**Naz Devi**

**Cell: 415-787-6954**

**San Francisco, CA 94012**

**Email:** [**nazdevijobs@gmail.com**](mailto:nazdevijobs@gmail.com)

**Professional Summary:**

* Fifteen (15) years of experience working in Data Architect/Engineer, AWS, S3, Data Lake, Azure, Cloud Solutions, Redshift, ETL/Pentaho, Informatica, SSIS Data Integration development in the areas of design, development, analysis, implementation and support of data warehousing tools ETL, Pentaho, Informatica, Big Data Hadoop/Spark, HIVE, MS SQL Server, Dashboard, Tableau, SQL server, Oracle, UNIX/Linux, PostgreSQL on client/server
* Developed in Pentaho, Informatica, Talend for Big Data projects, Dashboard Designs, and Data Visualizations. Utilized with Pentaho(PDI) Kettle, Informatica IDQ, PowerCenter, MDM, data management, data integration/quality, and data governance
* Managed Error Handling, Performance Tuning, Error Logging, clustering and High Availability
* Performed on Pentaho, Informatica, Talend Data Fabric ETL, Data warehouse concepts, Kimball, Star & Snowflake Schema, Fact/Dimension tables. Worked on Pentaho, Informatica, Talend/Informatica Integration Cloud Services, IaaS, PaaS, SaaS.
* Developed complex ETL jobs from various sources such as SQL server, PostgreSQL and other files and loaded into target databases using Pentaho, Informatica, Talend OS ETL tool. Created Big Data Hadoop/Talend Dashboard
* Created ETL/Pentaho, Informatica, Talend jobs both design and code to process data to target databases. Used MS SQL Server, Pentaho, Informatica, Talend Admin Console Job conductor to schedule ETL Jobs on daily, weekly, monthly and yearly basis
* Generated Data Quality Dashboard, KPI Reports ETL, SSRS, Dashboard Data Visualizations, SQL, Databases, Data Management, Data Warehousing, Data Governance concepts. Worked with Hadoop, Talend to architect, design and built Big Data solutions using Hive Hadoop, created dashboards, Data Visualizations
* Developed in Pentaho, Informatica, and Talend for Dashboards, Data Visualizations, Big Data Spark projects working on Tableau Engineer / Pentaho/ BI/Reporting platform to enable data access, analytic models, and data visualization.

**Technical Skills:**

MS SQL Server Database Administrator/Developer, BI, ETL, Pentaho, Informatica developer IDQ 9.1/9.5.1, Informatica Power Mart 6.x/5.x, Power Exchange, SQL Server Configuration, Replication, Virtualization.Visual Studio, C#, .Net, C++, VB, SDLC, Healthcare, HIPAA, PHI, SharePoint, PowerShell scripts, REST API, Open Data Protocol MS EXCEL, Access, Visio, Oracle, Java, Cogito, Star, Radar, Workbench, REST API, Open Data Protocol, AWS, Redshift, Data Lake, Azure, S3

JavaScript. SMS, SSRS, SSAS, SAAS, SSIS, ETL, Crystal Reports. NoSQL, PLSQL, MySQL, T-SQL, SQL queries, Transact-SQL, SQL Server architecture. Data Science, Data Analysis, data mining, Data Warehouse, Business Intelligence, Statistical analysis, concatenations, pivot tables, Table partitioning and archiving, Data cubes, Data marts, IaaS, PaaS, SaaS, Cloud Computing, Big Data. Optimize Stored Procedures, Indexing, Consistency Checks, performance tuning; SQL Server log shipping, SQL replication, scripting, Fine tune database, Function and trigger design and coding, Index implementation and maintenance, Clustering, Indexing. Random Forest, Machine Learning, Python, PowerPivot, PowerView Matlab/R, Ruby, Ruby on Rails, Agile, Waterfall, E-Commerce, Hadoop, SSAS OLAP cubes, Pentaho, Hadoop, PIG, Hive, Spark, Oracle. Tableau Architect, PHP, SQL Server Integration and Analytics Services, Data cubes, Data Science, Data Analysis, Data Warehouse Architect, mapping. E-Commerce, Hadoop, Big Data, MapReduce, Allscripts, R, HBase, Data modeling, HR Analytics, Data Integration architecture. OLTP, OLAP, database design, performance tuning and security model implementations.BI and analytic tools, Business Objects, QlikView, Tableau, COGNOS. Agile, Waterfall, Scrum development methodology, Web Services, Hyperion, OBIEE, Informatica. Healthcare, HIPPA, X12 EDI, Healthcare 835/837 Formats, PHI, HEDIS, Cerner, Med Epic Tapestry, Epic Cashe, FACETS, Epic Clarity, HIM, Meditech, TriZetto Reporting, Eclipsys, Allscripts, Cerner, Siemens and McKesson EMR, Epic systems, Epic Beaker/Labs, HIPPA, EDI, Revenue Cycle, HEDIS, SOX, Compliance

**Education:**

* **Master of Science**: Computer Sciences and Engineering, University of California, Los Angeles, CA
* **Bachelor of Science**: Computer Engineering & Management Information Systems, University of California, Los Angeles, CA

**Professional Work Experience:**

**L.A. Hospital/ LAC+USC Medical Center/ Data Tech Solutions, New York & CA**

**Data Engineer/ETL Developer**

**01/2017 – 01/2018**

**Hands on Tools:** MS SQL Server 2012/2014, Hadoop, Spark, Hive, Scala, Pentaho 6.1.0, IDQ 9.1/9.5.1Talend 6.x-6.4, Informatica MDM 9.X, SQL Server 2016, Hadoop 2.7.4, Hive, Spark 1.6.0

**Project Environment:** AWS, Redshift, ETL/Pentaho/Informatica/Talend, MS SQL Server, SSIS, MDM, Cassandra, Cloudera Hadoop, OLTP, SQL, Spark, Hadoop, SSAS OLAP Cubes, Hadoop, PIG, Hive, Spark, Oracle, PHI, Healthcare Data

