**RAJARAM K**

**SENIOR TALEND/BIGDATA CONSULTANT**

Email: **msiddiqui@bigdatadimension.net**

Desk: **+1 607-301-3282**



**SUMMARY**



* Senior Date Warehouse (ETL/BI) andBigdata Professional passionate to learn new concepts, tools, technologies, working collaboratively and exploring Data Architecture landscapes
* **17+ years of experience in IT** and **7 Years in DW, Big Data Ecosystem experience** in ingestion, storage, querying, processing and analysis of big data tools, AWS etc.,
* Expertise in Installation, Development and Administration activities
* Conceptual understanding and Hands on with Hadoop, MapReduce, HDFS, HBase, Zookeeper, Hive, Pig, Sqoop, Flow, Cassandra, Mongodb, Oozie, Spark, Pentaho and Talend etc.,
* **Big Data (Hadoop):**

1. In depth understanding/knowledge of Hadoop Architecture and various components such as

**HDFS, Job Tracker, Task Tracker, NameNode, DataNode and MapReduce concepts**

1. Experience in installation and configuration of Hadoop Ecosystem components: **Hive, HBase, Pig, Sqoop,**

**Etc.,**

1. Experience in building, maintaining multiple **Hadoop clusters** (prod, dev. etc.) of different sizes and configuration and setting up the **rack topology** for large clusters

* **SDLC/Project Lead:** 
  1. Experience in System Analysis, Design, coding, testing and implementations

1. Followed Test driven development of Agile, Water Fall models to produce high quality software
   * Project Management, Project Schedule, Project Tracking and Team Management
   * Expertise in analysing the problems and transferring the business concept to functional requirements
   * Managed to work in Dev, QA Movement and Production environments and expertise in SDLC and Agile model



**SKILLS**



|  |  |
| --- | --- |
| Domain | Energy and Petrochemicals, Manufacturing, Insurance, Financial Systems, Healthcare, Banking, Medical |
| Operating System | Windows 7/8/ XP/Vista, UNIX, LINUX, Eclipse |
| ETL Tools /Packages | TalendIntegration Suite 4.x, 5.x, 6.2,supporting BigData, DI, DQ, Data Profiling, Jenkins 1.4, Pentaho 4.5, SQL, DB2, Documentum 6.5, Informatica, Abinitio package  **Big Data Ecosystems**: Hadoop 2.4,MapReduce, HDFS, HBase, Hive, Pig, Sqoop, Oozie, Spark, Scala, Storm, Ambari, Impala  VirtualBox: Cloudera v5, Hortonworks 2.x, Pentaho Kettle 4.x, 5.x (Spoon, Kitchen, BA etc.,) |
| Databases(RDBMS,MPP, NoSQL) | Oracle 11g, MS SQL Server 2012,2008 R2, Teradata 14, RedShiftPostgreSql 9.4, MySQL, HBASE, Hive 13, MongoDB 2.2, Cassandra 2.1, MS Access |
| Programming Languages/File Formats | JSON, Avro, Parquet, Xml, XSLT, JQUERY, CSS, VB.Net, ASP.Net 4.0, C#.Net, JQuery, Ajax, Java Script, Visual Basic 6.0, IIS 7.0, MVC 3, AJAX,APIs |
| Scripting | Java, Python, Hdfs Commands, Linux, Unix Commands, VB Script |
| Reporting Tools | Pentaho BA Tool, Tableau, Spot Fire ,SSRS 2008R2 , BI, Business Objects XI, Crystal Reports XI, Pentaho Report Designer |
| Cloud Computing platforms | Hadoop 2.5, AWS EMR 4.0.0, EC2, S3, RDS, Redshift, Windows Azure, Cloudera, Hortonworks |
| Source Code Management | TFS, SVN,Vss 6.0 |
| Data Modelling, planning | MS Office, MS Project, MS Visio, Erwin |



**ROFESSIONAL EXPERIENCE**



|  |
| --- |
| **Ethicon, a Johnson & Johnson Company Somerville, NJ** |
| **Talend Architect/Admin 05/2016 – till date** |

