**Agile Process Document – How to start a Data Engineering Project**

**Definitions and Purpose**

Item – Initiative

Owner- Predefined

Purpose –

* High-level vision covering the desired and expected business goals for their delivery
* Decision made to create an initiative for each quarter of the year. So there will be 4 Initiatives per calendar year.

Sign-off – Initiative closes once related epics are completed and signed-off or once the quarter gets over, whichever is earlier. In the latter case, the incomplete epics get spilled over to the next initiative.

Item – Epic

Owner – Product Owner (PO)

Purpose –

* High-level definition of the workstream or business outcome delivering some value
* Product Owner to define the acceptance criteria at the epic level

Sign-off – Epic closes once related features are completed and signed-off or once the quarter gets over, whichever is earlier. In the latter case, the incomplete features get spilled over to a similar epic next quarter.

Item- Feature

Owner – Product Owner/ Business Analyst

Purpose –

* Definition of the requirements/ changes to be implemented
* May span multiple sprints but needs to get completed within a quarter
* MUST have an epic as a parent
* MUST include testable acceptance criteria
* Product Owners will be presented with the testing evidence, regression results by the Business Analysts and the Lead Engineers once the underlying stories are completed.
* For mapping documents, etc., sign-off to be given at the feature level by the PO post submission of proper documentation.

Sign-off – Product Owner will be giving the sign-off post reviewing the test evidences for all completed stories.

Item – Story

Owner – Lead Engineer/ QA Engineer/ Business Analyst (BA)

Purpose –

* Represent development unit of work leading to completion of the parent Feature
* MUST be completed in a Sprint
* MUST be sized with story points
* MUST have a feature as a parent
* On completion of a story, test evidences must be provided that tie to the acceptance criteria of the feature.

Sign-off – Sign-off is given by the BA (Business Analyst), Tech Lead, Lead Engineer, Test or QA Lead, Architect

Item – Bugs

Owner – Anyone

Purpose –

* To record a negative incident discovered during UAT or other testing.
* Stories must be created for fixing the bug, the bug itself is used for analysis only having a maximum of 1 story point.

Sign-off – Closed and documents updated at the Feature Level.

Definition of Ready and Definition of Done –

**For Features**, **Definition of Done** indicates conditions based on which we can call the feature to be complete and ready for Production or UAT implementation or successor work.

Example – “All items in acceptance criteria of User Stories must be met”

**For Stories, Definition of Ready** can comprise the following –

* User Story has been elaborated and is clear
* User Story has acceptance criteria
* Estimated and sized with story points.
* Story is achievable within a sprint
* Story is testable upon completion
* Any prerequisites that need to be completed prior to this story starting have already been met.
* Dependencies have been identified, and discussed with appropriate stakeholders

**Definition of Done-**

* Code is built without any error
* Unit Testing is complete
* Code review is complete
* ATF testing is passed
* Regression testing has completed
* Acceptance criteria have been met
* Signed off by the PO
* Confluence updated as required
* Ready to demo the progress

**Workflow – For implementing any Feature, the following are the standard stories that will be taken up by the development team (including Business Analysts, Architects, Technical Leads, Developers, and Test Engineers)**

