Effective Testing Process and Tools for Data Migration / Integration / DWH ETL Projects

Wayne Yaddow
Data Quality Analyst, Datagaps.com

Practitioner Track Organized by

IQ International

22nd Annual MIT International Conference on Information Quality
MIT ICIQ 2017
October 6 - 7, 2017. Hosted by the University of Arkansas at Little Rock

Effective Testing Processes and Tools for Data Migration/Integration & DWH ETL Projects

Agenda

- ETL testing issues and challenges
- The QA department solutions for ETL testing
- A delivery model framework
- Framework implementations
- Tools and methods to aid in your success
- Client / QAM collaborations for success
- Skills and experiences for ETL testing

DW/BI projects are expected to increase in number. Testing and data quality will play a major role in ensuring their success. The more prepared you are as a testing organization, the higher the success you will achieve in DW/BI programs.

Common ETL Testing Issues & Challenges

- Source data does not often meet requirements – “it’s dirty”
- Data requirements are not clearly defined - data rules for integrity, controls, security, availability and recoverability may be ill-defined
- Inadequate ETL design documents as input to test planning
- Huge, complex, source data inputs with high rates of updating
- Reductions in ETL test coverage due to complex organization of data
- Local testers lack skills for database and ETL testing (SQL, DWH architectures, data profiling)
- Business analysts, data analysts, DBA’s, ETL developers may be enlisted to test for lack of tester skills among QA teams

Ultimate Goals for ETL Testing

There is an exponentially increasing cost to businesses associated with finding defects late in the development lifecycle. Considering the importance of early detection, we list our primary goals for testing the ETL's.

Data completeness: Make certain that all expected data is loaded.

Data transformation: Ensuring that all data is transformed correctly according to business rules and/or design specifications.

Data quality: Ensuring that the ETL application correctly rejects, substitutes default values, corrects or ignores and reports invalid data.

Performance and scalability tests: Making sure that data loads and queries perform within expected time frames and that the technical architecture is scalable.

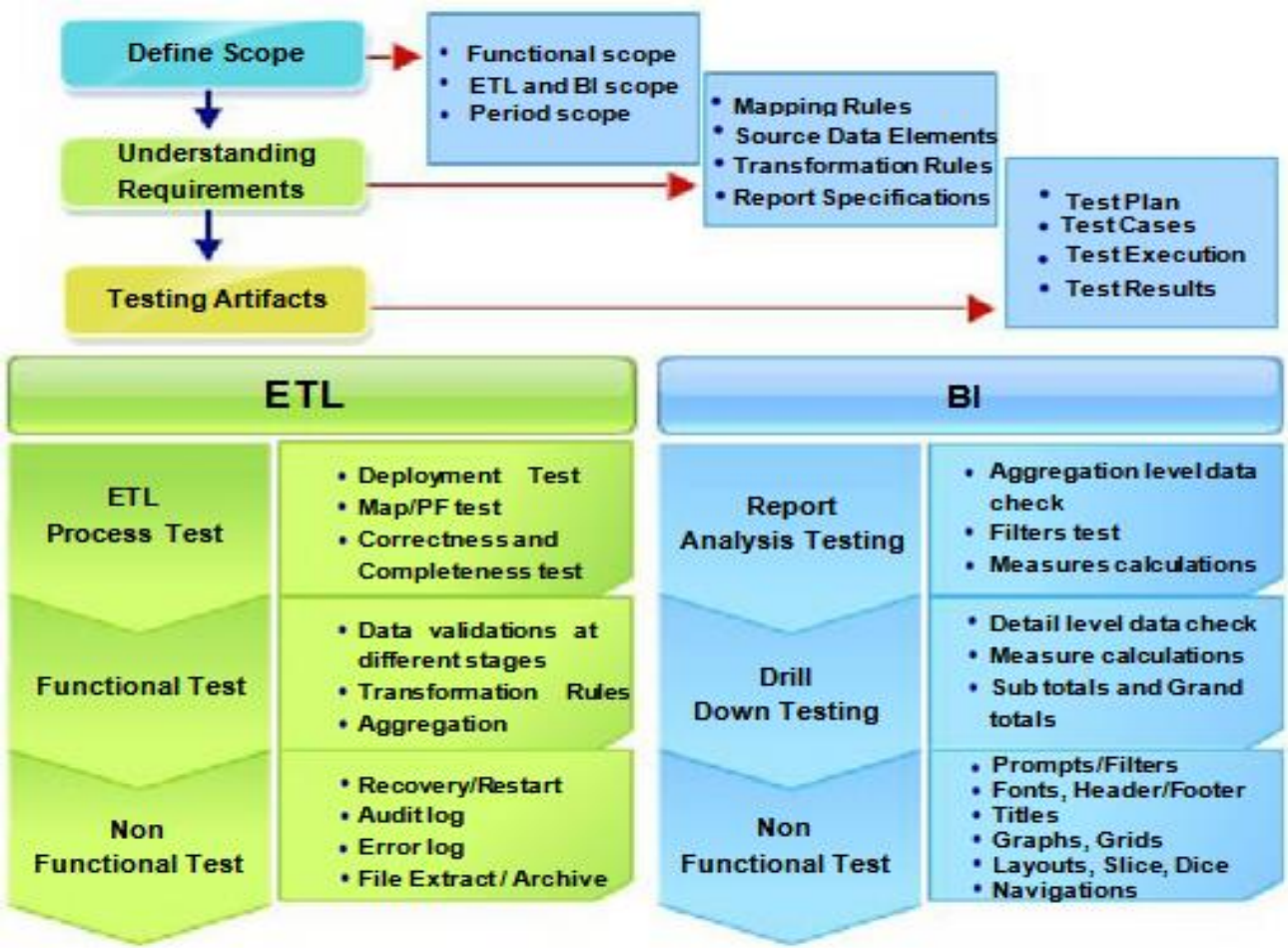
Integration testing: Ensuring that the ETL process functions well with other upstream and downstream processes.

