



TEAM: B

# Receipt & Invoice Digitizer

Milestone 4: Polishing & Integration

# Our Team



**Vaibhavi D.**

PPT Doc.



**Pranitha Piduru**

API design



**Goutham Mourya**

Template based Parsing



**Tisha Gurjar**

API testing



**Niharika**

Documentation



**Mayank Kumar**

Team Lead

# Problem Statement

Manual receipt and invoice management remains inefficient due to inconsistent formats, extraction inaccuracies, and limited data filtering capabilities.

There is a need to:

- Improve system accuracy
- Enhance search and filtering
- Optimize database performance
- Integrate with accounting or ERP systems

# **OBJECTIVES OF MILESTONE 4**

Milestone 4 focuses on enhancing system accuracy, usability, and integration readiness.

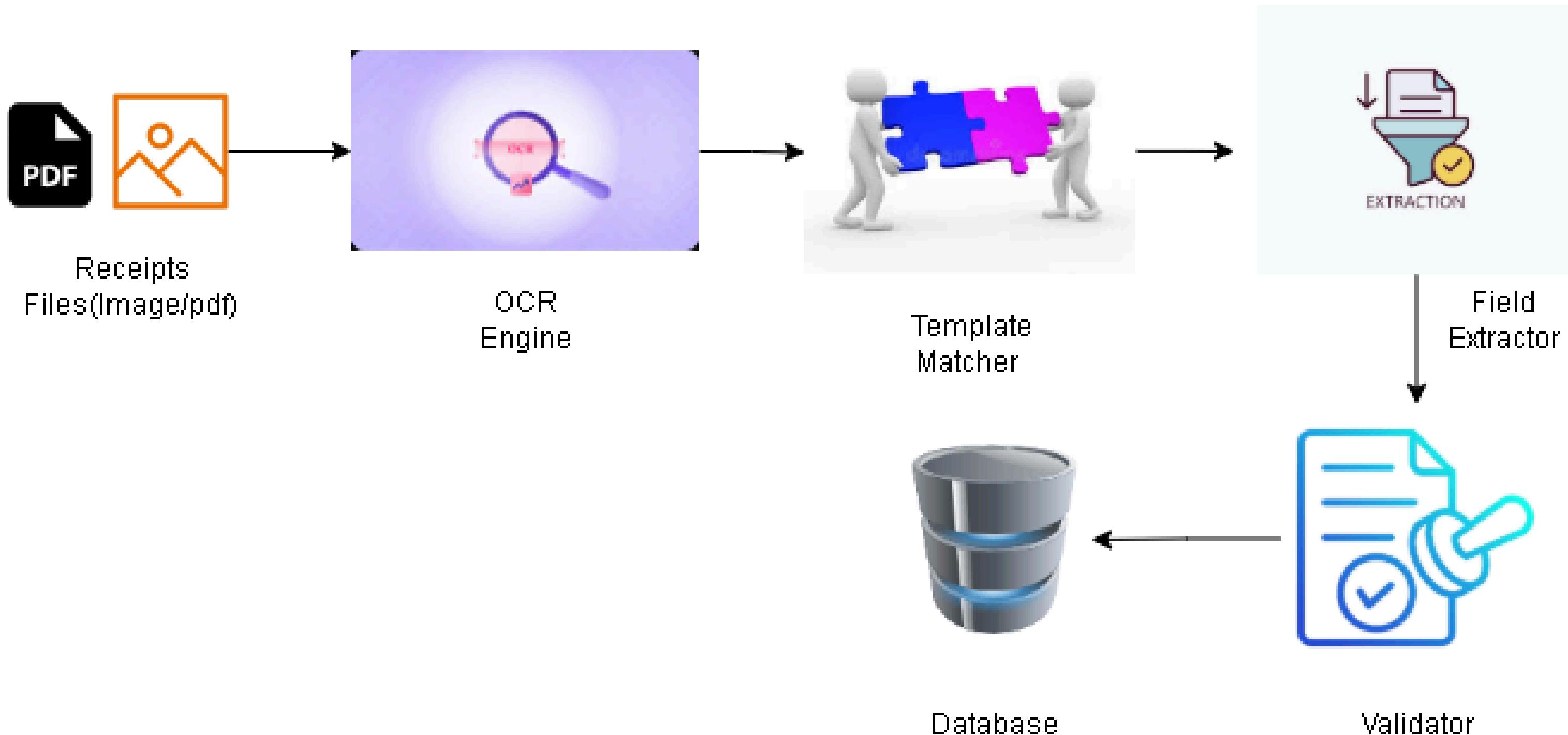


**The objectives are to:**

- ERP system integration
- Template-based parsing enhancement
- Advanced search & filtering
- Database query optimization
- Production deployment improvements

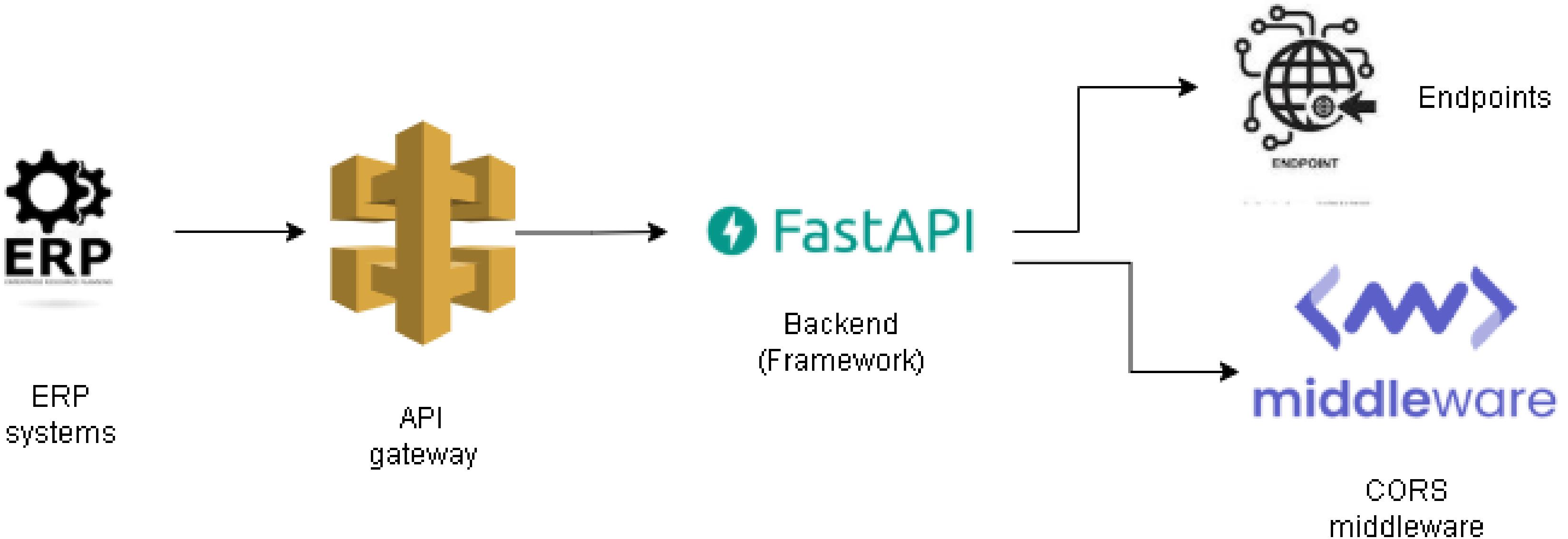
# WORKFLOW DIAGRAM

## Template based Parsing



# SYSTEM ARCHITECTURE

## ERP system Integration



# TEMPLATE-BASED PARSING ENGINE

Improves extraction accuracy for recurring vendor formats.

