

Title of paper: Agility in DW Testing

Narendra Parihar
Microsoft
Hyderabad, India
nparihar@microsoft.com

Abstract:

“Business Intelligence Solutions are Opportunity” – Ram, Director, Business Intelligence CoE, Microsoft, India.

There are many differences between the normal testing and Agile Testing which we need to understand. This paper tries to provide info on what is the current test practices followed in IT industry when it comes to testing. What are the challenges testers and test managers face and how they can overcome them with better planning and seeing the complete picture instead of targeting bricks? It also talks about some tried DW testing agility initiatives which can reduce lot of the testing work and increase productivity of the test team....result world class product!!!

Few key things covered in this paper are:-

- Hunt for customers, catch them
- Ask 5 Why's
- Cycle time to react to customer requirements
- Behave as multiple user
- International competition to GO LIVE
- New Business Opportunities need exploration of new concepts

Keywords:

DW Testing, Agility in Testing, Pro-activeness, Think, Business Intelligence

1 Introduction

Testing in datawarehouse (DW) is bit trick. Datawarehouse includes huge set of data, and complex aggregations logics. There are different pieces of a DW like reporting, ETL (extract, transform, load), data mining, data cleansing, staging servers, Cubes; adhoc-reporting and all them require a different testing strategies. Data Quality and freshness are the two key factors for success of any datawarehouse.

- How do you ensure Data Quality?
- How do you ensure Data freshness?
- Where to start testing a Datawarehouse?
- What test processes are applicable in a datawarehouse testing?
- What should be ratio of Dev: Test in a Datawarehouse?
- What kind of testing skills you need to be a good Datawarehouse tester?
- How much QA can contribute in improving overall customer experience of reports generated from a Datawarehouse?

There are many best practices which can help you improve and decide in all above aspects, and this paper tries to

provide insights into them. Next sections in this paper will cover how to overcome all challenges using agility....

1 QA resource, test estimates, and Number of test cycles

If you are working on a Datawarehouse project, first thing test managers/leads needs to ensure is to be part of all project meetings especially there is a misconception that why testers should attend design review's or code review meetings.

-Attend all project meetings

Any kind of meeting which involves discussion around technical design or Code reviews are the key hours where testers can contribute as well, understand logic (of course they should not get biased with it☺)

Testers can provide great inputs for example; you are in a design discussion which shows show sales of a particular product in all fiscal months of this year.

Now 80% of the time Development/design team will be discussing how to implement it, and test being involved can think in terms of what customer actually want to see using this report, and suggest something like show highest fiscal month with highest sale in GREEN color. Such small things when delivered to customers make them happy as they don't have export same data in excel to filter and figure out.

Agility is the key here; look for opportunities in project discussion to give key inputs.

Test resources are a debatable topic which often Project managers and test managers end up in meetings. Datawarehouse projects require more test resources than usual web applications or traditional projects.

Example: - There is a small change in a table, so for development its change at one place, but from testing perspective, that table is used in 10-15 different calculations differently, and 10 reports, and testers have at validate each and every place.

Ideal Dev: Test ratio may be 4:3 (based on my experience)

Test estimations are not easy overall in industry and same applies in DW projects.

You can overcome using this being proactive in your estimations:-

List down all risks dependencies, most critical thing in a DW project is your ETL runtimes. Jot down how much testing effort is required for a table/column/facts/dimensions update, delete, modify, additions and your estimations will be better.

Allocate good amount of time for ad-hoc testing in cycles.

No. of Drops are something normally Project manager will communicate to you. However it should be collective decision involving PM/Dev and Test to decide. Ideally there should be at least 3 drops to ensure that you have plans to mitigate risk of more bugs found or sloppy code or bad impact analysis done

If you can pull off in drop2 is good to have.

If you can pull off drop1 to Production is great!!! (Something really want to achieve in some project) –My Manager @ MS

2 Test Processes, Challenges

Prepare your team on testers, skilled to application/technology; go for trainings etc. if required before project kickoff. Test should be involved in requirements discussion and baseline meetings (at least test lead). Proactively review the requirement, that's where you will catch most BUGS and save your company dollars!!!

