

## Email

[jsyedbi@gmail.com](mailto:jsyedbi@gmail.com)

## Telephone

+1-703-679-7861

## Certifications

Domain Expertise  
(By Clients)

## -Education

- West University
- ULA University
- UNIAC University
- BPP University

## -Healthcare/Pharma

- Vertex Pharma
- Fidelis Healthcare
- United Healthcare
- MedAssurant
- MedRecovery
- EASAI

## -Financial/Banking

- Absa Bank
- Barclays Bank

## -Real Estate

- Onboard Analytics

-Advertising/Deal  
Management

- AOL
- Vartopia
- Weather Channel

## -Email Marketing

- Ebay
- GSI
- Edialog

## -Government

- USPTO
- NOAA

## -Retail

- P&G

## Professional Experience

Visionary, resourceful, productive, **Technical Leader(ETL, BI, DWH, MDM)** with more than 12+ years of specialty expertise in developing, supporting, testing and deploying Data Warehousing, MDM and Business Intelligence solutions in Healthcare, Pharma, Banking, Advertising, Education, Email Marketing & retail domains. Have thorough understandings of data analysis, data profiling, database development & requirement analysis which helped me gradually developed the Data Warehousing and Business Intelligence approaches to improve the performance & ROI by 50%-100% based on the company's existing data warehouse and business intelligence infrastructure.

My first philosophy is to make the best use of the tools and technologies already in place and to provide the valuable consulting and education services to enable clients to choose the better cost effective solutions, if required. As I have witnessed many poor project results simply because everyone expects the out-of-box features to satisfy all the requirements, and there is no one to deep dive into the tool and explore its full potential.

My second philosophy is to focus on leveraging the offerings of open source and cost effective BI/ETL tools to streamline the Data Integration process with low ownership cost. BI/DW world is very vendor-diversified, so productivity can be dramatically increased by gluing various tools together.

## Job Responsibilities

- **ETL Admin/DBA Tasks:** Deep experience in performing DBA/ETL (Talend, OWB, Informatica & Datastage) admin tasks. Involved in tool installations, configuration, adding users, managing roles, assigning privileges, memory management, storage estimation, tuning of ETL, BI and database routines.
- **Data Warehouse, BI & MDM Development:** Proven track record in planning, building, managing and deploying successful large-scale Data Warehouse, MDM, and decision support systems. Comfortable with both technical and functional applications of RDBMS, Data Mapping, Data management, Data transportation and Data Staging.
- **Cloud ETL Development:** Specialty expertise in ETL routines development in Amazon cloud using Talend, Informatica, Datastage, OWB, ODI, SSIS, and Pentaho. Extensively worked on the ETL mappings, analysis and documentation. Deep experience in creating fact, dimension, staging, ODS and other database objects like views, stored procedure, function, indexes, constraints with Oracle, Sql Server, Teradata, MySQL, Sybase & Netezza.
- **Requirement Gathering, QA & UAT:** Deep experience in Understanding business processes and gathering requirements from subject matter experts and writing test cases and smooth transition of code to QA.
- **Best Practices:** Deep experience in developing ETL, BI and Database development best practices like naming conventions, Source to Target mapping documents, Technical Specification, Bus Architecture Matrix etc.
- **Methodologies:** Expertise in implementing Ralph Kimball(Top Down), Bill Inmon(Bottom Up) and Agile methodologies, and also well versed with Dimensional Modeling (**Star, Snow Flake and Galaxy Schemas**) to build the Data warehouse.
- **SCD & CDC:** Specialty expertise in implementing and deploying complex ETL flows to track change records with SCD (type 1,2 and 3), database CDC routines with Replication Server, GoldenGate and Informatica Real.
- **Data Profiling and Data Quality:** Deep experience in data profiling, Data Architecture Policies, Standards, Governance and data cleansing with Talend Data Profiler, IBM Information Analyzer, IBM Quality Stage and Informatica data quality.
- **Error Handling & Tuning :** Expertise in developing ETL error handling, Batch processing and Data Lineage.

## Skills Profile

ETL Tools	<b>Informatica (9.x,8.x,7.x, 6.x),Talend(TIS &amp; TDQ) 3.2, 4.2, 5.0, OWB(11g,10g,9i), Datastage(7.5,8.0, 8.5),SSIS, Pentaho ETL, Sybase ETL, BO Data Integrator, ODI</b>
BI Reporting Tools	<b>OBIEE(10g), MicroStrategy(9.0,8.1, 8.0.2), OBIEE, Business Objects, Discoverer 9i</b>
Services/CRM	SoapUi, Webservices, Salesforce, Banner
Databases	Oracle 7.x/8/8i/9i/10g/11g, Sybase IQ/ASE, Ms-Sql Server 2005/2008, Netezza, MySQL, Teradata, Exadata, AWS, RDS, PostgresSql, Vertica, Greenplum
Tools	Information Analyzer, Informatica MDM, Talend MDM, Erwin, Visio, Power Designer
Languages	Sql, Pl-Sql, T-Sql, Xml
Methodologies	Ralph Kimball Methodology, Bill-Inmon Methodology, Agile Methodology
Operating Systems	Windows 95/98/NT4.0/ 2000,UNIX(Sun Solaris, Hp, AIX Linux), Ms-Dos

## Qualification, Certification/Training

- **Masters of Business Intelligence** from **Grantham University**, USA in 2010.
- Bachelor of Information Technology from Manipal University, India in 2002.

