

RAG Test Dataset — Billing Microservices

Generated: 2025-11-22 14:13:58 UTC

Purpose: Sample PDF containing Customer, Product and Billing data for Retrieval-Augmented Generation (RAG) experiments on MCP server.

How to use this PDF for RAG testing

Split the PDF into sections (Customer, Product, Billing) or by page if your pipeline supports PDF chunking. Use the tables and example JSON snippets as ground-truth documents to test retrieval and question-answering. Each invoice references customer_id and product_ids — good for entity linking tests. Fields are intentionally small-scale and clearly labeled to simplify debugging.

Customers

customer_id	name	email	country	since
CUST-001	Atlas Trading	accounts@atlas.example.com	Morocco	2020-03-15
CUST-002	Sahara Industries	contact@sahara.example.com	Morocco	2019-11-02
CUST-003	Mediterranean Foods	sales@medfoods.example.com	Spain	2021-07-21
CUST-004	Nordic Supplies	info@nordic.example.com	Sweden	2018-01-05
CUST-005	Oceanic Imports	procurement@oceanic.example.com	Portugal	2022-06-30

Products

product_id	name	sku	price	currency
PROD-001	Granular Fertilizer 25kg	GF-25	35.50	USD
PROD-002	Liquid Fertilizer 10L	LF-10	22.00	USD
PROD-003	Industrial Pump	PUMP-7	450.00	USD
PROD-004	Conveyor Belt Segment	CB-01	120.75	USD
PROD-005	Packaging Box (100pcs)	BOX-100	12.00	USD

Invoices (Billing)

invoice_id	customer_id	date	due_date	total_amount	currency	status
INV-1001	CUST-001	2025-07-12	2025-08-12	379.00	USD	PAID
INV-1002	CUST-003	2025-09-01	2025-09-30	450.00	USD	UNPAID
INV-1003	CUST-002	2025-10-05	2025-11-05	500.00	USD	PARTIALLY_PAID
INV-1004	CUST-004	2025-04-20	2025-05-20	362.25	USD	PAID
INV-1005	CUST-005	2025-11-10	2025-12-10	397.50	USD	UNPAID

Example JSON snippets (useful for indexing)

Customer JSON:

```
{ "customer_id": "CUST-001", "name": "Atlas Trading", "email": "accounts@atlas.example.com", "country": "Morocco", "since": "2020-03-15" }
```

Product JSON:

```
{ "product_id": "PROD-001", "name": "Granular Fertilizer 25kg", "sku": "GF-25", "price": 35.5, "currency": "USD" }
```

Invoice JSON:

```
{ "invoice_id": "INV-1001", "customer_id": "CUST-001", "date": "2025-07-12", "due_date": "2025-08-12", "items": [ { "product_id": "PROD-001", "qty": 10, "unit_price": 35.5 }, { "product_id": "PROD-005", "qty": 2, "unit_price": 12.0 } ], "currency": "USD", "status": "PAID" }
```

Notes:

This PDF is intentionally small and structured. For production-like testing, add more records and include edge cases: missing fields, large textual descriptions, attachments links, and multi-language content.