LinkedIn: Winlove Navarro
Github: Winlove – Course Projects

A Cloud Architecture Design Capstone Project

Professional Certification – Data Science

By:

Winlove Chryseis Navarro

2025

LinkedIn: Winlove Navarro
Github: Winlove – Course Projects

Table of Contents

LinkedIn: Winlove Navarro
Github: Winlove – Course Projects

Abstract

Fresh Deal represents a transformative digital marketplace designed to revolutionize agricultural commerce by directly connecting producers with wholesale buyers through a cloud-based data architecture. This capstone project presents the design and implementation of a scalable Microsoft Azure infrastructure that addresses critical challenges in the agricultural supply chain, including intermediary dependencies, product freshness preservation, and operational transparency. The platform integrates five primary data sources—Product Listing Database, E-commerce transactions, geolocation services, web tracking, and customer databases—through sophisticated batch and streaming ingestion pipelines. Utilizing an architecture with Bronze, Silver, and Gold tiers, the system progressively refines raw data into actionable business intelligence that powers four critical operational domains: shipping logistics, accounting systems, targeted marketing, and reporting & forecasting. The event-driven pipeline architecture incorporates robust failure handling mechanisms with exponential backoff retry logic, ensuring system reliability and data processing continuity. This integrated approach eliminates traditional distribution intermediaries, enhances producer visibility, optimizes logistics operations, and empowers stakeholders to make data-driven decisions. The resulting architecture demonstrates how modern cloud technologies can bridge traditional industry gaps, creating a transparent, efficient marketplace that benefits farmers, buyers, and the broader agricultural ecosystem while supporting sustainable and profitable farming practices.

I. Introduction

The agricultural industry continues to face significant challenges in connecting producers directly with buyers, maintaining product freshness, and ensuring operational transparency. Traditional distribution channels often involve multiple intermediaries, leading to delays, increased costs, and reduced quality. Fresh Deal addresses these issues by serving as a centralized digital marketplace where agricultural suppliers can post their products in real time, enabling wholesale buyers to browse, select, and purchase directly from producers.

At the core of Fresh Deal's efficiency is its scalable and reliable cloud-based data architecture, designed to integrate multiple data sources and sinks into a cohesive ecosystem which are then ingested through pipelines. These pipelines feed into specialized systems for accounting, shipping, targeted marketing, and real-time reporting, ensuring that every transaction and operational process is informed by accurate, up-to-date information.

This architecture leverages curation and aggregation pipelines to transform raw data into actionable insights, enabling demand forecasting, performance analysis, and personalized customer engagement. By integrating robust data management with streamlined logistics, Fresh

Deal not only enhances the visibility and market reach of local producers but also empowers stakeholders to make data-driven decisions.

II. FRESHDEAL

Fresh Deal is a next-generation digital platform built to directly connect agricultural suppliers and sellers with wholesale buyers, eliminating unnecessary middlemen and streamlining the farm-to-market process. Acting as a centralized marketplace, it brings supply and demand together in one transparent, efficient, and real-time environment.

- For Suppliers – Farmers and producers can easily list their agricultural products for sale, ensuring that new offerings are quickly added to the market pool. This speed not only maximizes product freshness but also expands the seller's reach to a wider network of buyers.
- For Buyers – Wholesale customers have access to a dynamic dashboard that reflects the market pool in real time. This allows them to browse, compare, and select products with up-to-the-minute availability and pricing information.
- Logistics – Once a purchase is confirmed, a trusted third-party delivery service takes over, collecting the products from the supplier and delivering them directly to the buyer. This streamlined logistics chain reduces delays, preserves product quality, and ensures a smooth purchasing experience.

III. Mission and Objectives

Mission

Our mission is to develop a scalable, reliable cloud architecture that streamlines every stage of the Fresh Deal platform—from product listing to delivery—while empowering farmers to sell directly to buyers. By removing unnecessary intermediaries and harnessing the power of advanced data systems, Fresh Deal aims to make agricultural commerce faster, more efficient, and more profitable for producers.

Objectives

- Enable Direct Farm-to-Buyer Deliveries – Minimize storage time, preserve peak freshness, and ensure products reach customers in optimal condition.
- Enhance Transparency and Traceability – Provide clear visibility into product origins, quality, and handling, fostering trust between farmers and buyers.
- Leverage Data for Smarter Decisions – Collect and analyze diverse datasets to optimize logistics, forecast demand, improve product offerings, and deliver performance insights that help farmers succeed.

