|  |  |
| --- | --- |
| **Contact Number** | **210 388 9396** |
| **Alternate Contact Number** | **210 388 9396** |
| **Current location of the candidate (City/State)** | **Austin, TX** |
| **Email Address** | [**kartheeks.hadoop@gmail.com**](mailto:kartheeks.hadoop@gmail.com) |
| **Skype ID** | **Kartheek Subramanyam** |

**Kartheek Amaravati Subramanyam**

Sr. Bigdata/ Python

**PROFESSIONAL SUMMARY:**

* Overall 8+ years of overall experience with strong emphasis on Design, Development, Implementation, Testing and Deployment of Software Applications.
* Over 5+ years of comprehensive IT experience inBigData and BigData Analytics, Hadoop, HDFS, MapReduce,YARN, Hadoop Ecosystem and Shell Scripting.
* 5+ years of development experience using Java, J2EE, JSP andServlets.
* Highly capable for processing large sets of Structured, Semi-structured and Unstructured datasets and supporting BigData applications.
* Hands on experience with Hadoop Ecosystem components likeMapReduce (Processing), HDFS (Storage), YARN, Sqoop, Pig, Hive, HBase, Oozie, ZooKeeper and Spark for data storage and analysis.
* Expertise intransferring data between a Hadoop ecosystem and structured data storage in a RDBMS such as MY SQL, Oracle, Teradata and DB2 using Sqoop.
* Experience in NoSQL databases likeMongo DB, HBase and Cassandra.
* Have excellent knowledge on Python Collections and Multi-Threading.
* Managed continuous maintenance and troubleshooting of Python Django projects
* Skilled experience in Python with proven expertise in using new tools and technical developments
* Experience in Apache Spark cluster and streams processing using Spark Streaming
* Worked on several python packages like numpy, scipy, pytables etc.
* Expertise in moving large amounts of log, streaming event data and Transactional data using Flume.
* Experience in developing MapReduce jobs in Java for data cleaning and preprocessing.
* Expertise in writing Pig Latin, HiveScripts and extended their functionality using User Defined Functions (UDF's).
* Expertise in handling structured arrangement of data within certain limits (Data Layout's) using Partitions and Bucketingin Hive.
* Expertise in preparing interactive Data Visualization's using Tableau Softwarefrom different sources.
* Collaboratedwith other developers to handle complicated issues related with deployment of Django based applications.
* Hands on experience in developing workflows that execute MapReduce, Sqoop, Pig, Hive and Shell scripts using Oozie.
* Experience working with Cloudera Hue Interfaceand Impala.
* Hands on experience developing Solr Indexes using MapReduceIndexer Tool.
* DevelopedSPARK CODE using SCALA and Spark-SQL/Streaming for faster testing and processing of data.
* Expertise in Object-oriented analysis and design (OOAD) like UML and use of various design patterns.
* Experience in Java, JSP, Servlets, EJB, Web Logic, Web Sphere, Hibernate, SpringJBoss, JDBC, RMI, Java Script, Ajax, Jquery, XML and HTML.
* Fluent with the core Java concepts like I/O, Multi-threading, Exceptions, RegEx, Data Structures and Serialization.
* Performed unit testing using JunitTesting Framework and Log4J to monitor the error logs.
* Good Knowledge of Python and Python Web Framework Django.
* Experienced with Python frameworks like Webapp2 and, Flask.
* Experience in process improvement, normalization/de-normalization, data extraction, cleansing and manipulation.
* Converting requirement specification, Source system understanding into Conceptual, Logical and physical Data Model, Data flow (DFD).
* Expertise in working with transactional databases like Oracle, SQL server, My SQL, and Db2.
* Expertise in developing SQL queries, Stored Procedures and excellent development experience with Agile Methodology.
* Ability to adapt to evolving technology, strong sense of responsibility and accomplishment.
* Excellent leadership, interpersonal, problem solving and time management skills.
* Excellent communication skills both Written (documentation) and Verbal(presentation).