**Contribution:**

* Worked as a Data Engineer to help form a cloud based big data engineering team to deliver platform automation and security.
* I possess excellent communication, verbal, written, presentation, strong interpersonal skills and willing train other team members.
* Built data workflows by using AWS EMR, Data Lake, Redshift, Hadoop, Spark, Spark SQL, Scala, and Python
* Designed, reviewed and fixed security vulnerabilities at network/subnet/security groups level
* Created security standardized templates including password management strategy and implementation
* Installed custom software and automated installation process
* Optimized Redshift database for optimal performance
* Expert knowledge of AWS EMR, Spark, Scala, Hadoop, HortonWorks, S3, RedShift
* Expert level knowledge of Linux/Unix, PowerShell, network security architecture and database tuning
* Strong communication and teaching skill to train teams with proven problem solving and critical thinking skills
* Used Pentaho/PDI, Kettle, Informatica, SSIS, Batch Data Analysis using Hive, SQL SSAS
* Worked with Informatica Cloud Data Integration to deliver accessible, trusted, and secured data to facilitate more valuable business decisions to identify competitive advantages to better service customers, and build an empowered workforce.
* Used Informatica Cloud Data Integration for global, distributed data warehouse and analytics projects.
* Worked with cloud data warehouses AWS Redshift, Azure SQL Data Warehouse, and Snowflake and Informatica Cloud Data Integration solutions to augment the performance, productivity, and extensive connectivity to cloud and on-premises sources.
* Worked with Informatica Cloud since it’s flexible and scalable transformations and advanced capabilities, to seamlessly integrate
* growing data volumes across disparate sources in the data warehouse using wizards, preconfigured templates, and out-of-the-box
* Worked with Healthcare Data/Claims 835 and 837 formats for analytical purposes, X12 Enterprise Data Integration(EDI), PHI
* Worked with X12 EDI standards for healthcare data, HEDIS, and HIPAA
* Worked with Big Data, Hadoop, Hive, Pig, Sqoop, Pentaho, Informatica
* Established, maintained, and enforced ETL architecture design principles, techniques, standards, and best practices
* Managed technical designs of ETL reference architectures to ensure high data quality, data integration performance, and error recovery/handling, optimize performance
* Reviewed/assessed existing ETL applications to update features, performance improvements, upgrades, and ongoing sustainment
* Conducted design reviews, code reviews, and performance tuning minimizing bottlenecks, maximizing performance
* Researched and recommended future improvements in ETL, Pentaho, and Informatica, and daily functions
* Worked with team to develop in-house knowledge repository for best practices, solution documentations, manuals, and procedures for user education
* Used Pentaho, Informatica, ETL, SSIS for data management, data integration, data quality, MDM, data governance
* Worked to architect, design and built Big Data solutions using Hive Hadoop for analysis and to build dashboards
* Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Used Hadoop, Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Loaded the data into Spark RDD and do in memory data Computation to generate the Output response.
* Worked with Hadoop, SSIS, SSAS OLAP cubes, Pentaho, Hadoop, PIG, Hive, Spark, Oracle, and MS SQL Server
* Implemented custom Error handling in ETL Pentaho Informatica jobs and worked on different methods of logging
* Hands-on data integration, data management/data warehousing experience, Pentaho, Informatica PowerCenter/IDQ development with direct development experience Informatica Business Glossary, Informatica Versions 10.x/9.1/8.x/7.x Source Analyzer, Mapping Designer, Mapplet, Transformations, Workflow Monitor, Workflow Manager, Informatica developer IDQ 9.1/9.5.1, Informatica Power Mart 6.x/5.x, Power Exchange, Power Connect, Power Analyzer Data Profiling and Data cleaning, Flat file system Fixed width, Delimiter , OLAP.
* Worked with Pentaho, Informatica MDM 9.X or 10.X, MDM Hub & IDD
* Performed Informatica Data Profiling with IDQ and Analyzer
* Performed analysis to identify data anomalies, data cleansing (ETL) and resolved data quality issues
* Developed IDQ using Joiner to configure & Develop Business Rules
* Performed data integration using Pentaho, Informatica, cross system joins for identifying duplicates and data anomalies
* Created IDQ Dashboards/KPI Metrics for reporting. Performed ETL and resolved data quality issues with analysis
* Extensively worked on Informatica IDE/IDQ. Involved in data profiling using IDQ (Analyst Tool) prior to data staging.
* Used Pentaho, Informatica IDQ's standardized plans for addresses and names clean ups.
* Worked on IDQ file configuration at user's machines and resolved the issues.
* Used IDQ to complete initial data profiling and removing duplicate data. And extensively worked on IDQ admin tasks IDQ Admin and IDQ developer.
* Implemented Informatica Business Glossary, Informatica Data Quality, MDM, PowerCenter, Informatica Analyst/Metadata Manager for glossary
* Performed on Pentaho ETL Data Conversion and Data Transformation, SF Data Modeling, Jitterbit.
* Worked in Product Development (PDP) Data Governance Office (DGO), Business Glossary.
* Responsible for developing Informatica Business Glossary solution based on functional and technical design specifications for business designs and technical requirements and develop a catalog structure in Informatica.
* Used BI and analytic tools, Business Objects, QlikView, Tableau, COGNOS, MicroStrategy, Netezza Agile, Waterfall, Scrum development methodology, Web Services, Hyperion, OBIEE, Informatica IDQ/PowerCenter
* Provided technical leadership and governance of the big data team and the implementation of solution architecture
* Managed the architecture dashboard design changes due to business requirements and other interface integration changes
* Provided an overall architect responsibilities including roadmaps, leadership, planning, technical innovation, security
* Designed, Layout, and Deployed Hadoop clusters in the cloud using Hadoop ecosystem & open Source platforms
* Configured and tuned production and development Hadoop environments with the various intermixing Hadoop components
* Provided End-to-end systems implementation such as data security and privacy concerns
* Designed and implemented geospatial big data ingestion, processing and delivery

**Wal-Mart Stores, GA, MI, NY, MA, WA, CA**

**Data Integration Pentaho/Big Data Hive Hadoop Developer/ETL BI DW Dashboard Developer**

**06/2016 – 01/2017**

**Hands on Tools:** Redshift, AWS, Data Lake, Azure, Spark, Scala, ETL/Informatica, Pentho, Talend 6.x-6.4, Hadoop 2.6.5, SQL Server 2014/2016, Oracle 11.2.0.4,Spark 1.6

**Project Environment:** ETL, Pig, Hive, HDFS, MapReduce, Sqoop, Storm, Spark, Kafka, Yarn, Oozie, Zookeeper, SQL, Hadoop, HiveQL, HQL.

