

Summary:

- ☐ Hands-on data scientist with strong machine learning concepts and techniques: Regularization, Boosting, Random Forests, Decision Trees, Bayesian models, Neural networks, Support Vector Machines (SVM)
- ☐ Built statistical models using supervised and unsupervised learning
- ☐ Built end-to-end data ETL applications for natural language processing.
- ☐ Strong hands-on skills in developing parallel distributed data processing and analytics applications in Java and Python, hosted on Amazon Web Services S3 EC2 and EMR clusters using PySpark, Hadoop, MapReduce, Hive, Cassandra, Hbase, Spark, Splunk, MongoDB.
- ☐ Developed Map Reduce jobs for deployment on large clusters of 300 nodes, configured and managed hadoop clusters in production.
- ☐ Self-starter with strong written and oral communication skills. Demonstrated success as a team player possessing the commitment and discipline to work well independently.

Technical Skills

- ☐ Language/Software: Python, Java,
- ☐ Text Mining - Elastic Search, Python NLTK, Beautiful Soup
- ☐ Data Collection frameworks – Python Scrapy
- ☐ Data Modeling Tools & Packages – R, SAS, @Risk, WEKA, KNIME, NumPy, MATLAB, JMP, WEKA
- ☐ BI Visualization - Tableau, Excel
- ☐ Operating System: Ubuntu, Red Hat Linux, Windows.
- ☐ Cloud Infrastructure Environment: Amazon Web Services EC2 EMR & S3 instances, IAM.
- ☐ NoSQL databases: Cassandra, MongoDB, ElasticSearch
- ☐ Tools: PyCharm, Eclipse
- ☐ Distributed/Parallel Processing: Cloudera Hadoop, Splunk, HDFS, Cassandra, Hive
- ☐ Statistical Modeling: XLStat, R, SAS
- ☐ Cloud Virtualization Frameworks: OpenNebula, VMWare vCloud, OpenStack, Chef, Puppet, Docker
- ☐ Source Control: Git, CVS
- ☐ GitHub Repository - <https://github.com/Saarus/>

Professional Experience:

Macys, Atlanta, GA (Pyramid Consulting)

03/18 – Till Date

- ☐ Working as part of implementation team of metrics collector system for Macys Data center
- ☐ Developing REST api generated data ingestion ETL pipeline using Python, PySpark and Cassandra. The system is built and tested on Ubuntu and RedHat linux instances.
- ☐ Using GitLab, Jira and Jenkins and Chef for continuous integration and deployment.

Coca-Cola, Atlanta, GA (CapGemini)

04/17 - 03/18

Principal data and cloud architect

- ☐ Worked as part of implementation team of Net sales record system for Coca-Cola bottler companies
- ☐ Developed automated scalable data ingestion ETL and analysis framework using Kubernetes, Docker, Azure services (HDInsights, Compute, AZURE SQL Managed service) Python (Numpy, Pandas, Scipy, PySpark) and Kafka.

- Used Git, Jira and Jenkins , SaltStack for continuous integration and deployment.

Federal Government, Remote, Suwanee, GA (assured-consulting)

08/15 – 4/17

- Worked as part of implementation team of Python based data scraping, archival storage and indexing of FAA data.
- Developed automated scalable data ingestion framework using Python (iPOPO, NLTK, Scrapy, BeautifulSoup, PySpark), Kafka, Docker, MongoDB, Elasticsearch, Splunk
- Used Git, Jira and Amazon web services cloud environment to spin EC2 instances and scale docker swarm clusters.

NOKIA (Here, a Nokia Company) Remote, Suwanee, GA (TCS)

07/13 to 08/15

Data Architect-Big Data (consultant)

- Worked as part of implementation team of Big data, Hadoop MapReduce ETL project market place transactions
- Responsible for porting the ETL process from in-house cloudera hadoop environment to Amazon Cloud EC2 & EMR Cluster environment , instance setup and data load in Amazon NoSQL distribution (DynamoDB),Cassandra
- Developed ETL pipeline for real time analytics using Splunk, Apache Spark, Storm, Kafka, Zookeeper, and Cassandra with production deployment on AWS in Java & Scala.
- Developed custom Cassandra (DataStax Enterprise Edition) applications to run Spark nodes for Machine learning and sentiment analysis modeling using Mahout and R libraries
- Created Archival Storage & Indexing previous questions & answers for Full Text Search utilizing features like Tokenization, Stemming, Filters, Analyzers, Fuzzy Matching

Environment: Java, Scala, Python (NLTK, Scrapy, Beautiful soup, PySpark), Ruby on Rails, Amazon web Services EC2, EMR & S3, AWS Beanstalk, Amazon NoSQL DynamoDB, MongoDB, ElasticSearch, JSON, Java, Hadoop Map Reduce, Hive, Cassandra, Zookeeper, Splunk, Apache Spark, Apache Storm, Kafka, Docker