1. Report Analysis – Analyzing existing reports from legacy systems to provide feedbacks or any inputs during the 2nd step of preparing or updating the mapping document. Output – Feedback for mapping document preparation.
2. Requirement Gathering/ Data Mapping & Updates – based on the feedback received during the Requirement analysis phases, the Business Analysts will design the mapping sheet (which in turn is needed by the developers to start the build phase). Output – Mapping Sheet approved by the Product Owner and ready for development work to begin.
   1. In case of pure technical stories where business analysis is not needed, The Tech Lead of Architect will gather the requirements and share the Business Requirement Document (BRD) with the Developers and QA(Test) Engineers.
3. Build Work and Unit testing in Dev (Analyzing/ Coding/ Implementing in the Development Environment)– The development work will start post obtaining the mapping sheet from the Business Analysts. The first story is to build the change/ increment and carrying out Unit Testing in the development environment.
4. UAT Execution/ Migration – The code once developed in the dev environment needs to be migrated to UAT environment as part of this story and the Unit test cases that were created and tested in Dev needs to be migrated and executed in UAT as well.
5. Functional Understanding and Manual Testing in Dev environment - As the Software Developers/ Data Engineers start on the development work mentioned in step #3 above, the QA Engineer needs to understand the change/ increment being implemented so that he/she can start creating regression scenarios to be run later using automation tools. The QA Engineer will also use this story to Manually test the code once the developer finishes the update, pushes to git and is carrying out Unit testing.
6. Creating Regression Test Scenarios – The QA Engineer, post understanding the change from Step #5 above will be creating Regression Testing Scenarios in the dev environment as part of the story.
7. Migrating and Testing Regression Scenarios in UAT – Post creating regression scenarios in the dev environment, the scenarios are migrated to UAT. Assuming the final code version is pushed to UAT by this time, the QA Engineer will execute all the regression scenarios in UAT and share the output with the Product Owner at the Feature Level.
   1. Assumption- The developer has made all changes to the code and has pushed his changes to UAT.
8. Review – Post the Regression results are available in UAT, the Business Analyst/ Tech Lead/ Architect will review the results as part of this User Story.
9. Product Owner (PO) Sign-Off – Post the Review, Product Owner will validate all the results and will approve the change to be deployed in the next deployment cycle, thereby closing the Feature.
10. For fixing any bug, all the above steps will be applicable with the only difference that the first story will be Bug Analysis (1 story point max) instead of Report Analysis.

**Story Pointing Guidelines**

Sprint Length is assumed to be 2 weeks- 10 working days

Maximum Story points per story in a sprint – 8

|  |  |
| --- | --- |
| **Story Point Sizing** | |
| Points | Max Days |
| 1 | 1 |
| 2 | 3 |
| 3 | 4 |
| 5 | 6 |
| 8 | 10 |

|  |  |
| --- | --- |
| **Feature Sizing** | |
| Size | Number of Sprints to complete |
| Small | 1 |
| Medium | 2 |
| Large | 3 |
| XL | 4 |

**How to Structure?**

**Example Usage** – Let’s have the following problem statement –

1. There is data residing in legacy systems such as Hadoop.
2. The architecture expects the data to be pushed to AWS S3, and later to Snowflake.
3. In total, data from 50 source tables are needed to be uploaded to S3 and subsequently to Snowflake.
4. Later, from the source tables at S3, data is loaded to Snowflake (called Landing stage).
5. Data Marts are required to be populated for end-consumption post transformations from landing to staging layer (intermediate) and then from staging to Gold Layer (ay) with the final marts at Gold Layer. For this example, let us consider only one Data mart needs to be created.
6. To conceptualize the data marts, Source (Landing Layer) to Target (Gold Layer) Mapping sheets need to be devised by the Business Analysts in discussion with the Data Product Owners and other Subject Matter Experts.
7. Post the mapping sheets are ready, Developers start the coding work and migrate the code to UAT.
8. Test Engineers run automated Regression Suits to test the code.
9. After development of all the features, the code is finally productionized
10. We assumed there is only one environment (Production). In practice, All work Sprint 3 onwards will be repeated for Dev, UAT and Production environments.

On the above Problem Statement, the following is likely to be the Epics/ Feature/ Story Structure –