Regression tests: Assuring that existing functionality remains intact each time a new release of code is completed

User-acceptance testing: Ensuring the solution meets users' current expectations and anticipates their future expectations

QA Department Solutions for ETL Testing

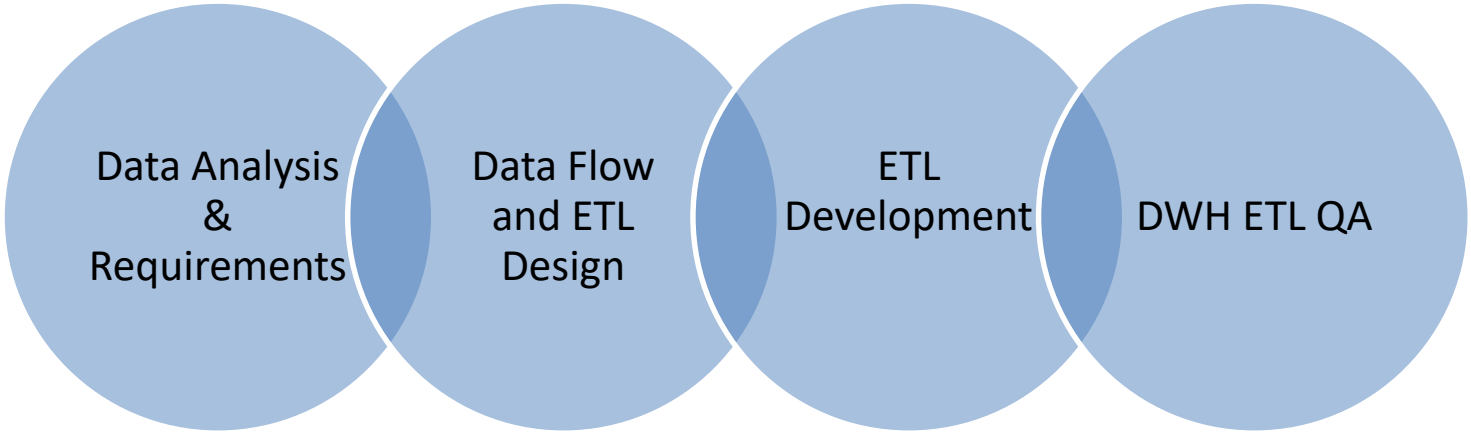
A QA solution includes testers with skills to perform complex test planning and test execution for data integration activities.