## Awards &amp; Achievements

- **White Paper: "Proactive Data Warehousing with Column Based Database"** has been published by B-Eye-Network and received five star rating from B-Eye-Network. Paper can be found at <http://www.b-eye-network.com/view/11135>
- Assisted IM managers in BARCLAYS/ABSA BANK to prepare ETL, Dimensional Modeling, Bus Architecture Standards and guidelines, which are now being followed by the Barclays bank Information management team.
- Extensive experience in handling big data projects and developed the complex ETL architecture to handle the massive data volumes and optimize the processes to streamline the data load.
- Overwhelming Commitment & Work Performance by Barclays Bank, Faritec & THS.

## Employment History

- Working for Infomatics Corp as Sr. Consultant from Apr 2012 – till now.
- Worked for Digipulse Technologies as Sr. Consultant from Feb 2011 – Apr 2012.
- Worked for MedAssurant as Data Warehouse Developer IV from March 2010—Feb 2011.
- Worked for PPC as Sr. Analyst from Apr 2007—March 2010.
- Worked for TCS Ltd as IT Analyst from June 2004 –April 2007.
- Worked for Anthesis Technologies as System Analyst from July 02 - May 04.

<b>Project Details</b>
------------------------

<b>Project :</b> Ebay/GSI/E-dialog PPI	<b>Role :</b> Technical Leader (ETL, DWH, MDM)	<b>Client :</b> GSI Commerce/E-Dialog	<b>Period :</b> Nov 2012- ongoing
--	--	---------------------------------------	-----------------------------------

**Project Overview:** This project is in development stage to standardized Edialog/Ebay email marketing Data Integration and Business Intelligence process to load the Data-Mart, Data Warehouse and BI Dashboards for the marketing department to implement new business strategies and which gives the effective information to the Senior Executives & Top Management for decision making. Purpose of this project is to build an Enterprise Data Warehouse to integrate inbound/outbound events used for email marketing to build suite of best-in-class core applications for GSI/Edialog clients. The project requires extensive Data Integration process to load the data into Enterprise Data Warehouse from 1000+ multi format data sources scattered in 450+ database schemas in 5+ legacy systems which makes up to 400+ million inbound/outbound events generated in a day. The project requires development of complex transformation routines for data profiling, data cleaning, and data transformation to streamline the Data Integration process. Eleven enterprise system applications and multiple work streams will be integrated into a new technology infrastructure to support the core business functions of the organization.

**Solution:** Implementing Informatica/Talend Data Integration/MDM and Microstrategy BI solution to streamline the Data Integration and Business Intelligence process. Developed the new end-end scalable ETL architecture, Implemented best practices, naming conventions and successfully developed data warehouse, data marts and third party vendor's loads with Talend/Informatica. Hadoop has been used extensively to store the massive amount of events data and Talend/Informatica Big data components have been used to read the data from Hadoop Hive.

