

Professional Summary:

- Around **11+ years** of experience in Information Technology that includes **Databricks, Snowflake Development, Cloud Data Engineering, Hadoop and its ecosystem Development, Spark Development, Hadoop Administration and Networking Automation Framework Development.**
- Implemented End-to-End data pipelines using **Databricks** and its components (**Databricks Repos, Databricks SQL warehousing, Unity Catalog, Databricks workflows and Delta Live tables**).
- Implemented common API using **pyspark** to ingest data from various sources into the Databricks delta tables.
- Having experience in designing and implementing **generic data ingestion** and transformation framework in **python** to ingest data into **snowflake** from different sources.
- Having experience in implementing generic **auditing, error handling and data quality framework** for ETL pipelines in **Python**.
- Having experience in scheduling, monitoring ETL pipelines using **Airflow**.
- Having good knowledge on GCP cloud components like **GCS buckets, Dataproc, Big Query and Composer**.
- Having good knowledge on Azure cloud environment and data analytics services like **ADLS, ADF, Azure Databricks, Azure Event Grid Azure Key Vault and logic apps**.
- Experience in **development & administration** of Big Data applications, **Batch processing and streaming applications** in secured (**AD/LDAP, Kerberos and Ranger**) environments.
- Experience in Big Data technologies like Hadoop framework and its eco-system such as **HDFS, Map Reduce, Yarn, HIVE, Pig, HBase, Flume-NG, OOZIE, Hue, Sqoop and Zookeeper**.
- Experience in implementing **ETL pipelines** in **Scala/Python** using **Spark libraries** like **Spark Core, Spark Streaming and Spark SQL**.
- Experience in implementing **streaming applications** in **Scala/Python** using **Flume-NG, Kafka, Spark Streaming and Hadoop API**.
- Experience in integration of **Kafka** with **Spark Streaming** to ingest click stream data from Adobe into **Hadoop (HDFS)**.
- Experience in integration of **Kafka** with **Spark Streaming** and ingest data into NOSQL database **HBase** using **Scala and Python**.
- Experience in ingesting data from sources like **Teradata, Oracle and MySQL** into **Hadoop (Hive)** and vice versa using **Diyotta ELT** application.
- Experienced in writing **custom UDFs** and using them in **Hive** and **Spark SQL** applications.
- Good working experience with **Hive** and **HBase Integration** using **HBase Storage Handlers**
- Experience in schedule and trigger **MR Jobs, Spark, Sqoop actions, Hive Actions, Pig Actions and Hadoop Streaming** jobs using scheduling tools like **CTRL+M, CRONTAB and Oozie**.
- Experience in managing and reviewing the **Hadoop log files** for **debugging** the running **Spark Jobs, Yarn jobs and Map Reduce Jobs**.
- Experience in monitoring the health of Hadoop clusters using **Ambari UI** and **Cloudera Manager UI**.
- Experience in importing and exporting data using **Sqoop** from **HDFS** to **Relational Database Systems (RDBMS)**, **SQL** and vice versa.
- 4-year experience in **Networking Domain, Implementing Automation frameworks in Python** for interacting with **Multi-Layer Switches** (Switching and Routing), **QLogic CNA 1G/10G cards** (Ethernet) and Cavium security processors (SSL, TLS and IPsec) and validate their functionality in **Linux** and **virtualized** environment
- Skilled in writing **Scripts** in **Scala** and **python** using **Spark, Spark Streaming and Spark SQL API**.
- Experience python libraries **Pyspark, Pyhive, Pydoop, Hdfs, Happybase, Hadoopy, pexpect, numpy, pandas** and other built in modules like **os, re, sys, time** etc
- Experience in loading data into the **Hive** with the help of **Hive JDBC Driver** using **JAVA JDBC API**.
- Experience in Implementing the **Restful Web service** to show the **Hbase** data in web using **URL**.
- Experience in interacting with network devices like **Emerson ATCA Multi-layer Switches** (Switching and Routing), **Qlogic 1G/10G converged network adapters** (Ethernet) and Cavium security processors.
- Experience in implementing **Automation Framework and libraries** in **Python** to validate the security protocol functionality (SSL, TLS, IPSEC, Cavium Driver functionality) of Cavium security processors in **Hadoop Environment, Linux** and **virtualized (qemu)** environment.
- Experience in Implementing **Automation Framework and libraries** in **Python** to validate NIC protocol functionalities like **Ethernet, IP Fragmentation, Bonding, MTU, VLAN, Network Partitioning (NPAR) and SR-IOV** etc on all linux distributions like **Redhat, Cent OS, Suse** and **VMware environment(VMWARE Esxi)** for **Qlogic 1G/10G Converged Network Adapters (CNA)**.
- Experience in all stages of **SDLC (Agile, Waterfall)**, writing **Technical Design documents, Development, and Testing**.