Epic – Quarter XX

1. Feature- Loading Source Tables (50) from Hadoop to AWS S3– Comprising Specific Tech Stories such as – Size - L
   1. Sprint 1
      1. Team Onboarding and Accesses – 8 Story Points – Ownership with Client/ Business Vertical – Admin and Sales
      2. Architecture Review – 8 Story Points – Joint Ownership- Client and Consultants – Consultants (Architects) and Client (Architects)
      3. POC on Connectivity (Hadoop to AWS S3) - 8 Story Points - Joint Ownership- Client and Consultants (Developers)
   2. Sprint 2
      1. Finalizing Onboarding and Accesses – 8 Story points - Ownership with Client/ Business Vertical- Admin and Sales
      2. Finalizing Architecture - 8 Story Points – Joint Ownership- Client (Architects) and Consultants (Architects)
      3. Load the data to AWS S3 tables (1-10) – 8 Story Points – Ownership – Consultants (Developers)
   3. Sprint 3
      1. Load data to AWS S3 tables (11-30) – 8 Story Points – Ownership – Consultants (Developers)
      2. Load data to AWS S3 tables (31-50) – 8 Story Points – Ownership – Consultants (Developers)
2. Feature - Loading Source Tables (50) from AWS S3 to Snowflake– Comprising Specific Tech Stories such as – Size -M
   1. Sprint 3
      1. POC on S3 to Snowflake Connectivity – 8 Story Points- Joint Ownership- Client and Consultants (Developers and Architect)
   2. Sprint 4
      1. Finalizing S3 to Snowflake Connectivity - 5 Story Points- Joint Ownership- Client and Consultants (Developers and Architect)
      2. Loading tables to Snowflake Landing Zone (1-20)-8 Story points – Ownership with Consultants (Developers)
      3. Loading tables to Snowflake Landing Zone (21-40)-8 Story points – Ownership with Consultants (Developers)
      4. Testing (Manual/Automated) to check data consistency across Hadoop and AWS S3 – 8 Story Points – Ownership – Consultants (QA Engineer)
   3. Sprint 5
      1. Loading tables to Snowflake Landing Zone (41-50)-5 Story points – Ownership with Consultants (Developers)
3. Feature – Commence Mapping from landing to Intermediate Staging Layer and Populate Staging Tables – Size - M
   1. Sprint 5 –
      1. Commence Mapping sheet updates from Landing to Staging Layer – 8 Story Points – Ownership – Consultants (Business Analysts) and Client
      2. POC - Staging table creation at Snowflake Staging layer – 8 Story Points – Ownership – Client and Consultants (Developers)
      3. Testing (Manual/Automated) to check data consistency across Hadoop and S3 – 5 Story Points – Ownership – Consultants (QA Engineer)
      4. Testing (Manual/Automated) to check data consistency across Snowflake Landing Zone and Snowflake Staging Layer – 3 Story Points – Ownership – Consultants (QA Engineer)
   2. Sprint 6 –
      1. Finalizing Mapping Sheet from Landing to Staging Layer – 5 Story Points – Ownership – Consultants (Business Analysts) and Client
      2. Populating Staging tables with data (tables 1-20)– 8 Story points – Ownership – Consultants (Developers)
      3. Populating Staging tables with data (tables 21-40)– 8 Story points – Ownership – Consultants (Developers)
      4. Testing (Manual/Automated) to check data consistency across Snowflake Landing and Snowflake Staging – 8 Story Points – Ownership – Consultants (QA Engineer)
   3. Sprint 7 –
      1. Populating Staging tables with data (tables 41-50)– 5 Story points – Ownership – Consultants (Developers)
4. Feature - Commence Mapping from Staging to Data Mart and Populate Final Mart Tables – Size - L
   1. Sprint 7 –
      1. Commence Mapping from Staging Area to Final Mart (Gold Layer) – 8 Story Points – Ownership- Client and Consultants (Business Analysts)
      2. POC – Create tables on Final Gold Layer – 8 Story Points – Ownership – Client and Consultants (Developers)
      3. Testing (Manual/Automated) to check data consistency across Snowflake Landing and Snowflake Staging – 3 Story Points – Ownership – Consultants (QA Engineer)
   2. Sprint 8 –
      1. Concluding Mapping from Staging Area to Final mart (Gold Layer) – 8 Story Points - Ownership- Client and Consultants (Business Analysts)
      2. Finalize – Creation of tables on Final Gold Layer – 8 Story Points – Ownership – Client and Consultants (Developers)
      3. Populate Final Mart Tables at Gold Layer – 8 Story Points – Ownership – Consultants (Developers)
      4. Testing- Functional Understanding and Manual Testing – 3 Story Points – Ownership – Consultants (QA Engineer)
   3. Sprint 9
      1. Populate Final Mart Tables at Gold Layer – 8 Story Points – Ownership – Consultants (Developers)
      2. Unit Testing - 8 Story Points – Ownership – Consultants (Developers)
      3. Testing- Functional Understanding and Manual Testing – 5 Story Points – Ownership – Consultants (QA Engineer)
      4. Testing- Creating Regression Scenarios – 3 Story Points – Ownership – Consultants (QA Engineer)
   4. Sprint 10
      1. Testing- Completing creation of Regression Scenarios – 8 Story Points – Ownership – Consultants (QA Engineer)
      2. Review – 8 Story Points – Ownership – Consultants (Business Analysts)
   5. Sprint 11
      1. Testing – Running Regression Scenarios and submitting Evidence – 8 Story Points – Ownership – Consultants (QA Engineer)
      2. PO Review (Product Owner)