

PROFESSIONAL SUMMARY:

- **12+ years of experience. 8+ years in Big Data, Analytics, Hadoop development.**
- **Expertise in Implementing CDP-pv , CDP-pc and HDP Clusters on Cloud and on-Premise.**
- **2+ Years in designing Architecture, providing solutions for Big Data clusters.**
- **In depth knowledge of Hadoop Architecture and various components such as Hadoop, Spark, Kafka, Storm, Cassandra and YARN / MapReduce programming paradigm and for working with Big Data to analyze large data sets efficiently.**
- **Hands on experience in working with Ecosystems like Hive, Pig, Sqoop, Map Reduce, and Flume.**
- **Hands-on experience in designing and implementing security like Kerberos, SSL, KNOX.**
- **Experience in importing and exporting tera-bytes of data using Sqoop from HDFS to Relational Database Systems and vice-versa.**
- **Experience in Designing Solution for activities such as Realtime and batch Processing of clusters using Apache, Cloudera and AWS.**
- **Experience in analyzing data using HiveQL, Pig Latin, and custom MapReduce programs in Java.**
- **Hands On experience in NoSQL databases like HBASE. Knowledge of job workflow scheduling and monitoring tools like Oozie and Zookeeper.**
- **Good experience in Python, UNIX and shell scripts.**
- **Involved in creating Hive tables, data loading and writing hive queries.**
- **Experience in NoSQL database MongoDB and Cassandra.**
- **Followed Test driven development of Agile and scrum Methodology to produce high quality software.**
- **Experience working in the Cloud environment like Amazon Web Services (AWS)**
- **Excellent work ethics, self-motivated, quick learner and team oriented. Continually provided value added services to the clients through thoughtful experience and excellent communication skills.**

TECHNICAL SKILLS:

Certifications	AWS Sol. Architect, HDPCD, HDPCA, HDP Spark, CDH5 and CDH4
Hadoop Vendor	Cloudera Data Platform (CDPpv and CDPpc), HortonWorks (HDP),
Databases	NoSQL, MySQL, MongoDB, Cassandra, Hbase
Hadoop eco Systems	Cloudera Manager/Apache Hadoop HDFS, NFS, Map reduce, Hbase, Pig, Hive, Sqoop, Flume, Oozie, Zookeeper, Hue, Hadoop Tokenization, Nagios, Ganglia
Microservices	Docker, Kubernetes
Cloud Provider	GCP, Azure, AWS
Security Tools	SSL, Kerberos, RANGER, Knox
Data Pipelining	Nifi, Streamsets
Technical Blogs	https://vikassri.in / https://bigdatum.wordpress.com

EDUCATION:

B-Tech in Information Technology, India.

PROFESSIONAL EXPERIENCE:

Hortonworks Data Platform, Bangalore

Dec'18 - Till Date

Role: Resident Architect

Client: Retails/bank/Chemical/E-Com

Being in consulting, I have worked for different clients which include finance/retail/chemical sectors. Some of them are world's leader in their own sector.

Responsibilities:

- Building a world class big data lake Using **CDP-pv** and **HDP 3.1/Cloudbreak** on cloud/on premises.
- **Upgrade to CDP-pv** from HDP2.6.5 using sidecar migration/in-place upgrade
- Implement **security on CDP-pv**, including **TLS/Data at Rest/ Knox / Ranger**.
- Requirement gathering and providing End to end solutions.
- Design of platforms involve **multi-tier architecture** with different technology stacks.
- Migration of project from on premise to clouds (**GCP/AZURE**)
- Worked on Installing and configuring **LLAP, Bigdata Stack** and helping developers for tuning **MapReduce jobs** for data cleaning and preprocessing.
- **Performance tuning** of the cluster as well as taking care of the problems and solving it.

Optimum Solution, Singapore

Apr'16 – Nov18

Role: Big Data Architect

Client: Standard Chartered Bank

SCB (known as Standard Chartered Bank) is one of the major British multinational banking and financial services company headquartered in London. Involved in creating a big data environment where we can take advantage of bigdata performance and implement its components to solve some of the issues like name screening/fraud detection etc.

Responsibilities:

- Involved with Solution Architect to build **EDMP (Enterprise data management platform)** Using HDP 2.4.
- Requirement gathering and providing End to end solutions.
- Design of platforms involve **multi-tier architecture** with different technology stacks.
- Involved in all phases of the **SDLC** including analysis, design, development, testing, and deployment of **Hadoop cluster**.
- Implementation of Applications on Big Data System (HDP2).
- Worked on Installing and configuring **MapReduce, HDFS** and developed multiple **MapReduce jobs in java** for data cleaning and preprocessing.
- Create / design the schema on hive and HBase according to requirement.

HCL, Noida
Role: Hadoop Consultant
Client: NBCU

Dec '14 – Apr'16

NBCU (known as National Broadcasting Company) is a major provider of Entertainment Contents, they have hundreds of websites and many channels. This project helps to identify the power of data analytics. We have designed and built the world class data warehouse using the big data Ecosystem. It processes over 20+ TB data per month coming into the organization from Weblogs, Advertisements, website views and from consumers through Different applications.