**Responsibilities:**

- o Performed complex DBA tasks on Oracle, Mysql, SqlServer, Netezza and PostgresSql by creating database, user, grants, partition, index, explain plan, session monitoring, hint, parallelism, CTAS, Rownum Pagination, backups, stats collection, sql tuning etc.
- o Developed end-end ETL architecture, job restartability mechanism, data traceability, Hashing algorithm, schema looping, database pushdown, Tuned the jobs with Common database connection sharing, ELT, partitioning, multi node parallelism, bulk loading, batch size, commit interval, cursor size and buffering mechanism.
- o Developed cleanup strategy to routinely clean the staging, auditing, partition manager, logging and stats tables.
- o Developed scalable and high performance ETL architecture to handle the load of 300+ million records in a day by extensively utilizing the partitions, sub partitions, CTAS, Batch load, Database Parallelism, Hints, With clause, Bulk Loader, Tmpump, Fastload etc.
- o Developed the Delta load strategy with GoldenGate to apply the CDC from 30+ legacy source systems.
- o Developed data capture strategy to fetch the data from 2000+ schemas scattered across 30+ legacy systems.
- o Developed transaction data warehouse load strategy to refresh the data warehouse in 15 minutes window as soon the data arrives in the source.
- o Developed the complex ETL architecture mechanism to run the Staging, Ods and Data warehouse loads in parallel by utilizing the Batches/MD5 mechanism.
- o Used MD5/Sha2/CRC32 algorithm to detect the SCD Type1/Type2 changes on big dimensions to avoid heavy row by row comparison.
- o Developed end-end ETL jobs, Data Quality and Data Profiling routines with Informatica/Talend Data Integration/MDM.
- o Developed the complex ETL jobs to load the data from multiple formats like xml, multiformat, databases, webservices, into Prep, Staging, Data marts & Data warehouse.
- o Developed complex ETL jobs using Talend/Informatica to push and pull the data to Webservices and Salesforce.
- o Developed complex data cleaning and data validation ETL job to clean address, phone no, claims and also to validate the record counts.
- o Developed the control file validation logic to validate the data loaded with the source data.
- o Developed Data Lineage and Batches mechanism in the ETL architecture and developed the joblet/ mapplet in the ETL tool.
- o Used Informatica Power Center 9.1 and used most of the Informatica functions extensively to build business rules to load data. Most of the transformations used were like the Source qualifier, Aggregators, lookups, Filters, Sorting, Source Qualifier, SQL, Joiner, Expression, Router, Reject Strategy, Update Strategy & Sequence generators.
- o Used parameters to handle the values of database connections, file path, data load history dynamically.
- o Developed Informatica session, task and workflow to execute the ETL jobs and load the data from heterogeneous sources like Salesforce ERP system, .csv, .xml, .excel, oracle, mysql and sqlserver.
- o Developed complex ETL jobs for Staging, Transposing, Reduce, and Deduplication & Standardization.
- o Developed the control flows to setup the jobs dependency and also applied the breakpoint logic to kick off the workflow where it fails earlier.
- o Developed dimensions & Fact loading. Applied SCD type1, type 2 and CDC to capture the changes.
- o Tuned the ETL jobs and STAR schemas to optimize load and query performance.
- o Developed Change data capture mechanism to capture the changes with hash code and Oracle Minus, Sql Server Except functionality.
- o Developed Metadata, Data Dictionary and Control database tables like Batches, Logcatcher, Statcatcher, Flowcatcher, Dataproviders, BatchProcess, DatabaseColumns, SourceToTarget, DataQualityRules, tparallelize etc.
- o Developed joblets for Batches, BatchProcesses, LogCatcher, Audit, ErrorHandling, Notification, FTP.
- o Developed complex ETL jobs to FTP the data from multiple vendors locations and automating the downstream jobs based on file arrival.
- o Performed Talend and Database administration tasks to create the project, creating users, giving permissions, managing the log reporting, job automation and running the jobs from web console.
- o Performed Customer Requirements Gathering, Requirements Analysis, Design, Development, Testing, End User Acceptance Presentation.
- o Developed best practices documents like Source File Definition, Source-to-Target Mapping, Fact & Dimension Bus Architecture Matrix, Capacity Estimate Matrix, Physical Design, and Logical Design, Production Review Checklist, Migration list, Turnover document.
- o Developed Meta Data Management, MDM, Data Quality and Data Profiling policies, standards and best practices.
- o Created technical and functional specifications, Dimensional Modeling & Database design using Visio & Erwin Models.

<b>Solution Environment</b>	Windows 2008, Linux, Oracle 10g Cloud, Sql Server 2008, Teradata, MySql, Salesforce, Hadoop, Hive, PostgresSql
<b>Tools</b>	Informatica 9.1, Talend Integration Suite 4.2/5.2, Micro strategy , SVN, Control-M, Golden Gate, Informatica MDM

<b>Project :</b> NOAA Data Integration	<b>Role :</b> Technical Leader (ETL, DWH, MDM)	<b>Client :</b> Apollo Global	<b>Period :</b> Feb 2012- Nov 2012
--	--	-------------------------------	------------------------------------

**Project Overview:** Apollo Global is a \$1 billion joint venture between Apollo Group and private equity firm The Carlyle Group formed in September, 2007 that owned University of Phoenix, BPP, ULA, UNIACC universities around the globe. The main purpose of this project is to integrate the data from the above listed universities into centralized operational data store and then build enterprise data warehouse to standardized Apollo universities Data Integration and Business Intelligence process to load the Data-Mart & Data Warehouse and to implement new business strategies and which gives the effective information to the Senior Executives & Top Management for decision making.

**Solution:** Developed comprehensive and scalable data integration process to migrate the ETL jobs developed in Informatica, Datastage, SSIS to Talend Data Integration and OBIEE solution to streamline the Data Integration and Business Intelligence process. Developed the new end-end ETL architecture, Implemented best practices, naming conventions and successfully developed data warehouse, data marts and third party vendor's loads with Talend.

