


























DWH_Project



















project epics

 Name	 progress	 DWH tasks	 roadblock
<u>requirement analysis</u>	1	<u>gather all the requirements for project</u>	
<u>design data architecture</u>	1	<u>choose data management approach, research and design layers, draw the architecture on draw.io</u>	
<u>project initialization</u>	1	<u>create GIT repo and design the structure, create db and schema, create detail project tasks on notion, define project naming convention, create azure account and set up env, create storage container in azure blob storage, upload files in storage container, configure the blob storage as per snowflake, set up snowflake env and configure, establish cloud storage and dwh connection</u>	
<u>bronze layer creation</u>	1	<u>Analysing: Source Systems, Validating: Data Completeness & Schema Checks, Coding: Data Ingestion, Document: Draw Data Flow (Draw.io), Commit Code in Git Repo, bring the data from cloud storage to internal snowflake storage</u>	<u>User character length limit (30) exceeded by string, Copy executed with 0 files processed, User character length limit (1) exceeded by string 'Female'</u>
<u>silver layer creation</u>	1	<u>Analysing: Explore & Understand Data, Document: Draw Data Integration (Draw.io), Coding: Data Cleansing, Validating: Data Correctness Checks,</u>	<u>data redundancy found in tables(DUPLICACY), UNWANTED SPACE IN VALUES, DATE COLUMNS</u>



 Name	 progress	 DWH tasks	 roadblock
		<u>Document: Extend Data Flow (Draw.io), Commit Code in Git Repo</u>	<u>TIMESTAMP WAS AN ISSUE.</u>
<u>gold layer creation</u>	1	<u>Analysing: Explore Business Objects, Coding: Data Integration, Validating: Data Integration Checks, Document: Draw Data Model of Star Schema (Draw.io), commit code in Git repo</u>	

DWH tasks

 project epics	 Name	 checklist
<u>design data architecture</u>	<u>choose data management approach</u>	
<u>design data architecture</u>	<u>research and design layers</u>	
<u>design data architecture</u>	<u>draw the architecture on draw.io</u>	
<u>project initialization</u>	<u>create GIT repo and design the structure</u>	
<u>project initialization</u>	<u>set up snowflake env and configure</u>	
<u>project initialization</u>	<u>create azure account and set up env.</u>	
<u>project initialization</u>	<u>define project naming convention</u>	
<u>project initialization</u>	<u>create detail project tasks on notion</u>	
<u>project initialization</u>	<u>create storage container in azure blob storage</u>	
<u>project initialization</u>	<u>upload files in storage container</u>	
<u>project initialization</u>	<u>configure the blob storage as per snowflake</u>	
<u>project initialization</u>	<u>establish cloud storage and dwh connection</u>	
<u>project initialization</u>	<u>create db and schema</u>	
<u>requirement analysis</u>	<u>gather all the requirements for project</u>	
<u>bronze layer creation</u>	<u>Analysing: Source Systems</u>	
<u>bronze layer creation</u>	<u>bring the data from cloud storage to internal snowflake storage</u>	

 project epics	 Name	 checklist
<u>bronze layer creation</u>	<u>Coding: Data Ingestion</u>	
<u>bronze layer creation</u>	<u>Validating: Data Completeness & Schema Checks</u>	
<u>bronze layer creation</u>	<u>Document: Draw Data Flow (Draw.io)</u>	
<u>bronze layer creation</u>	<u>Commit Code in Git Repo</u>	
<u>silver layer creation</u>	<u>Analysing: Explore & Understand Data</u>	
<u>silver layer creation</u>	<u>Document: Draw Data Integration (Draw.io)</u>	
<u>silver layer creation</u>	<u>Coding: Data Cleansing</u>	
<u>silver layer creation</u>	<u>Validating: Data Correctness Checks</u>	
<u>silver layer creation</u>	<u>Document: Extend Data Flow (Draw.io)</u>	
<u>silver layer creation</u>	<u>Commit Code in Git Repo</u>	
<u>gold layer creation</u>	<u>Analysing: Explore Business Objects</u>	
<u>gold layer creation</u>	<u>Coding: Data Integration</u>	
<u>gold layer creation</u>	<u>Validating: Data Integration Checks</u>	
<u>gold layer creation</u>	<u>Document: Draw Data Model of Star Schema (Draw.io)</u>	
<u>gold layer creation</u>	<u>commit code in Git repo</u>	

DATA CHECKS IN TRANSFORMATION

 Property	 error/issue
<u>silver layer creation</u>	<u>PRIMARY KEY DUPLICACY / NULL VALUES.</u>
	<u>RECORD COUNT CHECK</u>
	<u>UNWANTED SPACES IN STRINGS</u>
	<u>DATA STANDARDIZATION & CONSISTENCY</u>
	<u>REPLACE ALL NULL WITH DEFAULT VALUES</u>

Property	error/issue
	<u>CHECK FOR DATE COLUMNS(TIMESTAMP COLUMNS), THE END DATE SHOULD BE GREATER THAN THE START DATE</u>

roadblock

challenges	error/issue	resolution
<u>bronze layer creation</u>	<u>User character length limit (30) exceeded by string</u>	resolved by restructuring the table prd_info column prd_nm
<u>bronze layer creation</u>	<u>Copy executed with 0 files processed</u>	resolved with checking naming convention as files are case sensitive.
<u>bronze layer creation</u>	<u>User character length limit (1) exceeded by string 'Female'</u>	erp_cust_AZ12 restructured for column GEN
<u>silver layer creation</u>	<u>data redundancy found in tables(DUPLICATION)</u>	REMOVED BY KEEPING LATEST RECORD ONLY
<u>silver layer creation</u>	<u>UNWANTED SPACE IN VALUES</u>	TRIM FUNCTION
<u>silver layer creation</u>	<u>DATE COLUMNS TIMESTAMP WAS AN ISSUE.</u>	MANUALLY CHECK FOR RECORDS, THE START DATE SHOULD BE YOUNGER THAN THE END DATE & THUS , THE END DATE SHOULD BE YOUNGER THAN NEXT RECORD'S START DATE SO THE DATA DOESN'T OVERLAP. CONFIRM WITH THE SOURCE TEAM FOR THE SAME . USE LEAD() -1 FOR RESOLVING THE ISSUE.