



The following documentation is an electronically-submitted vendor response to an advertised solicitation from the *West Virginia Purchasing Bulletin* within the Vendor Self-Service portal at wvOASIS.gov. As part of the State of West Virginia's procurement process, and to maintain the transparency of the bid-opening process, this documentation submitted online is publicly posted by the West Virginia Purchasing Division at WVPurchasing.gov with any other vendor responses to this solicitation submitted to the Purchasing Division in hard copy format.

Header @ 3

List View

General Information | [Contact](#) | [Default Values](#) | [Discount](#) | [Document Information](#) | [Clarification Request](#)

Procurement Folder: 1379271


Procurement Type: Central Master Agreement

Vendor ID: VS0000029949 

Legal Name: COMPU VISION CONSULTING INC

Alias/DBA:

Total Bid: \$737,484.80

Response Date: 03/06/2024 

Response Time: 12:16

Responded By User ID: compuvision123 

First Name: Shaloo

Last Name: Mital

Email: certs@compuvis.com

Phone: 7324221500

SO Doc Code: CRFQ

SO Dept: 0210

SO Doc ID: ISC240000011

Published Date: 3/1/24

Close Date: 3/6/24

Close Time: 13:30

Status: Closed

Solicitation Description: Addendum No 1 -Professional Database Engineer (OT24116)

Total of Header Attachments: 3

Total of All Attachments: 3



Department of Administration
 Purchasing Division
 2019 Washington Street East
 Post Office Box 50130
 Charleston, WV 25305-0130

**State of West Virginia
 Solicitation Response**

Proc Folder: 1379271
Solicitation Description: Addendum No 1 -Professional Database Engineer (OT24116)
Proc Type: Central Master Agreement

Solicitation Closes	Solicitation Response	Version
2024-03-06 13:30	SR 0210 ESR03062400000004618	1

VENDOR
 VS0000029949
 COMPU VISION CONSULTING INC

Solicitation Number: CRFQ 0210 ISC2400000011
Total Bid: 737484.8000000000465661287307 **Response Date:** 2024-03-06 **Response Time:** 12:16:11
Comments:

FOR INFORMATION CONTACT THE BUYER
 Toby L Welch
 (304) 558-8802
 toby.l.welch@wv.gov

Vendor Signature X **FEIN#** **DATE**

All offers subject to all terms and conditions contained in this solicitation

Line	Comm Ln Desc	Qty	Unit Issue	Unit Price	Ln Total Or Contract Amount
1	Professional Database Engineer - YR1	2080.0000	HOUR	84.750000	176280.00

Comm Code	Manufacturer	Specification	Model #
80111609			

Commodity Line Comments:

Extended Description:

Fill out an hourly rate into the unit price field. The quantity amount is estimated for evaluation purposes only. The actual hours used may vary depending on usage.

Line	Comm Ln Desc	Qty	Unit Issue	Unit Price	Ln Total Or Contract Amount
2	Professional Database Engineer - Optional YR2	2080.0000	HOUR	87.290000	181563.20

Comm Code	Manufacturer	Specification	Model #
80111609			

Commodity Line Comments:

Extended Description:

Fill out an hourly rate into the unit price field. The quantity amount is estimated for evaluation purposes only. The actual hours used may vary depending on usage.

Line	Comm Ln Desc	Qty	Unit Issue	Unit Price	Ln Total Or Contract Amount
3	Professional Database Engineer - Optional YR3	2080.0000	HOUR	89.910000	187012.80

Comm Code	Manufacturer	Specification	Model #
80111609			

Commodity Line Comments:

Extended Description:

Fill out an hourly rate into the unit price field. The quantity amount is estimated for evaluation purposes only. The actual hours used may vary depending on usage.

Line	Comm Ln Desc	Qty	Unit Issue	Unit Price	Ln Total Or Contract Amount
4	Professional Database Engineer - Optional YR4	2080.0000	HOUR	92.610000	192628.80

Comm Code	Manufacturer	Specification	Model #
80111609			

Commodity Line Comments:

Extended Description:

Fill out an hourly rate into the unit price field. The quantity amount is estimated for evaluation purposes only. The actual hours used may vary depending on usage.

ADDENDUM ACKNOWLEDGEMENT FORM
SOLICITATION NO.: CRFQ ISC24*011

Instructions: Please acknowledge receipt of all addenda issued with this solicitation by completing this addendum acknowledgment form. Check the box next to each addendum received and sign below. Failure to acknowledge addenda may result in bid disqualification.

Acknowledgment: I hereby acknowledge receipt of the following addenda and have made the necessary revisions to my proposal, plans and/or specification, etc.