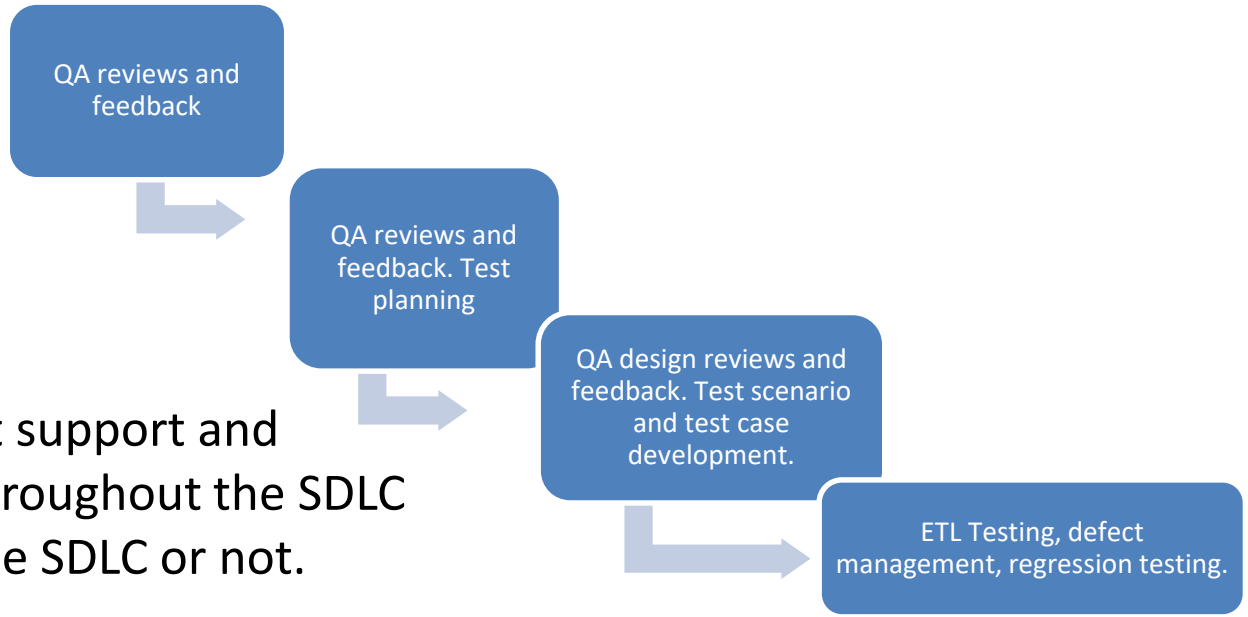


QA Skills, Process, Tooling for Success

Parameter	Requirements	Recommendations
Skills	As the skill set needed for testing at every stage of data movement is generally new, competency building correctly from understanding the technology, identifying testing needs specific to DW/BI is a first step to success.	Establish a DW/BI training lab with necessary infrastructure to support dedicated training and enablement. Establish training capsules specific to each testing type.
Test Process	By understanding the nuances of DW/BI, the project must build test process assets for every test life-cycle stage. Included are a customized test strategy covering DW/BI use cases, mining methods, reporting needs, estimation models, test data planning, test environment architecture, and non-functional testing methods and procedures.	Most testing processing asset will need to be created from scratch due to the particulars of data testing. Invest in a DW/BI data analyst and test process engineer to collaborate and spend time to build foundational QA elements.
Tooling	No overall test execution can be mostly manual. For each testing type, tooling must be established based on DW technology. For example, tooling for conversion of data transformation logic for verification purposes.	Develop an understanding of tooling needed to execute your testing based on the DW/BI technology stack. Seek out open source of vendor tool kits that suit your data testing needs.