**EDUCATIONAL QUALIFICATION:**

Bachelors in Electronics & Communication Engineering from Anna University, Chennai, India.

**TECHNICAL SKILLS:**

**Languages**: SQL, C, C++, Java, J2EE, Pig Latin, Hive.

**Big Data Technologies:** Hadoop, HDFS, MapReduce, Hive, Pig, HBase, Spark, Scala, Impala, kafka, Hue, Sqoop, Oozie, Flume, Zookeeper, Cassandra, Cloudera CDH5, Python,PySpark, Solrand Horton works.

**Databases:** Oracle, MySQL, SQL Server, DB2, Mongo DB,Teradata, HBase, Cassandra.

**Scripting and Query Languages:** UNIX Shell scripting, SQL and PL/SQL.

**Web Technologies:** JSP, Servlets, JavaBeans, JDBC, AWT, Swing, JSF, XML, CSS, HTML, XHTML, JavaScript, AJAX.

**Operating Systems:** Windows 7&8, UNIX, Linux, CentOS, Ubuntu.Tools Eclipse, Tableau, squirrel, Talend, Toad, SQL Server Studio, GIT, SVN, Concurrent Versions System (CVS).

**Reporting Tools:** Crystal Reports, SQL Server Reporting Services and Data Reports, Business Intelligence and Reporting Tool (BIRT)

**PROFESSIONAL EXPERIENCE:**

**Company:** Intel, Austin, TX. Nov’16-Present.

**Role:** Sr. Bigdata/ Python

**Responsibilities:**

• Responsible for understanding the requirements and implementing the security using AD Groups for the Dataset.

• Monitored all aspects of applications being developed to ensure they met standards.

• Involved Low level design for MR, Hive, Impala, Shell scripts to process data.

• Worked on ETL scripts to pull the data from DB2/Oracle Data Base into HDFS.

• Experience in utilizing Spark machine learning techniques implemented in Scala.

• Involved in POC development and unit testing using Spark and Scala.

• Created Partitioned Hive tables and worked on them using Hive.

• Installing and configuring Hive, Sqoop, Flume, Oozie on the Hadoop clusters.

• Involved in scheduling Oozie workflow engine to run multiple Hive and Pig jobs.

• Develop and implement Python/Django applications.

• Developed a process for the Batch ingestion of CSV Files, Sqoop from different sources and also generating views on the data source using Shell Scripting and Python.

• Integrated a shell script to create Collections/morphline, SolrIndexes on top of table directories using MapReduceIndexer Tool within Batch Ingestion Framework.

• Implemented partitioning, dynamic partitions and buckets in HIVE.

• Build and Release Management Built application logic using Python 2.7.

• Developed Hive Scripts to create the views and apply transformation logic in the Target Database.

• Developed Python Mapper and Reducer scripts and implemented them using Hadoop streaming

• Implement, installed deploy multi node MapR Hadoop Cluster, developed HA architecture for Hadoop Eco System components; develop puppet recipes for automation of Hadoop Installation and configuration of nodes.

• Involved in the design of Data Mart and Data Lake to provide faster insight into the Data.

• Expertise with MapR M3 Edition, M5 and M7 enterprise editions and using Pig, Hive, HCatalog, Flume, Sqoop, HBase, and MapR Control System; monitored cluster using tools (Ganglia and Nagios)

• Involved in using Stream Sets Data Collector tool and created Data Flows for one of the streaming application.

• Performedlog management using Chukwa and benchmarking and fine tuning cluster for optimum performance.

• Developed and implemented API services using Python in spark.

• Usermanagement with LDAP and setting up Kerberos authentication for cluster and creation of volumes for users and allocating quotas; established NFS gateways on data nodes/NFS mount on edge nodes

• Experienced in using Kafka as a data pipeline between JMS (Producer) and Spark Streaming Application (Consumer).

• Involved in the development of Spark Streaming application for one of the data source using Scala, Spark by applying the transformations.

• Skilled in using collections in Python for manipulating and looping through different user defined objects.

