**Ravi Teja Voleti**

**+1 (770) 742-8768**

**ravivoleti3@gmail.com**

* Extensiveexperience dealing with tools & technologies like **Talend**, **Spark, Big Data, ETL , Datastage**, **ESB,** **Informatica**, **Tableau, Python, Unix** etc and databases such as **Hive, Hbase, Postgres, Redshift, MS SQL** Server, **Oracle**, **MySQL,** Databases on Windows \ Unix platforms in **On premise** , Big Data & Cloud Environments **AWS** and **Azure** with expertise in analysis, design, development.
* Proficiency on normalized and dimensional modeling techniques , OLAP, OLTP, **Data modeling** using **Erwin**, Star Schema modeling, Snow-Flake modeling, Fact and Dimensions tables, Slowly Changing Dimensions **(SCDs)** for **Data Lake**, **Data Warehouse**, Data marts, Data Sources.
* Worked with Spark **Data Frames**, **Pyspark SQL,** HDFS, S3 **commands** etc and developed scripts , custom code with python, unix shell etc
* Created complex **DI , Big Data Batch Spark jobs, Event Based (AWS lambda, ActiveMQ Kafka** etc**) and Real Time Data Processing** usingStreaming jobs**, ESB routes** in Talend using components such as **tMap, tJoin, tReplicate, tParallelize, tJava, tJavaFlex, tAggregateRow, tDie, tWarn, tLogCatcher, cLog, tfileoutputparquet,tsqlrow**  etc
* Worked on Talend installation on remote servers with High Availability, **TAC** execution plans, job conductor tasks, triggers, building **CICD** pipelines with **Git, Jenkins, Nexus**, Bitbucket Integration and other advanced areas such as Talend **Metadata Bridge , Metaservlet API, Dynamic Schema, Data Mapper** , Routes etc
* Hands-On experience in **developing**, **testing** and expert in using the **troubleshooting** and **debugging** techniques to understand the errors in jobs\code, evaluating and solving the data issues of structured, unstructured , raw and transformed data, exception handling, auditing, enabling notification and alerting processes using **AWS SNS** , TAC Notifications , **SMTP** etc
* Extensive experience in database programming using **SQL,** complex queries that includes **Stored Procedures, Functions, Views** etc. Good at using necessary **tuning**, **partition** techniques and indexes to achieve the best performance results.
* Developed parallel and server jobs using different processing stages in Datastage like**Transformer, Aggregator, Lookup, Join, Sort, Copy, Merge, Funnel, CDC, Change Apply and Filter** ODBC, Shared containers and local containers, reusable mapping parameter sets etc
* Experience using all stages in **sequencers** and create parallel jobs with **RCP** and Multi instance.
* Loaded/extracted/parsed data from different file formats **parquet, .txt, delimited, fixed width, csv, orc, json, xml** files.
* Extracted/moved data and did multiple operations with source systems like **Amazon S3**, **Hive, ADLS, Azure (Event Hub, Queues),** **HDFS**, **APIs**, **Salesforce Objects**, **SFDC Platform Events, Salesforce Push topics** **Five9**, Apache **Kafka**, **ActiveMQ Topics & Queues**, **Back Office** and **SQL Server Databases etc**
* Actively involved in data collection, group discussions etc. In addition to my enthusiasm, I will bring to the position strong communication skills, flexibility, and the ability to work effectively and efficiently in a team and individually. I have also played a key role in handling the **long term project** development issues at **multiple levels**.

**EDUCATION**

* Masters in Computer Science Engineering (MS)

University of Central Missouri, Warrensburg, MO

**Certifications**

Certified Professional for Talend.

Certified Professional for Informatica.