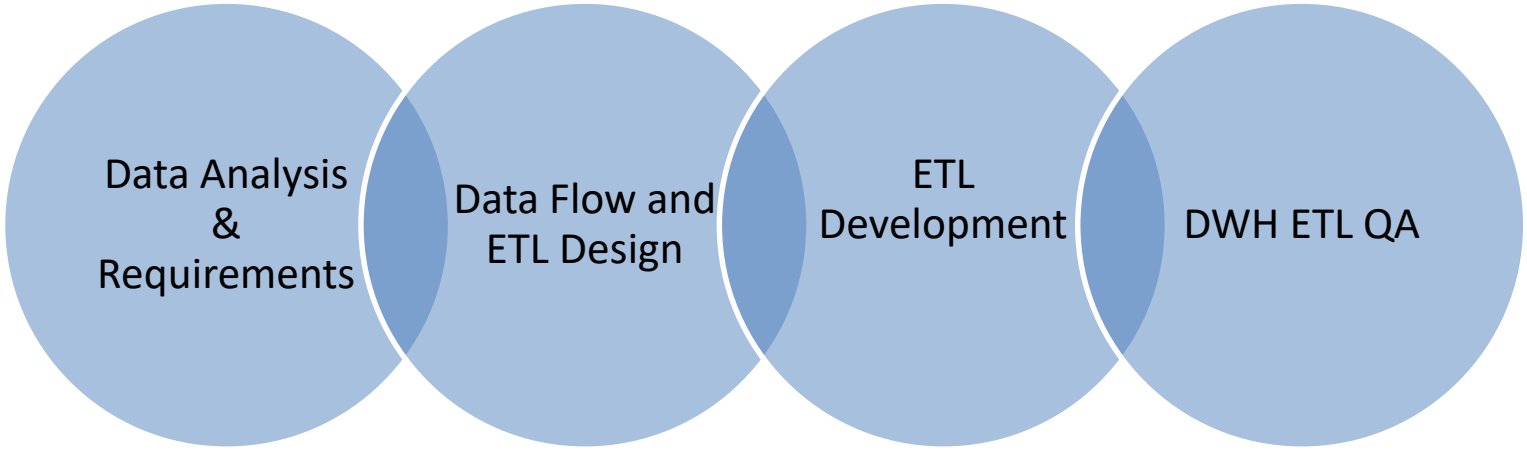
QA Framework ETL Testing Implementation





QA department support and participation throughout the SDLC whether an agile SDLC or not.

