

# Somdeep Dey

github.com/somdeep | sd2988@columbia.edu | 9792878549

10351 NE 10th St, Apt 2003 | Bellevue, WA, 98004

## EDUCATION

### **COLUMBIA UNIVERSITY | MASTERS IN COMPUTER SCIENCE**

Dec 2016 | New York, NY, USA • GPA: 3.88/4.0

Algorithms | Cloud and Big Data | Micro Services | OS | Database Systems Implementation | Computer Vision | PLT  
Graduate Course Assistant for Software Engineering, Introduction to Databases and Programming for Entrepreneurs

### **PUNE INSTITUTE OF COMPUTER TECHNOLOGY (PICT) | BE IN COMPUTER SCIENCE**

May 2015 | Pune, India • GPA: 3.6/4.0

Operating Systems | Compilers | Advanced DBMS | Algorithms | Software Architecture | Distributed Systems

## EXPERIENCE

### **MICROSOFT | SOFTWARE ENGINEER 2 | BING KNOWLEDGE GRAPH AND DESKTOP TABS**

Feb 2020 – Present | Seattle, WA

- Helped drive Tabs on Desktop for Bing as a unique Search Engine feature by building end-to-end, scalable data pipeline using Cosmos, Spark, C# and .NET to ship Tabs to 80 languages and markets, bringing in 40 million impressions, over 20 million DAU and 2 million new clicks daily.
- Led three-person effort to attain coverage improvements of 20% via harnessing Bing Knowledge Graph and designing Signature-based query recognition to shift away from static pattern-based triggering. Constructed data pipeline using Azure Batch scheduling and grew key metrics like user tab interaction stability from 60% to 99.993%. Developed distributed microservices-based metrics architecture to generate business insights periodically at scale on Power BI and Azure Synapse Analytics, that helps identify and drive new feature areas across the team and capture issues.
- Oversaw four-person effort dovetailing with multiple teams to fully migrate DesktopTabs to utilizing more efficient column-store based Knowledge Graph V2 - from designing microservices-based architecture to final data pipeline and drove this to completion backed by in-depth metrics, richer data sources and language globalisation.
- Leveraged Bing Index to crawl IPEDS and USNews sources and built end-to-end scalable, periodically-refreshed data extraction pipeline provisioned on Azure SQL Server and MongoDB on Azure Cosmos DB to generate Programs, Cost, Admissions and Rankings experiences for Global Universities on Bing.com, Mobile and Edge Chromium.

### **MICROSOFT | SOFTWARE ENGINEER | BING KNOWLEDGE GRAPH DATA PLATFORM**

Feb 2017 – Feb 2020 | Seattle, WA

- Improved batch and streaming pipeline processing over 10 petabytes of data each day, serving Bing, Office and Edge, using Cosmos, Spark, Hadoop, Java, C# and .NET hosted on Azure Data Lake, by achieving reduction in downstream XML payload size by 80% and dropping turnaround time by 50% to under 24 hours via intelligent understanding of user interest on Bing.com and constructing algorithms to identify, rank facts by importance using custom models and construct graph payloads using Depth First traversal at scale.
- Spearheaded Verified Facts effort to identify and persist static facts in high-availability low-latency column store with fully orchestrated microservices focused on identifying and taking away daily unwanted churn of over 500 million facts/20% of daily incremental load. Incorporated a service that automated data quality via statistical analysis to build 'churn timeline' concept and periodically pushed ratified data to a distributed, in-memory, graph processing engine.
- Shifted metrics framework at data ingestion and crawling level to event-driven, dynamically-scaling approach using Kusto and Azure Data Explorer for enriched dashboards. Added E2E integration testing and shipped platform maintenance, versioning snapshots and rollback mechanisms to maintain quality of releases via Azure DevOps and to reduce manual livesite interventions by developers on Bing.com.

