Prateek Garg

gargprateek.iitb@gmail.com | +91-9004697545
linkedin.com/in/gargprat

EDUCATION

IIT BOMBAY
M.Tech.
Computer Science & Engg.
July 2015 | India

SMVD UNIVERSITY
B.Tech.
Computer Science & Engg.
June 2012 | India
University Gold Medalist
Cum. GPA: 9.99 / 10

SKILLS

FRAMEWORKS
• Hadoop • Spark • Hive
• Solr • Kafka • Oozie • Neo4j

PROGRAMMING
Over 5000 lines:
• HQL • Python • Shell • Java

Over 1000 lines:
• C • C++ • Scala • LaTeX

LINKS

Github:// **gargprat**
Medium:// **gargprateek.iitb**

COURSEWORK

GRADUATE
Adv. Distributed Systems-Engg a Cloud
Implementation Techniques for RDMS
Algorithms and Complexity
Artificial Intelligence
Information Retrieval & Mining
Machine Learning (Audit)
(Teaching Asst)
System Administrator
Computer Programming and Utilization

UNDERGRADUATE
Relational DBMS
Theory of Computation
Computer Organization & Architecture
Engineering Mathematics
Artificial Intelligence
Neural Networks

EXPERIENCE

MORGAN STANLEY | Senior Software Developer
Aug 2015 - Present | Mumbai, India

- Designed & developed applications using **complex data processing pipelines, on terabyte scale data**, to curate client's demographics and historical financial information leveraging **Hadoop, Spark, Hive, Kafka & Solr**
- **Optimized complex Hive queries** using techniques including data storage formats, vectorization, join orders, and query rewriting to accommodate join skew over and above Hive's CBO
- Designed & implemented a **Spark** based framework to **partially update** Solr Index from data on HDFS
- Incorporated parallelism to enhance performance thereby allowing index update on 60+ attributes of **15+ million** documents in **under 10 minutes**
- Developed framework, using **Python**, around **Apache Solr** logs for insights & analysis on query patterns and query performance

INDIAN INSTITUTE OF SCIENCE | Summer Research Fellow
Summer 2011 & 2012

- Developed a quantitative steganalysis module to estimate amount of data hidden in image with modifications, as large as 10%, to DCT coefficients
- Implemented an automatic video based traffic analysis system leveraging optical flow for detection & tracking of vehicles driving through controlled area

PROJECTS

HADOOP DISTRIBUTED FILE SYSTEM
Prof. D.B. Phatak | IIT Bombay
Proposed and implemented a **block placement policy** to tackle bottlenecks due to optimistic assumptions on homogenity and data locality in Hadoop.

RAFT, A CONSENSUS ALGORITHM
Prof. Sriram Srinivasan | IIT Bombay
Implemented Raft, a consensus algorithm for managing replicated log in distributed systems using **Go**. Implementation includes leader election, log replication along with sustainability against dynamic crashing of functioning servers and addition of new servers.

TABLE PARITIONING FOR PARALLELISM
Prof. S. Sudarshan | IIT Bombay
Extended syntactic implementation of **PostgresSQL-9.2**, in context of parallel databases, to support table partitioning to handle multiple related tables. Tables on remote servers, linked to master table by referencing foreign key, re partitioned on-the-fly based on partitioning of master table.

AWARDS

2013	National	**AIR 90** (out of ~225000) in GATE - 2013
2012	University	**Chancellor's Gold Medal** - SMVD University
2012	University	**Full fee waiver** for entire B.Tech. based on merit