**Technical Skills**

**Technologies :** Spark2.3**,** AWS, Azure, Talend Real Time Big Data Platform 6.4.1,

Talend Open Studio 6.3, 6.2.6.1, Talend Platform for MDM 5.4.2,

ESB, IBM Infosphere Information Server Datastage 9.1, 8.7,

8.5, Ascential Datastage 8.1, Tableau 9.1,8.2

**Data Modeling :** Star-Schema Modeling, Snowflake-Schema Modeling,

Erwin Data Modeling

**Databases :** Oracle 11i,10g,9i/8i/7.x, MS SQL Server 6.5/7.0, 2000,

SQL Server 12.0/11.x, Postgresql, Pg Admin 9.6, 9.5, My SQL,

Amazon Redshift, Hive, Presto, Hbase

**Environments :** UNIX (Amazon EC2, Azure), Windows 10, 7/XP, EMR

**/Platforms**

**Scheduling :** Talend Administration Center (TAC),

Datastage Director, Unix (crontab), Windows (Task scheduler),

AWS Lambda

**Programming :** SQL, Unix Shell Scripting, C, C++, Python 3.6, 2.7, Java,

Windows Power Shell

**Languages**

**Other Tools :** TOAD for DB2, Quality

Center, Github, Bit Bucket, Sharepoint, Jira, OneDrive,

Amazon Work Docs, HUE, Dremio, Pycharm, Apache Ambari etc

**Professional Experience**

**Virgin Voyages, Atlanta, GA April 2018 – Till Date**

**(Employer: Capgemini America, Inc)**

**Senior Data Engineer (Talend \ Spark\ Big Data \ ESB\Tableau …)**

Virgin Voyages is a startup [cruise line](https://en.wikipedia.org/wiki/Cruise_line) company. IT services are provided from the scratch. This is a Data Lake Project in the cloud where the data is moved from different sources to AWS. Talend is used as a strategic tool in building the seaware application, recommendation engine and the customer 360 dashboards.

* Developed Talend Jobs & PySpark Code to move data from **Salesforce Objects to AWS S3**
* Used the joblets as reusable pieces of code
* Developed Talend Spark, Big Data Batch Jobs and custom **Pyspark SQL** code to move data between **Hive, Amazon Redshift, Hbase, Oracle, Postgresql etc**
* Developed ESB routes to move data between ActiveMQ Topics, ActiveMQ Queues, **Kafka**, **Salesforce Pushtopics, Salesforce Platform Events**
* Enabled **Durable Subscription & Client Acknowledgement** while reading data from ActiveMQ in Talend ESB Routes to prevent **data loss**
* Developed **Metaservlet API (**runtask, associatepregeneratedjob, gettaskIdbyname, getESBtaskstatus etc**) & Unix shell** scripts to update the TAC, to trigger jobs by polling script, to **automate the creation of Tac tasks**.
* Wrapper jobs are created to invoke java jars, spark scripts etc
* Performed various **pyspark data frames** operations such as **read data** from **parquet** , **filter , sort, insert** the data into hive, files , **print schema**, **create temporary tables** etc
* Deployed ESB routes using **kar** files in runtime containers **and ESB conductor**
* Integrated DI jobs with ESB routes (Apache Camel).
* Created **metadata driven framework** to load data to multiple tables using a single job
* Created **Tableau** dashboards, graphs, charts by connecting to oracle database etc
* Created **external** and **internal hive and redshift tables** on top of **AWS S3** and **HDFS** with **partitioning**
* Monitored the routes using karaf and ESB logs
* Implemented **Dynamic Schema** instead of developing multiple jobs
* Used components such as **cTimer, cConvertBodyto, cJMS,cMessagingendpoint,cKafka, tjmsinput, cMessageFilter, thttprequest** etc
* Enabled Email **Notifications on TAC** instead of adding extra components on jobs.
* Developed **Big Data Spark jobs** with SCD-2 logic in the AWS data lake (S3) and developed pyspark code to handle history data.
* Worked with Hive **Partitions** and ran hive queries with different engines such as **TEZ, Map Reduce , Spark**

**Environment:** Spark 2.3,Talend Real Time Big Data Platform 6.4.1, AWS, Windows 10, **My SQL,** Oracle, redshift, hbase, Talend Administration Center (TAC), Bit Bucket, DBeaver, **AWS Work Space,** My SQL Work bench, **HUE, Hbase ,Hive, ActiveMQ, ESB,** Tableau, Salesforce & salesforce Workbench, Postgresql.