IV. Planning Phase

The planning phase defines how data will move through the platform. This includes outlining the data diagram as the system's blueprint, identifying data sources that feed the platform, and determining the data sinks where processed information will be delivered for maximum value.

a. Data Diagram

The following data diagram illustrates Fresh Deal's information ecosystem, where multiple data sources converge to power business operations. Five primary data sources—Product Listing Database, E-commerce platform, geolocation services, web tracking systems, and Customer Database—continuously feed information into the central platform. This integrated data is then processed and distributed to four key operational areas: shipping logistics, accounting systems, targeted marketing campaigns, and reporting & forecasting analytics. This architecture demonstrates how Fresh Deal transforms raw data into actionable business intelligence across all operational domains.

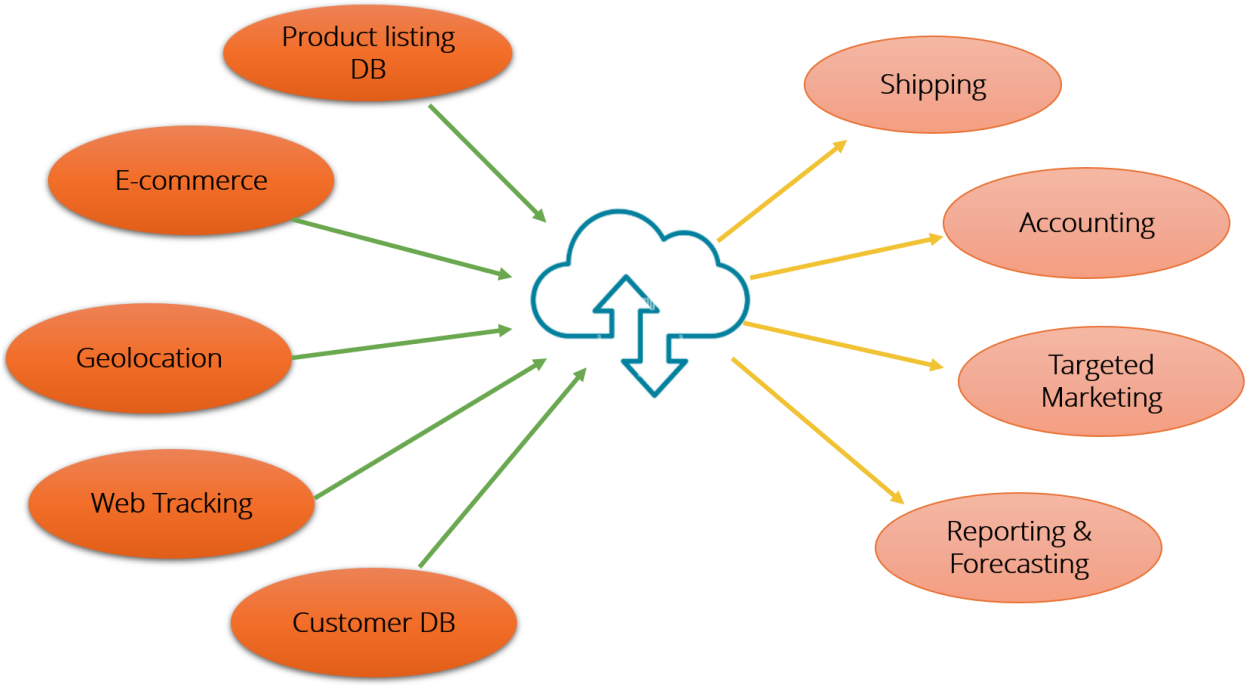

Figure 1: Data Diagram

b. Data Sources

Fresh Deal's data architecture relies on five primary input streams that collectively provide the foundation for all platform analytics and operational processes.

1. Product Listing Database serves as the central repository for supplier and seller product information within the platform's ecosystem.

   a. Nature: Structured
   b. Ingestion Type: Batch
   c. Usage: Product catalog, inventory tracking, pricing analysis
2. E-Commerce data stream captures real-time transactional activities across the platform recording critical business events including purchases, order fulfilments, and payment processing.
   a. Nature: Structured
   b. Ingestion Type: Streaming
   c. Usage: Accounting (sales analysis and revenue reporting)
3. Geolocation captures comprehensive information about customer location that primarily supports shipping operations by enabling real-time user location tracking which facilitates accurate delivery estimations.
   a. Nature: Structured and semi structured
   b. Ingestion Type: Batch
   c. Usage: Shipping (real-time user location)
4. Web Tracking captures user behaviour across Fresh Deal's platform, monitoring essential engagement metrics including page views, clicks, sessions, duration, and navigation paths.
   a. Nature: Semi structured
   b. Ingestion Type: Batch
   c. Usage: User behaviour analytics
5. Customer Database serves as Fresh Deal's comprehensive repository of customer information, containing essential data including customer identity, contact details, preferences, purchase history and behavioral patterns.
   a. Nature: Structured and semi structured
   b. Ingestion Type: Batch
   c. Usage: Marketing (segmentation, retention and personalization)

c. Data Sink

The following four data sinks represent the primary destinations where Fresh Deal's processed information is applied to drive business operations and strategic decision-making.

