**Raj Kumar**

**rkkw.kumar@gmail.com**

**+1(816)-682-6219**

# PROFESSIONAL SUMMARY:

* Overall 8+ years of proactive IT experience in **Analysis, Design, Development, Implementation,** and **Testing** of software applications which includes an accomplished almost 3 Years of experience in **Big Data, Development** and **Design** of Java based enterprise applications**.**
* Leveraged strong Skills in developing applications involving Big Data technologies like **Hadoop, Spark, ElasticSearch, MapReduce, Yarn, Flume, Hive, Pig, Kafka, Storm, Sqoop, HBase, Hortonworks, Cloudera, Ambari, Avro and Scala.**
* Strong experience on Hadoop working environment includes Map Reduce, **HDFS, HIVE, PIG, SQOOP, HBASE, Flume, Zookeeper** and **Oozie** for data storage and analysis.
* Hands on experience in importing and exporting data from different databases like **Oracle, Mysql Teradata** into **HDFS** and **Hive** using **Sqoop.**
* Experience in deploying and managing the different **Hadoop** components (HIVE, PIG, SPARK, SQOOP, FLUME, OOZIE, HBASE, ZOOKEEPER) using **Cloudera Manager** and **Hortonworks Ambari.**
* Strong knowledge on **Hadoop** **HDFS** architecture, **Map-Reduce (MRv1)** and **YARN(MRv2**) framework.
* Managing the **health of Cluster**, resolving performance related issues, coordinating with various parties for Infrastructure Support.
* Extensive experienced in running query using **Impala.**
* Extensive experience in collecting and storing stream data like log data in HDFS using Apache **Flume.**
* Extensively used **MapReduce Design Patterns** to solve complex MapReduce programs.
* Experience in extending Hive and Pig core functionality by writing custom **UDF**s like **UDAFs** and **UDTFs.**
* Experienced in **collecting** data from different sources and **cleaning** the data by using **Pig**.
* Good knowledge on **Data Modeling** and **Data warehouse** concepts.
* Experienced implementing **Security** mechanism for **Hive Data.**
* Extensively used ETL methodology for performing Data Profiling, Data Migration, Extraction, Transformation and loading using **Talend/SSIS** and designed data conversions from wide variety of source systems.
* Experience with **Hive Queries** Performance Tuning.
* Experienced with improving data cleansing process using **Pig Latin** operations, transformations and join operations**.**
* Experienced with performing CRUD operations using **HBase** Java Client API and Rest API.
* Extensive knowledge on NoSQL Databases like **HBase, Cassandra and MongoDB.**
* Experience in troubleshooting errors in **HBase Shell, Pig, Hive** and **MapReduce**.
* Good Knowledge in Hadoop **YARN** environment including **Spark**.
* Worked on loading the production log files data into HDFS using **FLUME**.
* Experience in integrating **Apache Kafka** with **Apache Storm** and created **Storm data pipelines** for real time processing.
* Experience in **Spark, Spark-SQL (SHARK), Spark Streaming and Spark Mlib.**
* Extensively created mappings in **Talend** using **tMap**, **tJoin**, **tReplicate**, **tParallelize**, **tJava**, **tJavarow**, **tDie, tAggregateRow**, **tWarn**, **tLogCatcher**, **tMysqlScd**, **tFilter**, **tGlobalmap** etc.
* Expertise in ETL tools for Extract, Transform and Load the data into warehouse like Informatica, Pentaho and Talend.
* Strong experience in Designing and Developing Business Intelligence solutions in **Data Warehousing/Decision** **Support** Systems using Informatica Power Center.
* Involved in deploying systems on Amazon Web Services Infrastructure services **EC2**, **S3**, **RDS**, **AWS**, and Cloud Formation.
* Experience in working **AMAZON WEB SERVICES** and **HADOOP IN CLOUD (EMR).**
* Experience in deploying Hadoop cluster on Public and Private Cloud Environment like using Amazon Web Services **(AWS) EC2**.
* Strong experience in database design, writing complex **SQL** queries & **stored procedures** using **PL/SQL**.
* Good knowledge in creating **PL/SQL Stored Procedures, Creating indexes, Packages, Functions, Triggers, Cursors** with Oracle (9i, 10g, 11g), and MySQL server.
* Experience in **Java, JSP, Servlets, EJB, Hibernate, Spring, Java Script, and HTML.**
* Experience in writing Shell Scripts (bash, SSH, Perl).
* Understanding of OLAP, OLTP, Business Intelligence and Data Warehousing concepts with emphasis on ETL and Business Reporting needs.
* Expertise in utilizing **Putty/Secured Shell Client tools** for interacting with UNIX operating system, Citrix Client, and VPN for remote server connectivity.
* Ability to adapt to evolving technology, strong sense of responsibility and accomplishment.
* Extensively worked with **Object Oriented Analysis, Design** (OOAD) and development of software using UML Methodology, good knowledge of **J2EE design patterns** and **Core Java Design Patterns.**
* Exposed into methodologies like **Scrum, Agile** and **Waterfall**.
* Strong analytical and Problem-solving skills.
* Excellent communication, verbal skills to interact with client, onsite-offshore coordination.

