose*

[Github Link](#)

- Toy-Compiler can be used for educational purpose. Helpful for those who need to design a compiler for a simple language. It has the complete compiler built in C language which includes lexer, parser and code generator built without using any libraries like Lex / Yacc

## OTHER TECHNICAL SKILLS

---

**Technologies:** Git, Docker, Microsoft Azure Cloud, Distributed Systems, Map Reduce, Micro-Service Architecture, App Development, Web Development, Formal Methods (TLA+)

**Languages:** C#, Python, Java, C/C++, JavaScript

**Libraries:** scikit-learn, pandas, NumPy, Matplotlib, spaCy

**Others:** LaTeX

## ACHIEVEMENTS

---

**2019:** Microsoft Azure Backup Star Award for Innovation

**2019:** Top 1.4% (among 100,000 students) in GATE 2019

**2015:** Indian National Academy Of Science Summer Research Fellowship for Mathematics

**2014:** Course Topper in Number Theory

**2013-14:** Awarded Merit Scholarship at BITS Pilani for 3 semesters

**2013:** Course Topper in Discrete Mathematics

**2013:** Qualified for 2013 ACM-ICPC Asia Amritapuri Regional, Team Name: Param Name

**2012:** Top 2% (among 500,000 students) in IIT-JEE 2012