•  Implemented Frameworks using Java and python to automate the ingestion flow.

• Wrote a Python module to connect and view the status of an Apache Cassandra instance.

• Developed a script in Scala to read all the Parquet Tables in a Database and parse them as Json files, another script to parse them as structured tables in Hive.

• Designed and Maintained Oozie workflows to manage the flow of jobs in the cluster.

• Configured Zookeeper for Cluster co-ordination services.

• Generated Python Django forms to record data of online users and used PyTest for writing test cases

• Developed a unit test script to read a Parquet file for testing PySpark on the cluster.

• Involved in exploration of new technologies like AWS, Apache Flink, and Apache NIFIetc which can increase the business value.

• Directed team meetings to check on progress and plan code directions

• Was responsible for quality of applications delivered for end clients.

• Held team meetings and prioritized work tasks.

**Environment:** Hadoop, HDFS, Hive, HBase, Zookeeper, Impala, Cloudera, Oracle, SQL Server, UNIX Shell Scripting, Flume, Scala, Spark, Sqoop, Python, kafka, PySpark.

**Company:** Mercedes-Benz Financial Services, Farmington Hills, MI. Feb'16 – Oct’16.

**Role:** Bigdata Hadoop Developer

**Responsibilities:**

• Responsible for gathering requirements from the business partners.

• Developed a process for Sqoopingdata from multiple sources like SQL Server, Oracle and Teradata.

• Responsible for creation of mapping document from source fields to destination fields mapping.

• Developed a shell script to create staging, landing tables with the same schema like the source and generate the properties which are used by Oozie jobs.

• Developed Oozie workflow’s for executing Sqoop and Hive actions.

• Worked with NoSQL databases like Hbase in creating Hbase tables to load large sets of semi structured data coming from various sources.

• Involved in building database Model, APIs and Views utilizing python, in order to build an interactive web based solution

• Performance optimizations on Spark/Scala. Diagnose and resolve performance issues.

• Responsible for developing Python wrapper scripts which will extract specific date range using Sqoop by passing custom properties required for the workflow.

• Developed scripts to run Oozie workflows, capture the logs of all jobs that run on cluster and create a metadata table which specifies the execution times of each job.

• Developed Hive scripts for performing transformation logic and also loading the data from staging zone to final landing zone.

• Developed monitoring and notification tools using Python.

• Worked on Parquet File format to get a better storage and performance for publish tables.

• Involved in loading transactional data into HDFS using Flume for Fraud Analytics.

• Developed Python utility to validate HDFS tables with source tables.

• Designed and developed UDF’S to extend the functionality in both PIG and HIVE.

• Import and Export of data using Sqoop between MySQL to HDFS on regular basis.

• Managed datasets using Panda data frames and MySQL, queried MYSQL database queries frompython using Python-MySQL connector and MySQL dB package to retrieve information.

• Developed and tested many features for dashboard using Python, Java, Bootstrap, CSS, JavaScript and JQuery.

• Responsible to check-in the developed code into Harvest for release management which is a part of CI/CD.

• Involved in using CA7 tool to setup dependencies at each level (Table Data, File and Time).

• Automated all the jobs for pulling data from FTP server to load data into Hive tables using Oozieworkflows.

• Involved in developing Spark code using Scala and Spark-SQL for faster testing and processing of data and exploring of optimizing it using Spark Context, Spark-SQL, Pair RDD's, Spark YARN.

• Migrating the needed data from Oracle, MySQL in to HDFS using Sqoop and importing various formats of flat files in to HDFS.

**Environment:** Hadoop, HDFS, Hive, HBase, Zookeeper, Oozie, Impala, Java(jdk1.6), Cloudera, Oracle, Teradata SQL Server, UNIX Shell Scripting, Flume, Scala, Spark, Sqoop, Python.

**Company:** Cybage Software Pvt. Ltd, HYD, India. Jan'12 - Dec'15.

**Role:** Hadoop Developer.

**Responsibilities:**

• Responsible for understanding the scope of the project and requirements gathering

• Used MapReduce to Index the large amount of data to easily access specific records.