* Good understanding on the JnJ domain and project
* Documentations and Presentations on Design, Testing, deployment etc.,
* Involved in Planning, Data Modeling and design DW ETL process
* Identify Potential Opportunities for implementations in DW by accessing RDBMS and BigData environments
* Involved in extending Pentaho Kettle Bigdata project and used different components in Pentaho like Database Lookup & Join, Generate rows, Calculator, Row normalizer &denormalizers, Java Script, Add constant, Add Sequence for business needs
* Developed the ETL mappings using XML, .csv, .txt, Avro, Json sources and also loading the data from these sources into relational tables with Pentaho, Talend, bigdata
* Written custom script logic and used in Pentaho transformations, mappings etc.,
* ETL design and development, Creation of the Pentaho ETL mappings, and workflows to implement the Business Logic.
* Design, develop and deploy end-to-end Data Integration solution.
* Analyse and developed scripts using Pig, Python language and executed
* Develop Adhoc reports using Tableau
* Created and managed internal, external tables using hive queries and enforced partitioning, buckets etc.,
* Implement **TalendETL** processes with **AWS clusters**
* Job Scheduling in Tidal
* File management using Linux, hdfs Commands and security enforcement
* Load data between RDBMS and AWS clusters (S3, Redshift)
* Implemented CDC feature, complex transformation and mappings etc., for incremental Load
* Hive tables to access S3 metadata
* Worked in a concrete Talend Framework along with pre-built templates
* Managing the jobs in TAC, Deployment plans, Monitor the Jobs, Troubleshooting etc.,
* Implementing Data Integration process with TalendBigData Integration Suite 6.2
* Ensure Data Quality, Data Cleanliness, Data Analysis, Best practices, Performance optimization etc.,
* Design, Develop and deploy jobs to load the data into Hadoop and Hive
* Create DDL, DML and validate DB scripts as per standards, performance measures, Optimizations, best practices for Hive and HBase tables etc.,
* Used Sqoop for import/export data between RDBMS and Hive, HBASE tables
* Developed Python scripts for custom requirement and PoC needs
* Writtten and executed Test scenarios
* Managed small Team

**Technology/ Environment: AWS Cloud, Hortonworks 2.2,** Red Hat Linux AS/EL 4.0,Pentaho Kettle, Pentaho BA, HBASE, Hive, Sqoop, Linux Commands, Putty,Talend 6.2, Tableau, Redshift, AWS S3, Python, Oracle 11g, etc.,

|  |  |
| --- | --- |
| **FirstTech Federal Credit Union Bank** | **MountainView, CA** |
| **Talend Admin/Architect** | **10/2015 – 03/2016** |

* Involved in Pentaho Kettle Bigdata project and used different components in Pentaho like Database Lookup & Join, Generate rows, Calculator, Row normalizer &denormalizers, Java Script, Add constant, Add Sequence.
* Developed the ETL mappings for XML, .csv, .txt sources and also loading the data from these sources into relational tables with Pentaho
* ETL design and development, Creation of the Pentaho ETL mappings, and workflows to implement the Business Logic.
* Understand Banking management and Billing System entities
* Created and managed internal, external tables using hive queries and enforced partitioning, buckets etc.,
* File management using Linux, hdfs Commands and security enforcement
* Custom java routines for achieving various business logics
* Create DDL, DML and validate DB scripts as per standards, performance measures, Optimizations, best practices for Hive and HBASE tables etc.,
* Handled various file formats like xml, .csv, .txt, Parquet, Avro, JSON etc., before loading data into HBASE
* Developed Custom code and UDF scripts using Python, Pig UDFs, Java API
* Ensure Data Quality, Data Cleanliness, Data Analysis, Performance optimization etc.,
* Design, Develop and deploy jobs to load the data into Hadoop and Hive
* Develop use cases and scripts using Python
* Writtten and executed Test scenarios
* Used Sqoop for import/export data between RDBMS and Hive, HBASE tables
* Used Apache Spark/Scala for streaming data from large datasets
* Manage and Review **Hadoop** log files
* Load incoming data and optimizations using Cassandra
* Hands on with executing Linux and HDFS Commands
* Involved in Installation of Hortonworks Hadoop and Talend for Big data in DW
* Design DDL, DML and validate DB scripts as per standards, performance measures, Optimizations, best practices etc.,
* Implemented CDC feature, complex transformation and mappings etc., for incremental Load
* Worked in a concrete Talend Framework along with pre-built templates
* Managing the jobs in TAC, Deployment plans, Monitor the Jobs, Troubleshooting etc.,
* Developed custom reports using Tableau

**Technology/ Environment: Windows Azure Cloud, Hortonworks 2.2,** Red Hat Linux AS/EL 4.0, Pentaho Kettle, Pentaho BA, Spark, Scala, HBASE, Hive, Oozie, Sqoop, hdfs Commands, Linux Commands,Puttyetc.,Talend 6.1, Big Data, Python, SQL Server 2012, MySQL, Oracle 11g, Tableau,

|  |  |
| --- | --- |
| **VISA Inc.,** | **FosterCity, CA** |
| **Sr. DW/BIGDATA Consultant** | **06/2015 – 09/2015** |

As a premier payments technology company, Visa’s global network connects thousands of financial institutions with millions of merchants and cardholders every day.