QA Framework ETL Testing Implementation

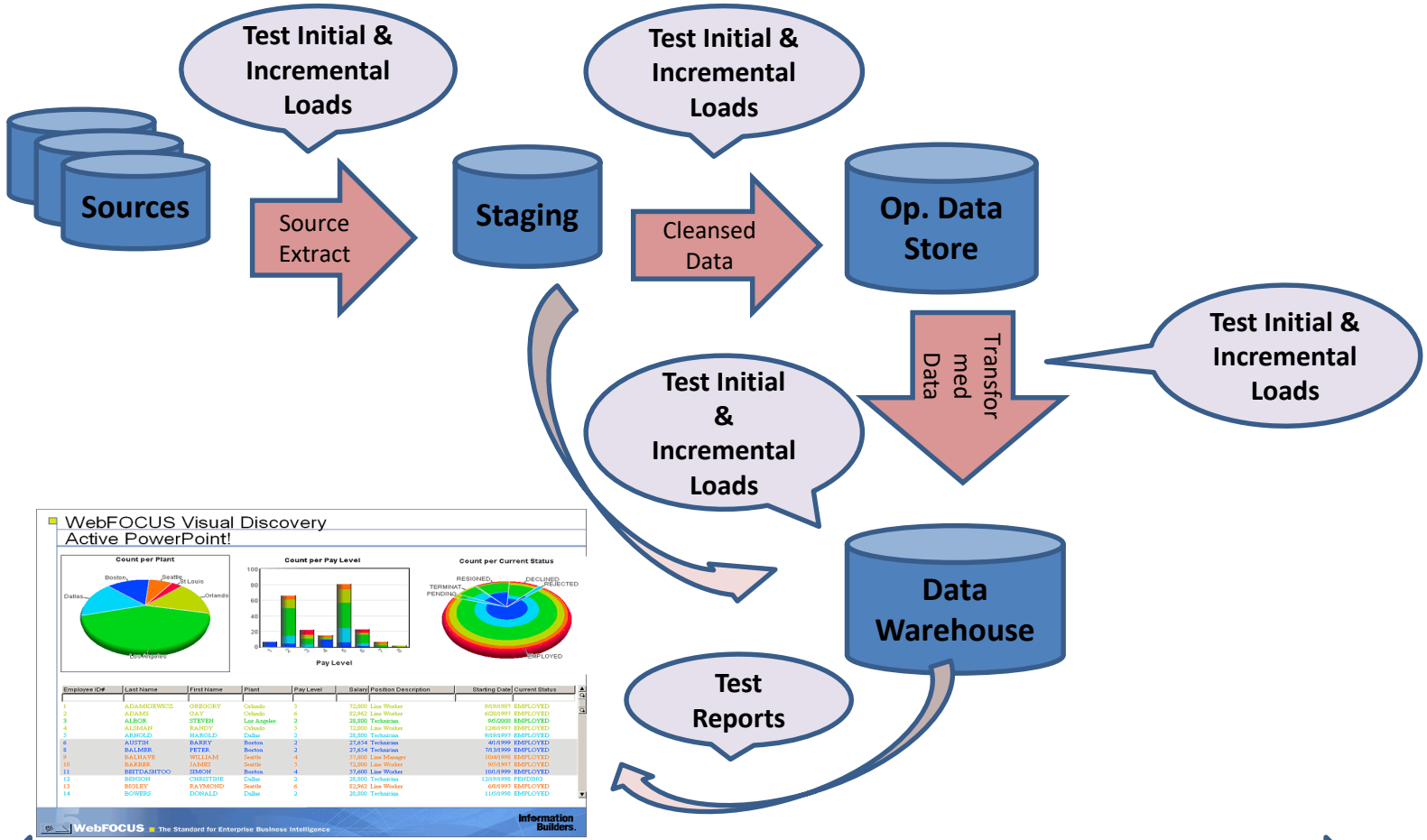


-> Develop QA approach and plan
-> Develop test scenario ideas
-> Develop test cases
-> Conduct reviews of test cases
-> Plan QA environment
-> Record test cases in defect mgt. tool
-> Conduct ETL tests
-> Test ETL and data fixes
-> Run regression tests
-> Develop test reports

QA tasks throughout the SDLC

QA department ETL Testing Solutions

The QA department ETL Testing Service assures end-to-end testing to verify every data extraction, each set of transformations and and all loads to the DWH.



Highlights of DWH / ETL Testing

ETL tests should follow guidelines listed in extensive test planning checklists that help avoid often over-looked tests. Each DWH table and associated column is tested after loads (ETL's) to staging, the DWH, and data marts.

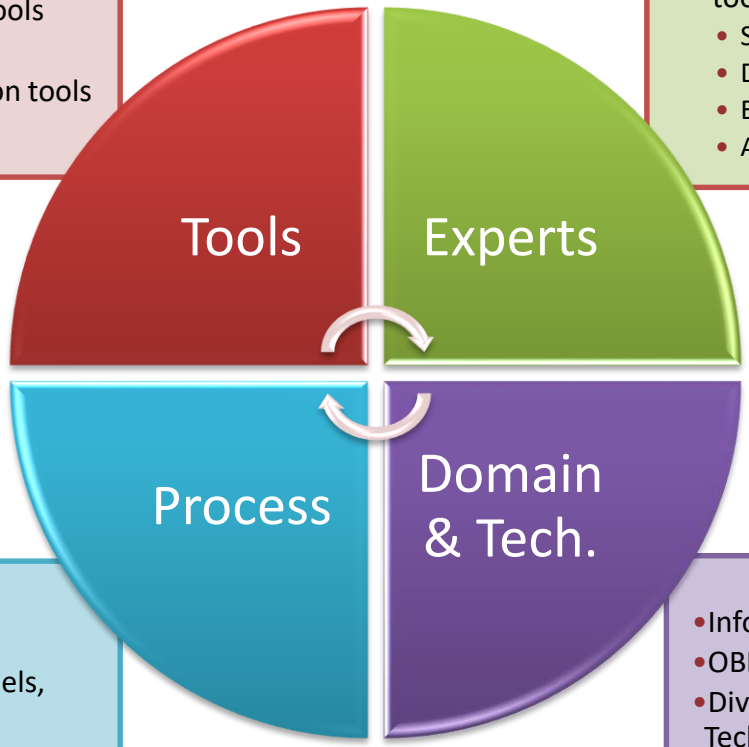
Plus…

- Validate source to target mappings
- "Profile" source & target data
- Validate views, packages, & schemas
- Verify performance SLA's
- Verify lookups for target values
- Verify default field values
- Verify relational DB keys

Important QA Tools, Expertise, Domain, Process

- Capability across all major industry tools
- Defect management tools
- Project Management
- Selection of automation tools

- Flexible pool of DWH & ETL experts
- Expertise in diverse industry tools...
  - SQL queries
  - DB editors
  - Excel, MS Access
  - Automation tools



- Agile, V-Model
- TPI, TMM
- Requirements, data models, mappings analysis
- Data integration testing
- Data profiling

- Informatica, SSIS, Oracle
- OBIEE, SSRS
- Diverse domains including Technology, Finance, Insurance, Banking

Test Tools and Methods

DB EDITORS

(Ex., Toad, SSMS, SQL Developer)

- Data profiling for value range & boundary analysis
- SQL queries
- Null field analysis
- Row counting
- Data type analysis
- Referential integrity analysis
- Distinct value analysis by field
- Duplicate data analysis
- Cardinality analysis
- Schema tests: Stored procedures, tables, views, package validation

MS Excel

- Data filtering for profile analysis
- Data value sampling
- Data type analysis

MS Access

Table and data analysis across schemas including most functionality of Excel and much more.

