**Jayahasan Bakthavatchalam**

**Professional summary**

* 7+ years of professional IT experience which includes more than 3 years of experience in Big Data ecosystem and related technologies.
* Experience in working with the Hadoop eco systems (**Pig, Hive, Oozie, Hbase, Flume, Zookeeper, Sqoop**)
* Worked in NoSQL databases like **HBase, Cassandra**, and **MongoDB**.
* Developed and implemented **MapReduce** programs for analyzing Big Data with different file formats like structured and unstructured data.
* Experience in importing and exporting data using **Sqoop** from HDFS to Relational Database Systems (RDBMS) and vice versa.
* Knowledge on several tools like **Spark, Kafka, Storm, Hue, Zookeeper, Solr** search engine
* Expertise in cleansing and analyzing data using **HiveQL**, **Pig Latin** and custom MapReduce programs in Java.
* Involved in importing streaming logs and aggregating the data to **HDFS** through **Flume**.
* Experience in writing custom **UDFs** for extending **Hive** functionalities.
* Expertise on basics of **R Programming**
* Scheduled workflows using **Oozie** with help of Oozie coordinator and Zookeeper.
* Ability to develop **Pig UDF'S** to pre-process the data for analysis.
* Knowledge on different file formats like **Avro, Parquet, JSON**
* Worked extensively with Data migration, Data cleansing, and ETL Processes features for data warehouses.
* Good experience working with Cloudera Distribution in standalone, pseudo and distributed mode.
* Used the **Agile**, **Waterfall** and **Scrum** methodology to develop the applications.
* Familiarity with **Amazon Web Services** to integrate with Hadoop and increase its efficiency.
* Having good experience on Core Java in implementing **OOP concepts**, **Multithreading**, **Collections, Exception handling**.
* Worked with **Horton works and Cloudera** distribution
* Designed and developed a web-based client using **Servlets, JSP, Java Script**, and XML using **Spring** MVC Framework.
* Experience in using in the integration tool like **Jenkins** to deploy the developed programs in production.
* Involved in the ETL process using **Ab Initio** tool to setup a data extraction from several databases.
* Experience in developing and deploying applications using Web Sphere Application Server, **Tomcat** and **Web Logic**.
* Expertise in debugging the applications and Unit Testing the application using **JUnit**.
* Implemented web services using **SOAP** and **REST**
* Experienced in analyzing business requirements and translating requirements into functional and technical design specifications using UML.
* Strong and effective problem-solving, analytical and interpersonal skills, besides being a valuable team player.

**Technical Expertise**

|  |  |
| --- | --- |
| Big Data Technologies: | HDFS, Hive, Map Reduce, Pig, Sqoop, Flume, Zookeeper, Oozie, YARN, Spark, Avro, Impala, Parquet, Hbase, Kafka, Tableau, Cascading |
| Programming Languages: | C, C++, Java, R, Scala |
| Scripting Languages | Python and Shell |
| Web Services | Restful, Soap, AWS |
| Application Server: | Websphere, Weblogic Server, Apache Tomcat. |
| Java Frameworks: | Spring, Hibernate, Struts |
| Development Methodologies | Agile, Waterfall and Scrum |
| Databases | Oracle, MySQL, Hbase, Cassandra, MongoDB |

**Capital One Bank, Richmond, Virginia September 2015 – Aug 2016**

**Hadoop Developer**

The EDS commercial project was acquiring enormous amount if data from several vendors regarding the commercial lands and retail prices. The primary object was to setup a flow to acquire data from several vendors to Hadoop and transform the data as required by the data analyst. We used a Cascading tool which is **ETL** Hadoop tool based on **Java** programming with **cascading API**

**Responsibilities:**

* Imported the retail and commercial data from various vendors into **HDFS** using EDE process and **Sqoop.**
* Designed the Cascading flow setup from the Edge node to the HDFS(Data lake)
* Created the **cascading** code to do several type of data transformations as required by the DA
* Used the **Hue** to create external **Hive** tables on the data in the data imported and on transformed data.
* Developed the code for removing or replacing the error fields in the data fields using **cascading**
* Created the **custom functions** for several data type conversions, handling the errors in the data provided by the vendor
* Developed the precise pom.xml importing the **maven** dependencies from the nexus repository to run the cascading and Hadoop jobs.
* Monitored the cascading flow using the **Driven** component to ensure the desired result was obtained
* Optimized a Capital one tool **Docs**, for importing the data and converting the data into parquet file format post validation.
* Involved in testing the tool **Spark** for exporting the data from HDFS to external database in POC
* Developed **Spark** programs to apply the transformations in the data using **Scala**
* Developed the **shell scripts** for automating the cascading jobs for Control Mschedule.
* Involved in testing the **AWS Redshift** to connecting with SQL database for testing and storing data in POC
* Developed **Hive** queries to analyze the data according to the customer rating Id for several projects
* Imported the data using **Sqoop** from the structured databases like MySQL.
* Implemented the **Kafka** integration with Spark to stream the real time data.
* Converted the raw files (CSV, TSV) to different file formats like **Parquet** and **Avro** with data type conversion using cascading
* Deployed the jars using the Integration tool **Jenkins**
* Developing a batch and stream processing using **Spark** with help of **Scala.**
* Implemented the transformations and actions in RDD using **Scala** for modifying the extracted data.
* Involved in writing the test cases for the cascading jobs using **Plunger** framework.
* Setting up the cascading environment and troubleshooting the environmental issues related to cascading.
* Assisted in creating and maintaining Technical documentation to launching HADOOP Clusters and even for executing Hive queries and Pig Scripts