Addendum Numbers Received:

(Check the box next to each addendum received)

- | | |
|--|--|
| <input checked="" type="checkbox"/> Addendum No. 1 | <input type="checkbox"/> Addendum No. 6 |
| <input type="checkbox"/> Addendum No. 2 | <input type="checkbox"/> Addendum No. 7 |
| <input type="checkbox"/> Addendum No. 3 | <input type="checkbox"/> Addendum No. 8 |
| <input type="checkbox"/> Addendum No. 4 | <input type="checkbox"/> Addendum No. 9 |
| <input type="checkbox"/> Addendum No. 5 | <input type="checkbox"/> Addendum No. 10 |

I understand that failure to confirm the receipt of addenda may be cause for rejection of this bid. I further understand that any verbal representation made or assumed to be made during any oral discussion held between Vendor's representatives and any state personnel is not binding. Only the information issued in writing and added to the specifications by an official addendum is binding.

Compu-Vision Consulting Inc.

Company

Authorized Signature

03/06/2024

Date



Department of Administration
 Purchasing Division
 2019 Washington Street East
 Post Office Box 50130
 Charleston, WV 25305-0130

State of West Virginia
 Centralized Request for Quote
 Service - Prof

Proc Folder: 1379271		Reason for Modification:	
Doc Description: Addendum No 1 -Professional Database Engineer (OT24116)		Addendum No 1 is issued to publish questions and answers.	
Proc Type: Central Master Agreement			
Date Issued	Solicitation Closes	Solicitation No	Version
2024-03-01	2024-03-06 13:30	CRFQ 0210 ISC2400000011	2

BID RECEIVING LOCATION

BID CLERK
 DEPARTMENT OF ADMINISTRATION
 PURCHASING DIVISION
 2019 WASHINGTON ST E
 CHARLESTON WV 25305
 US

VENDOR

Vendor Customer Code: VS0000029949
Vendor Name : Compu-Vision Consulting Inc.
Address :
Street : 2050 Route 27 Suite 202
City : North Brunswick
State : New Jersey **Country :** USA **Zip :** 08902
Principal Contact : Linda Lopes
Vendor Contact Phone: 732-422-1500 **Extension:** 212

FOR INFORMATION CONTACT THE BUYER

Toby L Welch
 (304) 558-8802
 toby.l.welch@wv.gov

Vendor Signature X **FEIN#** 23-2977235 **DATE** 03/06/2024

All offers subject to all terms and conditions contained in this solicitation

ADDITIONAL INFORMATION

Addendum No 1 is issued for the following reasons:

1) To publish a copy of vendor questions with the agency's response.

---no other changes---

INVOICE TO		SHIP TO	
DEPARTMENT OF ADMINISTRATION OFFICE OF TECHNOLOGY 1900 KANAWHA BLVD E, BLDG 5 10TH FLOOR CHARLESTON WV US		WV OFFICE OF TECHNOLOGY BLDG 5, 10TH FLOOR 1900 KANAWHA BLVD E CHARLESTON WV US	

Line	Comm Ln Desc	Qty	Unit Issue	Unit Price	Total Price
1	Professional Database Engineer - YR1	2080.00000	HOUR	\$ 84.75	\$ 176,280.00

Comm Code	Manufacturer	Specification	Model #
80111609			

Extended Description:

Fill out an hourly rate into the unit price field. The quantity amount is estimated for evaluation purposes only. The actual hours used may vary depending on usage.

INVOICE TO		SHIP TO	
DEPARTMENT OF ADMINISTRATION OFFICE OF TECHNOLOGY 1900 KANAWHA BLVD E, BLDG 5 10TH FLOOR CHARLESTON WV US		WV OFFICE OF TECHNOLOGY BLDG 5, 10TH FLOOR 1900 KANAWHA BLVD E CHARLESTON WV US	

Line	Comm Ln Desc	Qty	Unit Issue	Unit Price	Total Price
2	Professional Database Engineer - Optional YR2	2080.00000	HOUR	\$ 87.29	\$ 181,563.20

Comm Code	Manufacturer	Specification	Model #
80111609			

Extended Description:

Fill out an hourly rate into the unit price field. The quantity amount is estimated for evaluation purposes only. The actual hours used may vary depending on usage.

INVOICE TO		SHIP TO	
DEPARTMENT OF ADMINISTRATION OFFICE OF TECHNOLOGY 1900 KANAWHA BLVD E, BLDG 5 10TH FLOOR CHARLESTON WV US		WV OFFICE OF TECHNOLOGY BLDG 5, 10TH FLOOR 1900 KANAWHA BLVD E CHARLESTON WV US	

Line	Comm Ln Desc	Qty	Unit Issue	Unit Price	Total Price
3	Professional Database Engineer - Optional YR3	2080.00000	HOUR	\$ 89.91	\$ 187,012.80

Comm Code	Manufacturer	Specification	Model #
80111609			

Extended Description:

Fill out an hourly rate into the unit price field. The quantity amount is estimated for evaluation purposes only. The actual hours used may vary depending on usage.

