
PROFESSIONAL SUMMARY:

- **11+ years** of experience. **8+ years** in **Big Data, Analytics, Hadoop development.**
- Expertise in Implementing Bigdata on **Cloud and on Premise.**
- **2+ Years** in **Architect, providing solution and designing** of Bigdata 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 the 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** activities such as **realtime and batch** Processing of clusters using **Apache, Cloudera and AWS.**
- Excellent knowledge on Hadoop Architecture and ecosystems such as **HDFS, Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce.**
- Experience in analyzing data using **HiveQL, Pig Latin, and custom MapReduce programs in Java.**
- Hands On experience in **NoSQL database** like **HBASE.** Knowledge of job workflow scheduling and monitoring tools like **Oozie and Zookeeper.**
- Good experience in **Python, UNIX and shell scripts.**
- Experience in writing **UDFS in java for hive and pig.**
- 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) .**
- Good knowledge on integrating the **BI tools** like Talend with the **Hadoop stack** and extracting the required Data.
- Experience in working with the Different file formats like **TEXTFILE, AVRO, PARQUET and JSON.**
- Good knowledge on **Hadoop administration** activities such as installation and configuration of clusters using **Apache and Cloudera.**
- 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	HortonWorks (HDP), Cloudera , Pivotal
Databases	NoSQL, MySQL , MongoDB, Cassandra ,Hbase
Programming Languages	TCL , Core JAVA , Python, Shell scripting
Schema	Avro, Parquet, Orc
Hadoop echo Systems	Cloudera Manager/Apache Hadoop HDFS, NFS, Map reduce, Hbase, Pig, Hive, Sqoop, Flume, Oozie, Zookeeper, Hue, Hadoop Tokenization, Nagios, Ganglia,

	<i>Splunk</i>
<i>Real Times processing Tools</i>	<i>SPARK, STORM , FLUME , KAFKA</i>
<i>Cloud Provider</i>	<i>GCP, Azure, AWS</i>
<i>Security Tools</i>	<i>SSL, Kerberos, RANGER, Knox</i>
<i>Data Pipelineing</i>	<i>Nifi, Streamsets, Falcon</i>
<i>Virtualization</i>	<i>Docker/Mesos/Vmware</i>
<i>Technical Blogs</i>	<i>Bigdataboard.blogspot.in / bigdatum.wordpress.com</i>

EDUCATION:

B-Tech in Information Technology, India.

PROFESSIONAL EXPERIENCE:

Hortonworks Data Platform, Bangalore

Dec'18 - Till Date

Role: Resident Architect

Client: Walmart Labs

Walmart is one of the major retailer company headquartered in US. Involve in creating a bigdata environment where we can take advantages of bigdata performance and implement it components to solve some of the issues like product catalogs /fraud detection etc.

Responsibilities:

- *Building worldclass bigdata lake Using HDP 2.6 and HDP 3.1/Cloudbreak on cloud/on premises.*
- *Requirement gathering and providing End to end solutions.*
- *Designing of platform involve **multi tier architecture** with different technology stacks.*
- *Migration of project from on premise to clouds (**GCP/AZURE**)*
- *Implementation of Applications on Bigdata System (HDP2 and HDP3).*
- *Worked on Installing and configuring **LLAP, Bigdata Stack** and helping developer for tuning **MapReduce jobs** for data cleaning and preprocessing.*
- *Performance tuning of the cluster as well as taking the problem and solve 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. Involve in creating a bigdata environment where we can take advantages of bigdata performance and implement it components to solve some of the issues like name screening/fraud detection etc.

Responsibilities:

- *Involve with Solution Architect to building **EDMP(Enterprise data management platform)** Using HDP 2.4.*
- *Requirement gathering and providing End to end solutions.*

- Designing of platform 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 Bigdata 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.

PROFESSIONAL EXPERIENCE:

HCL,Noida

Dec '14 – Apr'16

Role: Hadoop Consultant

Client: NBCU

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. Here we are designing and building the world class datawarehouse using bigdata 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 of building Best Data ware house.
- 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.
- Imported and exported data into **HDFS and HIVE using Sqoop**
- Implemented the workflows using **Apache Oozie framework** to automate tasks.
- Improved performance using advanced joins in **Apache Pig and Apache Hive**.
- Involved in data ingestion into **HDFS using Sqoop and Flume** from various sources.
- Continuous monitoring and managing the **Hadoop cluster using PIVOTAL HD**.
- 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**.

Environment: Hadoop, Eclipse, Java, JDBC, Mongo db, Cloudera , Oozie, Python, Apache, Linux, UNIX Shell, Pig, HDFS, Hive, Hbase, MapReduce, Sqoop.

TCS, Gurgaon

APR'12– Dec '14

Role: Hadoop Consultant / Admin

Client:- APPLE Inc.

Description: Apple otherwise doesn't need introduction but let me say that is one of the biggest that designs, develops, and sells consumer electronics, computer software, and online services. Its best-known hardware products are the Mac personal computers, the iPod portable media player, the iPhone smartphone, the iPad tablet computer, and the Apple Watch smartwatch. 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 company. Basically it's a single project where all the application was 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.
- Developed Simple to complex Map/reduce streaming jobs using **Python** that are implemented using **Hive and Pig**.
- Written the **Map Reduce programs, Hive UDFs in Java** where the functionality is too complex.
- Extracted and updated the data into **MangoDB** using **Mongo import and export command line** utility interface.
- Experienced in implementing Security like **SSL, Kerberos, ACL's**.
- Got good experience with **NoSQL database**.

Environment: HDP 2.4 , Hadoop, HBase, Python, Sqoop, Pig, Hive, HDFS, mapreduce, Cloudera, Oozie, Flume.

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 site using its paytm Wallet. It conducted business under the name of one97 Communication. Purpose of this project was to implement the dataware house to cater the business requirements and to help in the growth of company. Upper management would be relied on it to take business decision. Performing analytics on top of the cluster see the customer focus and do different kind analysis.

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

Environment: Hadoop, Apache, Pig, Sqoop, Flume, Python, MapReduce, HDFS, Hive , Oozie, Java, Hue.

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 manages 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, IPAIRS, BCIRS, SOA, shell Scripting, Unix based tools, Oracle etc.

Vikas Srivastava