Azure Dataset Column & Service Reference

This document explains the dataset columns available in the billing and resources tables, and provides clear descriptions of Azure services found in the dataset. It is designed to support RAG-based Q&A for cost analysis and Azure knowledge.

# Billing Table (f\_billing.csv)

\*\*invoice\_month\*\* — Month of billing in YYYY-MM format (e.g., 2022-12). Used to filter and compare costs over time.

\*\*account\_id\*\* — Unique identifier for the billing account. Helps track usage and cost per account.

\*\*subscription\*\* — Azure subscription under which the resources are billed. Logical boundary for usage and access.

\*\*service\*\* — The Azure service generating the cost (e.g., Storage, Virtual Network).

\*\*resource\_group\*\* — Logical grouping of resources in a subscription. Useful for organizing projects or workloads.

\*\*resource\_id\*\* — Unique identifier for each billed resource. Used for granular tracking of costs.

\*\*region\*\* — Azure region where the resource is deployed (e.g., westeurope). Important for regional cost comparisons.

\*\*cost\*\* — Total cost incurred for that service/resource, usually usage\_qty × unit\_cost.

\*\*unit\_cost\*\* — Cost per unit of usage (e.g., per GB, per transaction). Critical for efficiency and pricing models.

\*\*usage\_qty\*\* — Quantity of resource consumed (e.g., GB stored, vCPU hours). Combined with unit\_cost for spend calc.

# Resources Table (resources.csv)

\*\*owner\*\* — Designated owner of the resource. Useful for accountability and cost ownership.

\*\*env\*\* — Environment label (dev, test, prod). Helps separate production vs. non-production spend.

\*\*tags\_json\*\* — JSON object with user-defined metadata tags (e.g., department, project, cost center). Flexible allocation.

# Azure Services (from dataset)

\*\*Storage\*\* — Durable object/blob/file storage for data at rest. Costs include capacity, transactions, and data egress.

\*\*Virtual Network\*\* — Logical isolation of resources with subnets, routes, and security groups. Costs may include peering and traffic.

\*\*Application Gateway\*\* — Layer‑7 load balancer with WAF; charged by capacity units, hours, and data processed.

\*\*Azure DNS\*\* — Authoritative DNS hosting. Billed per DNS zone and per DNS query volume.

\*\*Bandwidth\*\* — Data transfer (egress) out of Azure or across regions; billed per GB.

\*\*Content Delivery Network\*\* — Edge caching for static content; billed for data transfer and requests.