### **NEXTJUMP | SOFTWARE ENGINEERING INTERN**

June 2016 – August 2016 | New York, NY

Worked on the Onecart team using a combination of PHP and Go to build backend APIs and service points for the Onecart experience serving as the central point of search and payment for Next Jump's millions of users on the Perks at Work website.

## TECHNICAL SKILLS

C#, .NET, Cosmos Big Data, Data Engineering, Distributed Systems, ETL, XML, JSON, Java, Azure PowerBi, Kusto, Apache Spark, Hadoop, C, C++, Microsoft SQL Server, JSP, Razor, Node.js, JavaScript, Git, Typescript, Python, PHP, Linux, Go, XAMP, HTML, AngularJS, AWS Elastic Beanstalk, MongoDB, NoSQL, Tableau, Azure Key Vault, RESTful Services, JIRA, Google Workspace API.

# PROJECTS

## **STATBALL | SPORT STATS SCRAPING AND ANALYSIS ON .NET CORE**

Sep 2021 - Oct 2021 | Seattle, WA

Analysing soccer data using C# for intelligent player scouting, similarity analysis and mapping team data to player fit, replete with Tableau - generated visualizations.

## **FLOW | AUTOMATED FLOW FOR WEBSITE WITH PURCHASES, PAYMENTS AND CONTENT DELIVERY**

Apr 2020 - May 2020 | Seattle, WA

Constructed website for donations toward COVID relief via digital art purchases, using C#-based microservices, including real-time Gmail API integration for content delivery and Razorpay for international payments.

## **OPERATING SYSTEMS | COURSE PROJECTS**

Jan 2016 - May 2016 | Columbia University, New York, NY

Built projects including industrial-strength web server, basic file system, kernel key-value store and a full fledged kernel scheduler in a Linux environment.

## **FLATMATE.IO | ROOMMATE FINDER BASED ON MEAN STACK AND BLUEMIX**

Dec 2015 | Columbia University, New York, NY

Collaborative filtering using Facebook and LinkedIn data, thus aligning personal and professional matches. Auto-scaling messaging infrastructure to ensure users have real-time connectivity and face no downtime.

## **STUDENT-COURSE MANAGEMENT SYSTEM | K-12 SYSTEM USING MICRO SERVICES ARCHITECTURE**

Oct 2015 | Columbia University, New York, NY

Implemented with Node.js : z-axis scaling, live schema change for administrators, single point of reference API gateway, ref-integrity using RabbitMQ, auto instance creation on increased data and data partitioning on surnames.

## **TWEETGEO | JAVA WEB APP ON AWS ELASTIC BEANSTALK DEPICTING MAP WITH REAL TIME TWEETS**

Sept 2015 | Columbia University, New York, NY

Twitter Live API data streamed to map using WebSockets, depicting trend heatmap geographically. Sentiment analysis of tweets using multi-threaded worker pool, highlighting public opinion of events.

## **LINUX KERNEL CONTRIBUTION | XFS FILE SYSTEM DEFRAGMENTER UTILITIES IMPLEMENTED USING C**

Sep 2014 - May 2015 | Pune, India

- 'Spaceman': generated free space report of the overall filesystem, enabling freespace defragmentation.
- Made headway into partial defragmentation (portions of large files), hitherto untried across file systems.
- Dave Chinner, maintainer of XFS, reviewed all patches, features were included in the Linux Kernel (3.19).
- 'FaSteR Spaceman' discussed by experts at Linux Filesystem Storage and Memory Management Summit (LSFMM - 2015).

## **VIRTUAL LEARNING ENVIRONMENT | JAVA-BASED VLE HOSTING ASSIGNMENTS AND VIDEOS.**

Jan 2014 - May 2014 | PICT, Pune, India

Private cloud infrastructure on Eucalyptus for use only within PICT Computer Science department. Role based access control, discussion board for interaction, email updates, timetable export and import and adaptive tests to ensure one-stop solution.