Education:

- **Masters (MS)** in **Computer Science** with GPA **3.8/4.0** from University of Central Missouri, Kansas City, Missouri (Aug 2015-Dec 2016).
- **Bachelors (B.Tech)** with GPA **3.6/4.0** from Jawaharlal Nehru Technological University, Hyderabad, India (Aug 2006 – Jun 2010).

Technical Skills:

Operating Systems: **WINDOWS 98/2k/XP, UNIX, Linux, VMware ESXi.**

Programming Languages: **java, Scala, Python.**

Big Data Technologies: **Hadoop, HDFS, MapReduce, Yarn, Hadoop Streaming.**

Hadoop Ecosystem: **Hive, Pig, Hbase, Oozie, Flume-ng, Hue, Sqoop, Zookeeper, Ambari, Kafka.**

Spark Ecosystem: **Spark, Spark Streaming, Spark SQL.**

Cloud Technologies: **Azure, ADLS, ADF, Azure Databricks, Azure key vault, Azure event grid.**

Cloud DW: **Snowflake, Snow SQL, Snow Pipes**

Hadoop Distributions: **Apache Hadoop, Cloudera, Horton works.**

Security Components: **AD, LDAP, Kerberos, Ranger, SSL, TLS, IPSEC.**

Scripting Languages: **Python, Perl, TCL/Expect, Shell.**

Database Languages: **Sql, MySQL.**

Subversion Tools: **Clear case, Svn, Git, GitLab.**

Bug Tracking Tools: **Clear quest, Remedy, QC, Jira, Bugzilla.**

Hypervisors: **VMware ESXi, Kvm, Xen, qemu.**

Traffic Generators: **Spirent, Smartbits, IXIA, Scapy, Netperf, Iperf.**

Ethernet Functionalities: **Bonding, NPAR, SR-IOV, MTU, VLAN.**

L2 Protocols: **VLAN, STP, RSTP, MSTP, LACP, FP.**

L3 Protocols: **RIP, OSPF, BGP, EIGRP, IGMP, RIPng.**

L4 Protocols: **TCP, UDP.**

Other Tools: **Wireshark, Spirent, Hudson, Vncserver, Putty, Xshell, Tcpdump, Jenkin, secure CRT.**

Work Experience:

● Working with **T-Mobile** as a **Data Engineer** in Frisco, Texas

Apr 2020 – Till Date

1. Currently evaluating the Databricks features like DLT, Databricks Volumes and other in **AWS** environment.
2. Worked with sys ads and Platform team to enabling **SCIM and SSO** for **Databricks** from one login.
3. Implemented common API to ingest data from various sources into data bricks delta tables.
4. Evaluated the Data Quality modules like **Great Expectations** and integrated to ingestion framework.
5. Have been working on implementing **ETL pipelines** in Databricks using Pyspark API.
6. Have been working with business stakeholders, collect the requirements and convert them into technical requirements.
7. Involved in designing and implementation of TSM (**T-Mobile Snow SQL Module Wrapper**)
8. Involved in implementing configuration driven data ingestion framework (**TDI**) in Python to ingest data into **Snowflake** from various sources.
9. Implemented generic **auditing, error handling** and **data quality framework** for **TDI & TSM** in **Python**.
10. Implemented POC in **Python** to ingest the data into **Snowflake** from files (Structured and Unstructured data), Oracle and Teradata.
11. Involved in implementing and testing TSM capabilities like **Landing to Stage, Stage to Core, Run SQL** capabilities.
12. Implemented Spark Application in **Azure Databricks** using **Python (Pyspark API)** to migrate the data from on-prem Hadoop cluster to ADLS Gen 2 Storage.
13. Implemented **common utilities, metadata for auditing** in Data Bricks using Pyspark API for Ingestion and Transformation jobs.
14. Implemented a **Pyspark application** to extract the data from **Lithium API** call and store the data into the **Delta Lake tables**.
15. Implemented a Spark Application in Python to generate Promotional data by converting business logic into Spark data frame API.
16. Created a generic **ADF pipeline** to trigger the Azure Databricks Notebooks.
17. Work closely with business users and convert the business use cases into technical requirements.
18. Participate in code reviews and provide feed backs based on the reference architecture, naming standards