**Contribution:**

* Performed data manipulations using various Pentaho/ Talend components like tMap, tJavarow, tjava
* Worked as a Data Engineer to help form a cloud based big data engineering team to deliver platform automation and security.
* Built data workflows by using AWS EMR, Spark, Spark SQL, Scala, and Python
* Configured and installed tools with a highly available architecture
* Designed, reviewed and fixed security vulnerabilities at network/subnet/security groups level
* Created security standardized templates including password management strategy and implementation
* Installed custom software and automated installation process
* Optimized Redshift database for optimal performance
* Expert knowledge of AWS EMR, Spark, Scala, Hadoop, HortonWorks, S3, RedShift
* Expert level knowledge of Linux/Unix, PowerShell, network security architecture and database tuning
* Strong communication and teaching skill to train teams with proven problem solving and critical thinking skills
* Worked as a Data Engineer to help form a cloud based big data engineering team to deliver platform automation and security.
* Built data workflows by using AWS EMR, Spark, Spark SQL, Scala, and Python
* Configured and installed tools with a highly available architecture
* Designed, reviewed and fixed security vulnerabilities at network/subnet/security groups level
* Created security standardized templates including password management strategy and implementation
* Installed custom software and automated installation process
* Optimized Redshift database for optimal performance
* Expert knowledge of AWS EMR, Spark, Scala, Hadoop, HortonWorks, S3, RedShift
* Expert level knowledge of Linux/Unix, PowerShell, network security architecture and database tuning
* Strong communication and teaching skill to train teams with proven problem solving and critical thinking skills
* Worked on dashboard designs, architect, design and built Big Data solutions using Hive Hadoop.
* Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Worked with Informatica Cloud Data Integration to deliver accessible, trusted, and secured data to facilitate more valuable business decisions to identify competitive advantages to better service customers, and build an empowered workforce.
* Used Informatica Cloud Data Integration for global, distributed data warehouse and analytics projects.
* Worked with cloud data warehouses AWS Redshift, Azure SQL Data Warehouse, and Snowflake and Informatica Cloud Data Integration solutions to augment the performance, productivity, and extensive connectivity to cloud and on-premises sources.
* Worked with Informatica Cloud since it’s flexible and scalable transformations and advanced capabilities, to seamlessly integrate growing data volumes across disparate sources in the data warehouse using wizards, preconfigured templates, and out-of-the-box mappings
* Configured deployed and maintained multi-node Dev and Test Kafka Clusters.
* Developed Spark scripts by using Scala shell commands as per the requirement.
* Used Pentaho, Talend, Informatica, Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Worked on migrating Map Reduce programs into Spark transformations using Spark and Scala.
* Designed and developed POCs in Spark using Scala to compare the performance of Spark with Hive and SQL/Oracle.
* Worked with Data Modelling, PHP and JSon, writing SQL queries tOracleRow, tOracleInput, tOracleOutput, and tMSSQLInput
* Analyzed source data for quality data by using Talend Data Quality.
* Experienced using Pentaho, Hadoop, Talend Integration Suite and Talend Open Studio Troubleshoot data integration issues and bugs, created dashboards, analyze reasons for failure, implement optimal solutions, and revise procedures and documentation as needed. Informatics PowerCenter, Business Glossary,
* Performed Informatica Data Profiling with IDQ and Analyzer
* Generated Data Quality Dashboard, KPI Reports ETL, Dashboard Data Visualizations, SQL, Databases, Data Management, Data Warehousing, Data Governance concepts.
* Performed Informatica Data Profiling with IDQ and Analyzer
* Generated Data Quality Dashboard, KPI Reports
* Performed analysis to identify data anomalies, data cleansing(ETL) and resolved data quality issues
* Developed IDQ using Joiner to configure & Develop Business Rules
* Performed cross system joins for identifying duplicates and data anomalies
* Created IDQ Dashboards/KPI Metrics for reporting
* Worked on Hadoop, Talend, Dashboard projects with client/provider data, built new dimensional Data Warehouse
* Utilized BI tools, Cognos, Informatica PIM, Business Objects, MicroStrategy, Netezza for various projects
* Worked with Hadoop, Talend dashboard projects, Data Modeling, Data Extraction, Data Migration, Data warehousing and report generation using Informatica, Cognos and MicroStrategy
* Worked with client and provider analytics, developed new data marts for new and existing data warehouses.
* Worked with Informatica PowerCenter, Oracle, Dimensional Data Modeling, Healthcare/Payor Data
* Worked with ETL SSIS, ERWIN, PLSQL, Informatica PowerCenter, Data Analyst, EPIC, Facets, Informatica MDM, Informatica IDD, TOAD, salesforce

