**Asmitha Kasu**

asmitha.kasu@hotmail.com

908-731-0167



**PROFESSIONAL SUMMARY:**

* Over 6+ Years of IT experience, which includes around 3 years of work experience in BigData related technologies in domains like Retail, Insurance and Finance.
* Experience in Hadoop ecosystem and its components like **MapReduce, HDFS, HBase, Oozie, Hive, Sqoop, Pig, Impala, Flume, Spark.**
* Experience in working with Hadoop distributions like **Cloudera (CDH 5.x)** and **Hortonworks**.
* Good experience in optimizing **MapReduce** algorithms using **Mappers, Reducers, combiners** and **partitioners** to deliver best results for large datasets.
* Extensively used **Apache Sqoop** for efficiently importing bulk data from Relational databases (SQL,Oracle) to HDFS and vice-versa.
* Strong experience in collecting and storing large amounts of log data into HDFS using **Apache Flume**
* Experience in writing**PIG/Hive** scriptsand creating UDF’s for processing and analyzing large volume of data efficiently.
* Proficiency in writing **Pig** Latin scripts and working with **grunt** shells.
* Worked with multiple file Input Formats such as **TextFile, Parquet, ORC**.
* Used compression techniques with file formats like **gzip, LZO, Snappy** to leverage the storage in HDFS.
* Experience in scheduling Hadoop jobs using **Apache Oozie** workflows
* Good knowledge on **Impala** and **Apache Spark** for real time data analysis.
* Knowledge on NoSQL columnar databases like **HBase, MongoDB** and**Cassandra.**
* Good experience in integrating **Kafka-Spark streaming** for high efficiency throughput and reliability.
* Integrated**Hive** and **Spark SQL**for high performance.
* Strong experience in working with real time streaming data using tools like **Spark Streaming, Spark SQL**.
* Competence in using Spark **RDD**s, **Scala**for converting **Hive** and **SQL** queries to Spark transformations.
* Experience in designing and developing Enterprise Applications for J2EE platform using **Servlets, JSP, Struts, Spring, Hibernate and Web services.**
* Proficiency in programming with different Java IDE’s like **Eclipse**, **IntelliJ** and **NetBeans**.
* Experience in developing and designing **Web Services (SOAP and Restful Web services)**.
* Expertise in client-side design and validations using **HTML**, **AJAX, CSS, Java Script, JSP, Struts Tag Library.**
* Good understanding of RDBMS- Relational Database Management Systems and worked on different databases like **Oracle** and **SQL Server.**
* Skilled in using Amazon Web Services such as **S3, EMR and EC2** for storing, transferring, processing and setting up multimode cluster.
* Worked with relative ease on different working strategies like Agile, Waterfall and Scrum and methodologies.

**TECHINAL SKILLS:**

|  |  |
| --- | --- |
| **Big Data Ecosystem** | Mapreduce, HDFS, Pig, Hive, Hbase, Sqoop, Flume, Oozie, Spark, Impala, Spark, Scala, Cassandra |
| **Programming Languages** | C, C++, Java, Scala |
| **Methodologies** | Agile, Scrum |
| **Databases** | SQL server, Oracle |
| **Web Technologies** | HTML, XML, JavaScript |
| **Operating Systems** | Windows, Linux (Ubuntu),Mac OS |
| **Web Servers** | Apache Tomcat, Web Logic |
| **Other tools** | Eclipse, IntelliJ, Maven, SBT, GIT, SVN, JIRA, JUnit, Hue |

**EDUCATION:**

* Bachelors in Computer Science and Engineering

**PROFESSIONAL EXPERIENCE:**

**State Farm, Dallas, TX May 2015–Till present**

Hadoop/Big Data Developer

State Farm provides a wide range of Insurance Policies. Retail Insurance risk analysis team deals with risks faced by the Retail businesses like Economic Fluctuations, Federal Regulatory Changes, Fuel prices, Interest Rates, Unemployment rate etc. For this project, we pull the data from our websites, Call records, Social sites, Surveys etc. In this project, I worked with data which is retrieved from Surveys conducted by the company.

Responsibilities:

* Worked on importing and exporting data from **Oracle** into HDFS using **Sqoop** for analysis, visualization and to generate reports.
* Performing full and incremental imports and created **Sqoop jobs.**
* Implemented multiple **MapReduce** jobs for data processing.
* Wrote **Hive UDF** to parse the staged raw data to get the Hit Times of the claims from a specific branch for a insurance type code.
* Scheduled the jobs with **Oozie** workflow engine.
* Worked in collecting and aggregating large amounts of log data using **Apache Flume** and staging data in HDFS for further analysis.
* Created internal table, Externals tables in **Hive**, and merged the data sets using Hive joins. Involved in integration of **Hive** and **Hbase.**
* Worked with different Hadoop file formats like **ORC** and report via **Talent** approachand compression techniques like **Lzo**.
* Implemented **Spark** using Scala and **Spark SQL** for faster testing and processing of data.
* Used **Pig** as ETL tool to do transformations, event joins and some pre-aggregations before storing the data to HDFS.
* Wrote Java API for **Hbase** transactions on **Hbase** tables.
* Involved in **HDFS** maintenance and loading of structured and unstructured data.