INVOICE TO		SHIP TO	
DEPARTMENT OF ADMINISTRATION OFFICE OF TECHNOLOGY 1900 KANAWHA BLVD E, BLDG 5 10TH FLOOR CHARLESTON WV US		WV OFFICE OF TECHNOLOGY BLDG 5, 10TH FLOOR 1900 KANAWHA BLVD E CHARLESTON WV US	

Line	Comm Ln Desc	Qty	Unit Issue	Unit Price	Total Price
4	Professional Database Engineer - Optional YR4	2080.00000	HOUR	\$ 92.61	\$ 192,628.80

Comm Code	Manufacturer	Specification	Model #
80111609			

Extended Description:

Fill out an hourly rate into the unit price field. The quantity amount is estimated for evaluation purposes only. The actual hours used may vary depending on usage.

SCHEDULE OF EVENTS

<u>Line</u>	<u>Event</u>	<u>Event Date</u>
1	Questions are due by 3:00 p.m.	2024-02-28



March 06, 2024

Toby L Welch
2019 Washington Street, East
Charleston, WV 25305
Fax: (304) 558-3970
Email: Toby.L.Welch@wv.gov

Subject: Response to West Virginia Office of Technology for Solicitation #CRFQ-0210-ISC2400000011 Professional Database Engineer (OT24116)

Dear Toby,

Compu-Vision Consulting Inc. (CVC) is pleased to submit the proposal to the West Virginia Office of Technology for Solicitation #CRFQ-0210-ISC2400000011 Professional Database Engineer (OT24116).

CVC is a **State of West Virginia Central Master Agreement holder** with Vendor code # VS0000029949. CVC is a GSA Contract holder for contract# 47QTCA23D004L for SIN 54151S and 47QTCA23D004M for SIN 561320SBSA. We possess the expertise and capabilities to excel in the areas deemed necessary to support the WVOT.

CVC Proposed Candidate

CVC is proposing Divya Sri as a highly skilled and experienced candidate for this solicitation.

Divya Sri is a **Professional Database Engineer** with extensive experience spanning **over nine years in Information Technology**. Her expertise encompasses a wide array of technologies, including the Hadoop ecosystem, Spark, Kafka, HBase, MapReduce, Python, Scala, Pig, Impala, Sqoop, Oozie, Flume, Storm, and Big Data solutions using Lambda Architecture. She has a strong background in implementing Big Data solutions in both on-premise and cloud environments, including Amazon EMR and AWS services such as S3, Redshift, RDS, and Glue, as well as Azure Data Factory, Databricks, and Data Lake. Divya excels in designing, developing, and documenting data-related solutions, including ETL processes, data models, and analytics algorithms. She is proficient in Agile methodologies, experienced in various ETL and reporting tools and adept at working with different flavors of Hadoop.

Divya's skill set extends to database administration, performance tuning, CI/CD deployments, integrating Power BI reports, and utilizing additional technologies such as SQL Server 2008 R2, PowerShell, Microsoft Dynamics CRM, Microsoft Biztalk Server, Hyland OnBase, Google Workspace, IBM Data Studio, Ademero Content Central, Visual Studio, and Adobe ColdFusion. Her commitment to staying abreast of the latest technologies while delivering high-quality solutions and mentoring teams effectively underscores her versatility and proficiency in the field.

Divya is currently working as a **Senior Database Engineer for Amex, Phoenix (AZ)** since July 2023.

We welcome the opportunity to elaborate further on our proposal and we look forward to working with the team at WVOT to demonstrate our commitment to your success.

