

Hands-On Assignment with Case Study

Azure & Data Analytics

Business Scenario – “SmartGear Retail”

SmartGear is a mid-sized electronics retailer operating in four regions (North, South, East, West). Management suspects inventory imbalances: popular items run out in some regions while slow movers pile up in others. You’ve been hired as a Junior Cloud Data Engineer to build the first slice of their Azure analytics platform.

Your mission for Week 1 is to land raw point-of-sale (POS) data in Azure, generate a quick regional sales summary, and set up the plumbing that future analytics will build on.

Dataset

File: smartgear_sales.csv – 1,000 rows, 7 columns

Column	Type	Description
OrderID	int	Unique transaction ID
OrderDate	date	yyyy-MM-dd
Region	string	North / South / East / West
StoreID	int	Branch identifier
Product	string	Item name
Quantity	int	Units sold
UnitPrice	float	Price per unit

Tasks & Deliverables

#	Task	Case-Study Angle	Points	Deliverable
A	Provision Storage	Finance wants a secure place for raw POS dumps.	10	taskA.pdf – screenshots/CLI output showing: <ul style="list-style-type: none"> Resource group rg-smartgear-dev Storage account st<initials>smartgear Container raw with the CSV uploaded
B	Databricks Quick ETL	Ops need a one-pager on “Which region sells the most units?”	15	taskB.ipynb – notebook that: <ol style="list-style-type: none"> Mounts storage Reads the CSV with spark.read.csv Aggregate total Quantity by Region Saves result as region_qty.parquet in a curated container

C	Ad-hoc Insight via Synapse	The COO wants to spot-check the parquet file without waiting for a dedicated warehouse.	10	taskC.png – screenshot of a Synapse serverless SQL query returning the parquet's top 5 rows
D	Automate with Data Factory	IT asks for a repeatable nightly ingest of raw files to a landing zone.	10	taskD.json – pipeline run output for a Copy Data activity from raw/smartgear_sales.csv → landing/smartgear_sales_<date>.csv
E	Reflection Memo	You'll brief your manager on lessons learned.	5	reflection.md (150–200 words) answering: <ul style="list-style-type: none"> • Biggest surprise about Azure tooling • A challenge you hit and fix applied • When to choose Data Factory vs. Databricks notebooks

Grading Rubrics

Criterion	Max Pts	Key Evidence
Storage setups are correct & documented	10	Proper naming, upload shown
Databricks notebook runs & parquet saved	15	Code executes, and aggregation is correct
Synapse serverless query works	10	Linked storage, result visible
Data Factory pipeline succeeds	10	JSON shows a success status
Reflection quality	5	Meets length, thoughtful answers