**SASHI KIRAN KAATA**

|  |  |
| --- | --- |
| Contact Number | 484-278-1218 |
| Alternate Contact Number | 484-278-1218 |
| Current location | Edison, NJ |
| Email Address | Kiranhadoop9@gmail.com |
| Skype ID | Sashikiran90 |

**Professional Summary:**

Seven years of professional experience in IT and Enterprise Application Development in multiple industries,which includes hands on experience in Big Data/Hadoop ecosystem and related technologies.

3 years of comprehensive experience on Big Data Technologies.

* Excellent understanding / knowledge of Hadoop Architecture, its components Map Reduce & HDFS.Good understanding on various daemonprocesses like Job Tracker, Task Tracker, Name Node, Data Node
* Experience in building pipelines using different services in AWS for various batch & real-time data feeds
* Experience in writing streaming and batch applications using spark, Kafka and Kinesis
* Architected & implemented logging framework – ELK stack (Elastic Search, LogStash & Kibana) on AWS
* Experience in working with Hadoop/BigData storage and analytical frameworks over Amazon AWS Cloudusing tools like SSH, Putty.
* Experience in installing, configuring, and administrating Hadoop cluster for distributions like Cloudera and Horton works.
* Very good understanding of NOSQL databases like HBase, Mongo DB, Cassandra & DynamoDB
* Designed and implemented Cassandra, a NoSQL based database and associated RESTful web service that persists high-volume user profile data for vertical teams
* Experience in building large scale highly available Web Applications. Working knowledge of web services and other integration patterns
* Data modelling on AWS RDS Mysql database for the Hive Metastore and well versed in datanormalization techniques
* Setting up & managing EMR cluster on AWS along with monitoring tools like Ganglia, Cloudwatch metrics
* Extensive expeince in configuring& using YARN,Pig, Hive, Sqoop& Flume
* Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems and vice-versa
* Hands-on experience in Python scripting, Jenkin deployments and Linux shell scripting
* Experience in extending Hive and Pig core functionality by writing customUDFs
* Experience in analyzing data using HiveQL, Pig Latin and Map Reduce
* Knowledge in job work-flow scheduling and monitoring tools like oozie and Zookeeper
* Experience using data warehousing and ETL tools like Informatica and Pentaho
* Extensively used Python Pandas library for data analysis & pruning of datasets
* Excellent interpersonal and communication skills, creative, research-minded, technically competent and result-oriented with problem solving and leadership skills

**Certifications:**

Certified AWS Developer – Associate

Validity: July 2017 – July 2019

License # ELPNFXBKCJ111VG8

|  |
| --- |
|  |

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Bigdata Ecosystem | MapReduce, HDFS, Hive, Pig, HBase, Mongodb, Zookeeper, Sqoop, Oozie, Flume, Spark, Kafka, Storm |
| Java/J2EETechnologies | Java, J2EE, Servlets, JSP, XML, AJAX, SOAP, WSDL |
| SDLC Methodologies | Agile |
| Enterprise Frameworks | MVC, Struts 2/1, Hibernate 3, Spring 3/2.5/2 |
| Version and Source Control | SVN, Git |
| Programming Languages | C,Java, XML, Unix Shell scripting, SQL and PlSql, Scala, Python |
| Web Technologies | HTML, DHTML, XML, XSLT, JavaScript, CSS |
| Modeling Tools | UML on Rational Rose 4.0 |
| IDE Tools | Intellij, Eclipse, Netbeans |
| Webservices | WebLogic, WebSphere, JBoss |
| Databases | Oracle DB2, MS - SQL Server, MySQL, MS – Access, Apache Cassandra, DynamoDB |
| Frameworks | MVC, Struts, Log4J, Junit, Maven, WebServices. |
| Operating Systems | Windows, UNIX, Linux,Centos |

**Professional Work Experience:**

**Client: DTCC, NewYork March 2017 – Present**

**Implementation: Perficient Inc**

DTCC has got on-premise stock trade data that has to be migrated to AWS environment. Need of the hour is to build a self sustainable architecture on AWS to handle the data ingestion into Data Lake & Data Marts built on top of S3. API’s are built inorder to manage and access the operational metadata using AWS API Gateway. Environment: Spark on EMR, ELK Stack, AWS API Gateway, AWS Lambda, AWS S3, RDS, Jenkins & Python

**Responsibilities:**

* Built API’s for AWS S3, DynamoDB using AWS Gateway API service to access the metadata and data in general stored in S3 & DynamoDB
* Developed lambda function to be triggered for an S3 PUT action for the data ingestion workflow to start processing using Spark EMR
* Built Data Lake & Data Marts in S3 to cater the different needs of analytics in addition to storing the metadata
* Developed Spark application functionality to convert the file from Fixed length flat files to CSV and Parquet and push into Data Lake & Data Mart respectively
* Developed Spark applications using RDDs, DatFrame and Dataset APIs
* Built on-premise data pipelines using Kafka and Spark streaming using the feed from API streaming Gateway REST service
* Data modelling on AWS RDS Mysql database for the Hive Metastore and performed datanormalization over those tables
* Extensively used Python Pandas module for Data analysis, Data scarping and parsing
* Deployed applications using Jenkins framework integrating Git- version control with it
* Participated in production support on a regular basis to support the Analytics platform