Sincerely

Linda Lopes | Director - Government Relations
732-422-1500 | rfp@compuvis.com

Resume – Database Engineer – Divya Sri

Job Title: Database Engineer
Professional Summary
<ul style="list-style-type: none"> ▪ Extensive years of experience in Information Technology which includes Hands on experience in Hadoop ecosystem including Spark, Kafka, HBase, MapReduce, Python, Scala, Pig, Impala, Sqoop, Oozie, Flume, Storm, big data technologies and worked on Spark SQL, Spark Streaming and using Core Spark API to explore Spark features to build data pipelines. ▪ Experience in building bigdata solutions using Lambda Architecture using Cloudera distribution of Hadoop, Twitter Storm, Trident, MapReduce, Cascading, HIVE, PIG and Sqoop. ▪ Strong Knowledge and experience on implementing Big Data in Amazon Elastic MapReduce (Amazon EMR) for processing, managing Hadoop framework dynamically scalable Amazon EC2 instances, Lambda, SNS, SQS, AWS Glue, S3, RDS and Redshift. ▪ end-to-end implementation project experience on data processing pipeline using Hadoop (CDP) ecosystem, Data Lake, and Data Warehouse. ▪ Strong experience with AWS Cloud on data integration with Databricks, Apache Spark, Airflow, EMR, Glue, Kafka, Kinesis and Lambda in S3, Redshift, RDS, MongoDB/DynamoDB ecosystems. ▪ Responsible for the design and development of data related solutions (databases, ETL systems, code, scripts, data models, reports, documentation) and supporting various products. Extensive hands-on experience developing database queries, ETL, performance tuning and database unit tests in a relational database environment. ▪ Having experience in developing a data pipeline using Kafka to store data into HDFS and Experience in building Real-time Data Pipelines with Kafka Connect and Spark Streaming and developing end to end data processing pipelines that begin with receiving data using distributed messaging systems Kafka for persisting data into Hive. ▪ Experienced in development methodologies like Agile/Scrum. ▪ Excellent Experience in Designing, Developing, Documenting, Testing of ETL jobs and mappings in Server and Parallel jobs using Data Stage to populate tables in Data Warehouse and Data marts. ▪ Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing Data Mining, Data Acquisition, Data Preparation, Data Manipulation, Feature Engineering, Machine Learning Algorithms, Validation and Visualization and reporting solutions that scales across massive volume of structured and unstructured Data. ▪ Strong knowledge in working with ETL methods for data extraction, transformation, and loading in corporate- wide ETL Solutions and Data Warehouse tools for reporting and data analysis ▪ Experience with different ETL tool environments like SSIS, Informatica, and reporting tool environments like SQL Server Reporting Services, and Business Objects. ▪ Having good experience on all flavors of Hadoop (Cloudera, Hortonworks, and MapR etc.) and Hands on experience in AVRO and Parquet file format, Dynamic Partitions, Bucketing for best Practice and Performance improvement. ▪ Expert in designing Server jobs using various types of stages like Sequential file, ODBC, Hashed file, Aggregator, Transformer, Sort, Link Partitioner and Link Collector. ▪ Proficiency in Big Data Practices and Technologies like HDFS, MapReduce, Hive, Pig, HBase, Sqoop, Oozie, Flume, Spark, Kafka. ▪ Develop data pipeline to synchronize and aggregate data from various data sources. Build the query data platform and provide guidelines for partner team to use aggregated data. Publish standard data sets & schema. ▪ Experience with source control systems such as Git, Bitbucket, and Jenkins. Working experience in CI/CD deployments ▪ Good knowledge in RDBMS concepts (Oracle 112c/1g, MS SQL Server 2012) and strong SQL, query writing skills (by using Erwin & SQL Developer tools), Stored Procedures and Triggers and Experience in writing Complex SQL Queries involving multiple tables inner and outer joins. ▪ Excellent knowledge with Unit Testing, Regression Testing, Integration Testing, User Acceptance Testing, Production implementation and Maintenance. ▪ Working knowledge of Azure cloud components (HDInsight, Databricks, Data Lake, Blob Storage, Data Factory, Storage Explorer, SQL DB, SQL DWH, Cosmos DB).