**GAP Retail Stores, San Francisco, CA, Dallas, TX, Seattle, WA**

**Pentaho/Informatica/Talend Spark/Hive Hadoop Developer**

**February 2016-June 2016**

**Hands on Tools:** Pentaho, Informatica, Talend 5.6.3, Spark 1.6, SQL Server 2014, Hadoop 2.6.5, IDQ 9.6.1.

**Project Environment:** Spark API, Cloudera Hadoop YARN, Spark 1.6, Data Aggregation, Data frames, SQL

**Contribution:**

* Worked with Pentaho, ETL, SSIS, Talend Open Studio &Talend Enterprise platform for data management
* Managed Error Handling, Performance Tuning, Error Logging, clustering and High Availability in Talend
* Worked with Informatica Cloud Data Integration to deliver accessible, trusted, and secured data to facilitate more valuable business decisions to identify competitive advantages to better service customers, and build an empowered workforce.
* Worked as a Data Engineer to help form a cloud based big data engineering team to deliver platform automation and security.
* I possess excellent communication, verbal, written and presentation skills, and strong interpersonal skills and am willing train other team members.
* Built data workflows by using AWS EMR, Spark, Spark SQL, Scala, and Python
* Configured and installed tools with a highly available architecture
* Designed, reviewed and fixed security vulnerabilities at network/subnet/security groups level
* Created security standardized templates including password management strategy and implementation
* Installed custom software and automated installation process
* Optimized Redshift database for optimal performance
* Expert knowledge of AWS EMR, Spark, Scala, Hadoop, HortonWorks, S3, RedShift
* Expert level knowledge of Linux/Unix, PowerShell, network security architecture and database tuning
* Strong communication and teaching skill to train teams with proven problem solving and critical thinking skills
* Used Informatica Cloud Data Integration for global, distributed data warehouse and analytics projects.
* Worked with cloud data warehouses AWS Redshift, Azure SQL Data Warehouse, and Snowflake and Informatica Cloud Data Integration solutions to augment the performance, productivity, and extensive connectivity to cloud and on-premises sources.
* Worked with Informatica Cloud since it’s flexible and scalable transformations and advanced capabilities, to seamlessly integrate growing data volumes across disparate sources in the data warehouse using wizards, preconfigured templates, and out-of-the-box mappings
* Used Pentaho, Hadoop, Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Used Pentaho, Hadoop, Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Developed Scala scripts, UDFFs using Data frames/SQL and RDD/MapReduce in Spark 1.6 for Data Aggregation, queries and writing data back into OLTP system through Sqoop.
* Performed performance tuning of applications for setting right Batch Interval time, correct level of parallelism and tuning.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Developed Scala scripts, UDFFs using both Data frames/SQL and RDD/MapReduce in Spark 1.6 for Data Aggregation, queries writing data back into OLTP system through Sqoop.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Developed Scala scripts, UDFFs using both Data frames/SQL and RDD/MapReduce in Spark 1.6 for Data Aggregation, queries and writing data back into OLTP system through Sqoop.
* Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Configured deployed and maintained multi-node Dev and Test Kafka Clusters.
* Developed Spark scripts by using Scala shell commands as per the requirement.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Developed Scala scripts, UDFFs using both Data frames/SQL and RDD/MapReduce in Spark 1.6 for Data Aggregation, queries writing data back into OLTP system through Sqoop.
* Loaded the data into Spark RDD and do in memory data Computation to generate the Output response.
* Optimizing of existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's.
* Worked on migrating Map Reduce programs into Spark transformations using Spark and Scala.
* Worked with Hadoop/Hive/Big data to architect, design and build solutions to create dashboards/Data Visualizations
* Utilized Hadoop HiveQL(HQL) development and performance tuning on full lifecycle implementations.
* Designed and developed POCs in Spark using Scala to compare the performance of Spark with Hive and SQL/Oracle.
* Worked with ETL interfacing components of solution design and configuration activity in establishing solutions
* Performed cross system joins for identifying duplicates and data anomalies
* Created IDQ Dashboards/KPI Metrics Dashboards for reporting using Hadoop/Talend
* Performed ETL and resolved data quality issues with analysis
* Worked with Informatica Power Center Development/IDQ Output
* Developed, enhanced, supported, and integrated products and software solutions
* Deployed, maintained, managed product and software solutions for various clients
* Worked as a Tableau Engineer / Pentaho/ BI/Reporting platform to enable data access, analytic models, and visualization.
* Worked with Data Management processes for Reporting and Analytics
* Worked with Big Data technologies including Hadoop HDFS, MapReduce, Pig, Hbase, and Hive, Python, and SQL