1. Shipping operations function as a data sink where processed order fulfillment information is transformed into logistics management which is essential for the delivery function. The focus of this approach is to enable a timely and accurate order fulfillment.
2. Accounting system serves as a data sink where financial, transactional, and operational data is processed to support critical financial management functions. The purpose of this approach is to drive profitability analysis, enable effective cost control measures, and support accurate financial forecasting.

3. Targeted Marketing system functions as a strategic data sink that transforms customer intelligence into personalized engagement strategies. By leveraging this, it will optimize the campaign effectiveness, increase conversion rates and build strong customer relationships.
4. Reporting and Forecasting system transform processed information into actionable business intelligence through visual interfaces and analytical tools. The focus of this system is to provide real-time visibility into KPIs and business trends that enables stakeholders to make accurate data-driven decisions.

V. Architectural Diagram Design

The following architectural diagram presents a comprehensive view of Fresh Deal's system design, showcasing how data sources, processing components, and operational endpoints interconnect to create a unified business intelligence platform.

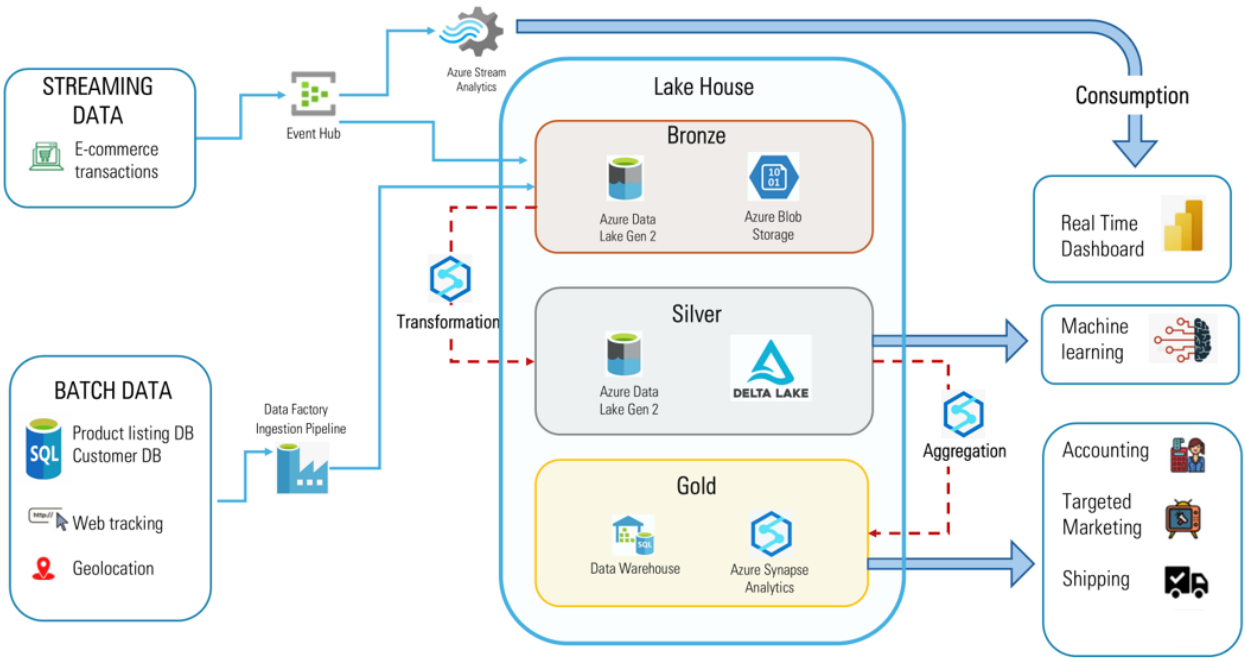

Figure 2: Fresh Deal's Architecture

It is a modern data lake architecture built on Microsoft Azure implementing an architecture pattern with bronze, silver, and gold tiers for progressive data refinement.

a. Data Ingestion

The system employs a dual-stream ingestion strategy to accommodate different data velocity requirements. Real-time streaming data, primarily consisting of e-commerce transactions, flows through Azure Event Hub with Azure Stream Analytics providing immediate processing capabilities for time-sensitive business events. Simultaneously, batch data processing handles

larger, less time-critical datasets including product listings, customer profiles, web tracking analytics, and geolocation information through Azure Data Factory's robust ingestion pipelines.

b. Lake House Architecture

The core processing infrastructure implements a three-tier medallion architecture within Azure's ecosystem which ensures data quality progressively improves while maintaining flexibility for various consumption patterns and analytical requirements.