Job Title: Database Engineer	
<ul style="list-style-type: none"> ▪ Experienced in building data pipelines using Azure Data Factory, Azure Databricks, and loading data to Azure Data Lake, Azure SQL Database, Azure SQL Data Warehouse, and controlling database access. ▪ Extensive experience with Azure services like HDInsight, Stream Analytics, Active Directory, Blob Storage, Cosmos DB, and Storage Explorer. ▪ Expertise in MS SQL Server Database and Performance tuning experience, Batch and distributed computing using ETL/ELT (Spark /SQL Server DWH) ▪ Expertise in Database Administrators in Enterprise Solution Services to design, execute, and maintain Microsoft SQL Server, Microsoft Dynamics, and Hyland OnBase based infrastructure. ▪ The primary responsibility of this Engineer includes developing and ensuring the effective operation of databases within the environment. Additionally, will create applications and reports that utilize these databases as necessary. Database Engineer will offer assistance for the infrastructure and systems developed, along with providing documentation and training to current employees on their specifics. ▪ Integrate Power BI reports into other applications using embedded analytics like Power BI service (SaaS), or by API automation. Also, one must be experienced in developing custom visuals for Power BI. ▪ Experience in developing Map Reduce Programs using Apache Hadoop for analyzing the big data as per the requirement. Practical understanding of the Data modeling (Dimensional & Relational) concepts like Star-Schema Modeling, Snowflake Schema Modeling, Fact and Dimension tables. ▪ Ability to review technical deliverables, mentor and drive technical teams to deliver quality products. ▪ Demonstrated ability to communicate and gather requirements, partner with Enterprise Architects, Business Users, Analysts, and development teams to deliver rapid iterations of complex solutions. ▪ Flexible and versatile to adapt to any new environment with a strong desire to keep pace with latest technologies. 	
Education/Affiliations	
<ul style="list-style-type: none"> ▪ Bachelor's in Electrical and Electronics Engineering - JNTUK, 2014 	
Skills	
<ul style="list-style-type: none"> ▪ Hadoop, MapReduce, Spark, HDFS, Sqoop, YARN, Oozie, Hive, Impala, Zookeeper, Apache Flume, Apache Airflow, Cloudera, HBase 	<ul style="list-style-type: none"> ▪ Hadoop, MapReduce, Spark, HDFS, Sqoop, YARN, Oozie, Hive, Impala, Zookeeper, Apache Flume, Apache Airflow, Cloudera, HBase
<ul style="list-style-type: none"> ▪ Microsoft PowerShell 4.0 or greater, Microsoft Dynamics CRM 2013, and Dynamics 365, Microsoft Dynamics GP 2018, Microsoft Biztalk Server, Visual Studio 2019 	<ul style="list-style-type: none"> ▪ Microsoft PowerShell 4.0 or greater, Microsoft Dynamics CRM 2013, and Dynamics 365, Microsoft Dynamics GP 2018, Microsoft Biztalk Server, Visual Studio 2019
<ul style="list-style-type: none"> ▪ C#.NET, ASP .NET, Entity Framework, ADO.NET 	<ul style="list-style-type: none"> ▪ C#.NET, ASP .NET, Entity Framework, ADO.NET
Roles and Responsibilities / Chronology of Professional Experience	
<p>Sr. Data Engineer Amex, Phoenix, Arizona</p>	<p>Jul 2023– Present</p>
<ul style="list-style-type: none"> ▪ Developed custom multi-threaded Java based ingestion jobs as well as Sqoop jobs for ingesting from FTP servers and data warehouses. ▪ Developed Scala based Spark applications for performing data cleansing, event enrichment, data aggregation, de-normalization and data preparation needed for machine learning and reporting teams to consume. ▪ Worked on troubleshooting spark application to make them more error tolerant. ▪ Experience in integrating Jenkins with various tools like Maven (Build tool), Git (Repository), SonarQube (code verification), Nexus (Artifactory) and implementing CI/CD automation for creating Jenkins pipelines programmatically architecting Jenkins Clusters, and scheduled builds day and 	

Job Title: Database Engineer

overnight to support development needs.

- Programmatically created CICD Pipelines in Jenkins using Groovy scripts, Jenkins file, integrating a variety of Enterprise tools and Testing Frameworks into Jenkins for fully automated pipelines to move code from Dev Workstations to all the way to Prod environment.
- Working on Docker Hub, Docker Swarm, Docker Container network, creating Image files primarily for middleware installations & domain configurations. Evaluated Kubernetes for Docker Container Orchestration.
- Experience on AWS cloud services such as EC2, S3, RDS, ELB, EBS, VPC, Route53, auto scaling groups, Cloud watch, Cloud Front, IAM to build configuration and troubleshooting for server migration from physical to cloud on various Amazon photos.
- Developed rest API's using python with flask and Django framework and done the integration of various data sources including Java, JDBC, RDBMS, Shell Scripting, Spreadsheets, and Text files.
- Analyzed SQL scripts and designed the solutions to implement using PySpark.
- Wrote Kafka producers to stream the data from external rest API to Kafka topics.
- Wrote Spark-Streaming applications to consume the data from Kafka topics and write the processed streams to HBase.
- Experienced in handling large datasets using Spark in Memory capabilities, using broadcasts variables in Spark, effective & efficient joins, transformations, and other capabilities.
- Architect & implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics).
- Worked extensively with Sqoop for importing data from Oracle.
- Batch scripts have been created to retrieve data from AWS S3 storage and to make appropriate transformations in Scala using the Spark framework.
- Involved in creating Hive tables, loading, and analysing data using hive scripts. Implemented Partitioning, Dynamic Partitions, Buckets in HIVE.
- Developed a python script to transfer data, REST API's and extract data from on-premises to AWS S3. Implemented Micro Services based Cloud Architecture using Spring Boot.
- Good experience with continuous Integration of application using Bamboo.
- Designed, documented operational problems by following standards and procedures using JIRA.
- Wrote Pig Scripts to generate Map Reduce jobs and performed ETL procedures on the data in HDFS.
- Developed Oozie work processes for planning and arranging the ETL cycle. Associated with composing Python scripts to computerize the way towards extricating weblogs utilizing Airflow DAGs.
- Using python, the ETL pipeline was developed and programmed to collect data from Redshift data warehouse.
- Used MongoDB to stored data in JSON format and developed and tested many features of dashboard using Python, Bootstrap, CSS, and JavaScript.
- Created workflows, mappings using Informatica ETL and worked with different transformations such as lookup, source qualifier, update strategy, router, sequence generator, aggregator, rank, stored procedure, filter, joiner, sorter.
- Worked on SSIS creating all the interfaces between front end application and SQL Server database, then from legacy database to SQL Server Database and vice versa.
- Good hands-on participation in the development and modification of SQL stored procedure techniques, functions, views, indexes, and triggers.
- Migrate data into RV Data Pipeline using DataBricks, Spark SQL and Scala.
- Used Databricks for encrypting data using server-side encryption.
- Used Delta Lake as it is an open-source data storage layer which delivers reliability to data lakes.
- Experience with Snowflake Virtual Warehouses.
- Responsible for ingesting large volumes of IOT data to Kafka.
- Wrote Kafka producers to stream the data from external rest APIs to Kafka topics.