**University of California, Los Angeles, CA**

**ETL Pentaho, Informatica, SSIS, Big Data Hadoop Developer**

**11/2014 – 01/2016**

**Hands on Tools:** Talend 5.4,Spark 1.6.0, SQL Server 2013, Oracle 11.2.0.4, ETL, Hadoop 2.4.1.

**Project Environment:** JSON, XML, Spark API, Spark-SQL, Data Frames, SQL/Oracle, Talend OS ETL, SQL server, PostgreSQL

**Contribution:**

* Utilized Talend open studio & Talend Enterprise platform for big data management,
* Worked with X12 EDI standards for healthcare data, HEDIS, and HIPAA
* Worked with Big Data, Hadoop, Hive, Pig, Sqoop, Pentaho, Informatica
* Established, maintained, and enforced ETL architecture design principles, techniques, standards, and best practices
* Managed technical designs of ETL reference architectures to ensure high data quality, data integration performance, and error recovery/handling, optimize performance
* Worked with Informatica Cloud Data Integration to deliver accessible, trusted, and secured data to facilitate more valuable business decisions to identify competitive advantages to better service customers, and build an empowered workforce.
* Used Informatica Cloud Data Integration for global, distributed data warehouse and analytics projects.
* Worked with cloud data warehouses AWS Redshift, Azure SQL Data Warehouse, and Snowflake and Informatica Cloud Data Integration solutions to augment the performance, productivity, and extensive connectivity to cloud and on-premises sources.
* Worked with Informatica Cloud since it’s flexible and scalable transformations and advanced capabilities, to seamlessly integrate growing data volumes across disparate sources in the data warehouse using wizards, preconfigured templates, and out-of-the-box mappings
* Reviewed/assessed existing ETL applications to update features, performance improvements, upgrades, and ongoing sustainment
* Conducted design reviews, code reviews, and performance tuning minimizing bottlenecks, maximizing performance
* Researched and recommended future improvements in ETL, Pentaho, and Informatica, and daily functions
* Worked with team to develop in-house knowledge repository for best practices, solution documentations and manuals
* Used Pentaho, Informatica, ETL, SSIS for data management, data integration, data quality, MDM, data governance
* Worked to architect, design and built Big Data solutions using Hive Hadoop for analysis and to build dashboards
* Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Used Hadoop, Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Loaded the data into Spark RDD and do in memory data Computation to generate the Output response.
* Worked with Hadoop, SSIS, SSAS OLAP cubes, Pentaho, Hadoop, PIG, Hive, Spark, Oracle, and MS SQL Server
* Implemented custom Error handling in ETL Pentaho Informatica jobs and worked on different methods of logging
* Hands-on data integration, data management/data warehousing experience, Pentaho, Informatica PowerCenter/IDQ development with direct development experience Informatica Business Glossary, Informatica Versions 10.x/9.1/8.x/7.x Source Analyzer, Mapping Designer, Mapplet, Transformations, Workflow Monitor, Workflow Manager, Informatica developer IDQ 9.1/9.5.1, Informatica Power Mart 6.x/5.x, Power Exchange, Power Connect, Power Analyzer Data Profiling and Data cleaning, Flat file system Fixed width, Delimiter , OLAP.
* Worked with Pentaho, Informatica MDM 9.X or 10.X, MDM Hub & IDD
* Performed Informatica Data Profiling with IDQ and Analyzer
* Performed analysis to identify data anomalies, data cleansing (ETL) and resolved data quality issues
* Developed IDQ using Joiner to configure & Develop Business Rules
* Performed data integration using Pentaho, Informatica, cross system joins for identifying duplicates and data anomalies
* Created IDQ Dashboards/KPI Metrics for reporting. Performed ETL and resolved data quality issues with analysis
* Extensively worked on Informatica IDE/IDQ. Involved in data profiling using IDQ (Analyst Tool) prior to data staging.
* Used Pentaho, Informatica IDQ's standardized plans for addresses and names clean ups.
* Worked on IDQ file configuration at user's machines and resolved the issues.
* Used IDQ to complete initial data profiling and removing duplicate data. And extensively worked on IDQ admin tasks and worked as both IDQ Admin and IDQ developer.
* Implemented Informatica Business Glossary, Informatica Data Quality, MDM, PowerCenter, Informatica Analyst/Metadata Manager for glossary
* Performed with Pentaho ETL Data Conversion and Data Transformation, SF Data Modeling, Jitterbit.
* Used Talend Integration Suite and Talend Open Studio Strong knowledge and Experience in using Informatica Power Center ETL
* Strong experience in Extraction, Transformation, loading (ETL) data from various sources into
* Data Warehouses and Data Marts using Informatica Power Center (Designer, Workflow Manager,
* Workflow Monitor, Metadata Manger).
* Performed data manipulations using various Talend components like tMap, tJavarow, tjava,
* tOracleRow, tOracleInput, tOracleOutput, tMSSQLInput and many more.
* Analyzing the source data to know the quality of data by using Talend Data Quality.
* Troubleshoot data integration issues and bugs, analyze reasons for failure, implement optimal solutions, and revise procedures
* Performed Migration projects to migrate data from data warehouses on Oracle/DB2 and migrated those to Netezza.
* Used SQL queries and other data analysis methods, as well as Talen Enterprise Data Quality
* Performed profiling and comparison of data used to make decisions regarding how to measure business rules quality of the data.
* Worked on TalendRTX ETL tool, develop jobs and scheduled jobs in Talend integration suite.
* Responsible for tuning ETL mappings, Workflows and underlying data model to optimize load and query performance.
* Developed Talend ESB services and deployed them on ESB servers on different instances.
* Implemented fast and efficient data acquisition using Big Data processing techniques and tools.
* Monitored and supported the Talend jobs scheduled through Talend Admin Center (TAC).
* Developed Oracle PL/SQL, DDLs, and Stored Procedures and worked on performance tuning
* Tuning of SQL Strong understanding of Dimensional Modeling, OLAP, Star, Snowflake Schema, Fact and Dimensional tables

**Covered California/ Health for California Insurance Center, San Francisco, CA**

**Talend Spark Hadoop Developer/ETL BI DW Dashboard Developer**

**01/2014- 10/2014**

**Hands on Tools:** Pentaho 5.x, Talend 5.3.0, Hadoop 2.4.1,Oracle 11.2.0.4, Informatica IDQ 9.6

**Project Environment:** ETL, AWS, S3, Redshift, Talend, Hadoop, Spark, Scala, Python, HiveQL, HQL, Data Visualizations Dashboards, Cloudera Hadoop YARN