DWH Test Automation Tools

- **Informatica:** Data Validation Option (DVO), Informatica Data Quality (IDQ)
- **RTTS:** QuerySurge
- **Datagaps:** ETL Validator, BI Validator
- **iCEDQ**
- Others to be identified through need

Test Automation is Challenging

Few test automation tools on the market because to be useful they should provide functions that…

- Verify that ETL scripts are transforming the data as per the expected logic based on the data mapping sheet regardless of source and target platforms
- Verify that detailed and aggregated data sets are created and are matching.
- Validate the business rules and data integrity
- Check cleansing of invalid records according to business rules
- Check all table / entity integrity
  - a. Referential Integrity
  - b. Domain Integrity
- Check for data consistency: Validity, Accuracy, Usability and Integrity of related data between applications
- Check for data redundancy
  - a. Duplicate Records
  - b. Repeated data records
- Check aggregate calculations

Collaborating Outside QA for Success

What QA Teams Need for Planning and Testing

Project Time Line / Expected Date for Test Completion	We need to know what you expect so we can create the team appropriate to meet those expectations. MS Project-like project plans that include testing timelines.
Project Understanding	Interview stakeholders to analyze project technical artifacts such as system, SDLC, architecture and technologies. Identify IT and business SME's who will support our efforts.
Business Requirements	These will be a primary means for us to understand the system and create appropriate test strategoes and test plans. This should include project "vision" documents, all business requirements that identify source data, reference data, and business intelligence reports.
Functional Requirements	These will be a primary means for us to understand the system and create appropriate test plans and test cases
Architecture Documents	This type of in-depth documentation is necessary for us to delve deep into the system and test it more thoroughly. These should include data dictionaries, data models, source to target mapping documents, ETL design, system and integration test plans.
Sample Test Documentation	If you have a preferred format for the Test Plan and Test Cases, we need that template or access to the test management system you want us to use.
Defect Report Format and Defect Tool	If you'd like defects reported in a specific manner, we need that format or template, or access to the system you want us to use. Otherwise, we can use a defect management of our own.

How the QA Team Collaborates for Your Success

QA department ETL Service Deliverables

Requirements Traceability Matrix	Complete, high level, detailed test case mapping to high and low level requirements
Test Plans	A comprehensive Test Plan. This high level plan lays out our understanding of your needs and the effort required to meet them. All necessary resources are explained and validated, as well as a timeline that meets your satisfaction.
Test Cases	A complete set of manual test cases that cover every facet of the functionality. These can be delivered in whatever format and tool you request.
Defect Reports	When testing begins, defect reports will be delivered in the manner that you request. That could be end of the day summary reports, constant entry into a online based application, emails, etc.
Status Reports with Metrics	Daily status reports are delivered at least once a day and sometimes twice a day depending on the aggressiveness of the schedule and your desires. The metrics provided will list the number of test cases created in each category and their prioritization.

QA department ETL Tester Skills / Experiences

- Firm knowledge of DWH and database concepts
- Advanced expertise with…
  - SQL queries
  - Stored procedures
  - DB and SQL editors
- Expert data profiling methods & tool skills
- Exceptional skills with MS Excel / Access for data analysis

QA department ETL Tester Skills / Experiences

- Understanding of data models, data mapping documents, ETL design and ETL coding. Participate in reviews
- Experience with multiple DB systems: Oracle, SQL Server, Sybase, DB2 etc.
- Ability to troubleshoot ETL tool and stored procedure sessions and workflows
- SharePoint
- Ability to perform adequate testing with huge volumes of data. Skill to select data samples
- Strong experiences with data-centric testing
- An understanding of database code deployment

A QA Service Model Proposition

- DW project supplement the QA team with ETL test experts when you need them; the test team can be quickly ramped up or down-sized
- DW team and QA department specialists can remain focused on what we each do best
- QA department engagements can validate whether existing testing and procedures meet expectations
- The QA department solution can dramatically improve data quality and reliability while accelerating time to production and reduce cost of development
- A dedicated QA team can likely speed up and improve test results: thorough test preparation and qualified testing staff.

Thank you

Wayne Yaddow, Datagaps
wyaddow@gmail.com
(914) 466-4066