Ramani's POV
QA Script Generation
Relational Data Projects
Sep 2023

Ramasubramaniam Ramani
https://www.linkedin.com/in/rramani44/

Data Project Observations

Focusing on Data centric projects with relational data stores– data lakes , migration , conversion

50% or more data movements involve no or minimal transformations - identifiers , dates , amounts get moved from layer to layer.

Summarizations - Rolled up dimensions

Business can test only a small percentage of the data during functional testing (UI , reports etc)
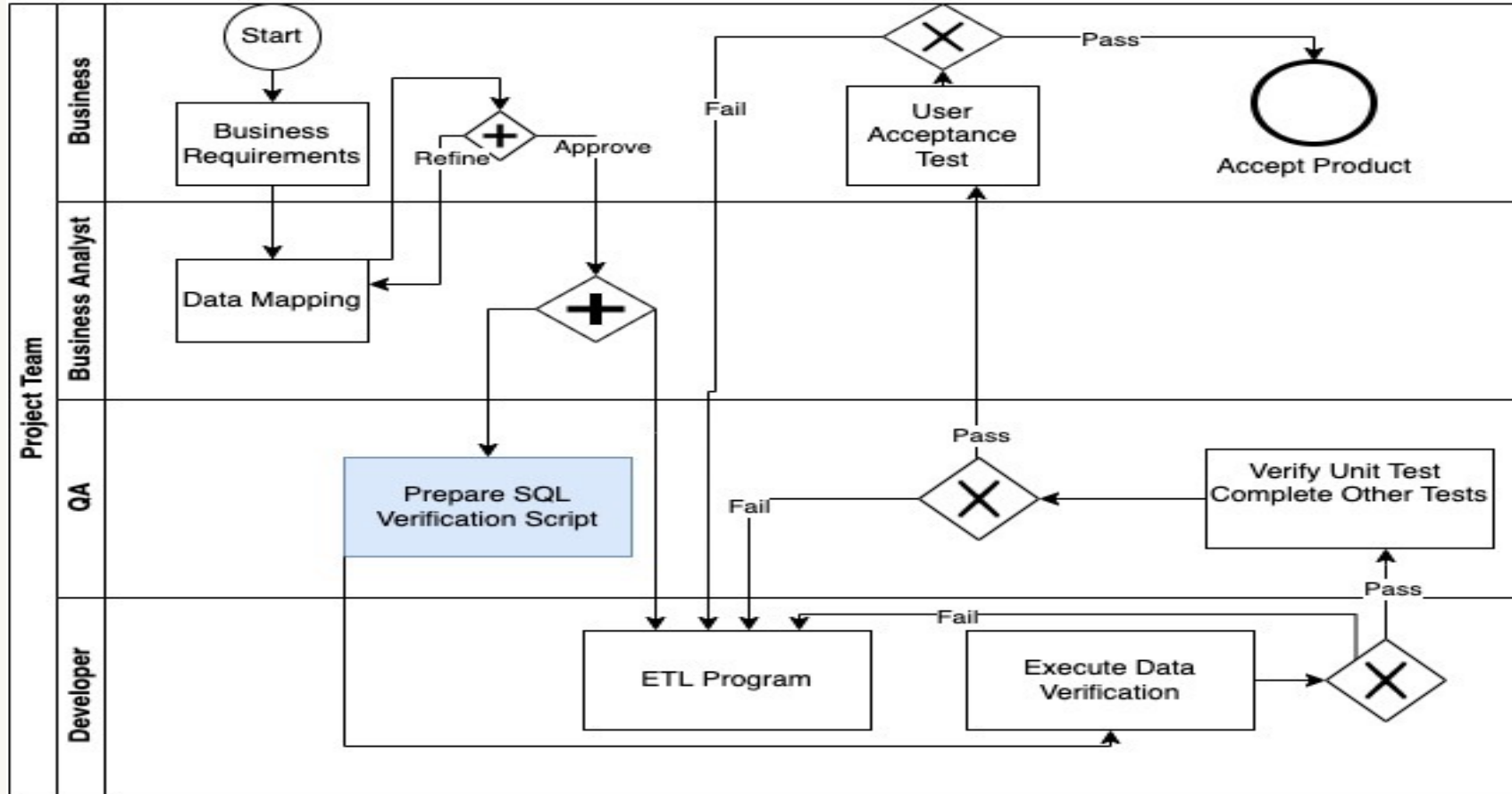
Technical team must verify the full dataset in the target – conformed (mapped) , complete , controlled (reconciled)

The QA team must certify the full dataset before business testing can begin

Illustrative Process

Data Project Delivery View

Annotation
The mapping must be approved by business. They might delegate to BA team
BA team works with business and technical teams to complete business and technical maps
Automate the SQL generation as much as possible. Currently this is very manual
Developer will unit test with a few records

Solution Features

Used with a relational database to execute the SQL

Excel macro generates SQL

Compares all rows and columns and reports differences

Straight Pull Defaults Functions

Summarizations

Foundation for Operations dashboard

Could be incorporated into a Devops pipeline

Estimated saving between ½ hour to ½ day per ETL pipeline

Sample Output

interfaceName	RRTEST1	GroupNumber		1	sourceTableName	poc.table2	targetTableName	poc.table1	sourceWhere	id <> 3		targetWhere	
sourceColumn	targetColumn	sourceTransformation	primaryKeyFlag										
id	id		Y										
myChar	myChar												
newColumn	myDefault	'ab'											
myinteger	myinteger		10										

interfaceName	groupNumber	s_id	t_id	s_myChar	t_myChar	s_newColumn	t_myDefault	s_myinteger	t_myinteger	d_t_myChar	d_t_myDefa	d_t_myinteger
RRTEST1	1	1	1	RR	RR	ab	1	1	1	0	1	0
RRTEST1	1	2	2	PR	PR	ab	1	2	2	0	1	0

Sample Output

interfaceName			RRSUM1	GroupNumber	1	sourceTableName	stateSales	targetTableName	groupsumma	sourceWhere	targetWhere	
sourceColumn			targetColumn	sourceTransformation	primaryKeyFlag / GroupBy							
stateCode			stateCode		Y							
totSalesAmount			salesAmountT	sum(salesAmount)								

interfaceName	groupNumber		s_stateCode	t_stateCode		s_totSalesAr	t_salesAmountTot	d_t_salesAmountTotal	
RRSUM1	1		TX	TX		1001	1000	1	