1. Bronze Layer (Raw)

The Bronze tier serves as the initial landing zone, storing raw, unprocessed data in Azure Data Lake Gen 2 and Blob Storage while maintaining data lineage and preserving original formats.

2. Silver Layer (Curation)

The Silver tier leverages Delta Lake technology to provide cleaned, validated, and structurally consistent datasets with ACID transaction capabilities and schema evolution support.

3. Gold Layer (Aggregation)

The Gold tier represents the pinnacle of data refinement, featuring a comprehensive data warehouse powered by Azure Synapse Analytics that houses business-ready, aggregated datasets optimized for analytical queries and reporting.

c. Data Consumption

The architecture supports diverse consumption patterns through multiple endpoints designed for specific business needs. Real-time dashboards provide immediate visibility into KPIs and operational metrics for executive decision-making. Advanced machine learning models consume refined datasets to generate predictive insights and automated recommendations. Accounting systems access financial and transactional data for regulatory compliance and financial reporting, while targeted marketing campaigns leverage customer behavior analytics and segmentation data for personalized engagement strategies.

VI. Data Pipeline

Fresh Deal's data pipeline infrastructure manages the end-to-end data journey, implementing automated workflows that ensure reliable data transformation, quality validation, and timely delivery to downstream systems.

a. Pipeline Design and Approach

Fresh Deal's data pipeline architecture employs an event-driven design that coordinates batch and streaming data processing workflows through automated triggers. The system operates via two primary pathways: a batch ingestion pipeline handling product listing, customer data, web

tracking, and geolocation information, and a streaming pipeline for real-time transaction data. Upon successful ingestion, automated events trigger the curation pipeline for data cleansing and transformation, which then automatically initiates the aggregation pipeline for final processing and consumption preparation.

