ABHILASH BHAKTARAHALLI SUBRAMANYA

213-573-9659 | abhilashbss@gmail.com | LinkedIn | Portfolio | Github | Fremont, CA

SKILLS

Programming Languages: Python, Bash, SQL, Java, Scala, GoLang, C/C++, Gremlin	**Cloud**: AWS, GCP
Data tools: Spark ecosystem, Flink, Airflow, Nifi, Kafka, Hive	**Databases**: Redshift, Redis, Bigquery, Postgres, Cassandra, Scylla
CI/CD: Jenkins, CircleCI, Harness	**Orchestration**: Kubernetes, Docker, YARN
Self-Taught tools: Kinesis, MySQL, Tableau, MongoDB	**Web Frameworks**: ReactJs, SpringBoot, Flask, Django, Javascript

WORK EXPERIENCE

Site Reliability Engineer	**Oct 2023-Present**
Tiktok USDS, Mountain View, USA
- Managing operations of YARN and Kubernetes clusters of size of ~11000 nodes and ~15 PB of memory in multiple datacenters
- Created a Python utility application employing Generators and Asyncio for efficient log processing of Spark applications
- Deployed and owned Spark, YARN and various other batch components on a new data center
- Technologies Used: YARN, Spark, Flink, Python

Data Engineer	**June 2023-Sept 2023**
Chartboost, USA
- Configured Nix shell environment for multiple python projects and crafted reproducible container images with Poetry and Nix packages for GitHub Action workflows
- Enhanced Airflow DAGs using custom context managers to ensure atomic Bigquery updates for data rate of 1B records/hr
- Technologies Used: Nix, Ruff, Actionlint, Pyright, SqlFluff, Black, Pylint, Airflow, BigQuery

Data Engineer	**Nov 2019-Jul 2021**
Zeotap India Pvt Ltd, Bengaluru, India
- Built data pipelines in spark for analytical and data delivery use cases on a **100 billion**-node-ID-Graph and orchestrated in Airflow
- Automated and Improved ScyllaDB recovery by *85%* (*2-3 hours* to *20 min*) through a multi-threaded Python application
- Created CI/CD workflows using CircleCI and Harness.io for machine learning applications
- Built and managed Spark data export pipelines for various clients driving a total annual revenue of ***2M+ USD***
  Technologies Used: Spark, Zeppelin, Flink, Kubernetes, Livy, Scylla, Janusgraph, Python, Scala, Dataproc, EMR, Apache Airflow, AWS, GCP, S3, Hive, HDFS, Harness, CircleCI, YARN

Data Engineer	**Jul 2018-Oct 2019**
Endurance International Group, Bengaluru, India
- Engineered scalable Kafka streaming and batch Nifi ETL pipelines for data replication from 10+ OLTP sources to Redshift
- Integrated Bamboo and Jira with Google Chat using Python for alerts, leading to an annualized cost savings of ***700,000+ USD***
- Set up and configured Datadog to monitor Kafka and Nifi data pipelines and alert on metrics related to performance
  Technologies Used: Apache Nifi, Kafka, Kafka-connect, PostgreSQL, Redshift, Redis, AWS, Java, Python, Azkaban, S3, Metabase, HDFS, Jenkins, Bamboo, Datadog

EDUCATION

University of Southern California - *Los Angeles, USA*	**Aug 2021-May 2023**
Masters of Science in Computer Science, CGPA - **3.91/4.00**

BMS College of Engineering - *Bengaluru, India*	**Sep 2014-May 2018**
Bachelor of Engineering in Computer Science, CGPA - **9.33/10.00**

PROFICIENCIES

Soft Skills Highlights: *Passion* - Conducted various learn and share sessions among teams | *Trustworthy & Reliable* - Became point-man of data systems for resolving cross-functional team's data requests efficiently and received various ***spot awards*** | *Innovative & Process Improvement* - Created various utility tools to handle big data job failures for a better triage process in team

PROJECTS

- **Weenix Kernel Development:** Iteratively developed operating system kernel in C. Implemented Processes, Threads, Mutexes, and Context switching system calls in phase one. In second phase, built Virtual File System to create/delete files and directories. Third phase involved implementation of fork system call and Virtual Memory management system

- **EDFS and Partition MapReduce Framework:** Using Python, Firebase, MongoDB created. Created HDFS-like file system that stores data partitions as DB tables. Provided Map/Reduce API to perform map and reduce operations on EDFS data