**Royal Caribbean Cruises Ltd, Atlanta, GA Oct 2017- April 2018**

**(Employer: Capgemini America, Inc)**

**Programmer \ Developer Lead (Talend, Spark etc)**

Royal Caribbean Cruises Ltd. is an American global cruise company. It is one of the world's largest cruise line operator. This is migration project from Informatica to Talend. Also this project is designed to integrate the Data warehouse into Azure cloud with the usage of Talend Real time Data Operations and Enterprise Service Bus which allows the client to have effective reporting environment

* Worked in the project right from the start and a part of every technical discussion.
* Used Talend, Spark, Azure to develop processes for extracting, cleansing, transforming, integrating and loading data into the data warehouse database and developed Talend Jobs as per the business requirement.
* Installed Components from **Talend Exchange** to Talend Studio
* Developed windows powershell scripts and **Spark Scripts** as per the requirement and integrated windows powershell script with Talend using tsystem component
* Developed Talend jobs which upload and get files from **ADLS** \ Blob
* **CICD** was implemented for talend to move the artifacts to **Nexus** from **Github**
* Used **TAC Metaservlet** concepts to implement parallelism.
* Developed Talend jobs to call **Oracle Stored Procedures**
* Created **Service Bus, Queues , Consumer groups, Event Hubs** in Azure Environment
* Developed **pyspark code** which performs the **SCD** operations.
* Developed a Talend job to read and consume messages from **Azure Queues, Topics etc**
* Designing, developing and deploying end-to-end Data Integration solution
* Extensively involved in various iterations for design/develop and testing Talend code.
* Used Components such as tpartitioner, tcollector, tjava, tjavarow,tlogrow, trowgenerator, tfileinput delimited etc
* Made use of **Talend Metadatabridge** to convert the Informatica jobs to Talend Jobs
* Wrote Stored Procedures and handled complex queries
* Developed Talend ESB routes to pull data from Azure Event Hubs, Queues etc
* Worked on Error handling techniques and tuning the ETL flow for better performance.
* Implemented Dimensions (SCD) Type1, Type 2 to capture the changes using Talend ETL Tool
* Worked Extensively on TAC (Admin Console) where we Schedule Jobs in Job Conductor

**Environment:** Spark 2.3,Talend Real Time Big Data Platform 6.4.1, Azure Cloud(ADLS\Blob), Windows Powershell, Python, Windows 10, Oracle, SQL Server , Talend Administration Center (TAC), Github, Bit Bucket, DBeaver, SQL Developer, Dremio, Apache Ambari

**Capital One, Richmond, VA May 2016 - Oct 2017**

**(Employer: CDI Solutions Inc)**

**Senior Data Engineer**

Capital One offers credit cards, checking and savings accounts, auto loans, rewards, and online banking services for individuals and businesses. The project was to design, develop and maintain a data warehouse to understand the client needs, undertake trend analysis, provide better customer service and assess current and future profitability.

Involved in gathering requirements from users and involved in modifying various technical Specifications in the Development phase to Maintenance Phase.

* Implemented Dimensions (**SCD) Type1, Type 2** to capture the changes using Talend ETL Tool and Pyspark
* Preparing technical Design Documents and mapping the data.
* Develop ETL mappings for various sources (.TXT, .CSV, JSON) and also load the data from these sources into relational tables with Talend Open Studio.
* Created Projects, jobs in Talend Open Studio
* Used the Talend Components like tmap,tparallelise,tmssqlinput,tpostgreslqoutput etc
* Worked on **Global variables, Context variables** and extensively used over 100+components in Talend to create jobs.
* Used **Git** as version controlling tool for checking in SQL, Unix , Python scripts.
* Worked on Error handling techniques and tuning the ETL flow for better performance.
* Worked on Kafka Message Broker for **Real time jobs**
* Used **shared DB connection** in Talend to either commit all or rollback data for all jobs
* Scheduled **Talend** jobs using **crontab** in **Amazon EC2**
* Loaded data into Staging, Hub and Data Marts.
* Developed **ETL** Loader using **Python** code.
* Developed Talend jobs to get data from **Amazon S3**
* Developed Real Time jobs using Talend using **tkafkainput** component
* Used **AWS Lambda** as a scheduler.
* Worked on SQL for extracting the data from the source.
* Experience working with Amazon Web Services (AWS) Server and worked effectively in Agile Environment
* Deployed solutions that maximize consistency and usability of the data.

