****

**Kushal Kumar**Sr. Hadoop Developer

**Summary:**

* Over 7 years in IT with 4 years experience in Big Data technologies such as Spark, **Horton works** and **Cloudera Hadoop distributions**
* Experience in analyzing data using **HiveQL, Pig Latin** and custom **MapReduce** programs in Java.
* Experience in importing and exporting data using **Sqoop** from HDFS to Relational Database Systems and vice-versa.
* Extending **Hive** and **Pig core** functionality by writing custom UDFs in **Python**
* Experience working with Elastic Mapreduce , S3.
* Architected, Designed and maintained high performing ELT/ETL Processes.
* Tuning, and Monitoring Hadoop jobs and clusters in a production environment.
* Managed and reviewed Hadoop log file.
* Participated in an Agile SDLC to deliver new cloud platform services and components.
* Developed and Maintained the Web Applications using the Web server Tomcat.
* Exceptional ability to learn new technologies and to deliver outputs in short deadlines.
* Having Experience on UNIX commands and Deployment of Applications in Server.
* Experience writing custom SQL queries and building dashboards in Tableau

**Certifications:**

* CCA Spark and Hadoop Developer CCA-175

**Technical Skills:**

|  |  |
| --- | --- |
| Hadoop | Hadoop 2.2, HDFS, MapReduce, Pig 0.8, Hive0.13, Sqoop 1.4.4, Spark 1.3 Zookeeper 3.4.5, Yarn, ,Scala,Impala,Kafka,Tez,Tableau,NoSql-Hbase, Cassandra. |
| Hadoop management & Security | Hortonworks Ambari, Cloudera Manager |
| Web Technologies | DHTML, HTML, XHTML, XML, XSL (XSLT, XPATH), XSD, CSS, JavaScript |
| Server SideScripting | UNIX Shell Scripting, Python Scripting |
| Database | Oracle 10g, Microsoft SQL Server, MySQL, DB2,Optima,Teradata Sql, ,RDBS. |
|  |  |
| Web Servers | Apache Tomcat 5.x, BEA Weblogic 8.x, IBM Websphere 6.0/5.1.1 |
| IDE | WSAD5.0, IRAD 6.0, Eclipse3.5, Dreamweaver 13.2.1 |
| OS/Platforms | Mac OS X 10.9.5,Windows2008/Vista/2003/XP/2000/NT,Linux(All major distributions), Unix. |
| Methodologies | Agile, UML, Design Patterns, SDLC |
|  |  |

**Education:**

* MS-IST, Wilmington University, Wilmington ,DE
* Bachelor of Engineering in Mechanical Engineering, Manipal University, India

**Professional Experience:**

**Staples Development Lab, Seattle, WA August 2015 – Sep 2016**

**Sr. Hadoop Developer**

**Description:** Staples EDS working with a third-party on a project where primary objective was to create a personalized experience for a user from the point at which customer received the email all the way until they checkout. Different data sources gathered data at varying levels of customer interaction. Collecting and aggregating the data in HDFS and building HIVE tables in order to calculate effectiveness of various segments using KPI’S (ex: performance of email, dollar values of sales for segment etc).

**Responsibilities:**

* Worked on **Hortonworks-HDP2.2** distribution of Hadoop
* Experience working with Teradata Studio,MS SQL, DB2 for identifying required tables and views to export into HDFS.
* Responsible for moving data from Teradata, MS SQL Server, DB2 to HDFS to development cluster for validation and cleansing.
* Responsible for doing cleansing and validations at HDFS, Teradata, and **Hive** table level.
* Experience designing and optimizing ETL workflows into Hadoop using **Informatica**.
* Writing **SQOOP statements** for one-time imports and scripts for incremental import to HDFS from Teradata, SQL SERVER, DB2.
* Cleansing and validating data in HDFS and exporting to Teradata by writing SQOOP export statements.
* Worked extensively with SSH, SFTP to move data into HDFS from third-party server.
* Responsible for moving data from Linux file system into HDFS.
* Worked on monitoring and troubleshooting the **Kafka-Storm-HDFS** data pipeline for real-time data ingestion in Datalake in HDFS.
* Extensive experience working with ETL of large datasets using **Pyspark, Scala** in **Spark** on HDFS
* Experience working with **Spark SQL** and creating **RDD**’s using **pyspark**.
* Working knowledge of **Dataframes** API on Spark.
* Developed HIVE tables on data using different SERDE’s ,storage formats and compression techniques.
* Writing HIVEQL queries for integrating different tables to create views to produce result set.
* Extensive experience tuning Hive queries using memory joins for faster execution and appropriating resources
* Worked on right join logic recursively to generate a high-level overview of tables for **Tableau dashboards.**
* Worked extensively with **Tableau** to produce dashboards.

**Environment:** Hadoop, MapReduce,Spark, HDFS, Hive, Oozie, Java (jdk1.6),eclipse, Kafka, HBase (NoSQL),Informatica, Sqoop, Pig.

**GE Capital, Stamford, CT Oct 2014 – Jul 2015**

**Sr. Hadoop Developer/Java**

**Description:** GE Capital’s internet-based Consumer credit application (ETail) processing and Client onboarding IT processes, Consumer-facing improvements include a streamlined application form that will have application pre-fill and back-fill capabilities and a new mobile enabled process for Consumers. GE will rely on data provided by Clients and 3rd Parties to optimize the application processing for Consumers and Clients. A new Consumer e-Quickscreen function will enable GE to interface directly with Consumers to offer pre-approved Credit. This application will internally deal with different source systems (FDR, Surveyor) to process the consumer information and will take a decision whether to Approve/Reject/Pending statues and will arrange the loan amount as per credit rate history for an Approved loans sites.

