**Hanoosh Surapaneni**

**Ph: 917-972-8717**

**Email:hanoosh959@gmail.com**

**Professional Summary:**

* Around 7+ years of IT experience in software Development in **Big Data Technologies** and Analytical Solutions with 3+ years’ experience in design, architecture, and data modeling*as* database developer and 2+ years of hands-on experience in development and design of Java and related frameworks.
* Around 3+years of work experience as Hadoop Developer with good knowledge of **Hadoop** framework, Hadoop Distributed file system and Parallel processing implementation.
* Experience in **Hadoop Ecosystems HDFS, Map Reduce, Hive, Pig, HBase**, **Sqoop.**
* Excellent understanding / knowledge of Hadoop architecture and various components such as **HDFS, Job Tracker, Task Tracker, Name Node, Data Node** and **Map Reduce** programming paradigm.
* Good Exposure on Apache Hadoop Map Reduce programming, Hive, PIG scripting andHDFS.
* Expertise in composing **MapReduce Pipelines** with many user-defined functions using **Apache Crunch**
* Experience working with **Apache SOLR** for indexing and querying
* Experience in managing and reviewing Hadoop log files.
* Hands on experience in Import/Export of data using Hadoop Data Management tool **SQOOP.**
* Strong experience in writing Map Reduce programs for Data Analysis. Hands on experience in writing custom partitioners for **Map Reduce.**
* Performed data analysis using **Hive** and **Pig**.
* Excellent understanding and knowledge of NOSQL databases like **Mongo DB, HBase, and Cassandra.**
* Experience with distributed systems, large-scale non-relational data stores, RDBMS, NoSQL map-reduce systems, data modeling, database performance, and multi-terabyte data warehouses.
* Experience in Software Development Life Cycle (Requirements Analysis, Design, Development, Testing, Deployment and Support).
* Hands on experience in application development using **Java**, **RDBMS**, and Linux shell scripting.
* Experience working with JAVA J2EE, JDBC, ODBC, JSP, Java Eclipse, Java Beans, EJB, Servlets.
* Experience in using IDEs like Eclipse,NetBeans and Maven  
  Development experience in DBMS like Oracle, MS SQL Server, Teradata and MYSQL.
* Strong knowledge of data warehousing, including Extract, Transform, and Load Processes.
* Hands on experience on writing Queries, Stored procedures, Functions and Triggers by using SQL.
* Support development, testing, and operations teams during new system deployments.
* Evaluate and propose new tools and technologies to meet the needs of the organization.
* An excellent team player and self-starter with good communication skills and proven abilities to finish tasks before target deadlines.

**Technical Skills:**

|  |  |
| --- | --- |
| **Hadoop ECO Systems** | Hadoop, MapReduce, HDFS, HBase, Hive, Pig, Sqoop,ZooKeeper. |
| **NO SQL** | Mongo DB, Cassandra |
| **Data Bases** | MS SQL Server 2000/2005/2008/2012, MY SQL, Oracle 9i/10g, MS access, Teradata TeradataV2R5 |
| **Languages** | Languages Java JDK1.4 1.5 1.6 (JDK 5 JDK 6), C/C++, SQL, Teradata SQL, PL/SQL. |
| **Operating Systems** | Windows Server 2000/2003/2008, Windows XP/Vista, Mac OS, UNIX ,LINUX |
| **Java Technologies** | Servlets, JavaBeans, JDBC, JNDI, JTA, JPA |
| **Frame Works** | Jakarta Struts 1.1, JUnit and JTest, LDAP. |
| **IDE’s & Utilities** | Eclipse, Maven, NetBeans. |
| **SQL Server Tools** | SQL Server Management Studio, Enterprise Manager, QueryAnalyser, Profiler, Export & Import (DTS). |
| **WebDev. Technologies** | ASP.NET, HTML,XML |
| **Testing Tools** | Bugzilla, QuickTestPro(QTP)9.2, Selenium, Quality Center, Test Link |

## Education: Bachelor of Engineering from CBIT Osmania University ,India

**Professional Experience:**

**Client: AT&T, Alpharetta, GA Feb 2014 – Till Date**

**Role: Hadoop Developer**

AT&T is an American multinational telecommunication corporation. It is the largest provider for both mobile and landline telephone service, and also provides broadband subscription television services. Being one of the largest telecommunication providers AT&T has huge customer data that can be analysed and taken advantage of. To consumer marketing professionals, data about the users of mobile network are highly valuable so that the US-based network operator is turning access to and collaboration on its data into a new business service. In order to ensure secure data sharing and at the same time easing access and use of data, good management of data is required which involves data aggregation from multiple sources. AT&T has created programmable interfaces to each of its data sets that

Ensure read-only access to the data.