**Environment:** Spark**,** Talend Open studio 6.3,6.2,6.1, UNIX (AWS server), SQL Server2012, Postgresql, Pg Admin 9.6,9.5, Windows 7, AWS Lambda, SNS, EC2, RDS, S3, Python 3.6,2.7, Github, Crontab, Kafka etc

**CDI Solutions Inc San Ramon, CA March 2016 - May2016**

**Senior ETL (Talend\Datastage) Developer**

CDI Solutions, Inc. is a technical consulting firm that specializes in Custom Data Integration (CDI) initiatives. CDI Solutions provides clients to not only avoid the common pitfalls but to be highly successful in their BI project endeavors. I was involved with all of the phases of the life cycle starting from gathering the requirements, integrating data in the systems and presenting a centralized data warehouse interface.

* Involved with business analysts to understand the business requirement specifications and implemented the ETL jobs using Datastage.
* Worked with Oracle SQL Developer while Implementing Unit Testing of ETL Talend Jobs.
* Design & Development of DataStage Jobs with Unit Testing.
* Extensively used **Datastage Director** for monitoring and debugging of jobs and sequences.
* Designed ETL jobs to identify and remove duplicate rows using remove duplicate stage.
* Extensively worked with surrogate key generator to generate warehouse keys before loading data into warehouse.
* Designed Datastage **sequencers** to specify Job execution order.
* Tuned Datastage jobs to obtain better performance.
* Unit Test Datastage Jobs in development including creating the appropriate test data.
* Proactively evaluated the quality and integrity of data required by the users to satisfy their business needs.
* Developed Datastage jobs using **RCP**(Runtime Column Propagation)
* Developed UNIX Shell Scripting to automate file manipulation, data loading procedures, Data Conversion and scheduling jobs.
* Self-documenting ETL Jobs to comply with Metadata standards.
* Developed and maintained Technical documentation and run books.
* Used Datastage as an ETL tool to extract data from sources systems, loaded the data into the SQL Server database
* Tuned Talend Jobs as per requirement

**Environment:** IBM Datastage 9.1, 8.7, 8.5, 8.1(Administrator, Designer, Director), Talend Open Studio 6.0.1, Oracle11g, *UNIX*, Windows 7.

**Ellie Mae, Pleasanton, CA Oct 2015 - March 2016**

**(Employer: CDI Solutions Inc)**

**Senior EDW Developer**

Ellie Mae, Inc. is a leading provider of innovative on-demand software solutions and services for the residential mortgage industry. Ellie Mae’s all-in-one Encompass mortgage management solution provides one system of record that allows banks, credit unions and mortgage lenders to originate and fund mortgages and improve compliance, loan quality and efficiency.

* Worked on Star Schema and Snowflake Schema.
* Proactively evaluated the quality and integrity of data required by the users to satisfy their business needs.
* Designed and Implemented the ETL process using Talend Platform for MDM (Enterprise Edition) to load the data from Source to Target Database.
* Used **Execution Plan** in TAC instead of building parent jobs
* Created **Users** on **TAC**
* Involved in Data Extraction from .xslx, .csv files using Talend by using Java as Backend Language.
* Develop ETL mappings for various sources and also load the data from these sources into relational tables with Talend Enterprise Edition
* Create ETL Mapping to pull data from source, apply transformations, and load data into target database
* Developed Source to Target Mapping documentation based on business requirement specifications. Worked on Error handling techniques and tuning the ETL flow for better performance.
* Convert business requirements into System Requirement Specifications (SRS)
* Extensively used Talend components tMap, tDie, tConvertType, tFlowMeter, tLogCatcher, tRowGenerator, tOracleInput, tOracleOutput, tfileList, tDelimited etc
* Scheduling the ETL mappings on daily, weekly, monthly and yearly basis.
* Used Dynamic Schema concept using Talend as a part of the requirement.
* Identified strengths and opportunities for data improvement.
* Performed Unit Testing and System Testing.