**Technical Skills**

|  |  |
| --- | --- |
| **Big Data Technologies** | HDFS, Sqoop, Flume, Map Reduce, Pig, Hive, Oozie, Spark, Kafka, Storm, Zookeeper, YARN, Impala. |
| **Scripting Languages** | Shell, Scala, Python |
| **Languages** | C, C++, C#, Java, Python, SQL, PL/SQL, PIG Latin, HiveQL, Unix Shell Scripting. |
| **Front End Technologies** | HTML, XHTML, CSS, XML, JavaScript, AJAX, Servlets, JSP. |
| **Java Frameworks** | MVC, Apache Struts2.0, Spring and Hibernate. |
| **Web Services** | SOAP (JAX-WS), WSDL, Apache CXF, Apache Axis SOA, Restful (JAX-RS), JMS. |
| **Application Servers** | Apache Tomcat 5.5/6.0/7, WebLogic Server 8x/9x/10x, WebSphere 5.1/6.0. |
| **Databases** | Oracle 9i/10g/11g, MySQL, MS SQL Server, Teradata |
| **NoSQL Databases** | HBase, Cassandra. |
| **IDE** | Eclipse, NetBeans. |
| **Operating Systems** | Linux, UNIX, Mac, Windows 7/8/10. |
| **ETL Tools** | Informatica, Talend. |
| **Reporting Tools** | Tableau, SAP Business Objects. |

**Employment History**

**Johnson & Johnson – Raritan, NJ Jan 2016 – Present**

**Role: Hadoop Developer**

**Responsibilities:**

* 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.
* Involved in **installing**, **configuring** and managing Hadoop Ecosystem components like **Spark**, **Hive**, **Pig**, **Sqoop**, **Kafka** and **Flume**.
* Imported data into **HDFS** and **Hive** using **Sqoop**.
* Used **Sqoop** to migrate data to and fro from **HDFS** and **MySQL** or Oracle and deployed **Hive** and **HBase** integration to perform **OLAP** operations on **HBase** data.
* Imported and exported terabytes of data using **Sqoop** from Relational Database Systems to **HDFS**.
* Involved in **Hive** performance optimizations like **partitioning, bucketing** and perform different types of **joins** on **Hive** tables and implementing **Hive** serdes like **JSON** and **Avro**.
* Installed and configured **Pig** for **ETL** jobs. Written **Pig** scripts with regular expression for data cleaning.
* Creating **Hive** external tables to store the **Pig** script output. Working on them for data analysis in order to meet the business requirements.
* Used **AutoSys** Scheduler system to automate the pipeline workflow and orchestrate the map reduces jobs that extract the data on a timely manner.
* Used different **Scala** Collection API’s to process the data in Spark Streaming.
* Developed **Spark** code using **Scala** for faster data processing using **RDD's** and Dataframe API.
* Developed **Spark** Programs for Batch and real time Processing.
* Developed **Spark** Programs for data analytics.
* Used Kafka to store events from various system and processed using **Spark Streaming** to perform near real time analytics.
* Designed Sources to **Targets mappings** from SQL Server, Excel/Flat files to Oracle using **Informatica Power Center**.
* Created Data Marts and loaded the data using Informatica Tool.
* Experienced in troubleshooting the errors in **ILM jobs** with Informatica offshore team, Followed and Maintained Policies and Guidelines for data movement adhering to Client standards using with ILM tool.
* Continuous monitoring and managing the Hadoop cluster through **Cloudera Manager**.
* Involved in loading data from Linux file system to **HDFS**.