- Analyze spatial relationships between fields (vendor name typically at top, total at bottom)
- Store coordinate ranges for common fields across similar receipts
- Build confidence scores based on historical accuracy
- Support both rigid templates (exact layouts) and flexible templates (relative positioning)

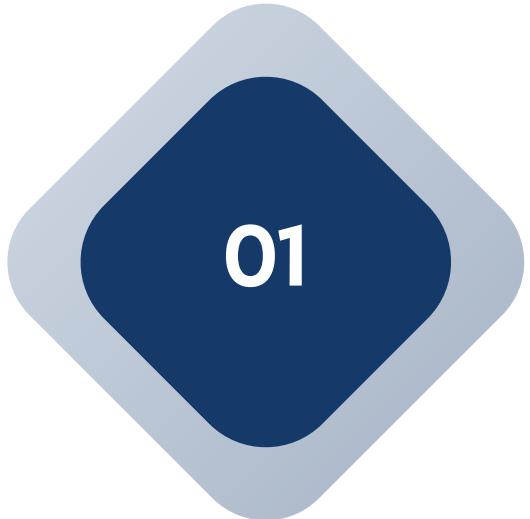
# RESULTS

Results: Average Confidence of  
items>80%

e?	Bill ID	Vendor	Date	Total (₹)	Tax (₹)	Subtotal (₹)	Category
□	832869	Unknown Vendor	2019-02-11	₹55.39	₹3.44	₹51.95	Food
□	1022	Chef Wang	2019-02-02	₹35.52	₹2.82	₹32.70	Food
□	24016	Longboards	2019-01-11	₹41.31	₹1.86	₹39.45	Food
□	168	WINGS & THINGS	2017-03-19	₹9.50	₹0.75	₹8.75	Food
□	3100 / 3525	CARL'S JR	2016-12-19	₹13.05	₹0.97	₹12.08	Food
□	Order 97	McDonald's Restaurant #3100	2013-02-27	₹7.27	₹0.63	₹6.64	Food

# ADVANCED SEARCH & FILTER MODULE

Enhances navigation of large transaction datasets.



01

Search by vendor /  
invoice number



02

Filter by date / amount /  
category



03

Multi-criteria filtering



04

Improved record  
navigation

# **DATABASE OPTIMIZATION**

Improves query performance  
and system responsiveness.

STRATEGIC INDEXING

QUERY CACHING

CONNECTION POOLING

REDUCED LATENCY

# ERP INTEGRATION LAYER

Enables communication between the digitizer and external ERP systems.

**ERPNext**, a open Source ERP system.

Supports ERPNext

API-based synchronization

Secure authentication

Maintains data integrity

# RESULTS

## Receipt Vault Analyzer API 1.0.0 OAS 3.1

[/openapi.json](#)

REST API for ERP integration and external data access

### default

<code>GET</code>	/ Read Root
<code>GET</code>	/api/v1/receipts Get Receipts
<code>GET</code>	/api/v1/receipts/{bill_id} Get Receipt
<code>POST</code>	/api/v1/erp/sync Sync To Erp
<code>POST</code>	/api/v1/ocr/process Process Image

## ERP system API'S design

### Schemas

`ERPExportResponse` > Expand all `object`

`HTTPValidationError` > Expand all `object`

`ReceiptBase` > Expand all `object`

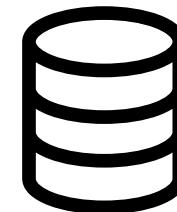
`ValidationErrorResponse` > Expand all `object`

# FRAMEWORKS & Libraries

Technologies supporting  
integration and optimization.



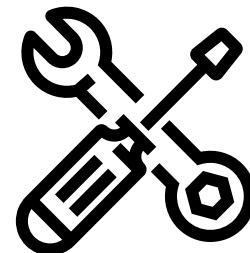
**Streamlit**



**SQLAlchemy ORM**



**pytesseract +  
PaddleOCR**



**Numpy + Pandas**



**ERPNext (Frappe)**



**Fastapi HTTP**

# SYSTEM IMPACT

Enhancements significantly improve scalability and usability.

FASTER DATA ACCESS

BETTER ACCURACY

IMPROVED RELIABILITY

INTEGRATION-READY  
ARCHITECTURE

# CONCLUSION

Milestone 4 successfully refines the Receipt & Invoice Digitizer by strengthening ERP integration, improving template-based accuracy, enabling advanced search capabilities, and optimizing database performance. The system is now prepared for enterprise-level deployment.

# THANK YOU

