Sunil Singh

GitHub: https://bit.ly/30ZIPT8
LinkedIn: https://bit.ly/2BegrCn
Email: sunil.hadoopdev@gmail.com
Mobile:425-458-8105

ABOUT ME

- Competitive software engineer with 11+ years' experience, possessed with solid professionalism & collaborative personality, with significant experience in software engineering related problem-solving using Java, Python & C++.
- Acquired experience in data engineering building pipelines, ETLs, schema modeling, data mining, data analysis, backend, web developments, micro-services, distributed computing, large-scale system design, networking latency optimization, distributed data storage, and cloud computing.

EDUCATION

University of Pune
- Master's in Computer Science

PROFESSIONAL EXPERIENCE

The Home Depot   **August 2020 to February 2021**
Senior Software Engineer

Module: Order Fulfillment   **Domain**: Retail

- Implementation effort on Re-designing existing efficient ways for consumption of event generated at different checkpoint of order life cycle, pushing, polling & peeking event from event queue defined in Kafka & Google PubSub, as multi-thread enabled microservices to gain higher performance when deployed in Google Cloud Run.
- Constructed module tool for capturing various missed or lapsed events for reprocessing based on recognizing pattern defined via Home-depot In-house algorithm, constraints & requirements.
- Implemented REST endpoints which provided the abstraction for ad-hoc query on elastic-search to ease the discovery in real-time for fulfillment Instructions for finance team.
- Engineering effort for testing, deploying & troubleshooting using Docker, Kubernetes, Jenkins, Google Cloud build, Google cloud watch while handling deployment in firm Infrastructure.
- Orchestrating & troubleshooting of various job scheduler defined as data pipeline using spark. And processing them on Hadoop ecosystems running in GCP.
- Designed the POCs for business delivery addressed the lapsed event audit storage on NoSQL Cassandra.

Java Spring Boot Microservices Event-Driven Architecture Python Kafka Cassandra Apache Spark Hadoop Apache Hive NoSQLBooster Parquet Postman Elasticsearch GCP IAM GitHub Git Docker Jenkins Kubernetes Google Cloud Platform Bootstrap NGINX

State Compensation Insurance Fund   **October 2019 to August 2020**
Data Engineer

Module: Data Management   **Domain**: Insurance & Claims

- Improvement effort related to the development of Online Payroll Application engine to process policy, claims & insurance related data for entire California state, Improved the source data loading from legacy data-warehouses into public AWS S3 & on-premises CHD HDFS.
- Addressed the business searching requirement using various constraints such as geocode, phone number, business name & alias, using apache solr text searching & indexed the data in HBase, improved the page ranking using the string distance calculation algorithm.
- Develop data pipelines optimized for scaling and troubleshoot potential platform issues.
- Ingested real-time streaming data from various claims, insurance & DMV data provider & Google Map API used this data to analyzie and monitored incoming update data to identify current deltas and detect anomalies in the stream (Kafka, Spark, Cassandra, Python).

Python Kafka Apache Spark Hadoop Apache Hive PySpark Apache Parquet Solr AWS S3 AWS EMR Git Jenkins Google Map API AWS CloudFront AWS DynamoDB AWS Kinesis Cloudera CDH

Google   **September 2018 to September 2019**
Software Engineer

Module: Google AdSense & Google Fiber   **Domain**: Internet

- Implemented couple of features for Google TV IPTV Platform which deliver optimized media content bandwidth for individual user's; increased 5% more customer satisfaction.
- Involved in implementation of migration of deprecated lookup system for advertisement data into the new MLT (Message Lookup Table) with the help of Dremel & JavaFlume MapReduce.
- Migrated RoSy API based implementation into the One Platform API within few modules which resulted in 21% faster processing time for call-flow based routing.
- Implemented a distributed algorithm that solved job resource starvation issues cause by configured quota limit. After deployment on the single region, all cell's reporting 98% lesser warnings related to quota limits.

Java C++ MapReduce Protobuf Stubby Goops Borg GSLB uberproxy GFS/Colossus ColumnIO/Capacitor sstable RecordIO Dremel

JP Morgan Chase   **January 2017 to September 2018**
Data Engineer

Module: CCB & CIB   **Domain**: Finance

- o Contributed implementation effort for Ethereum-based distributed ledger protocol which allow transferring digital assets on distributed ledger without leaking buyers, seller, trusted party & asset quantity; improved join-split performance from 49.8 sec to 44.6 sec on hyperthreaded machines.
- o Devised & coded global module to abstract explicit declaration of internal & external dependencies used by different applications; improved application development time 75% faster.
- o Develop Defined & Optimized blockchain latest requirement specification with benchmark demonstration using cosmos framework.
- o Architected & simplified spark streaming based transformation pipelines, they're consuming investment banking data from legacy warehouse Greenplum into Cassandra & HDFS (data lake); this enables data scientist team to analyze on CIB LOB data with approx. 0ms waiting time.

Python Java Kafka Apache Spark Hadoop Apache Hive PySpark Apache Parquet Solr AWS S3 AWS EMR Git Jenkins Google Map API AWS CloudFront AWS DynamoDB AWS Kinesis Cloudera CDH

AT&T Labs   **April 2016 to December 2016**
Software Engineer

Module: IoT R&D   **Domain**: Telecommunication

- o Developed features for Video Optimizer under the Statistics tab; this feature render real time visualization of user trace associated with peripheral usage & battery drainage; this increase 8% more user satisfaction due to visualization & enabled support for controlling peripheral.
- o Programmed global module to abstract explicit declaration of internal & external dependencies, which declared during construction phase of applications; improved development time 75% faster.

Java Spring boot Microservices Maven Kafka Apache Spark Apache Storm Solr HBase Hadoop Cloudera CDH AWS S3 AWS Lambda AWS Redshift AWS CloudFront

American Express   **November 2015 to March 2016**
Software Engineer

Module: Stream & Risk Management   **Domain**: Finance

- o Refactored around four plus legacy UI interface built on java applets & flash; cloned them via Angular.JS & decoupled backend as separate micro-service for payment network.
- o Implemented event driven transaction processing using immutable append only logs as system of record; this system fulfills two promises high availability, durable record storage, and a repeatable total order on those records; this provides almost approx. consistent response for write rate.

Java Spring boot Microservices Maven Kafka Apache Spark Apache Flink Flume Solr HBase Hive Hadoop MapR AWS S3 AWS Redshift AWS CloudFront

AT&T   **August 2013 to October 2015**
Software Engineer

Module: Stream & Risk Management   **Domain**: Finance

- o Collaborated development effort for CDO team for building 52+ batch mode data pipelines on retail sector warehouse to help data scientist for driving insight like store distribution growth, customer revenue & policy fluctuation pattern, optimal routing for 50,000+ field technicians.
- o Contributed development effort on enhancement of OADN-OS (Open Architecture Disaggregated Network Operating system); coded algorithm for transport layer routing protocol used in intra & inter-nodal path selection & tagging traffic IGP, MPLS, BGP-LU etc.

Java Spring boot Microservices Maven Kafka Apache SparkStreaming SparkSQL Flume Storm Cassandra HBase Hive Hadoop CDH EMR AWS S3 AWS Redshift AWS CloudFront

Supplementary Roles:
Developing python script.
Dockerized the applications deployed into AWS & Creation of Kubernetes resources.
Designed and implemented auto regional failover.
Using Cloud Watch for monitoring AWS or GCP cloud resources.
Initiating & setup application monitoring & troubleshooting.