sc3653@columbia.edu

linkedln.com/chandrasiddhartha
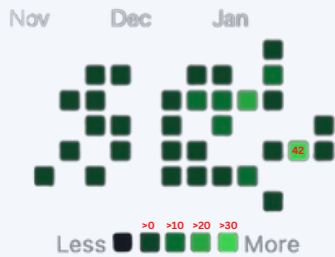
github.com/siddhartha–chandra

Siddhartha Chandra

TECHNOLOGIST

education   work

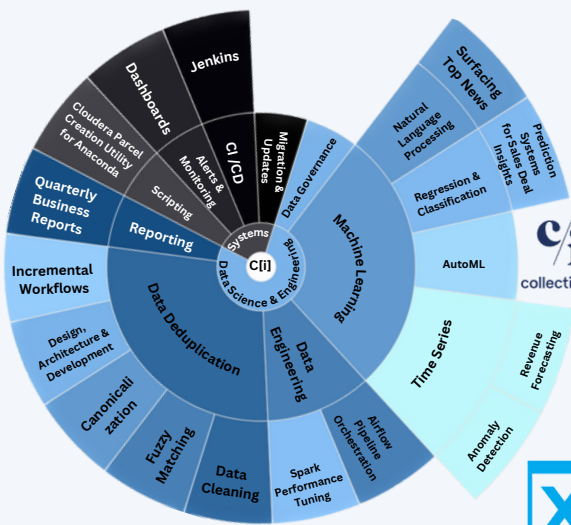Nov   Dec   Jan

42

Less >0 >10 >20 >30 More

MISSION

To create innovative, elegant and scalable solutions that eliminate productivity roadblocks and spark creativity, freeing up people's valuable time and empowering them to focus on what truly matters
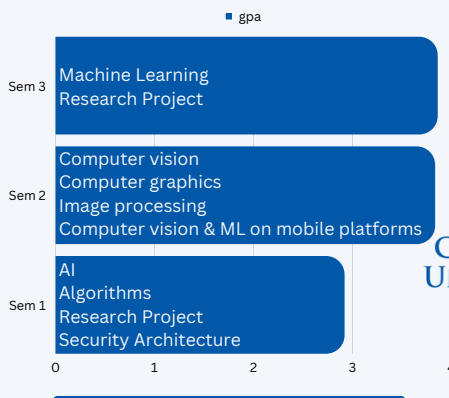
NOTABLES

- Authored a comprehensive study in code for comparing traditional & modern techniques for deduplication (2022)
- People mentored: 4
- Tech & behavioural interviews conducted: 10
- Company-wide hackathon winner (2020)
- Attended NE-Scala, Philadelphia (2016)
- Attended MongoDB World, NYC (2016)
- Presented a Tech talk @ SEAS, Columbia University (2016)
- SPOT award (quarterly) recipient for excellence in task execution (2011)
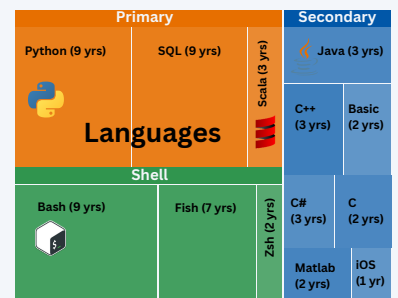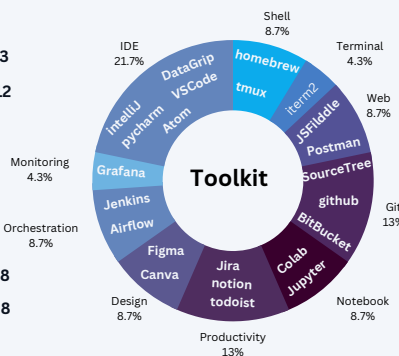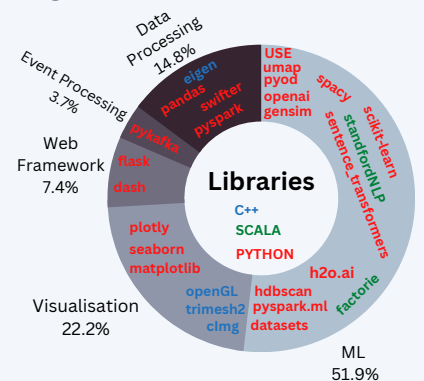- PRIMA award (annual) recognition for project execution (2010)

