SAI KIRAN NEELAKANTAM

Hillsboro, OR 97124 | 480-332-4188 | sai.neelakantam@outlook.com | https://www.linkedin.com/in/saineelakantam/

TECHNICAL SKILL SET:

Hadoop Ecosystem	: HDFS, MapReduce, Hive, Impala, Pig Latin, Sqoop, Oozie, YARN, Avro, Parquet
Apache Spark	: Spark Core, Dataframes, Spark SQL
Kafka Ecosystem	: Kafka Core, KConnect, KStreams, Schema Registry, REST Proxy
AWS Ecosystem	: S3, EMR, EC2, RDS, Athena, DMS, DP, SQS, SNS, IAM, Lambda, API gateway
DevOps Tools	: Docker, Terraform, CircleCI
Programming Languages	: Python, JAVA 8, SQL, PL/SQL, Shell Scripting, HTML5, XML, CSS3, JavaScript
Automation & Versioning Tools	: Airflow 1.8, Autosys, Oozie, Maven, SVN, GitHub
DBMS Tools	: Snowflake, MySQL, Oracle, Teradata

WORK EXPERIENCE:

Sr. Big Data Engineer – Nike Inc   **Feb'17 to Present**

- Evaluate, extract/transform data for analytic purpose within the context of Bigdata environment
- Design and develop efficient PySpark programs using cloud-based data platforms (Qubole, EMR) to extract/transform/load data in between various data warehouse applications.
- Built frameworks for various AWS technologies like S3, EMR, EC2, Athena, Datapipeline, DMS and SQS.
- Built terraform scripts, dockerfile and Circle CI jobs to automate the infrastructure deployment of the AWS services.
- Developed various microservice application using Spring Cloud framework and AWS services (SNS, SQS) to ingest real-time data from REST endpoints to data warehouse applications.
- Built data pipelines in Kafka to allow for data streaming and developed KStream application to transform the data
- Built custom producer and consumer for troubleshooting event tracking ecosystem. Audit records will be produced when a event is produced and then when a event is consumed.
- Developed complex data transformation logics in KStream application by building a dynamic transformation framework and built CI/CD pipelines to deploy them to KStream servers based on GitHub commits/branches.
- Developed various custom plugins/hooks/operators in Airflow which can be used in DAG's and other adhoc requirements.

Data Engineer – Bank of America   **Aug'15 to Feb'17**

- Migrated the existing data from Mainframes/Teradata/SQL Server to Hadoop and perform ETL operations on it.
- Optimized DQ framework using Spark (Dataframes, Spark Core) and decreased the runtime to meet the SLA of the job
- Involved in extracting the data from various sources into Hadoop HDFS for processing.
- Worked hands on with ETL process and ad-hoc requirements like developing UDF's to implement business logic in hadoop.
- Configured the daily workflow to manage independent jobs and to automate Pig, Hive and Sqoop etc. jobs.
- Used the Mainframe SerDe's and Avro SerDe's for serialization and de-serialization in hive to parse the contents.
- Created various shell scripts to automate the day-to-day activities and JIL's to automate the workflow in Autosys
- Worked on on-call production issues – scrubbing, resolve hive query issues, workaround for defects within SLA duration.

Big Data Engineer – Cloudwick Inc   **May'15 to Aug'15**

- Worked on setting up Hadoop cluster using Cloudera Manager and Hortonworks distributions, upgrading the versions of various components, cluster maintenance a performance tuning
- Analyzed the Big Data business requirements and transformed it into Hadoop centric technologies
- Developed various shell scripts to help us in day-to-day activities like memory utilization, CPU usage, storage and networks.
- Built various data pipelines using Oozie and perform complex ETL operations using Hive.
- Involved in initial design and planning phase of Spark adoption and provided support to multiple DE teams in their adoption

Software Graduate Intern – Intel Corporation   **May'14 to Nov'14**

- Developed and maintained an Automation application called ePDRD which performs various automation tools on the processor package design data to export and import data from various applications.
- Used various advanced data structures and algorithms to export/import data to an external program.
- Validated the datasets before exporting the data by compare 2 datasets based on predefined rules etc.
- Improved the efficiency and quality of package engineering design process.

CERTIFICATIONS:

- Cloudera certified Administrator for Apache Hadoop CDH 5
- O'Rielly Developer Certification for Apache Spark

EDUCATION:

Arizona State University, Tempe AZ, US	**Aug'13 to May'15**
Master of Science in Software Engineering	GPA: 3.7 / 4
Osmania University, Hyderabad TG, India	**Aug'08 to July'12**
Bachelor of Engineering in Information Technology	GPA: 3.41 / 4