Prepare test plan in detail, never assume in a test plan and have buffer in estimations and call out even a very unlikely risk(s).

Ideal time to start testcase writing is when coding a start, that's when you know all fields name and functions. Question every business logic in testcases. If a customer says I want to perform this workflow, ask 5 WHY's? Include all info that you figure out into testcases, one of the characteristics of an agile tester is he/she would write detailed testcases.

Think of automation during testcase writing itself. Capture testdata early in the cycle or create it. Sometimes you need business inputs in testdata so start the thread early.

Don't get biased with Code!!!

DW Testing challenges/Process in specific

Testing in Datawarehouse projects is a challenge as they will mostly have an ETL and reporting front ends, which differs a lot from traditional applications.

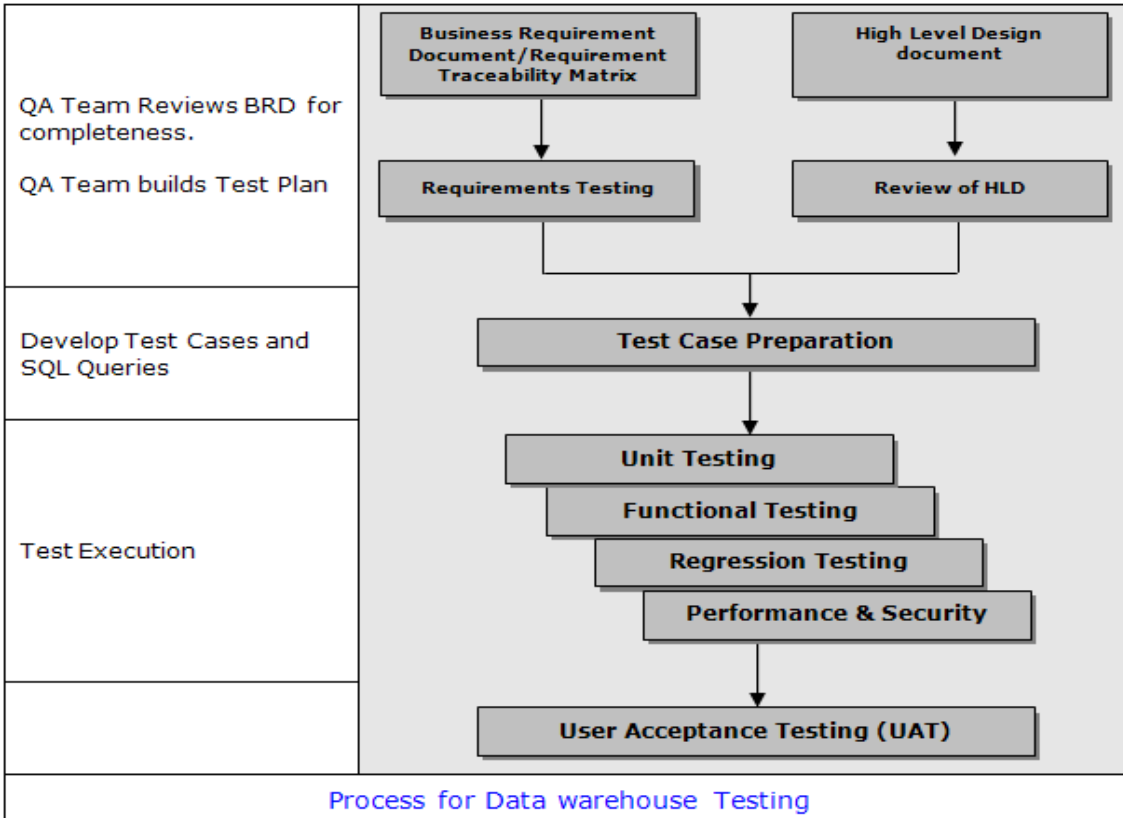
Data quality and freshness are the two key aspects of testing. The very reason users want to come to DW applications is to see reports, analyze data and take vital decisions.