Environment: MapReduce, Hive, HDFS, PIG, Sqoop, Flume, Oozie, Cassandra, Horton works distribution, Flume, Oracle, Linux, Python.

**L’Oreal, Berkely Heights, NJ Jan 2014 – April 2015**

Hadoop Consultant

L’Oreal is a French cosmetic company. It is one of the leading cosmetic companies in beauty industry. It extends its services across 130 countries. In L’Oreal all retail data sent by different sources is sourced into HDFS. In this project we analyze the sales data for enhancement of business growth.

Responsibilities:

* Implemented multiple **Map Reduce** Jobs in java for pre-processing.
* Developed efficient **Map Reduce** programs for filtering out the unstructured data.
* Developed **Hive** queries for data sampling and analysis to the analysts.
* Worked on creating **Hive** tables, loading with data and writing **Hive** queries that will run internally like MapR programs.
* Worked on developing **Pig** scripts.
* Used **Pig** as ETL tool to do transformations, even joins and some pre-aggregations before storing the data on to HDFS.
* Exported the analyzed data to the relational databases using **Sqoop** for visualization and to generate reports for the BI team.
* Developed custom aggregate functions in **Spark** and performed interactive querying using **Impala**.
* Worked on moving all log files generated from various sources into Hadoop HDFS using **Flume** for further processing.
* Extensive usage of **Maven** for building jar files of Map Reduce programs and deployed to cluster.
* Experienced in managing and reviewing Hadoop log files.
* For data ingestion **ETL Talend** jobs are maintained and developed efficiently.
* **Talend** Integration suite is used to develop and schedule jobs and also worked on **Talend ETL** tool.
* Used **Tableau** to visualize the top selling products, top areas where the products are being sold etc.

**Environment:** MS SQL Server, Oracle, PIG, Hive, Cloudera CDH5 distribution, HDFS, HBase, Sqoop, Impala, Java, Oozie, Tableau.

**Western Union, San Francisco, CA February 2013 – December 2013**

Hadoop Consultant

Western Union is an American multinational banking and financial services holding company with one of the largest banking division in USA. Project involved analysis of the customer behavior on the websites.

Responsibilities:

* Implemented Hadoop framework to capture user navigation across the application to validate the user interface and provide analytic feedback/result to the UI team
* Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using **Sqoop**.
* Performed analysis on the user navigation data by loading into HDFS and writing **MapReduce** jobs. The analysis provided inputs to front-end developers team.
* Extensively used **Pig** for validations.
* Worked on creating Hive tables and working on them using **HiveQL** and perform data analysis using Hive and Pig
* Worked in **HDFS** maintenance and administering it through Hadoop -Java API.
* Used **Flume** to collect, aggregate and store the web log data onto HDFS.
* Imported data using **Sqoop** to load data from MySQL to HDFS and Hive on regular basis.
* Automated all the jobs, for pulling data from FTP server to load data into **Hive** tables, using **Oozie** workflows.
* Supported **Map Reduce** Programs those are running on the cluster.

Environment: Hadoop, MapReduce, HDFS, Pig, Hive, HBase, Flume, Oozie, Java (jdk1.6), MySQL, Linux

**GGK Technologies, Hyderabad, India June2010-January2013**

Java Developer

Project: Health Care Services Corporation

**Health Care Service Corporation** (HCSC) is the largest customer-owned health insurance company in the United States, and an independent licensee of the Blue Cross and Blue Shield Association. Worked on enrollment system, which is an internal BCBS application. This application is used to process all the insurance applications received, check different conditions and see if the applicant is eligible for insurance or not. If the applicant is eligible the application goes into another system for further processing else the application gets struck and will be rejected.

Responsibilities:

* Worked on the Design, Coding, Unit Testing and Implementation of the web application.
* Involved in the complete **SDLC** software development life cycle of the application from requirement analysis to testing.
* Worked on design and development of presentation layer using **HTML, JSP, JavaScript and CSS**.
* Designed the screens and client/Server side validations for the applications using JavaScript and validation framework.
* Developed **SOAP** based requests for communicating with Web Services.
* Implemented HTTP Modules for different applications in **Struts** Framework that uses **Servlets, JSP, ActionForm, ActionClass and ActionMapping.**
* Developing web applications using **MVC Framework, Spring, Struts, Hibernate.**
* Analyzed and fixed defects in the Login application.
* Worked in creation of tables and indexes and wrote complex **SQL** queries.
* As per the project we also have another set of data in SAN. We used Amazon SDK to access Amazon S2 for file-based artifacts.
* Designed, customized and implemented the search feature for the application using application Id.
* Provided required technical support to the client in UAT phase.
* Involved in dynamic creation of error elements on demand when there is an error.

**Environment:** Java, J2EE, Struts, Servlets, SOAP web services, Spring, Hibernate, JavaScript, jQuery, Oracle, AJAX, JSPs, Servlets, Eclipse.