**Contribution:**

* Worked as a Data Engineer to help form a cloud based big data engineering team to deliver platform automation and security.
* Built data workflows by using AWS EMR, Spark, Spark SQL, Scala, and Python
* Configured and installed tools with a highly available architecture
* Designed, reviewed and fixed security vulnerabilities at network/subnet/security groups level
* Created security standardized templates including password management strategy and implementation
* Installed custom software and automated installation process
* Optimized Redshift database for optimal performance
* Expert knowledge of AWS EMR, Spark, Scala, Hadoop, HortonWorks, S3, RedShift
* Expert level knowledge of Linux/Unix, PowerShell, network security architecture and database tuning
* Strong communication and teaching skill to train teams with proven problem solving and critical thinking skills
* Performed in Extraction Transformation loading (ETL) data from various sources into Data Warehouses and Data Marts using
* Worked with X12 EDI standards for healthcare data, HEDIS, and HIPAA
* Worked with Big Data, Hadoop, Hive, Pig, Sqoop, Pentaho, Informatica
* Established, maintained, and enforced ETL architecture design principles, techniques, standards, and best practices
* Managed technical designs of ETL reference architectures to ensure high data quality, data integration performance, and error recovery/handling, optimize performance
* Worked with Informatica Cloud Data Integration to deliver accessible, trusted, and secured data to facilitate more valuable business decisions to identify competitive advantages to better service customers, and build an empowered workforce.
* Used Informatica Cloud Data Integration for global, distributed data warehouse and analytics projects.
* Worked with cloud data warehouses AWS Redshift, Azure SQL Data Warehouse, and Snowflake and Informatica Cloud Data Integration solutions to augment the performance, productivity, and extensive connectivity to cloud and on-premises sources.
* Worked with Informatica Cloud since it’s flexible and scalable transformations and advanced capabilities, to seamlessly integrate growing data volumes across disparate sources in the data warehouse using wizards, preconfigured templates, and out-of-the-box mappings
* Reviewed/assessed existing ETL applications to update features, performance improvements, upgrades, and ongoing sustainment
* Conducted design reviews, code reviews, and performance tuning minimizing bottlenecks, maximizing performance
* Researched and recommended future improvements in ETL, Pentaho, and Informatica, and daily functions
* Worked with team to develop in-house knowledge repository for best practices, solution documentations, manuals, and procedures for user education
* Used Pentaho, Informatica, ETL, SSIS for data management, data integration, data quality, MDM, data governance
* Worked to architect, design and built Big Data solutions using Hive Hadoop for analysis and to build dashboards
* Used Hadoop, Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Loaded the data into Spark RDD and do in memory data Computation to generate the Output response.
* Worked with Hadoop, SSIS, SSAS OLAP cubes, Pentaho, Hadoop, PIG, Hive, Spark, Oracle, and MS SQL Server
* Implemented custom Error handling in ETL Pentaho Informatica jobs and worked on different methods of logging
* Hands-on data integration, data management/data warehousing experience, Pentaho, Informatica PowerCenter/IDQ development with direct development experience Informatica Business Glossary, Informatica Versions 10.x/9.1/8.x/7.x Source Analyzer, Mapping Designer, Mapplet, Transformations, Workflow Monitor, Workflow Manager, Informatica developer IDQ 9.1/9.5.1, Informatica Power Mart 6.x/5.x, Power Exchange, Power Connect, Power Analyzer Data Profiling and Data cleaning, Flat file system Fixed width, Delimiter , OLAP.
* Worked with Pentaho, Informatica MDM 9.X or 10.X, MDM Hub & IDD
* Performed Informatica Data Profiling with IDQ and Analyzer
* Performed analysis to identify data anomalies, data cleansing (ETL) and resolved data quality issues
* Developed IDQ using Joiner to configure & Develop Business Rules
* Performed data integration using Pentaho, Informatica, cross system joins for identifying duplicates and data anomalies
* Created IDQ Dashboards/KPI Metrics for reporting. Performed ETL and resolved data quality issues with analysis
* Extensively worked on Informatica IDE/IDQ. Involved in data profiling using IDQ (Analyst Tool) prior to data staging.
* Used Pentaho, Informatica IDQ's standardized plans for addresses and names clean ups.
* Worked on IDQ file configuration at user's machines and resolved the issues.
* Used IDQ to complete initial data profiling and removing duplicate data. And extensively worked on IDQ admin tasks and worked as both IDQ Admin and IDQ developer.
* Implemented Informatica Business Glossary, Informatica Data Quality, MDM, PowerCenter, Informatica Analyst/Metadata
* Performed on Pentaho ETL Data Conversion and Data Transformation, SF Data Modeling, Jitterbit.
* Worked with client and provider analytics, developed new data marts for new and existing data warehouses.
* Worked with IDQ, Informatica PowerCenter, Oracle, Dimensional Data Modeling for Healthcare/Payor Data solutions
* Worked on data integration using ETL SSIS, developing data models using ERWIN, PLSQL, Informatica Data Analyst, EPIC, Facets, Informatica MDM, Informatica IDD, TOAD, salesforce.

**Nashville general hospital/Healthcare Providers, Nashville, TN**

**ETL Developer, Pentaho/Talend Big Data Hadoop Developer**

**01/2009 – 12/2013**

**Hands on Tools:** MS SQL Server, Azure, Data Lake, Redshift, S3, Hadoop, Spark, Pentaho 3.x/4.x, Talend 5.3, SQL Server 2007, Spark 1.3, Kafka 0.10, Hadoop 2.0.0, PostgreSQL 8.0