Timeline:
FEB 2023
NOV 2022
APR 2022
OCT 2016
SEP 2016
DEC 2014
OCT 2014
SEP 2013
MAY 2012
SEP 2008
JUL 2008
AUG 2008

Sunburst chart labels: Jenkins, Dashboards, Cloudera Parcel Creation Utility for Anaconda, Quarterly Business Reports, Incremental Workflows, Design, Architecture & Development, Canonicalization, Fuzzy Matching, Data Cleaning, Spark Performance Tuning, Airflow Pipeline Orchestration, Data Engineering, Data Deduplication, Data Science & Engineering, Systems, Reporting, Scripting, Alerts & Monitoring, CI /CD, Migration & Updates, Data Governance, Machine Learning, Natural Language Processing, Regression & Classification, AutoML, Time Series, Revenue Forecasting, Anomaly Detection, Surfacing Top News, Prediction Systems for Sales Deal Insights, C[i]

collective[i]

GPA chart

gpa

Sem 3 — Machine Learning, Research Project

Sem 2 — Computer vision, Computer graphics, Image processing, Computer vision & ML on mobile platforms

Sem 1 — AI, Algorithms, Research Project, Security Architecture

0   1   2   3   4

COLUMBIA UNIVERSITY

Infosys
POWERED BY INTELLECT
DRIVEN BY VALUES

% age chart

% age

75

50

25

0

Sem 1 Sem 2 Sem 3 Sem 4 Sem 5 Sem 6 Sem 7 Sem 8

Sem I: Applied Math(AM)–I, Applied Sciences(AS)–I,English Mechanics, Basic Electrical & Electronics English, Computer Programming–I, Basic workshop
Sem II: AM–II, AS–II, Comm Skills, English Drawing, Computer Programming–II, Basic workshop–II
Sem III: AM–III, Electronics Devices & Circuits,Electrical Network, Discrete Structures, Digital Logic Design & Applications, Foundations of I.T
Sem IV: AM–IV, Principles of Communication Engineering, Computer Organization & Architecture, Data Structures & Algorithms, Information Theory & coding, Industrial Economics & Management
Sem V: AM–V, Digital Communications, Automata Theory, Microprocessors, ComputerNetwork, Presentation & Communication Techniques, Computer Programming Lab
Sem VI: System Software & Operating Systems, Software Engineering, Object-Oriented Analysis & Design, Internet Technology & Applications, Database Systems, Digital Signal Processing
Sem VII: Computer Simulation & Modeling, Mobile Computing, Image Processing, Management Information Systems, Elective–I, Project-Steganography in images
Sem VIII: Data Warehousing & Mining, Multimedia Systems, Project Management, Elective–II, Project - Steganography in images

SKILLS

Databases treemap

NoSQL		Distributed
Mongodb (3 yrs)		Hive (5.5 yrs)
Elastic search (2.5 yrs)	Neo4j (0.7 yrs)	**Databases**
	Redis (0.2 yrs)	Time-series
		Influxdb (3.5 yrs)

Relational		Vector	
Maria DB (1 yr)	Postgres (0.2 yrs)	Weaviate (0.2 yrs)	NNext (0.2 yrs)

Libraries donut chart

Data Processing 14.8%
Event Processing 3.7%
Web Framework 7.4%
Visualisation 22.2%
ML 51.9%

Libraries: USE, umap, pyod, spacy, openai, gensim, eigen, swifter, pandas, pyspark, pykafka, flask, scikit-learn, standfordNLP, sentence_transformers, dash, plotly, seaborn, matplotlib, openGL, trimesh2, clmg, hdbscan, pyspark.ml, datasets, factorie, h2o.ai

C++ **SCALA** **PYTHON**

Toolkit donut chart

Shell 8.7%
Terminal 4.3%
Web 8.7%
Git 13%
Notebook 8.7%
Productivity 13%
Design 8.7%
Orchestration 8.7%
Monitoring 4.3%
IDE 21.7%