and best practices.

● Worked with **Sprint** as a **Data Engineer** in Overland Park, Kansas

Jan 2017 – Mar 2020

1. Implemented **Spark SQL application** to ingest the **Lithium** data (4 report files) into **Hive** with incremental load control logic.
2. Implemented the **Python API** to call the **Lithium REST API** to download chat logs every hour in csv format.
3. Implemented the **Spark application** in **Scala** to load the **Prospect** data received from **Audience Acuity (AA)** into the **Hive**.
4. Implemented **Python script** for pre-validation like **MD5 checksum, File Sizes, part files** etc and **post validation** of prospect data.
5. Ingesting the Data into Hadoop from different sources like **Teradata, Oracle, MySQL** using **Diyotta** ETL tool.
6. Implemented the **Spark SQL application** to process the Device history data from Bright Star into **Hive**.
7. Loaded the **FAST (Sprint Orders data)** in to Hadoop using **Diyotta tool** with control table (hive-Hbase integration tables) logic
8. Worked closely with the **business analysts** to convert the Business Requirements into Technical Requirements and prepared low and high level documentation
9. Rewritten **Hive queries** using **Spark SQL** for faster queries performance.
10. Written the **Python script** which will send the **HDFS utilization report** to the team based on Application.
11. Written the **Python script**, which captures all the data in EDL cluster resource information like **Application details, CPU, Cores, and Memory** etc. loaded into the data into a hive table.
12. Involved in **Setting up** the 350 node Hadoop Cluster using **HDP distribution(2.3.2,2.4,2.5,2.6.3)**
13. Helped the admin team in configuring the **Kerberos AD/LDAP authentication** to EDL cluster.
14. Involved in **Hadoop patch upgrades**, Major **HDP upgrades** along with the Admin team.
15. Written the **Python application** to test all the services like **Hive, beeline, Hbase, Kafka, Hdfs** etc after **upgrade as part of cycle 0 testing**.
16. Having experience in **Ranger** in giving access to users for HDFS, Hive and Kafka etc.
17. Having experience in creating Job docs and scheduling jobs with **CRON TAB** and **CTRI+M** scheduler.

● Worked with **Cavium Networks** as a **Software Engineer II** in Hyderabad, India

May 2014 – Aug 2015

1. Involved in setting up the 25 node **Hadoop cluster** from scratch using **Cloudera manager(CDH 5.0)**
2. Implemented the **spark streaming** job which integrates **Flume** and **Spark streaming** and loads the data into **Hbase** by processing the data with spark streaming using **Scala**.
3. Implemented the **spark streaming** application which gets the data from **Kafka** producer and loads the data into **Hbase** by processing the data with spark streaming.
4. Implemented the spark streaming application which will process the data from multiple files in Hdfs and store the data into Hbase.
5. Worked on **integration of Hive with Hbase** using Hbase Storage Handlers.
6. Involved in creating and scheduling the **MR Jobs, Sqoop actions, Hive Actions, and Hadoop Streaming jobs** using **Oozie**.
7. Experience in **debugging** and **monitoring** the Map Reduce jobs, Spark Jobs and health of Hadoop cluster in **Cloudera Manager UI, Hadoop Log files** and Job History server.
8. Implemented the **python script** to change the **XML data** into the **csv format** to load into the **Hive**.
9. Developed the **python script** to validate all the Hadoop and related services are running in the cluster
10. Developed the **python module**, which will perform all **Hbase functionalities**.
11. Developed the **python script** to delete the processed files from HDFS (Sqoop spool dir.), script will trigger using **CRON** job.
12. Developed the RESTFUL web service application to show the data in **Hbase** in the form of JSON in web using URL.
13. Involved in developing the java application to load the **Hive data** into the **MySQL** database using JDBC from hive warehouse directory.
14. Implemented **Automation framework** and **modules** in **python** to validate Cavium security processor **Driver functionality** in **Linux** and Virtualized environment.
15. Implemented Automation framework and test library to validate **SSL, TLS** and **IPSEC** on Cavium processors on both **Linux** and **Virtualized** environments.