Responsibilities:

- Handling Team of **6 members**, responsible for building Best Data warehouse.
- Providing **End to End solution** of Bigdata Problems, **Realtime** as well as **Batch** processing.
- Involved in all phases of the **SDLC** including analysis, design, development and deployment of **Hadoop cluster**.
- Worked on the proof-of-concept for **Apache Hadoop framework** initiation.
- Wrote **Hive UDFs** in **Java** where the functionality is too complex.
- Successfully integrated **Hive tables** and Different Application on top of hdfs for analytics.
- Develop **Automations Scripts** and designing data processing flow.
- Implemented the workflows using **Apache Oozie framework** to automate tasks.
- Involved in schema design for the application using **Hive and HBase**.
- Worked on analyzing Hadoop stack and different big data analytic tools including **Pig and Hive, HBase and Sqoop**.

TCS, Gurgaon
Role: Hadoop Consultant / Admin
Client: - APPLE Inc.

APR '12– Dec '14

Description: Apple otherwise doesn't need introduction but let me say that is one of the biggest product company that designs, develops, and sells consumer electronics, computer software, and online services. This Project helps to cater the requirements of all the application teams to store the data and perform the analytics to contribute into growth of the company. Basically, it's a single project where all the apple applications are deployed and persist their data.

Responsibilities:

- Extensive Experience is handling more **20 Cluster**, cluster size between **600-800 Nodes**.
- Implementing **Hortonwork distribution** on all the clusters.
- Installed and configured **Hadoop MapReduce, HDFS**, Developed multiple **MapReduce** jobs in **java** for data cleaning and preprocessing.
- Importing and exporting data into **HDFS** and **Hive** using **Sqoop**.
- Extracted files from **CouchDB** through **Sqoop** and placed in **HDFS** and processed.
- Installed **Oozie workflow engine** to run multiple **Hive and Pig** jobs which run independently with data availability.
- Involved in loading and transforming large sets of Structured, Semi-Structured and Unstructured data and analyzed them by running **Hive queries and Pig scripts**.
- **Monitoring and Debugging** Hadoop jobs/Applications running in production.
- Written the **Map Reduce programs, Hive UDFs in Java** where the functionality is too complex.
- Extracted and updated the data into **Mango DB** using **Mongo import and export command line** utility interface.
- Experienced in implementing Security like **SSL, Kerberos, ACL's**.

PAYTM.COM, Noida

JAN '10- APR '12

Role: Hadoop Developer/ Admin

Description: It is an Ecommerce company. It provides services allowing people to buy things online, they can do recharges and pay for different sites using its paytm Wallet. It conducted business under the name of one97 Communication. Purpose of this project was to implement the data warehouse to cater the business requirements and to help in the growth of the company. Upper management would be relied on to take business decisions.

Responsibilities:

- Worked extensively in creating **MapReduce jobs** to power data for search and aggregation.
- Used **Sqoop** to efficiently transfer data between databases and **HDFS** using **Flume** to stream the data from servers.
- Developed **Map Reduce** programs to cleanse the data in **HDFS** obtained from heterogeneous data sources to make it suitable for ingestion into **Hive schema** for analysis.
- The **Hive tables** created as per requirement were internal or external tables defined with appropriate static and dynamic partitions, intended for efficiency.
- Developed **Map Reduce jobs in java** for data cleansing and preprocessing.
- Develop testing scripts in **Python** and prepare test procedures, analyze test results data and suggest improvements of the system and software.
- Used Sqoop to import the data from **RDBMS to Hadoop Distributed File System (HDFS)** and later analyzed the imported data using **Hadoop Components**.
- Gained very good business knowledge on **claim processing, fraud suspect identification** etc.
- Assisted in **Cluster maintenance, Cluster Monitoring** and Troubleshooting, Manage and review data backups and log files

TECH MAHINDRA LTD, NOIDA

APR '09 – JAN '10

Role: DBA

Client: Bell Canada

Description: Bell Canada is Canada's largest telecommunications company, providing Mobile phone, TV, high speed and wireless Internet, and residential Home phone services. This project was to help to create the customer database and manage the services used by them, basically to cater and manage the whole database of bell customers along with their current services and usages.

Responsibilities:

- Used **IPAIRS** as an integrated framework on top of their databases.
- Used **SOA** as **ORM** tool for data persistence.
- Designed and developed **Automation Scripts, Implements patches, security**.
- Used **BCRIS** as Java IDE tool for creating **Jobs, DAG** and deployed the application.
- Developed server-side common utilities for the application and the front-end dynamic web pages using **JSP, JavaScript and HTML/DHTML**.

Environment: Java, PAIRS, BCIRS, SOA, shell Scripting, Unix based tools, Oracle etc.

Vikas Srivastava