The challenge lies in how to test for data in a DW which has more than 5k tables, complex business logics, and different logic for each view and then comes more unexplored part which is Cubes/complex reports with aggregates etc...

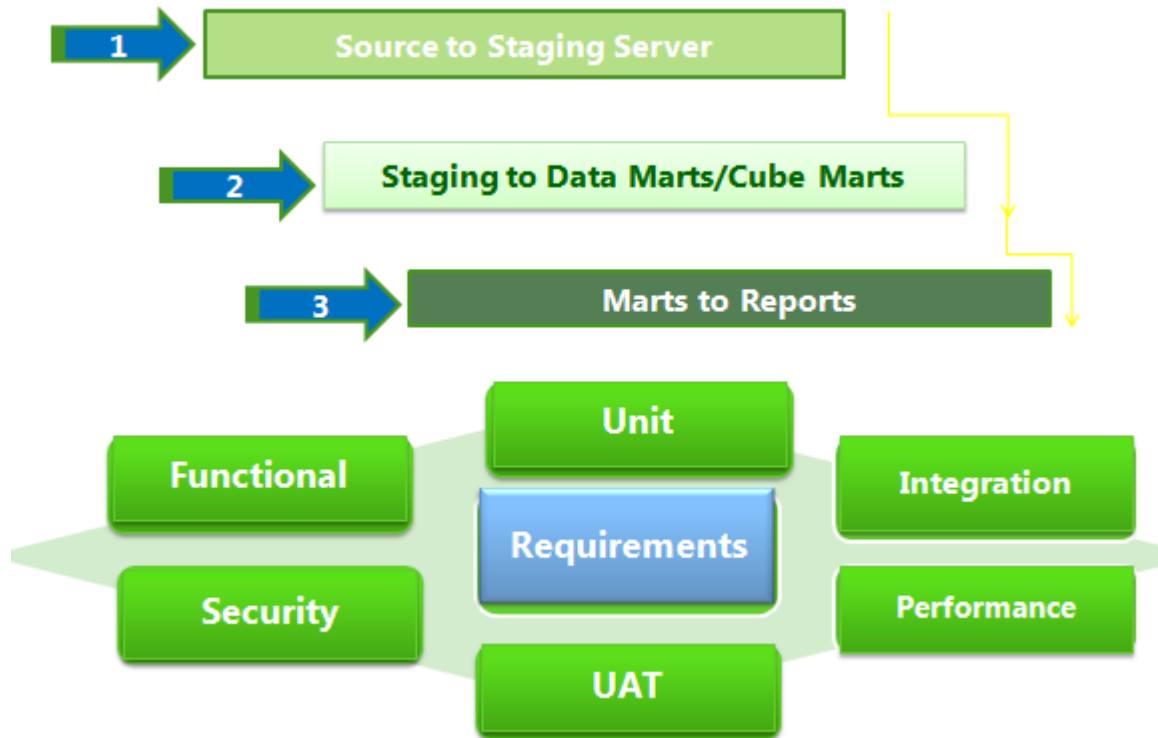
Some key points which can help you out here:-

- 1- Do your estimations with good buffer in a DW project, for example for a small change in column name, it's a small effort for development team, but for test team, that column might be used in 10 different places for joins, and it can break the entire application
- 2- Test using sampling technique, practically you can't compare 10 millions records of one table with other table, and usually there will be large number of such tables, write logical queries to test data
- 3- Test the data with data flow of the application at different stages
- 4- Automation is the mantra, use SP's and group most of the widely used SQL's
- 5- Understand the technical stuff involved in ETL's and test it

3 DW testing process and strategy



Test Strategy for a DW



3 Road from here

- Automate regression tests; don't do same stuff again and again!!!
- Make sure your estimations are correct
- Always plan for buffer/risks in DW testing
- See the bigger picture
- Train test resources to technology, DW testing is not all black box
- Think as Customers
- Improve, Improve, Improve.....

References

- DW/ETL Testing presentation by Narendra/Yatendra/Subho at Microsoft, 2008
- Wikipendia
- Janet Gregory and Lisa Crispin, Agile Testing book