Case Study

ETL/Data Warehouse Testing of a GIS Spatial Application

Your company information

Branch	Sydney, Australia
Website	

Contact name	Sapna Bhatia
Title	Business Manager
Phone number	+61 - 420983561
E-mail address	info@adactin.com ; sapna.bhatia@adactin.com

Solution group	Testing Services Group
Solution offering	ETL/Data Warehouse Testing
Project name or title	ETL/Data Warehouse Testing of a GIS Web based Application

Case Study for ETL/Data Warehouse Testing of a GIS Spatial Application

Client Profile

Client is a reputed organization which deals with various Planning and environmental aspects. Client is using geospatial technology in the delivery of planning and development services, such as the online lodgment and tracking of applications, viewing planning information on a web based interactive map and providing new ways for stakeholders to engage with the planning process. The tools and services which are developed by client, helps business and the community to access and transact with planning services from anywhere, any time. Client deals with the heaps of everyday and historical data from multiple sources and different departments which have their own standards for data. An ETL Tool (FME Workbench 2015) was employed in order to make a flawless integration between different data sources from different departments. Then the data needs to get upload to 2 clouds to make it appear on the Web Application.

Business Situation

This GIS Spatial Web based Application uses Google-maps Engine to help you quickly navigate the planning rules that apply to individual land parcels, or search for properties that match certain planning controls. It has been developed primarily for councils and professionals such as architects, builders, certifiers, developers and planners. This is a cloud based product which is developed on Azure & GME (Google Map Engine) cloud platform.

As Client deals with the heaps of everyday and historical data from multiple sources and different departments which have their own standards for data, an ETL/Data Warehouse and Security Testing was done by Adactin to make the Data Extraction, Transformation and loading from the multiple sources to the 2 clouds (GME and AZURE) smoothly without any data loss, truncation with the help of an ETL Tool as an integrator. This data upload happens on a monthly basis for the Web Application. So, the users of the GIS Spatial Web Application are able to see the application with the latest data. This Web Application has a lot of functionalities like Basic Search, Advanced Search, 46 different layers spread across whole NSW and many more.

The portal provides links to Environmental Planning Instruments – including LEPs or State planning policies - on a Govt. website, which remains the authoritative source.

Technical Situation

As the web application was built on a Cloud Platform and data coming everyday from the multiple sources and different departments in multiple formats, various ETL's with complicated logics were running at different stages to standardize the whole bunch of data into one standard format and then uploading that standardized data into the Google Map Engine Cloud and the Microsoft Azure Cloud.

Adactin provided a solution that helped to check the data from multiple sources , rectifying the data with inconsistent formats, find the incompatible and duplicate entries, loss of data while ETL process was running, find the wrongly entered data and enrich the data by verifying the ETL logic is properly working through FME Workbench and verify that the all the data is transformed from the multiple sources into a one standardized normalized data at one place without dropping any data, data loss, truncation, no null records are introduced etc.

Once all the data is transformed into one standardized format, then the client transforms that data onto the 2 clouds (GME and AZURE) via the FME Workbench 2015.Here again Adactin provided a solution to check the ETL logics are working fine and transferring all the whole bunch of data to both the clouds according to various requirements and rules without any data loss and truncation, without changing the data format and disturbing the Schemas.

Client was using multiple accounts for various tool and cloud accessing, here Adactin did the Security Testing as well to check that only the authorized accounts are able to access the only allowed services and vice versa. Data between both the clouds is interrelated so a unique Asset ID was provided so that both the clouds represent the data properly in the client Web based Solution which provides the various planning services of the council's data of NSW.Adatin also performed the Production Verification Test.

Various types of testing is conducted to test this product

TestingType	TestingDescription
Data Transformation testing	This testing was done mainly for the data which was externally injectedthroughthePythonScriptsoncethedataisinone standardizedformat.Forthis,Pythoncodeisimplementedtoextract thedataandconvertitintodesirableformat.Testingteamwas involved into testing of python code logic and workflows.
ETL/Data Warehouse Testing	In this module, client was pulling the data from multiple BAU sources, appliesETLlogic,convertsitintotargetdatabaseintoone standardizedformatandthenuploadsthesamedataintoGMEand Azurecloud.Testingteamwasinvolvedintotestingof FME workbenches ETL logic, database and data upload activities.
SecurityTesting	In this module, Security Testing was done to ensure that different user accounts have the correct rights / permissions in the respective Environmentsornot.NegativeTestingwasalsodonetoverifythat the user account doesn't have any rights where it is not expected to. This was done in multiple Environments.
Cross Browser andFunctional	This product is developed for community and Business, so Testing teamtestedthisproductonvariousOSandbrowsers.

Solution

Adactin proposed and implemented test process which was divided into 5 Phases.