**Role &Responsibilities:**

* Evaluated business requirements and prepared detailed specifications that follow project guidelines required to develop written programs.
* Responsible for building scalable distributed data solutions using **Hadoop**.
* Analysed large amounts of data sets to determine optimal way to aggregate and report on it.
* Developed Simple to complex Map reduce Jobs using Hive and Pig
* Optimized Map Reduce Jobs to use HDFS efficiently by using various compression mechanisms
* Handled importing of data from various data sources, performed transformations using Hive, MapReduce, loaded data into HDFS and Extracted the data from MySQL into HDFS using Sqoop
* Working with **Apache Crunch** library to write, test and run HADOOP MapReduce pipeline jobs
* Involved in joining and data aggregation using **Apache Crunch**
* Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team
* Extensively used **Pig** for data cleansing.
* Created partitioned tables in **Hive.**
* Managed and reviewed **Hadoop log** files.
* Involved in creating **Hive** tables, loading with data and writing hive queries which will run internally in MapReduce way.
* Used **Hive** to analyse the partitioned and bucketed data and compute various metrics for reporting.
* Installed and configured **Pig** and also written Pig Latin scripts.
* Developed **Pig Latin** scripts to extract the data from the web server output files to load into HDFS.
* Load and transform large sets of structured, semi structured and unstructured data
* Responsible to manage data coming from different sources
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required
* Involved in writing, testing, and running MapReduce pipelines using **Apache Crunch**

**Environment**: Hadoop, MapReduce, HDFS, Hive, Pig, Java, SQL, Apache Crunch, Sqoop, Java (jdk 1.6), Eclipse

**Client: SpotlightTicket Management, Calabasas, CA Jan 2013 – Dec 2013**

**Hadoop Developer**

The purpose of the project is to store terabytes of log information generated by the company websites and extract meaningful information out of it. The solution is based on the open source BigData s/w Hadoop. The data will be stored in Hadoop file system and processed using Map/Reduce jobs which includes getting the raw html data from the websites, process the html to obtain product and pricing information, extract various reports out of the product pricing information and export the information for further processing.

**Responsibilities:**

* Worked on analyzing Hadoop cluster and different big data analytic tools including Pig, Hbasedatabase and Sqoop.
* Responsible for building scalable distributed data solutions using Hadoop.
* Implemented nine nodes CDH3 Hadoop cluster on Red hat LINUX.
* Implemented  Apache Crunch library on top of map reduce and spark for data aggregation.
* Involved in loading data from LINUX file system to HDFS.
* Worked on installing cluster, commissioning & decommissioning of datanode, namenode recovery, capacity planning, and slots configuration.
* Created HBase tables to store variable data formats of PII data coming from different portfolios.
* Implemented a script to transmit sysprin information from Oracle toHbase using Sqoop.
* Implemented best income logic using Pig scripts and UDFs.
* Implemented test scripts to support test driven development and continuous integration.
* Worked on tuning the performance Pig queries.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Responsible to manage data coming from different sources.
* Involved in loading data from
* file system to HDFS.
* Load and transform large sets of structured, semi structured and unstructured data
* Cluster coordination services through Zookeeper.
* Experience in managing and reviewing Hadoop log files.
* Job management using Fair scheduler.
* Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team.
* Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshooting, manage and review data backups, manage and review Hadoop log files.
* Installed Oozie workflow engine to run multiple Hive and pig jobs.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Supported in setting up QA environment and updating configurations for implementing scripts with Pig and Sqoop.
* **Environment**: Hadoop, HDFS, Pig, Sqoop, HBase, Shell Scripting,Apache Crunch, Ubuntu, Linux Red Hat.