Environment: AWS, Azure, Agile, Jenkins, EMR, Spark, Hive, S3, Athena, Sqoop, Kafka, HBase, Redshift, ETL, Pig, Oozie, Spark Streaming, Docker, Kubernetes, Hue, Scala, Python, Apache NIFI, GIT, Micro Services, Snowflakes.

Data Engineer
Centene Corporation, Tampa, Florida

Aug 2021 to Jul 2023

Job Title: Database Engineer

- Involved in full Software Development Life Cycle (SDLC) - Business Requirements Analysis, preparation of Technical Design documents, Data Analysis, Logical and Physical database design, Coding, Testing, Implementing, and deploying to business users.
- Developed complex mappings using Informatica Power Center Designer to transform and load the data from various source systems like Oracle, Teradata, and Sybase into the final target database.
- Analyzed source data coming different sources like SQL Server tables, XML files and Flat files then transformed according to business rules using Informatica and loaded the data in to target tables.
- Designed and developed a number of complex mappings using various transformations like Source Qualifier, Aggregator, Router, Joiner, Union, Expression, Lookup, Filter, Update Strategy, Stored Procedure, Sequence Generator, etc.
- Involved in creating the Tables in Greenplum and loading the data through Alteryx for Global Audit Tracker.
- Analyzed large and critical datasets using HDFS, HBase, Hive, HQL, PIG, Sqoop and Zookeeper.
- Data Extraction, aggregations, and consolidation of Adobe data within AWS Glue using PySpark.
- Developed Python scripts to automate the ETL process using Apache Airflow and CRON scripts in the UNIX operating system as well.
- Worked on Google Cloud Platform (GCP) services like compute engine, cloud load balancing, cloud storage and cloud SQL.
- Developed data engineering and ETL python scripts for ingestion pipelines which run on AWS infrastructure setup of EMR, S3, Glue and Lambda.
- Changing the existing Data Models using Erwin for Enhancements to the existing Datawarehouse projects.
- Used Talend connectors integrated to Redshift - BI Development for multiple technical projects running in parallel.
- Performed Query Optimization with the help of explain plans, collect statistics, Primary and Secondary indexes. Used volatile table and derived queries for breaking up complex queries into simpler queries. Streamlined the scripts and shell scripts migration process on the UNIX box.
- Using g-cloud function with Python to load Data in to Big query for on arrival csv files in GCS bucket.
- Created iterative macro in Alteryx to send Json request and download Json response from webservice and analyze the response data.
- Migrated data from Transactional source systems to Redshift data warehouse using spark and AWS EMR.
- Experience in Google Cloud components, Google container builders and GCP client libraries.
- Supported various business teams with Data Mining and Reporting by writing complex SQLs using Basic and Advanced SQL including OLAP functions like Ranking, partitioning and windowing functions, Etc.
- Expertise in writing scripts for Data Extraction, Transformation and Loading of data from legacy systems to target data warehouse using BTEQ, FastLoad, MultiLoad, and T pump.
- Worked with EMR, S3 and EC2 services in AWS cloud and Migrating servers, databases, and applications from on premise to AWS.
- Tuning SQL queries using Explain analyzing the data distribution among AMPs and index usage, collect statistics, definition of indexes, revision of correlated sub queries, usage of Hash functions, etc...
- Developed shell scripts for job automation, which will generate the log file for every job.
- Extensively used spark SQL and Data frames API in building spark applications.
- Written complex SQLs using joins, sub queries and correlated sub queries. Expertise in SQL Queries for cross verification of data.
- Extensively worked on performance tuning of Informatica and IDQ mappings.
- Creating, maintain, support, repair, customizing System & Splunk applications, search queries and dashboards.
- Experience on data profiling & various data quality rules development using Informatica Data Quality (IDQ).
- Create new UNIX scripts to automate and to handle different file processing, editing and execution sequences with shell scripting by using basic Unix commands and 'awk', 'sed' editing languages.
- Experience in cloud versioning technologies like GitHub.