Home Depot, Smyrna, GA (TCS)

12/12-07/13

Architect-Big Data

- ☐ Led implementation team of Big data, Hadoop MapReduce project for revenue optimizer and pricing mark down engine, business analytics using R, Java, Apache Hadoop, Cassandra, Hive HQL and predictive analytical modeling.
- ☐ Developed custom Cassandra applications for real time retail data segmentation and price optimization algorithms.
- ☐ Responsible and accountable for the coordinated management of multiple clients related analytics projects directed toward strategic business and other organizational objectives.
- ☐ Build credibility, establish rapport, and maintain communication with stakeholders at multiple levels.
- ☐ Maintain continuous alignment of program scope with strategic business objectives, and make recommendations to modify the program to enhance effectiveness toward the business result or strategic intent.
- ☐ Configured and managed hadoop clusters using scalable open source cloud frameworks for optimal utilization of available nodes.
- ☐ Coach, mentor and lead personnel within a technical team environment.

Environment: Java, Python, Erlang, Ruby on rails, Cloudera Hadoop, Map Reduce, Cassandra, Hive, SAS, R, MongoDB, OpenStack, Amazon web Services, Chef, Puppet, Nagios, Storm, Zookeeper

AT&T Peachtree, GA (Collabera)

1/12-11/12

End-to-End Architect-Mobile Gaming API Platform

- ☐ Drive Solution Architectures for high performance and very highly scalable platforms.
- ☐ Collaborate with key Telecommunication Architecture on definition of delivery of solution.
- ☐ Developed and deployed mapreduce jobs in 6 hadoop clusters across 3 data centers. Total data nodes in hadoop cluster: ~133. Total capacity in terabytes: ~750TB. Total distinct business jobs running on these clusters: ~ 125. Total hadoop map-reduce jobs running daily across all hadoop clusters: ~1750. Configured Clusters on top of scalable OpenStack cloud infrastructure.
- ☐ Develop RESTful Java Web services using JAX-RS, EJB3, JPA Hibernate
- ☐ iOS and Android utility apps and games using Java Android SDK, HTML5, PhoneGap and native gaming engine Unity to call ATT's RESTful API's such as speech, Tropo, Phono etc.
- ☐ Participate in Joint Application Design/Requirements sessions
- ☐ Documentation: Architectural solutions to all phases of delivery including:-Initial end-to-end solution (based on Marketing Business Requirements), High Level Architecture Design, Content and Diagrams: -Activity -Context -Use Cases -Sequence -Deployment -Logical Architecture - Physical Architecture -Architectural budgets and costs based on hardware, software, licenses, maintenance, and development resources.

Environment: Java, Oracle ADF, Hadoop, Map Reduce, Lucene, Solr, JAX-RS, EJB3, JPA, Hibernate, iOS, TCP/IP, HTTP, SOAP, REST, XCAP, XDMS, XML, XSD, Cloud computing, Eclipse, Enterprise Architect, HTML5, PhoneGap, OpenStack, Chef, Puppet, Nagios

CherryBelly LLC, Suwanee, GA

03/11-01/12

Game Designer,Infrastructure Lead Architect/Team Lead for Facebook Gaming Framework

- ☐ Cherry Belly provides adults with lighthearted, engaging social media games with socially responsible messages. Game player base is monetized by placing advertisements as part of the storylines as well as selling virtual goods.
- ☐ Provisioned and managed Amazon Cloud servers for initial launch.
- ☐ Setup private cloud infrastructure using OpenStack to provide game developers a scalable, elastic shared data center to host gaming applications with optimal and cost effective hosting environment.
- ☐ Designed and developed Custom Rule Based Gaming Engine portal using Liferay portal framework, Java Server Faces, Spring and Hibernate.
- ☐ Designed and developed Single Sign On Interface with Google+, Facebook & MySpace.
- ☐ Developed custom pluggable architecture for in-game advertisement display and payment portal to accept credit and debit card transactions as well as Facebook credits.
- ☐ Designed predictive models using in-game player activity, Google analytics and Facebook insight data, Facebook fan pages and twitter followership to maximize key performance indicators for brands and achieve monetization in social media space.

Environment: Windows, AIX, PHP, core Java 5, JSF, Lucene, Solr, Mahout. LifeRay, Javascript, JQuery, Maven, Tomcat, Spring, Hibernate, XML, Facebook Graph API, Mahout, Amazon Web Services EC2, OpenStack, Chef, Puppet, Nagios, Apache Hadoop MapReduce

Prior Experience available on request.**Educational Qualification:**

B.E, Electrical & Electronics, Manipal Institute of Technology, India
Graduated 1995

MBA, Emory University, Atlanta.
Graduated December 2013