* Understand Revenue management and Billing System entities
* Integrated with systems using DB2, Abinitio, Oracle Systems, Hadoop Systems
* Worked using AWS Cloud EC2 and data processing using Redshift MPP
* Developed UDF scripts using Python, PIG
* Used Oozie job schedulers and Sqoop scripts for loading data
* Access and manage data using Collections from MongoDB
* Ensure Load balancing, Scalability, Performance Optimization, Data Quality etc., in databases, Jobs, queries from Databases and Big Data tools
* Developed custom code using Java API, UDFs
* Created and managed Partitioned tabled enforced with indexing, buckets etc., in Hive
* Managing the jobs in TAC, Deployment plans, Monitor the Jobs, Troubleshooting etc.,
* Develop the ETL mappings for XML, .csv, .txt sources and also loading the data from these sources into relational tables
* Implemented Incremental load for RDBMS, Big Data and flat files
* Created HBASE, Hive tables as per standards and maintained for loading data operations
* Parsed data formats JSON, Avro, Parquets, XML etc., and loaded into Hive, HBASE
* Ingest large amounts of incoming data, and in parallel writing materializations for query support and Managed using Cassandra
* Ensure Data Quality, Data Cleanliness, Data Analysis
* Work with the team in optimizing **Map/Reduce** programs
* Design and developed Talend Framework along with pre-built templates

**Technology/ Environment: AWS, RedShift, Cloudera,** Red Hat Linux AS/EL 4.0, HBASE, Hive, Abinitio, ETL tool Talend 6.1, Big Data, DB2, Java, Python, My SQL, MongoDB, Cassandra

|  |  |
| --- | --- |
| **Aetna Company** | **Omaha,NE** |
| ***Big Data Analyst, Architect*** | ***09/2014 – 05/2015*** |

Aetna is transforming health care to create healthier communities, a healthier nation and a healthier world. Let us show you how we’re focused on helping to create a healthier you.Founded in 1853 in Hartford, CT, Aetna is committed to providing individuals, employers, health care professionals, producers and others with innovative benefits, products and services. Discover more about our organization.

* As Big Data Analyst worked with the business users in finding patterns within structured and unstructured data involving different formats of data and data structures
* Involved in Installation of Cloudera 4x distribution Hadoop and Talend for data integration
* Involved in Pentaho Kettle Bigdata POC project and used different components in Pentaho like Database Lookup & Join, Generate rows, Calculator, Row normalizer &denormalizers, Java Script, Add constant, Add Sequence.
* Developed the ETL mappings for XML, .csv, .txt sources and also loading the data from these sources into relational tables with Pentaho
* ETL design and development, Creation of the Pentaho ETL mappings, and workflows to implement the Business Logic.
* Involved in ETL jobs migration from Pentaho ETL to Talend ETL
* Installing **Hadoop** ecosystem for Revenue management and Call center management
* Load the data from Teradata tables into **Hive** Tables
* Implement **ETL** processes with **Hadoop**, **MapReduce**, **Pig**, **Hive**
* Installation and configuration of Hadoop ecosystem like **HBase, Flume, Sqoop, Impala**
* Experimented with **Spark** and **Scala** with the Hadoop cluster
* Manage and Review **Hadoop log files.**
* Develop **Hadoop** monitoring processes (capacity, performance, consistency) to assure processing issues are identified and resolved swiftly.
* Address the performance tuning of **Hadoop** ELT processes against very large data set work directly with statistically on implementing solutions involving predictive analytics
* Work with the team in building **Map/Reduce** programs
* Cluster co-ordination services through **ZooKeeper**
* Job management using **Fair Scheduler**
* Developed **Pig** Latin scripts to extract the data from the web server output files to load into **HDFS**
* Developed the **Pig** UDF’S to pre-process the data for analysis
* Develop **Hive** queries for the claim analysts to identify the trends
* Developed workflow in **Oozie** to automate the tasks of loading the data into HDFS and pre-processing with **Pig**
* Implementation of TTU’s and TPT’s on new Talend infrastructure (v 5.4.2)
* Implement Big Data technology based on customer’s business requirements design and document comprehensive technical architecture
* Suggested changes to interfaces between Hadoop ecosystem and external systems, based on the requirements integration customization

**Technology/ Environment:** Pentaho Kettle, Pentaho BA, Red Hat Linux AS/EL 4.0, CDH3, Cloudera Manager, HBase, flume, pig, ZooKeeper,Oozie, Scala, Hive, and ETL tool (Talend), UNIX, Linux, Ubuntu, Tableau

BigDataPoC at IHG, to analyze 20 to 30GB of the website data collected in the log files for all the visitors logging or visiting the website. It spawns recommendations and tips for guests and registered users. These recommendations are responsible for an increase in cross selling of financial products.

