Big Data Course

Capstone Project
Work Breakdown Structure

For students

Samsung Innovation Campus

Capstone Project Work Breakdown Structure (WBS)

Today	7/30/2025
Project Team Name	Group 01
Project Topic	Real-Time Big Data Processing with PySpark: NYC Taxi Trip Analysis

Task	W1					W2					W3					W4				
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20
WEEK 01																				
Week 1.1. Project Kick-off & Detailed Planning																				
1.1 Review Project Guide & Define Scope	x																			
1.2 Formalize Roles & Responsibilities	x																			
1.3. Set up Project Structure (Folders & Git)		x																		
Week 1.2. Environment Setup & Data Acquisition																				
2.1. Install Required Libraries (requirements.txt)	x																			
2.2. Install & Configure PySpark		x																		
2.3. Download Sample Dataset (NYC Taxi Parquet)																				
Week 1.3. Initial Data Exploration (EDA)																				
3.1. Create Data Exploration Notebook		x	x	x																
3.2. Load Data & Analyze Schema				x																
3.3. Perform Basic Statistical Summary					x															
Week 1.4. Reporting & Final Check																				
4.1. Compile Weekly Progress Report (.docx)					x															
4.2. Design Weekly Progress Slides (.pptx)					x															
4.3. Final Team Review & Submit					x															
WEEK 02																				
Week 2.1. Bronze Layer Implementation																				
1.1. Develop Bronze Layer Script (bronze_layer.py)						x														
1.2. Run & Validate Bronze Layer Output						x														
Week 2.2. Silver Layer Implementation																				
2.1. Develop Silver Layer Script (silver_layer.py)							x													
2.2. Implement Feature Engineering Logic								x												
2.3. Run & Validate Silver Layer Output								x												
Week 2.3. Gold Layer Implementation																				

Task								9	10	11	12	13					
3.1. Design Business Aggregations								x									
3.2. Develop Gold Layer Script (gold_layer.py)								x	x								
3.3. Run & Validate Gold Layer Output									x								
Week 2.4. Reporting & Final Check																	
4.1. Compile Weekly Progress Report (.docx)									x								
4.2. Design Weekly Progress Slides (.pptx)									x								
4.3. Final Team Review & Submit									x								
WEEK 03																	
Week 3.1. Data Quality & Advanced Gold																	
1.1. Develop Data Quality Framework (data_quality.py)										x							
1.2. Integrate & Validate DQ Framework Output										x							
1.3. Develop Gold-to-CSV Export Script										x							
Week 3.2. Analytics & Visualization																	
2.1. Develop Analytics Dashboard Notebook											x						
2.2. Create & Refine Visualizations (Plots)											x						
2.3. Run & Validate Output											x						
Week 3.3. Main Pipeline Orchestration																	
3.1. Develop Main Pipeline Script (main_pipeline.py)												x					
3.2. Integrate Batch Layers (Bronze, Silver, Gold)												x					

Task																				
3.3. Run & Validate Output													x							
Week 3.4. Reporting & Final Check																				
4.1. Compile Weekly Progress Report (.docx)														x						
4.2. Design Weekly Progress Slides (.pptx)															x					
4.3. Final Team Review & Submit															x					
WEEK 04																				WEEK 04
Week 4.1. Streaming Pipeline Implementation																				
1.1. Design Streaming Logic & Schema															x					
1.2. Develop Streaming Pipeline Script															x					
1.3. Integrate Streaming into Main Pipeline																x				
1.4. Test & Validate Streaming Output																x				
Week 4.2. Finalization & Documentation																				
2.1. Final End-to-End Pipeline Testing																	x			
2.2. Update README.md with Full Instructions																	x			
2.3. Compile Final Report (.docx)																	x			
Week 4.3. Final Presentation & Submission																				
3.1. Prepare Final Presentation Slides (.pptx)																		x		
3.2. Prepare & Rehearse Live Demo																		x		
3.3. Final Project Submission																				x
3.4. Final Presentation																				x