

Pranay Ratan

# Data Migration Specialist

## Technical Challenge

## Migration Plan



01

## Data Cleanup Plan



02

## Post Migration QA & Debugging



03

## High-Volume Optimization



04

## Live Issue Response & Communication



05

## My Migration



06

# Migration Plan

- » *Inventory All Objects:*  
Products, customers, orders, discounts, etc.
- » *Dependency Mapping:*  
Magento  $\leftrightarrow$  ERP  $\leftrightarrow$  CRM
- » *Downtime Planning:*  
Confirm timing (e.g., Fri 10 PM  $\rightarrow$  Sat 4 AM UTC)
- » *Success Criteria:*
  - Record count match:  $\pm 0.1\%$
  - Financial totals:  $\pm 0.01\%$
  - No API 5xx errors
  - SEO & URL slugs preserved

Layer	Stack / Tool
Extraction	Python, MySQL connector, Magento API
Staging & Transform	dbt, Pandas, PostgreSQL, AWS S3
Load	Shopify REST & Bulk GraphQL APIs
Orchestration	Prefect 2.0 / Apache Airflow
Validation	Great Expectations, pytest, custom scripts
Monitoring	Datadog, Grafana, Slack alerts
Rollback	Shopify exports, RDS snapshots

## » Pre-Cutover Workflow

*T - 7 days to T - 1 hour*

- Extract: SQL dumps and API pulls
- Transform: Flatten EAV schema to Shopify format
- Clean: Use Pandas and dbt
- Enrich: Add tags, metafields, and URL redirects
- Load: Sandbox test via GraphQL Bulk API
- QA: Great Expectations, spot checks, checksum comparisons

## » Delta Migration

*T - 1 hour to Cutover*

- Capture last-minute changes via timestamps and webhooks
- Minimal ETL for deltas (updated records only)
- Ensure up-to-date sync before final cutover

## » Cutover Execution

### *T O – Downtime Window*

- Put Magento in maintenance mode
- Final delta data upload to Shopify
- Trigger ERP/CRM sync
- DNS/CDN switch to Shopify storefront
- Validate everything before reopening store

## ➤ ERP/CRM Sync Strategy

- ID Mapping: Magento  $\leftrightarrow$  Shopify cross-reference tables
- Inventory: Shopify webhooks  $\rightarrow$  AWS Lambda  $\rightarrow$  ERP update
- Orders: Past = read-only, future = real-time API sync
- Customers: Nightly GraphQL syncs to ERP

- Validation & Rollback
- Automated QA:
  - Record count
  - Financial totals
  - Sample hash matches
  - Performance Test: Simulate high user load
- Rollback Plan:
  - Trigger on API error >2%
  - DNS rollback to Magento
  - Shopify store import revert if needed

# Data Cleanup Plan

- » *Python Libraries:*  *pandas, numpy, regex, phonenumbers, openpyxl*
- » *Script Workflow:*



# Post-Migration QA & Debugging

- **Validation Tools:**  
pandas, SQL, Shopify admin
- **Checkpoints:**
  - Compare Magento vs Shopify data (orders, inventory, prices)
  - Match records using SKUs and customer IDs
  - Visual inspection and automated financial checks
- **Fixing Sync Issues:**
  - Targeted inventory updates via SKU match
  - Order history validation through data logs
  - Real-time or manual fixes for discrepancies
  -
- **Price Correction:**
  - Use CSV bulk re-import or Shopify Admin edits

# High-Volume Optimization

➤ **Goal:**  
Complete migration within 6-hour window

➤ **Strategy:**

- Use Shopify Bulk API for high-speed uploads
- Batch data to maximize throughput
- Limit parallel requests to avoid hitting rate limits

➤ **Resilience:**

- Implement logging for failed records
- Auto-retry mechanisms with delay and logging

# Live Issue Response & Communication

➤ **Scenario:**  
Wrong prices live on store

➤ **Immediate Actions:**

- Lock store temporarily if severe
- Identify and isolate affected SKUs
- Deploy hotfix via bulk CSV upload

➤ **Internal Slack Message:**  
"Hey team – live store has incorrect prices. Fix in progress. ETA 30 mins. Will confirm once done."

➤ **Post-Fix:**

- Revalidate products
- Document issue & mitigation steps

# My Migration Experience

- **Project:** Clinical trial dataset consolidation
- **Problem:** Inconsistent, incomplete data across multiple sources
- **Tools Used:**
  - Python (pandas, requests, BeautifulSoup, regex)
  - Custom web scrapers and data normalizers
- **Approach:**
  - Automated scraping, parsing, deduplication
  - Validation with test cases and pandas
- **Outcome:** Unified, clean dataset  
Enabled high-quality analysis and insights