**Client: Comcast, Philadelphia, PA Sep 2015 – Dec 2016**

**Role: Developer/Analyst (3 Projects)**

**Fairmount(Project 1): June 2016 – Dec 2016**

This is a data syndication Platform between Comcast & its Partners Cox & Shaw. The idea here is to build a sustainable & failproof cloud solution for the data transfer and development by leveraging the AWS cloud services. Environment: AWS-S3, Ec2, Firehose, Kinesis, Kafka, Mesos, Cassandra, Mesos

**Responsibilities:**

* Developed pipelines using Pentaho to migrate the data from Oracle to DynamoDB
* Designed solutions which integrate AWS cloud, EC2, S3 storage, Kinesis, networking and security
* Supported an application called DropWizard to securely transfer the data to the partners via Kafka
* Worked on Amazon Lambda, Firehose & kinesis for data transfer between source side to the destination
* Have set up Splunk forwarders for a critical application to constantly check the health of the application and create alerts based off that info
* Tested out Amazon Kinesis load intake capabilities by coding a Synthetic data generator to generate the data as prescribed in Size and # of records per second
* Experience with working on Cassandra database for an application called Casper (Cassandra Scrapper) deployed on Mesos

**Watermark Billing (Project 2): Jan 2016 – May 2016**

This is a data syndication Project between Comcast & its clients Cox & Shaw. Data is delivered to the clients for billing and reconciliation purposes. Environment: AWS-S3, Ec2, Firehose, Kinesis, Kafka, Spark Sql & OracleDB

**Responsibilities:**

* Delivery of user data to the Finance team – Watermark, meeting the quality compliance, data enrichments and data governance
* Used spark SQL to grab the data from Oracle and publish that to Finance SFTP servers
* Assistance in data reconciliation between Comcast and its clients
* Handled Cross-Platform communication among different teams of Comcast for any data related disputes
* Developed data pipelines to transfer the data to the partners Cox & Shaw

**Cloud Bridge(Project 3): Sep 2015 – Dec 2015**

A cloud platform to bridge the gap between different cloud infrastructures using a common API. Focus is to move the data between different clouds, ensuring the data quality and completeness with implementation of proper logging infrastructure. Environment: AWS – EC2, VPC, LoadBalancer, CloudWatch

**Responsibilities:**

* Architected and implemented the logging infrastructure – ELK stack (Elastic Search, LogStash & Kibana) to consume, parse & visualize the logs on AWS
* Good experience in writing filters for LogStash
* Experience in visualizing the real-time data for a Spark Streaming application using Kibana
* Experience with source code control (svn, git), build automation (Jenkins) tools
* Experience in automating the jobs using StoneBranch & was on-call & production support

**Client: Charter Communication, St louis, MO Dec 2014 to Aug 2015**

**Role: Hadoop Developer**

**Responsibilities:**

* Co-ordinated with business customers to gather business requirements
* Extensively involved in Design phase and delivered Design documents
* Good working knowledge of YARN Map Reduce V2 and its daemons Resource Manager and Application Master.
* Experience in loading data from Teradata database into HDFS using SQOOP
* Experience in moving large amount of log data from different servers
* Experience in writing transformer/mapping Map-Reduce pipelines
* Implemented test scripts to support test driven development and continuous integration.
* Developed oozie workflow for scheduling & orchestrating the ETL process
* Involved in Unit testing and delivered Unit test plans and results documents using JUnit and MRUnit
* Performed data analysis with HBase using Hive External tables
* Exported the analyzed data to HBase using Sqoop and to generate reports for the BI team.
* As a POC we have installed Cassandra on the Amazon EC2
* Created tables, inserted data and executed various Cassandra Query language (CQL 3) commands on tables using cqlsh
* Used DataStax OpsCentre, a web based tool to monitor and to simplify administration tasks.

**Environment:**Cloudera,Python, MapReduce, Cassandra, AWS, HDFS, Hive, Pig, Hue, Spark, Oozie, Core Java, UNIX scripting, Eclipse, Hbase, Flume, Informatica, SQL\*PLUS, Toad, Putty, UNIX Shell Scripting, Linux 5, YARN, Eclipse, Java SDK 1.6