**Client: Coremetrics Inc, San Mateo, CA Jan 2012- Dec 2012**

**Hadoop Developer**

**Responsibilities:**

* Involved in review of functional and non-functional requirements.
* Facilitated knowledge transfer sessions.
* Installed and configured Hadoop Mapreduce, HDFS, Developed multiple MapReduce jobs in java for data cleaning and pre-processing.
* Importing and exporting data into HDFS and Hive using Sqoop.
* Experienced in defining job flows.
* Experienced in managing and reviewing Hadoop log files.
* Extracted files from CouchDB through Sqoop and placed in HDFS and processed.
* Experienced in running Hadoop streaming jobs to process terabytes of xml format data.
* Load and transform large sets of structured, semi structured and unstructured data.
* Responsible to manage data coming from different sources.
* Got good experience with NOSQL database.
* Supported Map Reduce Programs those are running on the cluster.
* Involved in loading data from UNIX file system to HDFS.
* Installed and configured Hive and also written Hive UDFs.
* Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.
* Gained very good business knowledge on health insurance, claim processing, fraud suspect identification, appeals process etc.
* Developed a custom File System plug in for Hadoop so it can access files on Data Platform.
* This plugin allows Hadoop MapReduce programs, HBase, Pig and Hive to work unmodified and access files directly.
* Designed and implemented Mapreduce-based large-scale parallel relation-learning system
* Extracted feeds form social media sites such as Facebook, Twitter using Python scripts.
* Setup and benchmarked Hadoop/HBase clusters for internal use
* Setup Hadoop cluster on Amazon EC2 using whirr for POC.
* Wrote recommendation engine using mahout.
* **Environment:** Java 6, Eclipse, Oracle 10g, Sub Version, Hadoop, Hive, HBase, Linux, , MapReduce, HDFS,Hive, Java (JDK 1.6), Hadoop Distribution of HortonWorks, Cloudera, MapReduce, DataStax, IBM DataStag 8.1, Oracle 11g / 10g, PL/SQL, SQL\*PLUS, Toad 9.6, Windows NT, UNIX Shell Scripting.

**Laserfiche, Long Beach, CA Mar 2010 - Dec 2011**

**J2EE Developer**

**Responsibilities:**

* Involved in designing the application and prepared Use case diagrams, class diagrams, sequence diagrams.
* Developed Servlets and JSP based on MVC pattern using Struts Action framework.
* Used Tiles for setting the header, footer and navigation and Apache Validator Framework for Form validation.
* Using Resource and Properties files for i18n support.
* Involved in writing Hibernate queries and Hibernate specific configuration and mapping files.
* Used Log4J logging framework to write Log messages with various levels.
* Involved in fixing bugs and minor enhancements for the front-end modules.
* Used JUnit framework for writing Test Classes.
* Used Ant for starting up the application server in various modes.
* Used Clear Case for version control.
* Used SDLC Life Cycle

Environment: Java JDK1.4, EJB2.x, Hibernate 2.x, Jakarta Struts 1.2, JSP, Servlet, JavaScript, MS SQL

Server 7.0, Eclipse3.x, Websphere 6, Ant, Windows XP, Unix, Excel Macro Development.

**SPL WorldGroup Inc, San Francisco, CA Apr 2009 - Feb 2010**

**Java Developer**

**Responsibilities:**

* Involved in Requirement Analysis, Development and Documentation.
* Used MVC architecture (Jakarta Struts framework) for Web tier.
* Participation in developing form-beans and action mappings required for struts implementation and validation framework using struts.
* Development of front-end screens with JSP Using Eclipse.
* Involved in Development of Medical Records module. Responsible for development of the functionality using Struts and EJB components.
* Coding for DAO Objects using JDBC (using DAO pattern)
* XML and XSDs are used to define data formats.
* Implemented J2EE design patterns value object singleton, DAO for the presentation tier, business tier and Integration Tier layers of the project.
* Involved in Bug fixing and functionality enhancements.
* Designed and developed excellent Logging Mechanism for each order process using Log4J.
* Involved in writing Oracle SQL Queries.
* Involved in Check-in and Checkout process using CVS.
* Developed additional functionality in the software as per business requirements.
* Involved in requirement analysis and complete development of client side code.
* Followed Sun standard coding and documentation standards.
* Participation in project planning with business analysts and team members to analyze the Business requirements and translated business requirements into working software.
* Developed software application modules using disciplined software development process.
* Environment: Java, J2EE, JSP, EJB, ANT, STRUTS1.2, Log4J, Weblogic 7.0, JDBC, MyEclipse, Windows
* XP, CVS, Oracle.

**Client: Tata Indicom, India July 07 – Jan 09**

**Role: SQL Server Developer**

Tata Indicom is India's leading international telecom service provider. It is a part of the Tata Group. It started as a successor to the erstwhile Overseas Communication Services, and went on to become the premier provider of international voice and data services.

**Role &Responsibilities:**

* Actively involved in different stages of Project Life Cycle.
* Documented the data flow and the relationships between various entities.
* Actively participated in gathering of User Requirement and System Specification.
* Created new Database logical and Physical Design to fit the new business requirement and implemented the same using SQL Server.
* Created Clustered and Non-Clustered Indexes for improved performance.
* Created Tables, Views and Indexes on the Database, Roles andmaintained Database Users.
* Followed and maintained standards and best practices in database development.
* Provided assistance to development teams on Tuning Data, Indexes and Queries.
* Developed new Stored Procedures, Functions, and Triggers.
* Implemented Backup and Recovery of the databases.
* Actively participated in User Acceptance Testing, and Debugging of the system.

**Environment**: Windows 2000 adv. server, Windows 2000/XP, MS SQL Server 2000, IIS, MS Visual Studio.

**References:** Available upon Request.