Toolkit: DataGrip, VSCode, Atom, IntelliJ, Pycharm, homebrew, tmux, iterm2, JSFiddle, Postman, SourceTree, github, BitBucket, Colab, Jupyter, Jira, notion, todoist, Canva, Figma, Airflow, Jenkins, Grafana

Languages treemap

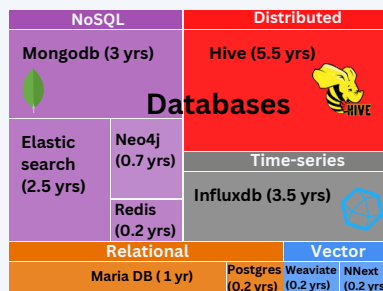| Primary			Secondary
Python (9 yrs)	SQL (9 yrs)	Scala (3 yrs)	Java (3 yrs)
Languages			C++ (3 yrs) / Basic (2 yrs)
Shell			
Bash (9 yrs)	Fish (7 yrs)	Zsh (2 yrs)	C# (3 yrs) / C (2 yrs)
			Matlab (2 yrs) / iOS (1 yr)

INTERESTS

NNext.ai
Lead ML Engineer
[Nov 2022 - Jan 2023]

- Productionized approach for transforming unstructured data into vector embeddings, followed by dimensionality reduction using techniques like PCA, UMAP & t-SNE. Coded & structured the entire functionality from scratch & surfaced it as API's, which could be tested from Postman. Used plotly to test visualisation.
- Built a module to detect outliers and misclassified points in vector space
- Prototyped a feature to empower sellers to query & view product reviews as points in a 3-D space
- Authored a library for database communications. Supported databases include Postgres, Weaviate & NNext DB

Collective[i], New York City
Data Science Engineer
[Oct 2016 - Apr 2022]

- **Incremental unification workflow**
  Created a light workflow for unifying person & organisation entities; resulted in **10x** improvement in speed
- **Prediction Systems for deal pipeline insights(Python)**
  Co-authored system for predicting deal win probability, close time & customer revenue based on deal pipeline. This included dashboard monitoring, Airfow orchestration, legacy code transition & technical-debt payoff
- **Automl architecture (Python)**
  Prototyped a democratized ML architecture (automl) using h2o.ai for predicting deal pipeline insights
- **Quarterly business report generator (Python)**
  Created a python script for generating quarterly business reports for the client success team. Reports would be generated from queries & sent to stakeholders via email; Simplified UI for inputting parameters using Jenkins
- **Active learning approach for Person profile unification (Python)**
  Created pipeline for de-duplicating people records in the System, achieving compression of **10%**. Added visual time-based dashboards to capture metrics of processes
- **Parcel creation utility for Anaconda distribution (Python)**
  Developed a utility that installs python packages and downloadable tars & then converts the setup into a Cloudera parcel for easy distribution across machines using Cloudera's parcel manager
- **Word2vec implementation for getting top news for sales reps (Python)**
  Trained a word2vec model for getting top news articles based on cosine similarity for sales representatives

x.ai, New York City
Data Scientist
[Oct 2014 - Sep 2016]

- **Person Entity Detection (Scala)**
  Created data model & system for detecting Person entities using Semantic-role-labelling & anophor resolution
- **Dashboard management (Scala)**
  Authored & maintained dashboards for measuring annotation quality, model performance & company KPI's
- **Email-parser (Python & Scala)**
  Created a parser for segmenting emails. Refined it into a linear chain conditional-random-fields (CRF) model. This boosted existing F1 score by **15%**

Infosys Ltd, Chandigarh, India
Sr. Systems Engineer
[Sep 2008 - May 2012]

- High priority support and issue fixing on CA Siteminder suite of security products
- Administrator and maintainer of the website (ASP .net) designed for tracking team progress
- Certified security agent builds on App/Web Servers on various platforms (Windows, AIX, Linux, Solaris, z/OS)
- Worked on implementation of i18n in a client product. Delivered the project in very tight timelines