**Client: Amplify,Austin, TX Oct 2013 to Nov 2014**

**Role: Hadoop Developer**

**Responsibilities:**

* Installed & maintained cloudera Hadoop distribution
* Worked on analyzing Hadoop cluster and different big data analytic tools including Pig, Hive, HBase & Sqoop
* Involved in loading the data from Linux file system to HDFS
* Exported the analyzed data to the relational databases using sqoop for virtualization and to generate reports for the BI team
* Involved in running Hadoop jobs for processing millions of records of text data.
* Supported in setting up QA environment and updating configurations for implementing scripts with pigand sqoop
* Involved in defining job flows, managing and reviewing log files.
* Monitored workload, job performance and capacity planning using Cloud era Manager.
* Installed Oozie workflow engine to run multiple Map Reduce, Hive and Pig jobs.
* Implemented Map Reduce programs on log data to transform into structured way to find user information.
* Responsible for loading and transforming large sets of structured, semi structured and unstructured data.
* Collected the log data from web servers and integrated into HDFS using Flume.
* Responsible to manage data coming from different sources.
* Extracted files from HBase and placed into HDFS using Sqoop and pre-process the data for analysis.
* Gained experience with NoSQL database.
* Created and maintained Technical documentation for launching HADOOP Clusters and for executing Hive queries and Pig Scripts.

**Environment:** Cloudera,JDK1.5,J2EE1.4,Struts1.3,JSP, AWS, Spring,Servlets2.5,WebSphere6.1,HTML,XML,JavaScript, Hadoop, HDFS, Pig, Hive, MapReduce, HBase, Sqoop, Python, Oozie, Ganglia,Flume, Toad, Putty, Windows NT, UNIX Shell Scripting,Linux,SQL

**Client: Lease Plan USA - Alpharetta, GA Oct 2012 to Sep 2013**

**Role: Java/Database developer**

**(Development focus of Java/J2EE based applications.)**

**Responsibilities:**

* Designed and developed Struts like MVC 2 Web framework using the front-controller design pattern, which is used successfully in several production systems.
* Spearheaded the "Quick Wins" project by working very closely with the business and end users to improve the current website's ranking from being 23rd to 6th in just 3 months.
* Normalized Oracle database, conforming to design concepts and best practices.
* Resolved product complications at customer sites and funneled the insights to the development and deployment teams to adopt long term product development strategy with minimal roadblocks.
* Convinced business usersand analysts with alternative solutions that are more robust and simpler to implement from technical perspective while satisfying the functional requirements from the business perspective.
* Applied design patterns and OO design concepts to improve the existing Java/JEE based code base.
* Identified and fixed transactional issues due to incorrect exception handling and concurrency issues due to unsynchronized block of code.

**Environment:**Java1.2/1.3,Swing,Applet,Servlet,JSP,customtags,JNDI,JDBC,XML,XSL,DTD,HTML, CSS, Java Script, Oracle, DB2, PL/SQL, Web logic, JUnit, Log4J and CVS

**Client: FedEx Corporation, Memphis, TN Nov 2011 to Sep2012**

**Role: Java/J2EE Developer**

**Responsibilities:**

* Involved in various phases of Software Development Life Cycle (SDLC) as design development and unit testing.
* Developed and deployed UI layer logics of sites using JSP, XML, JavaScript, HTML/DHTML, and Ajax.
* CSS and JavaScript were used to build rich internet pages.
* Agile Scrum Methodology been followed for the development process.
* Designed different design specifications for application development that includes front-end, back-end using design patterns.
* Developed proto-type test screens in HTML and JavaScript.
* Involved in developing JSP for client data presentation and, data validation on the client side with in the forms.
* Developed the application by using the Spring MVC framework
* Spring IOC being used to inject the parameter values for the Dynamic parameters.
* Developed JUnit testing framework for Unit level testing.

**Environment:**SpringMVC,Oracle11gJ2EE,Java,JDBC,Servlets,JSP,XML,DesignPatterns,CSS,HTML, JavaScript 1.2, JUnit, Apache Tomcat, My SQL Server 2008.

**Client: Xcel Energy, India Oct 2010 – Jul 2011**

**Role: Java Developer**

**Responsibilities:**

* Involved in software development life cycle (SDLC) of the project (analysis, design, development, testing and implementation).
* Used MS-Visio for analysis & design flow and monitored the initial applications prototype development for the project.
* Developed Web Services to get data from the external system in terms of .txt file to load into the database.
* Developed DTS/SSIS packages to load employee details into row Mony tables of the SQL server for further processing.
* Wrote Stored Procedures, functions and complex SQL queries for database operations.
* Used Java Scripts to perform client side validations.
* Worked on performance tuning of queries.
* Developed reports using Crystal Reports reporting tool.

**Environment:**Java, J2EE, JSP, Servlets, .Net Framework 2.0, ASP.NET, C#, VB.NET, ADO.NET, Oracle9i, SQL Server 2005, T-SQL/PL-SQL, HTML, XML, Web Services, JavaScript, Windows 2000, IIS, Tomcat, Visual Source Safe (VSS), Remedy and Crystal Reports.

**Education:-**

Bachelor’s - Computer Science, JNTU, - 2011.

Masters - Computer Science - Missouri Institute Science and Technology - Rolla, MO 2014