**Responsibilities:**

- o Performed complex DBA tasks on Oracle, Mysql, SqlServer, Netezza by creating database, user, grants, partition, index, explain plan,

	<ul style="list-style-type: none"> <li>session monitoring, hint, parallelism, CTAS, Rownum Pagination, backups, stats collection, sql tuning etc</li> <li>Developed the Delta load strategy with GoldenGate to apply the CDC from 100+ universities source systems.</li> <li>Developed data capture strategy to fetch the data from 5000+ schemas scattered across 100+ legacy systems.</li> <li>Developed transaction data warehouse load strategy to refresh the data warehouse in 30 minutes window as soon the data arrives in the source.</li> <li>Developed the complex ETL architecture mechanism to run the Staging, Ods and Data warehouse loads in parallel by utilizing the Batches/MD5 mechanism.</li> <li>Used MD5/Sha2/CRC32 algorithm to detect the SCD Type1/Type2 changes on big dimensions to avoid heavy row by row comparison.</li> <li>Developed end-end ETL jobs, Data Quality and Data Profiling routines with Informatica/Talend Data Integration/MDM.</li> <li>Developed the complex ETL jobs to load the data from multiple formats like xml, multiformat, databases, webservices, into Prep, Staging, Data marts &amp; Data warehouse.</li> <li>Developed end-end ETL architecture, job restartability mechanism, data traceability, Hashing algorithm, schema looping, database pushdown, Tuned the jobs with Common database connection sharing, ELT, partitioning, multi node parallelism, bulk loading, batch size, commit interval, cursor size and buffering mechanism.</li> <li>Developed cleanup strategy to routinely clean the staging, auditing, partition manager, logging and stats tables.</li> <li>Developed scalable and high performance ETL architecture to handle the load of 300+ million records in a day by extensively utilizing the partitions, sub partitions, CTAS, Batch load, Database Parallelism, Hints, With clause, Bulk Loader, Tump, Fastload etc.</li> <li>Developed complex ETL jobs using Talend/Informatica/DataStage/SSIS to push and pull the data to Webservices and Salesforce.</li> <li>Developed complex data cleaning and data validation ETL job to clean address, phone no, claims and also to validate the record counts.</li> <li>Developed the control file validation logic to validate the data loaded with the source data.</li> <li>Developed Data Lineage and Batches mechanism in the ETL architecture and developed the joblet/ mapplet in the ETL tool.</li> <li>Used Informatica Power Center 9.1 and used most of the Informatica functions extensively to build business rules to load data. Most of the transformations used were like the Source qualifier, Aggregators, lookups, Filters, Sorting, Source Qualifier, SQL, Joiner, Expression, Router, Reject Strategy, Update Strategy &amp; Sequence generators.</li> <li>Used parameters to handle the values of database connections, file path, data load history dynamically.</li> <li>Developed Informatica session, task and workflow to execute the ETL jobs and load the data from heterogeneous sources like Salesforce ERP system, .csv, .xml, .excel, oracle, mysql and sqlserver.</li> <li>Developed complex ETL jobs for Staging, Transposing, Reduce, and Deduplication &amp; Standardization.</li> <li>Developed the control flows to setup the jobs dependency and also applied the breakpoint logic to kick off the workflow where it fails earlier.</li> <li>Developed dimensions &amp; Fact loading. Applied SCD type1, type 2 and CDC to capture the changes.</li> <li>Tuned the ETL jobs and STAR schemas to optimize load and query performance.</li> <li>Developed Change data capture mechanism to capture the changes with hash code and Oracle Minus, Sql Server Except functionality.</li> <li>Developed Metadata, Data Dictionary and Control database tables like Batches, Logcatcher, Statcatcher, Flowcatcher, Dataproviders, BatchProcess, DatabaseColumns, SourceToTarget, DataQualityRules, tparallelize etc.</li> <li>Developed joblets for Batches, BatchProcesses, LogCatcher, Audit, ErrorHandling, Notification, FTP.</li> <li>Developed complex ETL jobs to FTP the data from multiple vendors locations and automating the downstream jobs based on file arrival.</li> <li>Performed Talend and Database administration tasks to create the project, creating users, giving permissions, managing the log reporting, job automation and running the jobs from web console.</li> <li>Performed Customer Requirements Gathering, Requirements Analysis, Design, Development, Testing, End User Acceptance Presentation.</li> <li>Developed best practices documents like Source File Definition, Source-to-Target Mapping, Fact &amp; Dimension Bus Architecture Matrix, Capacity Estimate Matrix, Physical Design, and Logical Design, Production Review Checklist, Migration list, Turnover document.</li> <li>Developed Meta Data Management, MDM, Data Quality and Data Profiling policies, standards and best practices.</li> <li>Created technical and functional specifications, Dimensional Modeling &amp; Database design using Visio &amp; Erwin Models.</li> </ul>
<b>Solution Environment</b>	Windows 2008, Linux, Oracle 10g Cloud, Sql Server 2008, Netezza, MySql, Salesforce, Hadoop, Hive, Hbase, Banner
<b>Tools</b>	Informatica 8.6.1, Talend Integration Suite 4.2/5.0, OBIEE, Control-M, DataStage 8.5, SSIS, Informatica MDM