• Utilized Apache Hadoop ecosystem tools like HDFS, Hive and Pig for large datasets analysis.

• Worked with administrator to set up and monitor the Hadoop cluster.

• DevelopedMapReduce ETL in Java/Pig anddata validation using HIVE.

• Worked on Hive by creating external and internal tables, loading it with data and writing Hive queries.

• Supported MapReduce Programs which are running on the cluster.

• Involved in HDFS maintenance and loading of structured and unstructured data.

• Developed MapReduce programs to perform data filtering for unstructured data.

• Designed the application by implementing Struts Framework based on MVCArchitecture.

• Designed and developed the front end using JSP, HTML and JavaScript and JQuery.

• Developed framework for data processing using Design patterns, Java, XML.

• Implemented J2EE standards, MVC2 architecture using Struts Framework.

• Implementing Servlets, JSP and Ajax to design the user interface.

• Used JSP, Java Script, HTML5, and CSS for manipulating, validating, customizing, error messages to the User Interface.

• Used the light weight container of the Spring Framework to provide architectural flexibility for Inversion of Controller (IOC).

• Used SpringIOC for dependency injection to Hibernate and Spring Frameworks.

• Designed and developed Session beans to implement the Business logic.

• Developed EJB components that are deployed on Web logic Application Server.

• Written unit tests using Junit Framework and Logging is done using Log4J Framework.

• Used Html, CSS, JavaScript and JQuery to develop front end pages.

• Designed and developed various configuration files for Hibernate mappings.

• Designed and Developed SQL queries and Stored Procedures.

• Used XML, XSLT, XPATH to extract data from Web Services output XML

• Extensively used JavaScript, JQuery and AJAX for client-side validation.

• Used ANT scripts to fetch, build, and deploy application to development environment.

• Developed Web Services for sending and getting data from different applications using SOAP messages.

• Actively involved in code reviews and bug fixing.

• Applied CSS (Cascading style Sheets) for entire site for standardization of the site.

• Offshore co-ordination and User acceptance testing support.

**Environment:**Java 5.0, Struts, Spring 2.0, Hibernate 3.2, WebLogic 7.0, Eclipse 3.3, Oracle 10g, Junit 4.2,Maven, Windows XP,J2EE, JSP, JDBC, Hibernate, spring, HTML, XMLCSS, JavaScript andJQuery.

**Company:** MM Imagine Technologies, Chennai, India.June’09 – Dec’11.

**Role:** Software Programmer.

**Responsibilities**:

• Involved in the analysis & design of the application using Rational Rose.

• Developed the various action classes to handle the requests and responses.

• Designed and created Java Objects, JSP pages, JSF, JavaBeans and Servlets to achieve various business functionalities. Created validation methods using JavaScript and Backing Beans.

• Involved in writing client side validations using JavaScript, CSS.

• Involved in the design of the Referential Data Service module to interface with various databases using JDBC.

• Used Hibernate framework to persist the employee work hours to the database.

• Developed classes and interface with underlying web services layer.

• Prepared documentation and participated in preparing user's manual for the application.

• Prepared Use Cases, Business Process Models and Data flow diagrams, User Interface models.

• Gathered & analyzed requirements for EAuto, designed process flow diagrams.

• Defined business processes related to the project and provided technical direction to development workgroup.

• Analyzed the legacy and the Financial Data Warehouse.

• Participated in Data base design sessions, Database normalization meetings.

• Managed Change Request Management and Defect Management.

• Managed UAT testing and developed test strategies, test plans, reviewed QA test plans for appropriate test coverage.

• Involved in Developing JSP's, action classes, form beans, response beans, EJB's.

• Extensively used XML to code configuration files.

• Developed PL/SQL stored procedures, triggers.

• Performed functional, integration, system and validation testing.

**Environment:**Java, J2EE,JSP, JCL,DB2, Struts, SQL, PL/DSQL, Eclipse, Oracle, Windows XP, HTML, CSS, JavaScript, and XML.