Job Title: Database Engineer

- Integrate Collibra with Data Lake using Collibra connect API.
- Design and Develop ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift.
- Create firewall rules to access Google Data proc from other machines.
- Write Scala program for spark transformation in Dataproc.
- Providing technical support and guidance to the offshore team to address complex business problems.

Environment: Informatica Power Center 9.5, AWS Glue, Talend, Google Cloud Platform(GCP), PostgreSQL Server, Python, Oracle, Teradata, CRON, Unix Shel Scripting, SQL, Erwin, AWS Redshift, GitHub, EMR

Big Data/Hadoop Developer

CITI Bank, Charlotte, NC

Mar 2018 to Aug 2021

- Involved in analyzing business requirements and prepared detailed specifications that follow project guidelines required for project development.
- Involved in Data Ingestion using Sqoop/Flume
- Worked with Flume in bringing click stream data from front facing application logs.
- Used Kafka functionalities like distribution, partition, replicated commit log service for messaging systems by maintaining feeds.
- Involved in writing Spark applications using Scala/Java.
- Experience in using Kafka as a messaging system to implement real-time Streaming solutions using Spark Streaming
- Developed SQOOP scripts to migrate data from Oracle to Big data Environment.
- Involved in Analysis, Design, and Implementation/translation of Business User requirements.
- Worked on collection of large sets using Python scripting, PySpark, Spark SQL
- Worked on large sets of Structured and Unstructured data.
- Actively involved in designing and developing data ingestion, aggregation, and integration in the Hadoop environment.
- Developed Sqoop scripts to import export data from relational sources and handled incremental loading on the customer, transaction data by date.
- Extensively worked with Avro and Parquet files and converted the data from either format Parsed Semi Structured JSON data and converted to Parquet using Data Frames in Spark.
- Performed data analysis and data profiling using complex SQL queries on various source systems including Oracle 10g/11g and SQL Server 2012.
- Identified inconsistencies in data collected from different sources.
- Participated in requirement gathering and worked closely with the architect in designing and modeling.
- Designed object model, data model, tables, constraints, necessary stored procedures, functions, triggers, and packages for Oracle Database.
- Wrote Spark applications for Data validation, cleansing, transformations and custom aggregations.
- Imported data from various sources into Spark RDD for processing.
- Developed custom aggregate functions using Spark SQL and performed interactive querying.
- Worked on installing cluster, commissioning & decommissioning of Data node, Name node high availability, capacity planning, and slots configuration.
- Developed Spark applications for the entire batch processing by using PySpark.
- Worked on Spark Data sources, Spark Data Frames, Spark SQL and Streaming using Scala.
- Good knowledge in creating data frames using Spark SQL.
- Migrated Hive QL queries on structured into Spark SQL to improve performance.
- Developed pyspark/Spark SQL scripts to analyze various customer behaviors.
- Extensively worked with Pyspark/Spark SQL for data cleansing and generating Data Frames and RDDs.
- Developed Spark code using python for pyspark, Scala and Spark-SQL for faster testing and processing of data.
- Used Spark API over Hortonworks Hadoop YARN to perform analytics on data in Hive.
- Implemented existing Hive script in Spark Scala for better performance.

Job Title: Database Engineer

- Explored with Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, Spark Streaming, Data Frame, pair RDD's, Spark YARN.
- Used Python for data validation and analysis purposes.
- Involved in creating Hive tables with Dynamic Partitions and Buckets, loading various formats of data like Avro, Parquet into the tables and analyzed data using HiveQL.
- Created Hive, HBase tables and HBase integrated Hive tables as per the design using ORC file format and Snappy compression.
- Performed Hive Query Optimization for better performance.
- Built reusable Hive UDF libraries for business requirements which enabled users to use these UDFs in Hive Querying.
- Involved in loading data into Cassandra NoSQL Database.
- Worked hands on with ETL processes. Handled imported data from various data sources, performed transformations.
- Visualized the results using Tableau dashboards and the Python Seaborn libraries were used for Data interpretation in deployment.
- Experienced in determining trends and significant data relationships Analyzing using advanced Statistical Methods.
- Took personal responsibility for meeting deadlines and delivering high quality work and create POCs to demonstrate new technologies including Jupyter Notebooks, PySpark.
- Strived to continually improve existing methodologies, processes, and deliverable templates.
- Worked with business owners/stakeholders to assess Risk impact, provided solutions to business owners.

Environment: Hortonworks, Hadoop, Big Data, HDFS, MapReduce, Sqoop, Oozie, NiFi, Python, SQL server, Oracle, HBase, Hive, Impala, Pig, Sqoop, Tableau, NoSQL, Unix/Linux, Spark, PySpark, Notebooks. Google Workspace, Google Sheets, Google App Sheets.

