

Hi guys, Following up on our discussion about building a lead generation pipeline using Meta's Ad Library, I've put together this technical scope for the automation project you'll be working on with Muhmd Farhan.

Project Overview

We're building a two-stage automation pipeline that identifies businesses actively advertising on Meta platforms and enriches their data for outreach. This approach targets companies with proven marketing budgets, making them higher-quality leads.

Technical Architecture

Stage 1: Meta Ad Library Extraction

Primary Options:

1. **Apify Actors** - Check these specific actors:
 - "Facebook Ads Library Scraper" by alexey
 - "Meta Ad Library Scraper" by microworkers
 - Compare their input/output capabilities against our requirements
2. **Custom Solution** - If existing tools don't meet our needs:
 - Build using Meta Ad Library API (https://developers.facebook.com/docs/marketing-api/reference/ads_archive/)
 - Python script with rate limiting and error handling
 - Deploy on AWS Lambda or similar for scalability

Required Implementation:

- Input validation for location, keywords, and active status
- Pagination handling for large result sets
- Data deduplication (companies may have multiple ads)
- Export to CSV with proper encoding

Stage 2: Data Enrichment Pipeline

Recommended Stack:

1. **Primary enrichment:** Clay.com (best balance of accuracy and API flexibility)
2. **Fallback options:**
 - Apollo.io for B2B data
 - Clearbit for company information
 - Custom Facebook page scraper using Apify/ScrapingBee for direct extraction

Enrichment Logic:

For each company:

1. Try Clay API with Facebook URL
2. If no results, try with company name + location
3. If still missing data, queue for manual scraping
4. Validate phone numbers (especially for WhatsApp compatibility)

Implementation Phases

Phase 1 :

- Evaluate and test Apify actors
- Set up API keys and authentication
- Build proof of concept with 50-100 companies

Phase 2:

- Integrate enrichment APIs
- Build error handling and retry logic
- Create monitoring dashboard

Phase 3:

- Full automation deployment
- Documentation and handover
- Performance optimization

Technical Considerations

1. **Rate Limits:** Meta API has strict limits - plan for 200 requests/hour
2. **Data Quality:** Implement validation for:
 - Email format (RFC 5322 compliant)
 - Phone number format (E.164 standard)
 - Deduplication based on company name fuzzy matching
3. **Cost Optimization:**
 - Cache enrichment results to avoid duplicate API calls
 - Batch process during off-peak hours
 - Monitor API usage to stay within budget

Let me know if you need any clarification or have questions about the technical approach. We should be able to have a working prototype by end of next week.

P.S. - Make sure to implement proper data retention policies given we're handling business contact information. Legal has approved 90-day retention for uncontacted leads.