16. Developed **Python modules** to test the **offload functionality** in the Hadoop environment.
17. Developed **Python modules** to test the data compression and decompression offloading to **Cavium security processor functionality** under Hadoop Environment.
18. Good working experience with **Jira** ticketing system for project development.
19. Documentation of User Requirement Specifications and System Requirement Specification

● Worked with **QLogic** as an **Associate Engineer** in Pune, India

Mar 2013-Apr 2014

1. Setting-up **Hadoop cluster** environment with the **Hadoop** stack components using **Apache Hadoop** and **Cloud era Manager** with Dell Power edge servers.
2. Made the setup to get the **network server logs** from **NFS** using **Flume** to Load the data into the **Hadoop**.
3. Involved in loading the data into the **Hadoop** cluster using **Hadoop** file system commands.
4. Configuring **Flume** over the **Server.log** file to automate the transfer process.
5. Developed a **Map Reduce Jobs** based on business logic to extract the required info from the Log file using **Python, Pig, and Java**.
6. Written the **Apache PIG scripts** to process the HDFS data.
7. Created **Hive external tables** to store the processed results in a tabular format.
8. Involved in integrating the **Hive external table to Hbase**.
9. Involved in creating and configuring the workflow jobs and triggering those jobs using **Oozie**.
10. Good working experience with **Remedy** ticketing system for project development.

● Worked with **QLogic** as an **Associate Engineer** in Pune, India

Sep 2012-Mar2013

1. Test topology design and build up test topologies
3. Installation and troubleshooting of **RHEL, SUSE and ESX/ESXi 4.x/5.x on HP, DELL servers**.
4. Involved in **designing and developing automation framework** for testing the **Qlogic NIC functionality on linux and VM ware environment Python**.
5. Sanity Testing of 10GB Ethernet Drivers (which includes Features like **VLAN, Jumbo Frame, NIC Teaming, TCP Offload, vmotion**) for VMware ESX/ESXi 4.x/5.x.
6. Feature Testing **Vlan, bonding, IPalias and MTU, Check Interrupt modes, Aggregation, Pause Frame, Checksum, Card sanity, Ping with Various Size packets, Multicast and LSO**.
7. Stress Testing, Cold Boot and Warm boot during traffic, Unplumbed and plumb interface. Module load and unload, Heavy traffic during all these operations.

● Worked with **Emerson** as a **Software Engineer I** in Hyderabad, India

Oct 2010 – Sep 2012

1. Test topology design and build up test topologies
2. Involved **Test Library and Automation Framework** implementation using **python**.
3. Involved in Test cases Design, Setting up the Network Test topology Designing.
4. Involved in Test library Development and Automation for L2 protocols like **STP, RSTP, MSTP, FP, RATE LIMITING, LACP, VLAN, BRIDGE and QinQ**.
5. Involved in Test library Development and Automation for L3 protocols like **RIP, OSPF, Static Routing, and IGMP**.
6. Test plan execution on a daily basis for L2 Features like **STP, RSTP, LACP, SA, FP, RATE-LIMITING, VLAN, BRIDGE and QinQ**.
7. Test plan execution on a daily basis for L3 Features like **Static Routing, RIP, OSPF, IGMP and Route-Redistribution**.