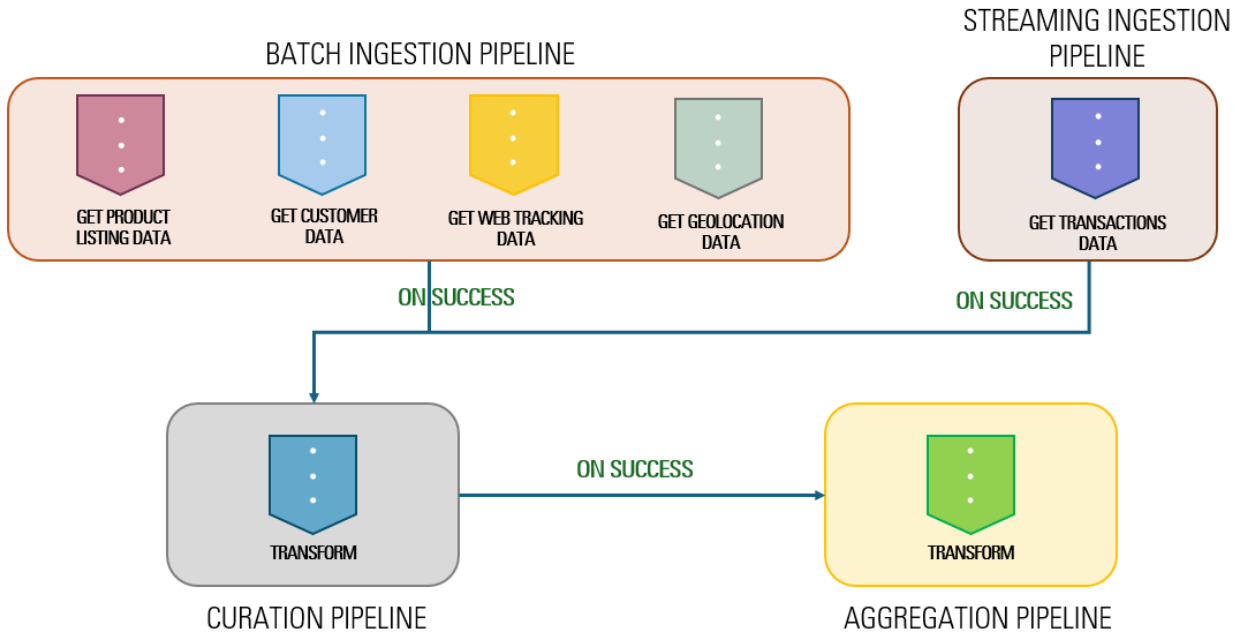

Figure 3: Pipeline Design

b. Pipeline Failure Handling

Fresh Deal implements a pipeline failure handling mechanism designed to ensure system reliability and minimize data processing disruptions. The system employs retry logic with exponential backoff, where the wait time between each retry attempt progressively increases to avoid overwhelming system resources and allow temporary issues to resolve naturally. When a pipeline task fails during execution, the system automatically initiates up to three retry attempts with increasing intervals between each attempt. If the pipeline successfully completes during any retry cycle, execution proceeds normally to completion. However, if all three retry attempts are exhausted without success, the system triggers an automated alert notification to operations teams, officially marking the pipeline as failed.

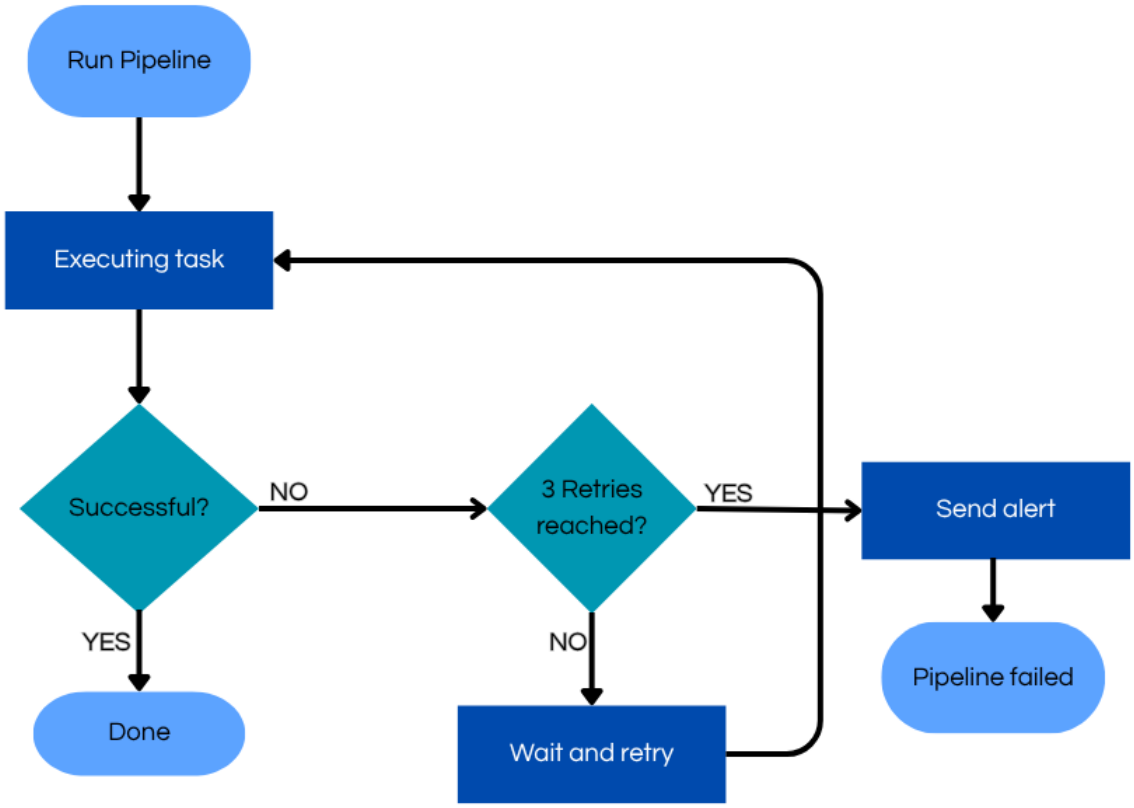

Figure 4: Pipeline Failure Handling

VII. Conclusion

The implementation of Fresh Deal's cloud architecture delivers transformative benefits by bringing transparency and efficiency to the agricultural marketplace. This data infrastructure significantly enhances the visibility of local producers, enabling them to showcase products and reach broader markets while facilitating optimized logistics operations that maintain product freshness throughout the supply chain. For buyers, the integrated system provides real-time market insights and comprehensive product information that supports informed purchasing decisions. Most importantly, this data-driven foundation empowers all departments—shipping, accounting, marketing, and strategic planning—to make evidence-based decisions that improve operational efficiency and enhance customer satisfaction.