**Environment:**  AWS, Cloudera Distribution CDH, Hadoop 2.4.1, HDFS, Hive 0.13, Pig 0.10 HBase 0.94.21, Sqoop, Flume, AutoSys, Oozie, Spark Streaming, Spark SQL, JSON, Java, Scala, Informatica, IMS, Agile Methodology

**Celgene – Warrren, NJ Feb 2015 – Dec 2015**

**Role: Hadoop Developer**

**Responsibilities:**

* Extensively worked on **Hive, Pig, Map Reduce, Sqoop, HBase, Oozie** in an optimized way of distributed processing.
* Designed and Developed Reference table engine frameworks on **Talend** using Hadoop tools such as **HDFS**, **Hive**, **Hbase** **Mapreduce**.
* Used different components in **Talend** like **tMap**, **tFlowToIterate**, **tLogcatcher**, **tNormalise**, **tFilelist**, **tHdfsinput**, **tFileOutputDelimited**, **tUnique**, **tfilterRow**, **thiveload**, **tflowmetercatcher**.
* Used **HCatalog** for accessing **Hive** tables through various applications.
* Experience on **Talend** components like **transformation**, file **processing,** java components, **Unix**, DB related and logging framework.
* Reduced Teradata space used by **optimizing tables** - adding compression where appropriate and ensuring optimum column definitions.
* Created custom UDF java routines in Talend using core java.
* Worked on Hbase API to connect to Hbase and perform filters on tables in **Talend.**
* Worked closely with System Analyst and Architects to design and develop **Talend** jobs to fit the business requirement.
* Worked with **Parquet, Avro** Data Serialization system to work with **JSON** data formats.
* Developed **Pig Latin** scripts to extract the data from the web server output files to load into **HDFS.**
* Develop **Pig UDFs** to pre-process the data for analysis.
* Extensively Used **Sqoop** to import/export data between **RDBMS** and **Hive** tables, incremental imports and created **Sqoop jobs** for last saved value.
* Streamlined Hadoop Jobs and workflow operations using **Oozie** workflow and scheduled through **AUTOSYS** on a monthly basic.
* Designed and developed sqoop jobs for **data ingestion from Teradata** database to HDFS. Merge multiple ingested files in Hadoop using pig.
* Involved in cluster co-ordination services using **Zookeeper.**
* Gather requirements and design of data warehouse and data mart entities.
* Populate or refresh Teradata tables using **Fast load, Multi load & Fast export** utilities for user acceptance testing and loading history data into Teradata.
* Conducted peer design and code reviews and extensive documentation of standards, best practices, and **ETL** procedures.

**Environment:** Hadoop, HDFS, Pig, Hive, Sqoop, Flume, Oozie, AutoSys, Core Java, Talend, HBase Cloudera Distribution, Teradata, Linux, Maven, SBT, Web Services**.**

**Michaels Stores Incorporation-Plano, MN Dec 2013– Jan 2015**

**Role: Hadoop Developer**

**Responsibilities:**

* Involved in data ingestion into **HDFS** using **Sqoop** from variety of sources using the connectors like jdbc and import parameters.
* Developed **MapReduce** programs using Java to parse the raw data, populate staging tables and store the refined data in partitioned tables in **HDFS.**
* Transformed raw data from several data sources into baseline data by developing Pig scripts and loaded the data into **HBase** tables.
* Designed **HBase** schema to avoid Hot spotting and exposed the data from **HBase** tables to REST API on UI.
* Helped market analysts by creating **Hive queries** to spot the emerging trends by comparing fresh data with **HDFS** reference tables and historical metrics.
* Worked with **Parquet, Avro** Data Serialization system to work with **JSON** data formats.
* Defined the job flows in **Oozie** to automate the process of data loading into the **HDFS** and **Pig.**
* Performed various performance optimizations like using distributed cache for small datasets, Partitions, **bucketing** in hive and Reduce Side joins in Mapreduce.
* Extensively Used **Sqoop** to import/export data between **RDBMS** and **Hive** tables, incremental imports and created **Sqoop jobs** for last saved value
* Performed storing of Solr indexes in HDFS.
* Responsible for **Creating Workflows** and Sessions using Informatica workflow manager and Monitor the workflow run and statistic properties on Informatica Workflow Monitor.
* Involved in complete **SDLC** life cycle of big data project that includes requirement analysis, design, coding, testing and production.

**Environment:** HDFS, Mapreduce, Horton Works, Pig, Hive, HBase, Solr, Sqoop, Oozie, Linux, Java, GIT, Jenkins, Nifi, Agile Methodology.