**Project Environment:** Talend, Spark RDD, Test Kafka Clusters, Spark API, JSON, Spark-SQL, Data Frames.

**Contribution:**

* Performed data integration using ETL/Pentaho/Informatica/Talend Open Studio Integration Suite.
* Used Pentaho/PDI, Kettle, Informatica, SSIS, Batch Data Analysis using Hive, SQL SSAS
* Worked with Healthcare Data/Claims 835 and 837 formats for analytical purposes, X12 Enterprise Data Integration(EDI), PHI
* Worked with X12 EDI standards for healthcare data, HEDIS, and HIPAA
* Worked with Informatica Cloud Data Integration to deliver accessible, trusted, and secured data to facilitate more valuable business decisions to identify competitive advantages to better service customers, and build an empowered workforce.
* Used Informatica Cloud Data Integration for global, distributed data warehouse and analytics projects.
* Worked with cloud data warehouses AWS Redshift, Azure SQL Data Warehouse, and Snowflake and Informatica Cloud Data Integration solutions to augment the performance, productivity, and extensive connectivity to cloud and on-premises sources.
* Worked with Informatica Cloud since it’s flexible and scalable transformations and advanced capabilities, to seamlessly integrate growing data volumes across disparate sources in the data warehouse using wizards, preconfigured templates, and out-of-the-box mappings
* Worked with Big Data, Hadoop, Hive, Pig, Sqoop, Pentaho, Informatica
* Established, maintained, and enforced ETL architecture design principles, techniques, standards, and best practices
* Managed technical designs of ETL reference architectures to ensure high data quality, data integration performance, and error recovery/handling,
* Worked with X12 EDI standards for healthcare data, HEDIS, and HIPAA
* Worked with Big Data, Hadoop, Hive, Pig, Sqoop, Pentaho, Informatica
* Established, maintained, and enforced ETL architecture design principles, techniques, standards, and best practices
* Managed technical designs of ETL reference architectures to ensure high data quality, data integration performance, and error recovery/handling, optimize performance
* Reviewed/assessed existing ETL applications to update features, performance improvements, upgrades, and ongoing sustainment
* Conducted design reviews, code reviews, and performance tuning minimizing bottlenecks, maximizing performance
* Researched and recommended future improvements in ETL, Pentaho, and Informatica, and daily functions
* Worked with team to develop in-house knowledge repository for best practices, solution documentations and manuals
* Used Pentaho, Informatica, ETL, SSIS for data management, data integration, data quality, MDM, data governance
* Worked to architect, design and built Big Data solutions using Hive Hadoop for analysis and to build dashboards
* Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Used Hadoop, Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Loaded the data into Spark RDD and do in memory data Computation to generate the Output response.
* Worked with Hadoop, SSIS, SSAS OLAP cubes, Pentaho, Hadoop, PIG, Hive, Spark, Oracle, and MS SQL Server
* Implemented custom Error handling in ETL Pentaho Informatica jobs and worked on different methods of logging
* Hands-on data integration, data management/data warehousing experience, Pentaho, Informatica PowerCenter/IDQ development with direct development experience Informatica Business Glossary, Informatica Versions 10.x/9.1/8.x/7.x Source Analyzer, Mapping Designer, Mapplet, Transformations, Workflow Monitor, Workflow Manager, Informatica developer IDQ 9.1/9.5.1, Informatica Power Mart 6.x/5.x, Power Exchange, Power Connect, Power Analyzer Data Profiling and Data cleaning, Flat file system Fixed width, Delimiter , OLAP.
* Worked with Pentaho, Informatica MDM 9.X or 10.X, MDM Hub & IDD
* Performed Informatica Data Profiling with IDQ and Analyzer
* Performed analysis to identify data anomalies, data cleansing (ETL) and resolved data quality issues
* Developed IDQ using Joiner to configure & Develop Business Rules
* Performed data integration using Pentaho, Informatica, cross system joins for identifying duplicates and data anomalies
* Created IDQ Dashboards/KPI Metrics for reporting. Performed ETL and resolved data quality issues with analysis
* Extensively worked on Informatica IDE/IDQ. Involved in data profiling using IDQ (Analyst Tool) prior to data staging.
* Used Pentaho, Informatica IDQ's standardized plans for addresses and names clean ups.
* Worked on IDQ file configuration at user's machines and resolved the issues.
* Used IDQ to complete initial data profiling and removing duplicate data. And extensively worked on IDQ admin tasks IDQ Admin and IDQ developer.
* Implemented Informatica Business Glossary, Informatica Data Quality, MDM, PowerCenter, Informatica Analyst/Metadata
* Performed on Pentaho ETL Data Conversion and Data Transformation, SF Data Modeling, Jitterbit.
* Reviewed/assessed existing ETL applications to update features, performance improvements, upgrades, and ongoing sustainment
* Conducted design reviews, code reviews, and performance tuning minimizing bottlenecks, maximizing performance
* Researched and recommended future improvements in ETL, Pentaho, and Informatica, and daily functions
* Developed in-house knowledge repository for best practices, solution documentations, manuals, and procedures for education
* Used Pentaho, Informatica, ETL, SSIS for data management, data integration, data quality, MDM, data governance
* Worked to architect, design and built Big Data solutions using Hive Hadoop for analysis and to build dashboards
* Used Hadoop, Hive, Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Worked with Informatica Cloud Data Integration to deliver accessible, trusted, and secured data to facilitate more valuable business decisions to identify competitive advantages to better service customers, and build an empowered workforce.
* Used Informatica Cloud Data Integration for global, distributed data warehouse and analytics projects.
* Worked with cloud data warehouses AWS Redshift, Azure SQL Data Warehouse, and Snowflake and Informatica Cloud Data Integration solutions to augment the performance, productivity, and extensive connectivity to cloud and on-premises sources.
* Worked with Informatica Cloud since it’s flexible and scalable transformations and advanced capabilities, to seamlessly integrate growing data volumes across disparate sources in the data warehouse using wizards, preconfigured templates, and out-of-the-box mappings
* Used Hadoop, Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Loaded the data into Spark RDD and do in memory data Computation to generate the Output response.
* Worked with Hadoop, SSAS OLAP cubes, Pentaho, Hadoop, PIG, Hive, Spark, Oracle, and MS SQL Server
* Implemented custom Error handling in ETL Pentaho Informatica jobs and worked on different methods of logging
* Hands-on data integration, data management/data warehousing experience, Informatica PowerCenter/IDQ development with direct development experience Informatica Business Glossary, Informatica Versions 10.x/9.1/8.x/7.x Source Analyzer, Mapping Designer, Mapplet, Transformations, Workflow Monitor, Workflow Manager, Informatica developer IDQ 9.1/9.5.1, Informatica Power Mart 6.x/5.x, Power Exchange, Power Connect, Power Analyzer Data Profiling and Data cleaning, Flat file system Fixed
* Worked with Informatica MDM 9.X or 10.X, MDM Hub & IDD
* Performed Informatica Data Profiling with IDQ and Analyzer
* Performed analysis to identify data anomalies, data cleansing (ETL) and resolved data quality issues
* Developed IDQ using Joiner to configure & Develop Business Rules
* Performed cross system joins for identifying duplicates and data anomalies
* Created IDQ, geospatial Dashboards/KPI Metrics for reporting. Performed ETL and resolved data quality issues with analysis
* Extensively worked on Informatica IDE/IDQ. Involved in data profiling using IDQ (Analyst Tool) prior to data staging.
* Used IDQ's standardized plans for addresses and names clean ups.
* Worked on IDQ file configuration at user's machines and resolved the issues.
* Used IDQ to complete initial data profiling and removing duplicate data. And extensively worked on IDQ admin tasks IDQ Admin and IDQ developer.
* Implemented Informatica Business Glossary, Informatica Data Quality, MDM, PowerCenter, Informatica Analyst/Metadata
* Performed on Pentaho ETL Data Conversion and Data Transformation, SF Data Modeling, Jitterbit.
* Worked in Product Development (PDP) Data Governance Office (DGO), Business Glossary.
* Responsible for developing Informatica Business Glossary solution based on functional and technical design specifications for business designs and technical requirements and develop a catalog structure in Informatica.
* Worked with Hadoop, Star Schema, Dimension and Fact data models for dashboards, data visualizations projects
* Worked to architect, design and built Big Data solutions using Hive Hadoop.
* Designed and developed POCs in Spark using Scala to compare the performance of Spark with Hive and SQL/Oracle.
* Worked on Pentaho, Informatica, SSIS, SQL queries worked on the design, development and testing mappings.
* Created dashboards, data visualizations, ETL job infrastructure using Talend Open Studio, Hadoop, Informatica
* Managed Error Handling, Performance Tuning, Error Logging, clustering and High Availability in Talend
* Worked with Business Analysts to correlate business requirements to domain entities and data elements
* Managed ETL interfacing components, dashboard designs, solution design and configuration activity in establishing solutions
* Monitored the daily runs, weekly runs and AdHoc runs to develop dashboards, load data into the target systems.
* Created test plans, test data for extraction and transformation processes and resolved data issues following the data standards.
* Used Talend, Hadoop, IDQ tool for profiling, applying rules and develop mappings to move data from source to target systems.
* Developed dashboards, Transformations, Mapplets and Mappings using Informatica Designer to implement business logic.
* Presented dashboard design architectures to the various stakeholders, customers, servers, Network, Security and other teams.
* Provided technical leadership and governance of the big data team and the implementation of solution architecture
* Managed the architecture dashboard design changes due to business requirements and other interface integration changes
* Provided an overall architect responsibilities including roadmaps, leadership, planning, technical innovation, security
* Designed, Layout, and Deployed Hadoop clusters in the cloud using Hadoop ecosystem & open Source platforms
* Configured and tuned production and development Hadoop environments with the various intermixing Hadoop components
* Provided End-to-end systems implementation such as data security and privacy concerns
* Designed and implemented geospatial big data ingestion, processing and delivery
* Provided cloud-computing infrastructure solutions on Amazon Web Services AWS - EC2, VPCs, S3, IAM
* Involved in the administration, configuration management, monitoring, debugging, and performance tuning, technical resolution on Hadoop applications suit, Hadoop platform, MapReduce, Hive, HBase, Spark, Flume, Oozie, Tez, Ambari, Kafka, Pig, Storm, Falcon, Atlas, Scoop, NFS, WebFDS, Hue, Knox, Ranger, Impala, and ZooKeeper
* Worked with star, snowflake schemas, indexing, aggregate tables, dimension tables, constraints, keys, and fact tables