ETL/Data Warehouse Testing of a GIS Web based Application

Phase-1 – Test Plan

- Liaise and consult Client test team to gather complete ETL and Functional requirements

- Planning for Data Management and Transformation testing
- Testing team spent time with SME's to understand end to end business processes and business need for the ETL's
- Creation of high level test scenarios

Phase-2 – ETL Test Cases Design/Enhance

- Design of detailed test cases for system and integration ETL testing for the data Transformation
- Tools and Software's set up for the ETL data transformation, migration testing
- Test data Setup

Phase-3 – Execution Cycle-1 – ETL/Data warehouse Te sting

- o ToolsandSoftware'ssetup
- o Testdatasetupintestenvironment
- o ExecutionofETLtestcasesstagewisedescribedbelow
- o LoggingofbugsinBugManagementtool
- o Publishingoftestresults

1. Stage 1(Business Rules) – In this Stage, testing was done to verify the whole ETL data transformation was done by following the various requirements and the Business rules.
2. Stage 2 (BAU_DATA) – In this Stage, Testing was done to verify the whole bunch of extracted data from multiple BAU sources .Data Cleaning Testing was done at this stage.
3. Stage 3 (ETL_INTERNAL) – In this Stage, ETL logics were tested to verify that all inconsistent data formats from multiple BAU sources are working as expected to convert inconsistent data formats into one standardized format and are transformed completely without any data loss truncation.
4. Stage 4 (R1STD_DATA) - In this Stage, Data Migration, Transformation was done to verify that all the data which was pulled from the multiple BAU's is completely transformed into one standardized data format without any data loss, truncation.
5. Stage 5 (PCO_Application Injection) – In this Stage, Testing was done to check the whole data where the external PCO injections were inserted for the links to be provided in the web application by making the use of SQL Queries.
6. Stage 6 (ETL_EXTERNAL) – In this Stage, ETL logics were tested to verify that all the standardized data is transformed onto the GME and Azure Clouds via the ETL tool without any data loss, truncation etc.
7. Stage 7 (GME_DATA) - In this Stage, ETL Testing was done to verify that all the standardized data was transformed, migrate to the GME cloud via ETL tool(FME Workbench 2015) without any data loss, truncation, schema change etc

8. Stage 8 (AZURE_DATA) In this Stage, ETL Testing was done to verify that all the standardized data was transformed, migrate to the Azure cloud via ETL tool(FME Workbench 2015) without any data loss, truncation, schema change etc

9. Stage 9 (SECURITY TESTING) - In this Stage, Security Testing was done to ensure that different user accounts have the correct rights / permissions in the respective Environments or not. Negative Testing was also done to verify that the user account doesn't have any rights where it is not expected to. This was done in multiple Environments.

10. Stage 10 (PRODUCTION VERIFICATION TESTING) - In this Stage, Production Verification Testing was done between Standardised Source Data to GME Cloud and AZURE Cloud, via ETL Tool (FME Workbench 2015).
So, 2 paths were getting tested:

Standardised Source Data->ETL TOOL(FME DESKTOP)->GME Cloud
Standardised Source Data->ETL TOOL(FME DESKTOP)->AZURE Cloud

Phase-4 – Regression Testing Cycle for the Front En d GIS Web based Application

- Test data setup in test environment
- Execution of Regression test cases for the application
- Parallel teams to test Cross Browser and OS/Devices Compatibility Testing
- Publishing of test results

Non Functional Testing Phases

Phase-5 –Security Testing

- Test Environment, Production Environment and users setup for Information security Testing
- Design and Execution of Test scenarios
- Publishing of test results

Benefits

Find below benefits of technical solution proposed to the client:

- Teamloggedmorethan300issuesaspartoftestingphasesleadingtoimprovingthe quality of the application.
- Allkeybusinessruleswerethoroughlytestedwithinthelimitedtimeframetomakesurethat they work correctly. No Production issues recorded for those areas.
- Suggestionstoimprovetheflowofbusinessprocessweremadewiththebusinessteam,to create a better application with the clean and latest NSW data for all the upcoming monthly data uploads.

- Suggestionstoimproveinfrastructureofapplicationtoavoiddowntime Fore.g.Traffic Manager & File configuration server should have at-least 2 instances, Data types and length should be consistent on different environments
- Effectivebugtrackingprocessensuredthatopenbugscouldbeeasilytrackedandfixed.

Products and Tools we used

Tool	Usage
Microsoft Excel SQL Developer SSMS(SQL	Test case creation and execution
Server Management Studio)2014 GME(Google	Building and execution of SQL Queries
Map Engine) ArcGIS Server 10.2, ArcCatalog	Data verification and validation
FME Workbench 2015 Mantis	Spatial Data verification and validation
	Data verification and validation
	Data validation and ETL logic testing
	Defect Management and Reporting

Assistance provided by client resources

- Assistance provided by client SME's in understanding business requirements and rules for all the ETL logics.
- Client SME's conducted knowledge sharing sessions on different modules of data Transformation
- Client's IT team helped in getting access to the various tools used like FME ,GME, AZURE,SSMS 2014,ArcGIS 10.2 cloud environment etc.
- Client's Development team assisted in bug fixing.
- Client's Project Management and Test Management teams assisted in test coordination with business users and development team.