**CGI-Hyderabad Feb 2012 – Oct 2013**

**Java/J2EE & PL/SQL Developer**

**Responsibilities:**

* Developed the application using Struts Framework that leverages classical Model View Layer (MVC) architecture. UML diagrams like use cases, class diagrams, interaction diagrams (sequence and collaboration) and activity diagrams were used.
* Involved in study of User Requirement Specification. Requested insufficient information and helped clearing ambiguity in requirements document.
* Used JSP, Servlets, and HTML5 to create web interfaces. Developed Java Beans and used custom tag libraries for embedding dynamic into JSP pages.
* Involved in development of User Interface using **JSP, ETL, JSTS, CSS, AJAX, Java Script, HTML,** and **Tag libraries**.
* Involved in design, develop and implementation of the application using Spring and j2EE framework.
* Implemented **Session Facade Design Pattern.**
* Defined required mappings in xml that are supported by MVC architecture and spring framework.
* Developed system preferences UI screens using JSP custom TagLibs, JSP templates, JavaScript, CSS and HTML.
* Used AJAXfor developing asynchronous web applications on client side.
* Involved in front-end and back-end development and enhancements using JSP’s, Servlets, Java Script, AJAX.
* Involved in implementation of modules related to file comparison received from clients using Spring Framework.
* Involved in Unit Testing and Bug-Fixing and achieved the maximum code coverage using JUNIT test cases.
* Involved in **XML access** to create the pages in portal content and assign the themes and skins to portlets and portlets to pages.
* Involved in performance tuning, scalability of the application and production support.
* Requirement Analysis, System Design, Preparing Functional Specifications, Preparing Design Specifications, and Development.
* Create, develop, modify and maintain database **objects, PL/SQL packages, functions, stored procedures, triggers, views,** and **materialized views** to extract data from different sources.
* Optimized existing packages for enhanced performance and inclusion of new features of latest database versions.
* Extracted data from various location and load them into the oracle table using **SQL\*LOADER**.
* Extensively worked on Schema Migration, Database Program Migration and Data Migration from Sybase to Oracle by using Oracle SQL Developer.
* Developed **PL/SQL Procedures** and **UNIX Scripts** for Automation of UNIX jobs and running files in batch mode.
* Extensively used the advanced features of PL/SQL like Records, Tables, Object types and Dynamic SQL.
* Created Mail Scripts in **UNIX** to send notification on successful completion or failure of various Interfaces and Jobs, with Customized Subject and Text.
* Developed batch process to handle large data sets and managed them in an **UNIX** environment.
* Assisted teams in their application(s) **design review, code review** and make sure the team follows the recommendations/guidelines.

**Environment:** Java/J2EE, Oracle 10g, SQL, PL/SQL, JSP, EJB, WebLogic 8.0, HTML, AJAX, Java Script, JDBC, XML, JMS, JUnit, log4j, MyEclipse 6.0**.**

**Zentek, Hyderabad July 2009 – Jan 2012**

**Java Developer**

**Roles and Responsibilities:**

* Involved in the analysis, design, and development and testing phases of Software Development Lifecycle (SDLC) using agile development methodology.
* Involved in business requirement gathering and technical specifications.
* 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.
* Presentation components in JSP pages are built using ICE faces tag libraries.
* ICE Faces libraries are used in all presentation pages like Search/Inquiry and data collection pages.
* Used **EJB**s (Session beans) to implement the business logic, **JMS** for communication for sending updates to various other applications and **MDB** for routing priority requests.
* All the Business logic in all the modules is written in core Java.
* GUI was developed using **JSP, AJAX** and JavaScript, spring framework. Involved in the Development of **Spring** Framework Controllers.
* Configured the URL mappings and bean classes using Springapp-servlet.xml. Sybase was the database and Mybatis was used.
* Developed **Mybatis** in Data Access Layer to access and update information in the database.
* Worked with Flied level engineers and teams to make the product more user-friendly. Performed testing for GUI and back end.
* Wrote Web Services using **SOAP** for sending and getting data from the external interface.
* Used **XSL/XSLT** for transforming and displaying reports Developed Schemas for **XML**.
* Involved in writing the **ANT** scripts to build and deploy the application. Developed a web-based reporting for monitoring system with **HTML** and Tiles using Struts framework.
* Used Design patterns such as Business delegate, Service locator, Model View Controller, Session, DAO.

**Environment:** JAVA multithreading, collections, SQL, PHP, Sybase, Eclipse, JavaScript, WebSphere, JBOSS, HTML5, DHTML, CSS, XML, Log4j, ANT, STRUTS 1.3.8, JUNIT, JSP.