SQL Server Developer

Global Edge Software, India

Apr 2015 to Dec 2017

- Coordinated with front-end application developers for implementing database architecture and design.
- Used various transformations in SSIS dataflow, control flow using for loop containers, and fuzzy lookups.
- Develop parameterized reports, caching reports, sub reports and Ad Hoc reports using SSRS.
- Implemented error handling and utilized event handlers for automated notifications using SSIS.
- Write Complex SQL Queries to generate Reports based on the business requirement.
- Redesigned the SSIS packages from the legacy DTS packages.
- Execute SSIS package include a master package which include number of child packages.
- Supporting ETL (Extract Transform and Load) for fetching data from multiple systems to single Data Warehouse.
- Created complex Ad-Hoc reports, Sub reports, linked reports related to State compliance reporting. Used custom code in SSRS for row color, visibility, and masking.
- Developed Full Analysis Cycle Project and created packages for extracting data from OLTP to OLAP. Created Multi-Dimensional Expression (MDX) scripts for OLAP data cubes.
- Involved in Designing, Developing and Testing of the ETL (Extract, Transformation and Load) strategy to populate the data from Heterogeneous data sources (SQL Server, Flat Files, Excel source files, XML files etc.)
- Performed different kinds of transformations like Lookup Transformations, Merge Joins, Derived Columns, Merge Join, Conditional Split, and Data Conversion with Multiple Data Flow tasks.
- Bulk data migration using Bulk Insert from flat files.
- Created Package Configurations, Event Handlers for On Error, On Pre/Post execution. Designed Complex Packages with Error Handling and Package Logging that stores the Logging results in SQL Tables and log files
- Performance tuning and testing on stored procedures, indexes using performance tool like SQL Profiler and DETA.
- Import/Export and successfully migrated the Server usage MS Access database data to SQL Server

Job Title: Database Engineer

using ETL/SSIS. Created dynamic and customized packages to support future changes. Scheduled the same packages by creating the corresponding job tasks.

- Formatted the reports using the global variables and expressions in SSRS, deployed the generated reports on to the Reporting server.

Environment: MS SQL Server 2022, MS SQL Reporting Server 2022, Visual Studio 2019, T-SQL, MS Excel, Microsoft SQL Server Integration Services (SSIS), Microsoft SQL Server Reporting Services (SSRS), Rally bug tracking and SVN, MS PowerShell 4.0, Microsoft Dynamics CRM 2013, and Dynamics 365, Microsoft Dynamics GP 2018.

Data Analyst

Avantel Softech Limited, India

May 2014 to Apr2015

- Coordinated with front-end application developers for implementing database architecture and design.
- Used various transformations in SSIS dataflow, control flow using for loop containers, and fuzzy lookups.
- Develop parameterized reports, caching reports, sub reports and Ad Hoc reports using SSRS.
- Implemented error handling and utilized event handlers for automated notifications using SSIS.
- Write Complex SQL Queries to generate Reports based on the business requirement.
- Redesigned the SSIS packages from the legacy DTS packages.
- Execute SSIS package include a master package which include number of child packages.
- Supporting ETL (Extract Transform and Load) for fetching data from multiple systems to single Data Warehouse.
- Created complex Ad-Hoc reports, Sub reports, linked reports related to State compliance reporting. Used custom code in SSRS for row color, visibility, and masking.
- Developed Full Analysis Cycle Project and created packages for extracting data from OLTP to OLAP. Created Multi-Dimensional Expression (MDX) scripts for OLAP data cubes.
- Involved in Designing, Developing and Testing of the ETL (Extract, Transformation and Load) strategy to populate the data from Heterogeneous data sources (SQL Server, Flat Files, Excel source files, XML files etc.)
- Performed different kinds of transformations like Lookup Transformations, Merge Joins, Derived Columns, Merge Join, Conditional Split, and Data Conversion with Multiple Data Flow tasks.
- Bulk data migration using Bulk Insert from flat files.
- Created Package Configurations, Event Handlers for On Error, On Pre/Post execution. Designed Complex Packages with Error Handling and Package Logging that stores the Logging results in SQL Tables and log files
- Performance tuning and testing on stored procedures, indexes using performance tool like SQL Profiler and DETA.
- Import/Export and successfully migrated the Server usage MS Access database data to SQL Server using ETL/SSIS. Created dynamic and customized packages to support future changes. Scheduled the same packages by creating the corresponding job tasks.
- Formatted the reports using the global variables and expressions in SSRS, deployed the generated reports on to the Reporting server.

Environment: MS SQL Server 2022, MS SQL Reporting Server 2022, Visual Studio 2019, T-SQL, MS Excel, Microsoft SQL Server Integration Services (SSIS), Microsoft SQL Server Reporting Services (SSRS), Rally bug tracking and SVN, MS PowerShell 4.0, Microsoft Dynamics CRM 2013, and Dynamics 365, Microsoft Dynamics GP 2018.