**Environment:** MapReduce, HDFS Sqoop, Cascading, LINUX, Shell, Hadoop, Spark, Hive, AWS RedShift, Hadoop Cluster, Jenkins

**Independent Health Plans, Buffalo, NY September 2014 – September 2015**

**Hadoop Developer**

Independent Health Plans is a not-for-profit health plan that continually aims to provide our community with innovative health-related products and services, which enable affordable access to quality health care.

**Responsibilities:**

* Processed data into **HDFS**, analyzed the claims data of the customer using **MapReduce** produce summary results from Hadoop to downstream systems
* Imported the insurance data of the customer from various systems/sources (like MySQL) into **HDFS** using **Sqoop.**
* Implemented **Hive** queries for Data cleaning and sorting in the HDFS.
* Scheduled an workflow to import the weekly transactions in the revenue department from RDBMS database using **Oozie**
* Applied **Hive** quires to perform data analysis on **HBase** estimating the per annum claims by the customers
* Developed the existing Hive and Pig scripts.
* Developed **MapReduce** pipeline jobs to process the data and create necessary HFiles and loading the HFiles into HBase for faster access without taking performance hit
* Created components like **Hive UDFs** for missing functionality in HIVE for analytics.
* Involved in implementing the **Pig** scripts during the ETL process.
* Imported the log data from different servers into **HDFS** using **Flume** and developed **MapReduce** programs for analyzing the data
* Hands on experience with NoSQL databases like **Cassandra** for POC (proof of concept) in storing data.
* Performed CRUD operations in Cassandra table with the data.
* Used different file formats like **Text files, Sequence Files, Avro**
* Cluster co-ordination services through **Zookeeper**
* Assisted in creating and maintaining Technical documentation to launching HADOOP Clusters and even for executing Hive queries and Pig Scripts

**Environment:** MapReduce, HDFS Sqoop, Flume, LINUX, Oozie, Hadoop, Pig, Hive, Hbase, Cassandra, Hadoop Cluster,

**Corning Inc., NY Oct 2013 - August 2014**

**Hadoop Developer.**

Corning Incorporated is the world leader in specialty glass and ceramic. Due to the increasing number of customers, there raised a need for consolidating the data for better customer management. I have worked in consolidating the data from various databases into Hadoop. The data cleaning is performed through MapReduce and processed using Hive to aggregate the data sources. Then Pig Latin was used to categorize the customers using Pig Latin. Analysts used Hive for real time accessing of the processed data.

**Responsibilities:**

* Used **Sqoop** to extract data from Oracle SQL server and MySQL databases to **HDFS**
* Developed workflows in **Oozie** for business requirements to extract the data using **Sqoop**
* Used **Hive** and **Impala** to query the data in **HBase**
* Written multiple MapReduce programs in java for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV and other compressed file formats
* **Hive** scripts were written in **Hive QL** to de-normalize and aggregate the data
* Used **Solr** for querying and searching the **Hbase DB**
* Optimized the existing Hive and Pig Scripts
* Created **external table** using **Hive** to perform analysis in **HDFS**
* Involved in loading data from **UNIX file system** to HDFS
* Designed workflows by scheduling Hive processes for Log file data, which is streamed into HDFS using **Flume**
* Involved in developing the **Python** scripts for the jobs scheduled.
* Involved in implementing a query to search the clients using their respective fields using **Neo4j**
* Developed schemas to handle reporting requirements using **Tableau**
* Involved with a team who worked on NoSQL databases like **MongoDB** for POC (proof of concept) in storing documents using **GridFs**.
* Have deep and thorough understanding of ETL tools and how they can be applied in a Big Data environment
* Worked with application teams to install operating systems, Hadoop updates, patches, version upgrades as required.

**Environment:** Hadoop, HortonWorks, Hive QL, Hive, HBase, Sqoop, Solr, Flume, Tableau, Impala, Oozie, MYSQL, Oracle SQL, Java, Unix Shell, Pig Latin.

**Grand Bank of Texas,Dallas, TX February 2012 - Aug 2013**

**Java/Hadoop Developer**

CAA is a finance-based application, used by OAM (Online Account Management) to submit the credit card applications. CAA is Server side component, which is a service process for OAM. OAM and CAA were developed using J2EE architecture.

