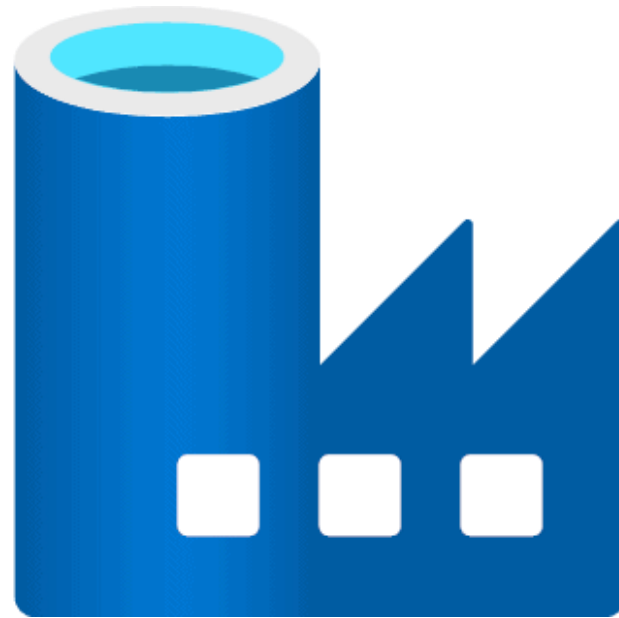


For Absolute Beginners

Azure Data Factory



What is ADF



Azure Data Factory is a cloud-based ETL (Extract, Transform, Load) and data integration service offered by Microsoft.



It allows users to create and schedule data-driven workflows (pipelines) to move and transform data from various sources.

Top-level concepts



Pipelines



Activities



Datasets



Linked services



Data Flows



Integration
Runtimes

Pipeline



FARM ANALOGY

The Complete Harvest Plan

A detailed daily schedule that coordinates everything:
"At 6 AM, send Truck #1 to Apple Farm, pick up 100 crates, sort at Warehouse A, deliver to Store #5 by 2 PM."



ADF CONCEPT

The complete workflow that orchestrates all data movement activities in a logical sequence with dependencies, error handling, and conditional logic.

Contract Hire Example:

Pipeline named "ContractHireETL" that runs daily at 3 AM to process new CSV files.

Linked Services



FARM ANALOGY

Business Contracts & Phone Numbers

Signed agreements with farms ("Green Valley Farm: Call 555-1234, Gate Code: A123") and stores ("City Mart: Manager John, Dock #5").



ADF CONCEPT

Connection information and credentials to external systems. Contains authentication details, connection strings, and access permissions.

Contract Hire Example:

"BlobStorageConnection" (storage account + access key) and "SQLDatabaseConnection" (server + database + credentials).

Datasets

FARM ANALOGY

Product Catalogs & Inventory Lists

Detailed specifications: "Red Apples: Size Large, Box of 50, Location: Section B, Row 3" or "Store Shelf: Produce Section, 4 shelves, capacity 200 items."

ADF CONCEPT

Data structure definitions that describe the shape, location, and format of your data. Includes schema, file paths, and table structures.

Contract Hire Example:

"ContractHireCSV" dataset defining CSV structure (EmployeeID, StartDate, Rate, etc.) and "ContractHireTable" for SQL destination.

Activities

FARM ANALOGY

Specific Tasks

Individual actions like "Pick 100 apple crates", "Wash and sort apples", "Load truck", "Drive to store", "Check quality", "Send email if problems."

ADF CONCEPT

Individual units of work within a pipeline. Include Copy Data, Data Flow, Stored Procedure, Web Activity, If/Else conditions, and loops.

Contract Hire Example:

Copy Activity to move CSV data, Data Flow to validate dates and rates, Email notification if errors found.

Triggers

FARM ANALOGY

Alarm Clocks & Signals

"Start harvest every day at 6 AM" (schedule), "When farmer calls saying fruit is ready" (event), or "When warehouse is 80% full" (condition).

ADF CONCEPT

Automated mechanisms that start pipeline execution. Can be schedule-based, event-driven (file arrival), or triggered by external systems.

Contract Hire Example:

Daily trigger at 3 AM, or event trigger when new CSV files are uploaded to blob storage.

Integration Runtime



FARM ANALOGY

The Truck Fleet

Different types of vehicles: local delivery trucks for nearby farms, refrigerated trucks for delicate produce, heavy trucks for bulk transport, special trucks with security clearance.



ADF CONCEPT

The compute infrastructure that executes data integration activities. Azure IR for cloud, Self-hosted IR for on-premises, SSIS IR for legacy packages.

Contract Hire Example:

Azure Integration Runtime for cloud-to-cloud data movement between Blob Storage and Azure SQL Database.

Data Flows

FARM ANALOGY

Processing Assembly Line

Detailed steps in the warehouse: "Station 1: Remove damaged fruit, Station 2: Sort by size, Station 3: Package in boxes, Station 4: Label and weigh."

ADF CONCEPT

Visual data transformation designer for complex ETL logic. Provides drag-and-drop interface for joins, aggregations, filtering, and data quality rules.

Contract Hire Example:

Data flow to validate contract dates, standardize department codes, calculate tenure, and filter out test records.

Monitoring & Logging



FARM ANALOGY

GPS Tracking & Delivery Reports

Real-time updates: "Truck left farm at 8 AM, arrived warehouse 9:30 AM, sorted 50 crates, 5 damaged, delivered to store at 1 PM."



ADF CONCEPT

Real-time visibility into pipeline execution, performance metrics, error details, data lineage, and historical run information through Azure Monitor.

Contract Hire Example:

Track pipeline execution times, number of records processed, validation errors, and set up alerts for failures.