<b>Project :</b> Vartopia Data Integration	<b>Role :</b> Technical Leader (ETL, DWH, MDM)	<b>Client :</b> Vartopia	<b>Period :</b> Feb 2012- July 2012
<p><b>Project Overview:</b> This project is in development stage to standardized Vartopia deal management Data Integration and Business Intelligence process to load the Data-Mart, Data Warehouse and BI Ad targeting Dashboards for the marketing department to implement new business strategies and which gives the effective information to the Senior Executives &amp; Top Management for decision making.</p> <p><b>Solution:</b> Implementing Informatica &amp; Talend Data Integration and Pentaho BI solution to streamline the Data Integration and Business Intelligence process. Developed the new end-end ETL architecture, Implemented best practices, naming conventions and successfully developed data warehouse, data marts and third party vendor's loads with Talend.</p> <p><b>Responsibilities:</b></p> <ul style="list-style-type: none"> <li>Performed complex DBA tasks on Oracle, Mysql, SqlServer, Netezza and PostgresSql by creating database, user, grants, partition, index, explain plan, session monitoring, hint, parallelism, CTAS, Rownum Pagination, backups, stats collection, sql tuning etc</li> <li>Developed end-end ETL architecture, job restartability mechanism, data traceability, Hashing algorithm, schema looping, database pushdown, Tuned the jobs with Common database connection sharing, ELT, partitioning, multi node parallelism, bulk loading, batch size, commit interval, cursor size and buffering mechanism.</li> <li>Developed end-end ETL jobs, Data Quality and Data Profiling routines with Informatica &amp; Talend Data Integration.</li> <li>Developed the complex ETL jobs to load the data from multiple formats like xml, multiformat, databases, webservices, into Prep, Staging, Data marts &amp; Data warehouse.</li> <li>Used Informatica Power Center and used most of the Informatica functions extensively to build business rules to load data. Most of the transformations used were like the Source qualifier, Aggregators, lookups, Filters, Sorting, Source Qualifier, SQL, Joiner, Expression, Router, Reject Strategy, Update Strategy &amp; Sequence generators.</li> <li>Used parameters to handle the values of database connections, file path, data load history dynamically.</li> <li>Developed Informatica session, task and workflow to execute the ETL jobs and load the data from heterogeneous sources like Salesforce ERP system, .csv, .xml, .excel, oracle, mysql and sqlserver.</li> <li>Used Informatica Power Center and used most of the Informatica functions extensively to build business rules to load data. Most of the transformations used were like the Source qualifier, Aggregators, lookups, Filters, Sorting, Source Qualifier, SQL, Joiner, Expression, Router, Reject Strategy, Update Strategy &amp; Sequence generators.</li> <li>Used parameters to handle the values of database connections, file path, data load history dynamically.</li> <li>Developed Informatica session, task and workflow to execute the ETL jobs and load the data from heterogeneous sources like Salesforce ERP system, .csv, .xml, .excel, oracle, mysql and sqlserver.</li> </ul>			
<b>Solution Environment</b>	Windows 2008, Linux, Oracle 10g Cloud, Sql Server 2008, Teradata, MySql, Salesforce, PostgresSql		
<b>Tools</b>	Informatica 8.6.1, Talend Integration Suite 4.2/5.0, Pentaho BI, , SVN, Control-M, Repository Manager, Informatica MDM		
<b>Project :</b> Weather Channel -Data Integration & BI	<b>Role :</b> Technical Leader (ETL, DWH, MDM)	<b>Client :</b> The Weather Channel	<b>Period :</b> Oct 2011- Feb2012



<b>Project Overview:</b> This project was developed to standardized TWC Data Integration and Business Intelligence process to load the Data-Mart, Data Warehouse and BI Ad targeting Dashboards for the marketing department to implement new business strategies and which gives the effective information to the Senior Executives & Top Management for decision making. This project involved transforming and developing the reports with OBIEE on a massive 500 billion records throughput on daily basis. The initial stage of the project required to extract the unstructured data from Hadoop and Google DFP servers and convert them into structured formats.	
<b>Solution:</b> Implemented Informatica 8.6, MDM, Pentaho, Talend Integration Suite 4.2 and OBIEE solution to streamline the Data Integration and Business Intelligence process. Developed the new end-end ETL architecture, Implemented best practices, naming conventions and successfully developed data warehouse, data marts and third party vendor's loads with Talend/Informatica. Hadoop, Hive and Hbase have been used extensively to do massive data load and data standardization. Talend Big data has been used extensively to pull the data from Hive/Hbase.	
<b>Responsibilities:</b> <ul style="list-style-type: none"> <li>Performed complex DBA tasks on Oracle, Mysql, SqlServer, Netezza and PostgresSql by creating database, user, grants, partition, index, explain plan, session monitoring, hint, parallelism, CTAS, Rownum Pagination, backups, stats collection, sql tuning etc</li> <li>Developed end-end ETL architecture, job restartability mechanism, data traceability, Hashing algorithm, schema looping, database pushdown, Tuned the jobs with Common database connection sharing, ELT, partitioning, multi node parallelism, bulk loading, batch size, commit interval, cursor size and buffering mechanism.</li> <li>Developed end-end ETL jobs, Data Quality and Data Profiling routines with Talend Integration Suite 4.2 and Informatica 8.6.</li> <li>Developed the complex ETL jobs to load the data from multiple formats like xml, multiformat, databases, webservices, into Prep, Staging, Data marts &amp; Data warehouse.</li> </ul>	
<b>Solution Environment</b>	Windows 2008, Linux, Oracle 10g, Sql Server 2008, Netezza, MySql, Hadoop, Hive, Hbase, PostgresSql
<b>Tools</b>	Informatica 8.6, Pentaho, Talend Integration Suite 4.2, Talend MDM , OBIEE, SVN, Control-M, Talend MDM