**Responsibilities:**

* Responsible for data gathering from multiple sources like Teradata, Oracle, SQL server etc.
* Responsible for doing validations and cleansing the data as well worked with Pentaho Data Integration (PDI) for data injection with flat files and JSON.
* Finding the right joins logics and create valuable data sets for further data analysis. Architecture design and develop the whole application to ingest and process high volume mainframe data into Hadoop infrastructure using Hadoop map-reduce.
* Design and develop customized business rule framework to implement business logic using hive, **pig UDF**
* Functions in **Python**
* Experienced in working with various kinds of data sources such as Teradata and Oracle. Successfully loaded files to HDFS from Teradata, and loaded from HDFS to **HIVE**.
* Experienced in using Zookeeper and **OOZIE** Operational Services for coordinating the cluster and scheduling workflows.
* Experienced in working with **Elastic MapReduce** (EMR)
* Analysis of **XML** and log files.
* Supported **Map Reduce** Programs which are running on the cluster. Involved in loading data from UNIX file system to HDFS.
* Exported the analyzed data to the relational databases using **SQOOP** for visualization and to generate reports for the BI team.
* Evaluated suitability of **Hadoop** and its ecosystem to the above project and implementing / validating with various proof of concept (POC) applications to eventually adopt them to benefit from the **Big Data Hadoop** initiative
* Working knowledge and proficiency in AWS EC2, S3 and EMR and testing feasibility of implementing Cloud Services in existing infrastructure
* Worked on transferring Data from Hadoop to S3
* Maintain System integrity of all sub-components related to Hadoop.

**Environment:** Apache Hadoop, Mapreduce, Pig, Sqoop, Hive, Impala, Oozie, Hbase.

**Xerox Healthcare, Brooklyn, NY Oct 2013 – September 2014**

**Hadoop Developer/Java**

**Description:** The assignment comprised of integrating the Health Enterprise portal with Fame Legacy System. Also to implement Single Sign-on (SSO), Security and Member eligibility information.

**Responsibilities:**

* Experience in developing solutions to analyze large data sets efficiently
* Developed **Map Reduce** application to find out the useful metrics from the data. Did a thorough testing in local mode and distributed mode found bugs with the code and ensured 100% issue free delivery to production.
* Expert level understanding of Map Reduce internals, including shuffling and partitioning. The bottlenecks in performance of a map reduce program.
* Created Hive external tables and managed tables, designed data models in hive.
* Implemented business logic using Pig scripts
* Finding the right joins logics and creates valuable data sets for further data analysis.
* Worked extensively on **Pig** and **hive**.
* Responsible to develop custom **udf’s** in pig, hive.
* Developed multiple **MapReduce** jobs in Java for data cleaning and processing.
* Responsible for building scalable distributed data solutions using **Hadoop.**
* Worked hands on with **ETL** process.
* Handled importing of data from various data sources, performed transformations using Hive, **MapReduce,** and loaded data into HDFS.
* Extracted the data from Teradata into **HDFS using Sqoop.**
* Exported the patterns analyzed back into Teradata using **Sqoop.**
* Developed Hive queries to process the data and generate the data cubes for visualizing.
* Implemented oracle as database to store the data and gained exposure to various database objects like tables, stored procedures, functions, and triggers using SQL, PL/SQL.

**Environment:** Java, J2EE, JavaScript, Struts, Spring, Maven, GIT Hibernate, SQL/PLSQL, Web Services,Unix,Linux, Hadoop, MapReduce, HDFS, Hive, Oozie, Java (jdk1.6),eclipse, Cloudera, HBase (NoSQL), Sqoop, Pig.

**IBS InfoWeb Private Limited, India Nov 2009– Dec 2012**

**Application Developer**

**Description:** Worked on a comprehensive reverse bidding pharmaceuticals portal, where the consumer is allowed to make the final decision on the price at which he intends to buy the medicine. This portal provides a common place for consumers, prescribers, providers and pharmacists enabling easy commerce.

**Responsibilities:**

* Developed the user interface with HTML, JavaScript, JSP and Tag Libraries using Struts
* Designed and developed the application using various design patterns, such as session façade, business delegate and service locator
* Developed authentication and authorization prototype using Axis-wsse (used as SOAP/WSS4J)
* Developed custom logging that logs application specific details about ERAGUI
* Configured Internationalization using resource bundles on JSP pages
* Developed Stateless Session beans provide a client's view of the application's business logic
* Developed functional and unit testing framework like Test Driven Development in different modules using JUNIT, Solved several key issues by improving code as well as business processes and integrated with ANT build Tool
* Developed Middleware Support for Data-flow Distributionin Web Services Composition
* Implemented Java Collection framework and Exception handling framework in middle Tier modules
* Configured open source tools like Log4J, commons BeanUtils, commons Digester in the application
* Implemented oracle as database to store the data and gained exposure to various database objects like tables, stored procedures, functions, and triggers using SQL, PL/SQL.

**Environment:** Java, J2EE, JavaScript, Struts, Maven., GIT, Spring, Hibernate, SQL/PLSQL, Web Services,Unix.