* Worked with Business Analysts and developed Use cases, developed POCs for complex use cases
* Setup Hadoop instances and clusters on Amazon **EC2**
* Created Oracle stored procedures for generating **Sqoop** scripts to extract data
* Implemented **Oozie** workflows
* Developed Apache Pig scripts using HCatalog for manipulating Hive data
* Developed **Hive** scripts for loading data from external files Designed and implemented Hadoop MapReduce jobs in Java
* Created Hive queries that helped market analysts spot emerging trends in the visitors to the website by comparing fresh data with EDW reference tables and historical metrics

***Technology/* Environment:**LINUX, Hive, PIG Latin, Oozie, Scoop, Hadoop, HDFS

|  |  |
| --- | --- |
| **Honda ETL Financial Systems** | **Torrance, CA** |
| ***Big Data Developer*** | ***01/2014 – 09/2014*** |

In October 1946, Soichiro Honda established the Honda Technical Research Institute in Hamamatsu, Japan, to develop and produce small 2-cycle motorbike engines. Two years later, Honda Motor Company, Ltd. was born, and in 1959 Honda opened its first storefront in Los Angeles with six industrious employees. Honda opened first U.S. plant in 1979 and have evolved into a company that directly employs nearly 25,000. More than 100,000 workers are employed at authorized Honda automobile, motorcycle and power-equipment dealerships in the United States. Tens of thousands of additional Americans are employed by more than nearly 600 U.S. suppliers from which Honda purchases parts and materials.

* Developing ‘As Is’ architecture to understand and conceptualize technology alignment of Business Intelligence and Big Data analytical needs
* Streamlining ‘Governance activities’ for Big Data, Installing Hadoop Cluster and to load the data into Hadoop & Hive tables
* Gained familiarity with Apache Hadoop (principal open-source Cloud Computing platform for storing, managing, processing and analyzing Big Data) and Experience in building applications from Hadoop technology stack
* Responsible in Installation and configuration of Cloudera distribution Hadoop 2, 3, NameNode, Secondary NameNode, Job Tracker, Task Trackers and DataNodes
* Load log data into **HDFS** using **Flume.**
* Created scripts using **Sqoop** to export data into **HDFS** and **Hive**
* Developed multiple **Pig** jobs for data cleaning and preprocessing
* Assisted with data capacity planning and node forecasting
* Scheduling cron jobs for file system check using **fsck** and also running the **balancer** for uniform load on the nodes
* Data scrubbing and processing with **Oozie**
* Expertise in Hadoop cluster task like **Adding and Removing Nodes** without any effect to running jobs and data.
* Install, Configure and Manage of **Hadoop** Cluster spanning multiple racks using automated tools like **puppet**
* Manage and review Hadoop Log files
* Set up automated 24x7x365 **monitoring** and escalation infrastructure for Hadoop cluster using **Nagios** and **Ganglia**

***Technology/ Environment:***: Hadoop, Hbase, Hive, Sqoop, Flume, Windows 2003, MS SQL Server 2000, ASP.Net,Visual Basic 6.0, Business Objects and ETL tool (Talend), UNIX, Linux, Ubuntu, eclipse, Java, Tableau

**Chevron CorporationBakersfield, CADWH Developer*05/2012 – 12/2013***

Chevron is one of the world’s leading integrated energy companies. The success is driven by our people and their commitment to get results the right way—by operating responsibly, executing with excellence, applying innovative technologies and capturing new opportunities for profitable growth. They are involved in virtually every facet of the energy industry. They explore for, produce and transport crude oil and natural gas; refine, market and distribute transportation fuels and lubricants; manufacture and sell petrochemical products; generate power and produce geothermal energy; provide renewable energy and energy efficiency solutions; and develop the energy resources of the future, including research into advanced biofuels.