<b>Project :</b> Vertex-Data Integration	<b>Role :</b> Technical Leader (ETL, DWH, MDM)	<b>Client :</b> Vertex Pharma, Boston	<b>Period :</b> Feb 2011-Sep2011
<b>Project Overview:</b> This project is developed to standardized Vertex Data Integration process to load the Data-Mart, Data Warehouse and BI Dashboards for various departments of the organization which contains integrated information to implement new business strategies and which gives the effective information to the Senior Executives & Top Management for decision making. Vertex Pharma has over 80+ million drug, member, claims, cases & carriers data coming from third party vendors like Cegedim, IMS, WCB etc and have over 100 different database designs and over 700 database instances with duplicate tables in Oracle, Mysql, Teradata and Sql Server which makes Data Integration and Data Warehouse development a nightmare with customized solution of (Procedures, SSIS Packages) to load the data efficiently and without manual intervention. Because of poor data quality and redundant sql code the client was missing the deliverables SLA's.			
<b>Solution:</b> Implemented Datastage 8.0, Talend Integration Suite 3.2 and MDM solution to streamline the Data Integration process. Developed the new end-end ETL architecture, Implemented best practices, naming conventions and successfully developed data warehouse, data marts and third party vendor's loads with Talend. Data is coming from Oracle, Sql Server, Xml, multiformat flat files.			
<b>Solution Environment</b>	Windows 2008, Linux, Oracle 10g, Sql Server 2008, Teradata, MySql		
<b>Tools</b>	Talend Integration Suite 3.2, Talend MDM , MicroStrategy, Datastage 8.0, SVN, Control-M, Informatica MDM		

<b>Project :</b> MedAssurant-Data Integration	<b>Role :</b> Lead ETL Developer & Architect	<b>Client :</b> MedAssurant, Bowie-MD	<b>Period :</b> Mar2010-Feb2011
<b>Project Overview:</b> MedAssurant has over 60+ million member data and have over 78 different database designs and over 1700 database instances in the organization which makes Data Integration and Data Warehouse development a nightmare with customized solution (Custom Sql + SSIS + Informatica 8.6) to streamline the multi formats, redundant ETL code, and poor data standards into unified standard format with the existing Data Integration architecture. The company has grown significantly and was facing increasing data quality, data integrity, data scalability, data security challenges, all of which reduce the efficiency and increase costs, legal and financial risks. The client has taken over Catalyst system in Atlanta which was using Informatica 8.6 for Data Integration. So the migration of Catalyst systems Data Integration process was also added in streamlining the unified Data Integration process.			
<b>Solution:</b> Initially involved in the ETL tool evaluation and pilot ETL project with Talend Open Studio 3.2. Implemented IBM Information Server Suite (DataStage, QualityStage, Information Analyzer) and MDM solution to address the MedAssurant challenges and to streamline the Data Integration process. Developed the new end-end ETL architecture and successfully developed data warehouse and data marts with the new standards and architecture. Data was coming from Oracle, Sql Server and flat file sources. ETL routines are used extensively to create full load jobs and delta load jobs. DataStage ETL routines are used extensively to migrate the data from Oracle, Sql Server/ Flat file sources to ODS database and then to EDW and Datamarts. IBM Information server suite is used for data profiling, data cleaning and end-end Medassurant Data Integration.			
<b>Project Phases:</b> This project is spanned into 4 phases. <ul style="list-style-type: none"> <li><b>Migration Stage:</b> Iport (Native ETL), SSIS, Talend 3.2 &amp; Informatica ETL → IBM Datastage.</li> <li>Oracle Facets, Sql Server Facets, Source Files → Data Profiler → Data Quality → Datastage ETL → Staging → ODS</li> <li>Sql Server ODS → Datastage ETL → QualityStage → Teradata EDW</li> <li>Sql Server EDW → Deduplication &amp; Standardization → MDM → Netezza Data Marts → OBIEE BI Reports</li> </ul>			
<b>Solution Environment</b>	Windows 2008, Linux, Oracle 10g, Sql Server 2008, Teradata, Netezza		
<b>Tools</b>	Datastage 8.5, Talend 3.2, QualityStage, Information Analyzer, Informatica 8.6, SSIS, Informatica MDM, OBIEE		