**Environment:** Talend Platform for MDM 5.4.2 (Enterprise Edition), Talend Adminstration Center (TAC), SQL Server 2012, Windows 7,Tableau 8.2,Unix, sharepoint

**CDI Solutions Inc, San Ramon, CA Aug 2015 – Oct 2015**

**ETL (Talend\Datastage) Developer**

CDI Solutions, Inc. is a technical consulting firm that specializes in Custom Data Integration (CDI) initiatives. CDI Solutions provides clients to not only avoid the common pitfalls but to be highly successful in their BI project endeavors. Datastage extracts data from the incoming sources, processes them according to the business requirements and loads them into the staging tables. Also it provides an integrated solution that enables storage, delivery, and management of network data and content to achieve business goals

* Communicated with business people, understood their needs and incorporated them in a functional document.
* Drafted technical design documents from the functional sheets
* Extensively involved in various iterations for design/develop and testing Datastage, Talend code.
* Developed data extractions, transformations and routines.
* Checked the database connectivity using Datastage Designer in higher environment before first cycle run process begins.
* Deployed solutions that maximize consistency and usability of the Datastage jobs and data.
* Developed jobs in Parallel Extender PX using different stages like Transformer, Aggregator, Lookup, Join, Merge, Modify, Remove Duplicate, Oracle Stage, Sort, Peek, Row Generator, Column Generator, Sequential File, and Data Set.
* Designed Datastage sequences to specify Job execution order.
* Deployed the Talend jobs and scheduled the jobs for a batch run.
* Loaded data into staging area.
* Tuned Datastage and Talend jobs to obtain better performance by using various tuning techniques.
* Unit tested Datastage Jobs in development including creating the appropriate test data.
* Identified strengths and opportunities for data improvement.
* Implemented Slowly Changing Dimension (SCD) as per business requirement for saving history using Data Stage, Talend for dimension tables in Star Schema.
* Supported Datastage environment whenever Datastage jobs got aborted.

**Environment:** IBM Information server 9.1, 8.7, 8.5, 8.1(Administrator, Designer, Director), Talend Open Studio 6.0.1, Oracle11g, Unix, Windows 7

**Aparna Group, Hyderabad, India Aug 2013 – July 2014**

**Graduate Engineer Trainee**

Aparna Group is a group of companies dealing with real estate, construction, building materials etc. This project is to provide the leadership with the estimates of sales, expenses, revenue forecasts etc to benefit administration, human resource and services for projects of any size. Mainly involved in designing, developing and maintaining two major types of decision support systems, HR and financial. I was involved with all of the phases of the life cycle starting from gathering the requirements for designing DS systems, integrating data in the systems and presenting a centralized data warehouse interface.

* Understanding existing business model and customer requirements.
* Designed the target schema definition and Extraction, Transformation and Loading (ETL)
* using datastage.
* Mapped data items from Source Systems to the Target System.
* Developed Talend Jobs as per the business requirement.
* Used various common Talend transformations components like tMap, tDie,
* tConvertType, tFlowMeter, tLogCatcher, tRowGenerator, tSetGlobalVar, tHashInput, tHashOutput etc.
* Used contexts to pass parameters to the job.
* Used the Datastage designer to develop processes for extracting, cleansing, transforming,
* integrating and loading data into the data warehouse database.
* Worked on programs for scheduling data loading and transformations using Datastage
* from legacy system to Oracle using Datastage Loader
* Used the Datastage director and its run-time engine to schedule running the solution,
* testing and debugging its components and monitored the resulting executable versions (on
* an ad hoc or scheduled basis).
* Used Datastage manager for importing metadata from jobs, new job categories and created

new data elements.

**Environment**: Talend Open Studio, Datastage (Manager, Designer, Director), Oracle9i, SQL Server 2000, Windows, Unix