* Interaction with IT team of the client for the purpose of requirement gathering, understanding requirements.
* Implementing Data Integration process with Talend Integration Suite 3.2
* Designing, developing and deploying end-to-end Data Integration solution.
* Developed Talend ETL jobs to push the data into Talend MDM and develop the jobs to extract the data from MDM.
* Used components in Talend like tMap, tmssqlinput, tmssqloutput, tFiledelimited, tfileoutputdelimited, tmssqloutputbulkexec, tUnique, tFlowToIterate, tIntervalmatch, tLogcatcher, tFlowmetercatcher, tFilelist, tAggregate, tSort, tMDMInput, tMDMOutput, tFilterRow.
* Developing the ETL mappings for XML, .csv, .txt sources and also loading the data from these sources into relational tables with Talend ETL
* Created source to target mappings, edit rules and validation, transformations, and business rules. Analyzed client requirements and designed the ETL mapping.
* Well acquainted with Performance Tuning of sources, targets, mapping and sessions to overcome the bottlenecks in mappings
* Created transformations and mappings

***Technology/ Environment:***: Talend 5.X, Oracle 10g,SQL Server,DB2,T-SQL,Sybase,MSVisio, Microsoft Office 2000, UNIX (AIX), Windows 2000/XP, Toad, BO XI

**Chevron CorporationBakersfield, CA**

|  |  |
| --- | --- |
| ***ETL & BI Developer*** | ***11/2011 –04/2012*** |

Chevron is one of the world’s leading integrated energy companies. The success is driven by our people and their commitment to get results the right way—by operating responsibly, executing with excellence, applying innovative technologies and capturing new opportunities for profitable growth. They are involved in virtually every facet of the energy industry. They explore for, produce and transport crude oil and natural gas; refine, market and distribute transportation fuels and lubricants; manufacture and sell petrochemical products; generate power and produce geothermal energy; provide renewable energy and energy efficiency solutions; and develop the energy resources of the future, including research into advanced biofuels.

* Interaction with IT team of the client for the purpose of requirement gathering, understanding requirements.
* Implementing Data Integration process with Talend Integration Suite 3.2
* Designing, developing and deploying end-to-end Data Integration solution.
* Developed Talend ETL jobs to push the data into Talend MDM and develop the jobs to extract the data from MDM.
* Used different components in talend like tmap, tmssqlinput, tmssqloutput, tfiledelimitede, tfileoutputdelimited, tmssqloutputbulkexec, tunique, tFlowToIterate, tintervalmatch, tlogcatcher, tflowmetercatcher, tfilelist, taggregate, tsort, tMDMInput, tMDMOutput, tFilterRow.

***Technology/ Environment:*** *Talend, Oracle 10g, SQL Server 2008, SSIS,TOAD, BO XI*

|  |  |
| --- | --- |
| ***Larsen and Toubro Infotech Ltd.,*** | ***Chennai, India*** |
| **Team Lead** | ***05/2005 – 10/2011*** |

* Prepared strategy to implement ETL code in production and load historical as well as incremental loads data in production
* Fine-tuned existing Talend maps for performance optimization.
* Made adjustments in Data Model and SQL scripts to create and alter tables.
* Involved in identifying and verifying Tables, Table Partitioning, Indexes and Constraints after the database migration
* Efficient Documentation was done for all phases like Analysis, design, development, testing and maintenance
* Monitoring and assigning Incidents to Support Team
* Involved in Project Management Activities
* Attend Change Management Review Meetings and responsible for RFC maintenance and getting Approval
* Prepare and share Incident Management Report and Issue Tracking Reports
* for follow up and Closure
* Active involvement in Analysis, Design, Coding & Bug Fixes
* Analysis on lifecycle management, document categorization, retention policies
* Involvement in Documentum configuration, creating workflows
* Ensure usability, scalability and robustness of the developed prototype
* Prepare Ad hoc Reports
* Designed and normalized database
* Prepared test cases and performed testing of the system
* Involved in implementing database concepts
* Designed and developed web pages as per the specification and tested the system
* Involved in gathering requirements and documented as per business processes
* Prepared Project Plans and tracked their execution
* Customized the tools based on business requirements
* Allocated work for the team members and ensured maximum utilization of resources

***Technology/ Environment:*** *VB.NET, ASP.NET, C#.NET,*SSIS, SQL Server, Oracle, BI, Crystal Reports, Documentum, SharePoint



**EMPLOYMENT HISTORY**



Bigdata Dimension Inc 08/2014 - Till Date Architect/Admin

Larsen and Toubro Infotech Ltd 04/2005 - 08/2014 Team Lead

ZyanInfotech 02/2002 - 03/2005 Software Engineer



**EDUCATIONAL QUALIFICATION**

|  |  |
| --- | --- |
| Education & Credentials | Bachelor of Engineering in Electronics and Communications stream from Madurai Kamaraj University, INDIA |

.