<b>Project :</b> EISAI Integration DWH	<b>Role :</b> Lead ETL Developer	<b>Customer :</b> EISAI, Woodcliffake-NJ	<b>Period :</b> Jan 2010 – Mar 2010
<b>Project Overview:</b> Developed complex ETL, Data profiling and Data quality processes to feed the EISAI Data warehouse & Data-Mart for various departments of the organization which contains integrated information to implement new business strategies and which gives the effective information to the Senior Executives & Top Management for decision making on the basis of raw data conversion in to EISAI business rules. EISAI has over 50 million claim, member, pharmacy and enrollment data which goes through complex name & address data cleaning, fixing postal codes with ISO standards, and matching & linking of the data to remove duplicate records. Informatica Data Integration suite is used to do the data quality, data profiling and ETL was used to extract the data from all the sources to load into Staging, ODS, EDW, Data Marts, Aggregate Tables, External Table modules and transpose the data and the push the data into Microstrategy reporting layer.			
<b>Solution Environment</b>	Windows 2008, Linux, Oracle 10g, Sql Server 2008, Teradata		
<b>Tools</b>	Talend 3.2, OWB 10g, Oracle, Sql-Server 2008, 2008, Kalido MDM, Kalido DIW, Mysql, Teradata, SSIS, MicroStrategy 9.0		

<b>Project :</b> NOAA EDW	<b>Role :</b> Sr.ETL Developer & Architect	<b>Customer :</b> NOAA, SilverSpring-MD	<b>Period :</b> Dec 2008 – Jan2010
<b>Project Overview:</b> NOAA collected data for water current, tides, water level and weather for bad weather forecasting, Tsunami alerts and help the shipping company in measuring the water levels and water currents. These data was segregated across multiple database instances with redundant table. So, to address these issues I have developed the operational Data store to streamline and standardize multi format sources into unified data store and to develop the data warehouse, datamarts to develop reports for better predictive analysis and forecasting to give the effective information to the			

Senior Executives & Top Management of the organization and also helping different departments and predictive analysis.	
<b>Solution Environment</b>	Windows 2008, Linux, Sybase ASE, Sql Server 2008, Sybase IQ
<b>Tools</b>	Pentaho ETL, Talend Integration Suite 3.2, Datastage 8.1, Unix, Windows, Oracle, OBIEE

<b>Project :</b> Fidelis Care – IQ EDW	<b>Role :</b> Lead ETL Developer & Architect	<b>Customer :</b> Fidelis, Buffalo-NY	<b>Period :</b> Nov 2007 – Nov 2008
<p><b>Project Overview:</b> This project was developed to create Fidelis care Data warehouse, Data-Mart and BI Dashboards for various departments of the organization which contains integrated information to implement new business strategies and which gives the effective information to the Senior Executives &amp; Top Management for decision making on the basis of raw data conversion in to Fidelis care business rules. Fidelis has over 80+ million member, claims, pharmacy, providers, enrollment, practitioners and government data and have over 150 different database designs and over 1000 database instances in the organization which makes ETL development a difficult task with sql stored procedure. Because of poor data quality and redundant sql code the client was missing the deliverables SLA's.</p> <p><b>Solution:</b> Successfully implemented Data Integration solution with IBM Datastage to standardize the ETL process. Developed the new end-end ETL architecture and successfully developed the ETL, Data profiling and Data quality jobs for ODS, Staging, EDW and data marts modules with the new standards and architecture. Data is coming from Sybase ASE and Sybase IQ databases. ETL routines are used extensively to create Full load jobs and Delta load jobs. Datastage ETL routines are used extensively to migrate the data from Oracle Database to Sybase ODS database and then to Sybase IQ EDW. IBM BI suite is used for data profiling, data cleaning and ETL development of IQ Data marts. Data stage is used extensively to develop ETL flows, transformations, Facts &amp; Dimensional loading for Data mart development. Business Object reporting tool is used to develop the Data marts reports and Dashboard.</p> <p><b>Project Phases:</b> This project is spanned into 4 phases.</p> <ul style="list-style-type: none"> <li>• Oracle Facets – IBM Information Analyzer(Data Profiling) – Sybase ODS.</li> <li>• Sybase ODS – DataStage ETL – Sybase IQ EDW,</li> <li>• Sybase IQ EDW – Data stage ETL – Sybase IQ Data marts.</li> <li>• Sybase IQ Data marts – Business Object (Reporting).</li> </ul>			
<b>Solution Environment</b>	Windows XP, Linux, AIX, Sybase IQ, Oracle 10g, Sybase ASE		
<b>Tools</b>	IBM Information Server, Data stage 8.1, Information Analyzer, Business Objects		