**Responsibilities:**

* Involved in review of functional and non-functional requirements.
* Installed and configured Hadoop **MapReduce** and **HDFS**.
* Acquired good understanding and experience of NoSQL databases such as **HBase** and **Cassandra**.
* Installed and configured **Hive** and also implemented various business requirements by writing **Hive UDFs**.
* Loaded home mortgage data from the existing DWH tables (SQL Server) to HDFS using **Scoop**.
* Wrote **Hive** Queries to have a consolidated view of the mortgage and retail data.
* Extensively worked on user interface for few modules using **HTML, JSP's, JavaScript**.
* Generated Business Logic using **servlets**, Session beans and deployed them on Web logic server.
* Created complex **SQL queries** and stored procedures.
* Developed the XML schema and Web services for the data support and structures.
* Implemented the Web service client for login verification, credit reports and applicant information using web service.
* Responsible for managing data coming from different sources.
* Used **Hibernate ORM** framework with spring framework for data persistence and transaction management.
* Wrote test cases in **JUnit** for unit testing of classes.
* Deployed the application in production with **Shell scripts**
* Extensively worked with **JDBC** programs using **Oracle** and **MySQL** databases and developed SQL and PL/SQL for Oracle to process the data.
* Provided technical support for production environments resolving the issues, analyzing the defects, providing and implementing the solution defects.
* Built and deployed Java application into multiple Unix based environments and produced both unit and functional test results along with release notes.
* Analyzed the banking and existing system requirements and validated them to suit J2EE architecture.
* Designed the process flow between front-end and server side components

**Environment:** JDK, J2EE, JMS, XML, Servlets, JSP, CVS, HTML, Oracle 8i, UML, and Tomcat.

### BB&T Corporation- Salem, NC July 2011 – January 2012

**IT Programmer**

**Mortgage Tracker**

BB&T mobilized the mortgage tracker program to keep track of the existence of the loan and the encumbrance through the granting of a mortgage which secures the loan. It provided crucial information about the size of the loan, maturity of the loan, interest rate and method of paying off the loan. It provided condition on the foreclosure for any non-paid mortgaged property. It calculated the periodic payments on annuity basis. It also provided the refinancing option for the qualified customer based on their payment history with this financial institution. My responsibility was to develop a relational database system and to check the full functionality at runtime environment.

**Responsibilities**

* Implemented OO in analysis, design, and development and application deployment.
* Developed using J2EE and relational database like Oracle
* Involved in the detailed design of user interface and business layers using JSP, Custom tags and JavaScript
* Deploy server components on Tomcat Web Server
* Tested the application functionality with JUNIT
* Log4j was used for logging messages of the application.
* Responsible for software design, coding, testing and documentation
* Defect tracking and fast fixing of bugs, which arise in test & production instances.

**Environment**: JSP, JavaScript, Oracle, Log4J, Eclipse, Tomcat5, XML, Junit, SVN, ANT

**Lexmark Inc., India May 2010 - April 2011**

**Java Developer**

The aim of the project is to develop Resource Action Management tool which is available to assist with the implementation of Resource Actions in the U.S. This is a Tool which will be basically used by the Workforce Management Personnel. Tool functionality varies based on a user's role. Some users will have view access only. Other users will have both view access and update capability. The tool is an intranet Web Application and the site help provides instructions for using the tool, resource action process steps, and user roles.

**Responsibilities:**

* Access to this site is provided for authorized users.
* Coding using **Java, JSP, and HTML**.
* Developed front end validations using **JavaScript** and developed design and layouts of **JSPs** and custom taglibs for all JSPs.
* Participated in planning and development of UML diagrams like Use Case Diagrams, Object Diagrams, Class Diagrams and Sequence Diagrams to represent the detail design phase.
* Implemented several Test cases using **Junit.**
* Implemented the **Log4J** logging component from Apache into the Application.
* Made Builds and deployed the same onto Common development test Environment, which is a Web sphere Application server Environment to verify its functional requirements.

**Environment:** Java, J2EE, Tomcat, JSP and Struts Framework, Eclipse, SQL and Oracle.

**Blue Dart Couriers, India January2009 -March 2010**

**Jr. Java Developer**

Blue Dart is South Asia's leading integrated air express carrier and premium logistics-services provider. This Corporation is a market leader in transportation, information, and logistics solutions. Blue Dart ‘Smart’ tool was developed to help the customers of Blue Dart to plan their shipment and delivery date. Customers can access this tool from Web as well as using the B2B interface, which is an XML interface.

**Responsibilities:**

* Requirements gathering, Conceptual Design, Analysis, Detail Design, Development, System Testing and User Acceptance Testing.
* Used **JDBC** for database connectivity.
* Front End-user Interface Design, Prototype was developed using **HTML, CSS**.
* Responsible for managing the several **switches** in the infrastructure
* Involved in creating
* Developed complex **SQL** queries to insert, update MySQL database. Used JDBC to invoke Stored Procedures.
* **MySQL** as the backend database.

**Environment:** JSP, Servlets, Switches (networking), HTML, JavaScript, MySQL

**Education:**

Bachelors of Engineering, India.