**California Health Systems /California Association of Hospitals & Health Systems, Los Angeles, CA & Sacramento, CA**

**ETL Data Integration/BI Architect Developer**

**08/2000 – 01/2009**

**Hands on Tools:**  Pentaho 3.x, IDQ 6.x, SQL Server 2000/2003/2007.

**Environment:** IDQ, Power Center ETL, MS SQL Server, Data Warehouse, SQL, Unix, SSIS, SSAS, SSRS.

**Contribution:**

* Managed Error Handling, Performance Tuning, Error Logging, clustering and High Availability in Talend
* Worked with Business Analysts to correlate business requirements to domain entities and data elements
* Performed data integration Pentaho, ETL SSIS, developer dashboard designs SSRS, interfacing components of solution design and configuration activity in establishing solutions
* Analyzed and performed data integrations, Pentaho,SSIS, SSAS, SSRS, created dashboards, data visualizations, etc.
* Worked on analytical/geospatial dashboard reports using SSRS, provided AdHoc Reports, worked on Data Visualizations reports and worked closely with the analytics team/data scientists
* Created Pentaho, SSIS, ETL job infrastructures, dashboard reports and Data visualizations
* Worked on dashboard reports, data integration, ETL components
* Worked on improving the performance of data integration jobs, data visualizations, geospatial dashboard KPI reports
* Monitored the daily runs, weekly runs and AdHoc dashboard reports, runs to load data into the target systems.
* Created test plans, test data for extraction and transformation processes and resolved data issues following data standards
* Created dashboard, KPIs, using IDQ tool for reports, applying rules and developed mappings to move data from source to target
* Developed Transformations, Dashboards, Mapplets and Mappings using Informatica Designer to implement business
* Performed Data analysis to ensure accuracy and integrity of data in the context of Business functionality.
* Developed Dimensional Modeling, Dashboards, OLAP, Star, Snowflake Schema, Fact and Dimensional tables and DW concepts.
* Developed, refined and scaled data management and analytics procedures, systems, workflows.
* Worked in the design and development of solutions for large volumes of data for dashboard, data visualizations,KPI reports
* Responsible for creating and maintaining Customer Intelligence Analytics Data Warehouse and Data Modeling.
* Created Source to Target Mappings and Facilitated Data Warehouse Model Reviews.
* Developed Customer Data Integration (CDI) participating in Data Modeling, JADs, Data Mapping and review sessions, source to target mappings, creating Business Conceptual Models, Logical Data Models and Physical Models.
* Facilitated Model Review and Geospatial Dashboard, Mapping Sessions
* Used SQL skills, querying large complex data sets and performance analysis.
* Worked in modelling, managing, scaling, performance tuning of high volume OLTP, OLAP and data warehouse environments.