<b>Project :</b> P&G Kalido DIW & BI	<b>Role :</b> Lead ETL Developer	<b>Customer :</b> P&G, Cincinnati-Ohio	<b>Period :</b> Apr 2007 – Nov 2007
<p><b>Project Overview:</b> This system was developed to create P&amp;G Data warehouse &amp; Data-Mart for various departments of the organization which contains integrated information to implement new business strategies and which gives the effective information to the Senior Executives &amp; Top Management for decision making on the basis of raw data conversion in to P&amp;G business rules. Data was coming in tabular, spreadsheet and xml formats. Informatica Data Integration suite is used to do the data quality, data profiling and ETL was used to extract the data from all the sources excluding finance and transpose the data and the push the data into MicroStrategy reporting layer.</p> <p><b>Project Phases</b> Project was spanned across 6 phases.</p> <ul style="list-style-type: none"> <li>• Oracle, Files- Informatica ETL - Oracle Staging</li> <li>• Oracle Staging – Informatica ETL Transposing- Oracle ODS – Informatica ETL – Oracle EDW -- Informatica ETL – Oracle Data Marts</li> </ul>			
<b>Solution Environment</b>	Windows XP, Linux, Sun, Oracle 10g, Sql Server, Teradata		
<b>Tools</b>	Informatica 8.1, Informatica MDM, MicroStrategy 8		

<b>Project :</b> AOL BI Implementation & Maintenance	<b>Role :</b> Sr. ETL Developer	<b>Customer :</b> AOL, Dulles-VA	<b>Period :</b> July 2006 – April 2007
<p><b>Project Overview:</b> This system was developed to create American Online Data warehouse &amp; Data-Mart for various departments of the organization which contains integrated information to implement new business strategies and which gives the effective information to the Senior Executives &amp; Top Management for decision making on the basis of raw data conversion in to AOL business rules. This project also involves migration of data from Redbrick db to NZ db. Data is coming in spreadsheet in .dat format and relational table format. Ab-Initio ETL was used extensively in migrating Redbrick and Oracle routines to Netezza db. MicroStrategy was used for web based reporting and Dashboards. Used Ralph Kimball dimensional modeling methodologies.</p>			
<b>Technologies</b>	Ab-Initio, OBIEE, Windows XP, Linux, Sun, Oracle, Netezza, Redbrick		

<b>Project :</b> Barclays BSC & Bi Details	<b>Role :</b> Sr. ETL Developer	<b>Customer :</b> Barclays Bank, Johannesburg-SA	<b>Period :</b> June 2004 – July 2006
<p><b>Project Overview:</b> This system was developed to create BARCLAYS/ABSA Group SMS Balance Scorecard Data warehouse &amp; Data-Mart for Resource, Process, Customer Relation &amp; Finance dept which contains integrated information to implement new business strategies and which gives the effective information to the Senior Executives &amp; Top Management for decision making on the basis of GSMS Data analysis. Data is coming in spreadsheet in .dat format and from other business units in xml and relational table format. Project is spanned across 6 phases. ETL-Front-End Phase, ETL-Back-End Phase, ETL-Data-Warehouse Phase, ETL-Data Mart Phase, ETL-Data-Mart Input Table Phase and ETL-Report development phase with Balance Score Card &amp; Discoverer. OWB is used extensively for extracting, transporting and loading data from Flat files to Staging and from staging to Data Warehouse and from Data Warehouse to business data marts and from Data Marts to Data-Mart Input Tables and from Data-Mart Input Tables to Balance Score Card Input Tables. Oracle Warehouse Builder Dimension modeling has been implemented to use Oracle OLAP Services extensively.</p>			
<b>Technologies</b>	OWB 10g, SSIS, OBIEE, Discoverer, Sql Server 2005, XML, Windows XP, Linux, Sun, Oracle, Teradata		

<b>Project :</b> THS OIS	<b>Role :</b> Sr. ETL Developer	<b>Customer :</b> The Home Store	<b>Period :</b> Aug 2002 – May 2004
<p><b>Project Overview:</b> The main aim of this project was to create a data warehouse which contains integrated information of THS business units, consolidated sales and distribution of products and gives the effective information to the Senior Executives &amp; Top Management for decision making and data analysis for market &amp; trend analysis. THS have total 100 home furnishing and grocery supermarkets malls all over India and have 10 warehouses. Data of all the stores was store in centralized database. Based on that Data, Data warehouse was created to generate various reports to show the sales information in various regions using Business Objects.</p>			
<b>Technologies</b>	Informatica 5.1 ,Business objects, Windows NT, Oracle 8i, Oracle 9iAS		