BRIGHTGROVE

Fast, safe, reliable and adaptable
business partner to help your business evolve

CASE STUDY

Amazon web data processing tool PoC creation for an e-commerce company

Location *Sweden*	***Industry*** *E-commerce*
Tenure *from 2022*	***Delivery centers*** *Ukraine*
Tech stack *Node.js, ReactJS, Python, LAION, BERT language model, Google Cloud, Firestore, BigQuery*	***Team size*** *8 specialists: Product Owner, Full-stack Tech Lead, Data Engineer, Computer Vision Engineer, NLP Engineer, UX Designer, ML Solutions Architect, Delivery Manager*

CHALLENGE

The client needed a smart solution that could analyze thousands of Amazon marketplace products to identify optimization opportunities. The system needed to process multiple data points including product images, descriptions, reviews, and market performance metrics to generate actionable recommendations for sellers.

Faced several critical challenges:

- Manual product optimization was becoming unsustainable with their growing portfolio (500+ products across multiple categories)
- Inconsistent decision-making in product updates due to lack of data-driven insights
- 20–30% of products were underperforming due to suboptimal listing attributes
- Limited ability to identify and react to market trends in real-time
- Need to scale operations while maintaining quality of recommendations

HOW WE HELPED

PoC creation

The Brightgrove team was fully responsible for the E2E creation of a PoC to help the client identify if the offered technical solution could cover his business requirements. This PoC involved the following phases:

- Focus on core functionality validation
- Use managed services to reduce operational overhead
- Implement basic monitoring
- Limited but functional error handling

Taxonomy and categories implementation

We developed an AI-powered solution that analyzes product data, identifies trends, and provides recommendations to optimize listings and boost sales. For instance, if a seller offers black 0.5L water bottles, the system analyzes similar products across marketplaces, identifies trends in consumer preferences, and might recommend optimizing to green 0.75L bottles based on market data and customer sentiment analysis.

Language models and image-to-taxonomy encoding

For the PoC and further development, we needed to research language models capable of encoding text data—product descriptions and other metadata—to taxonomy representation.

We needed to evaluate various language models by several criteria:

- Model size
- Performance
- Easy to adapt to new data (fine-tuning)

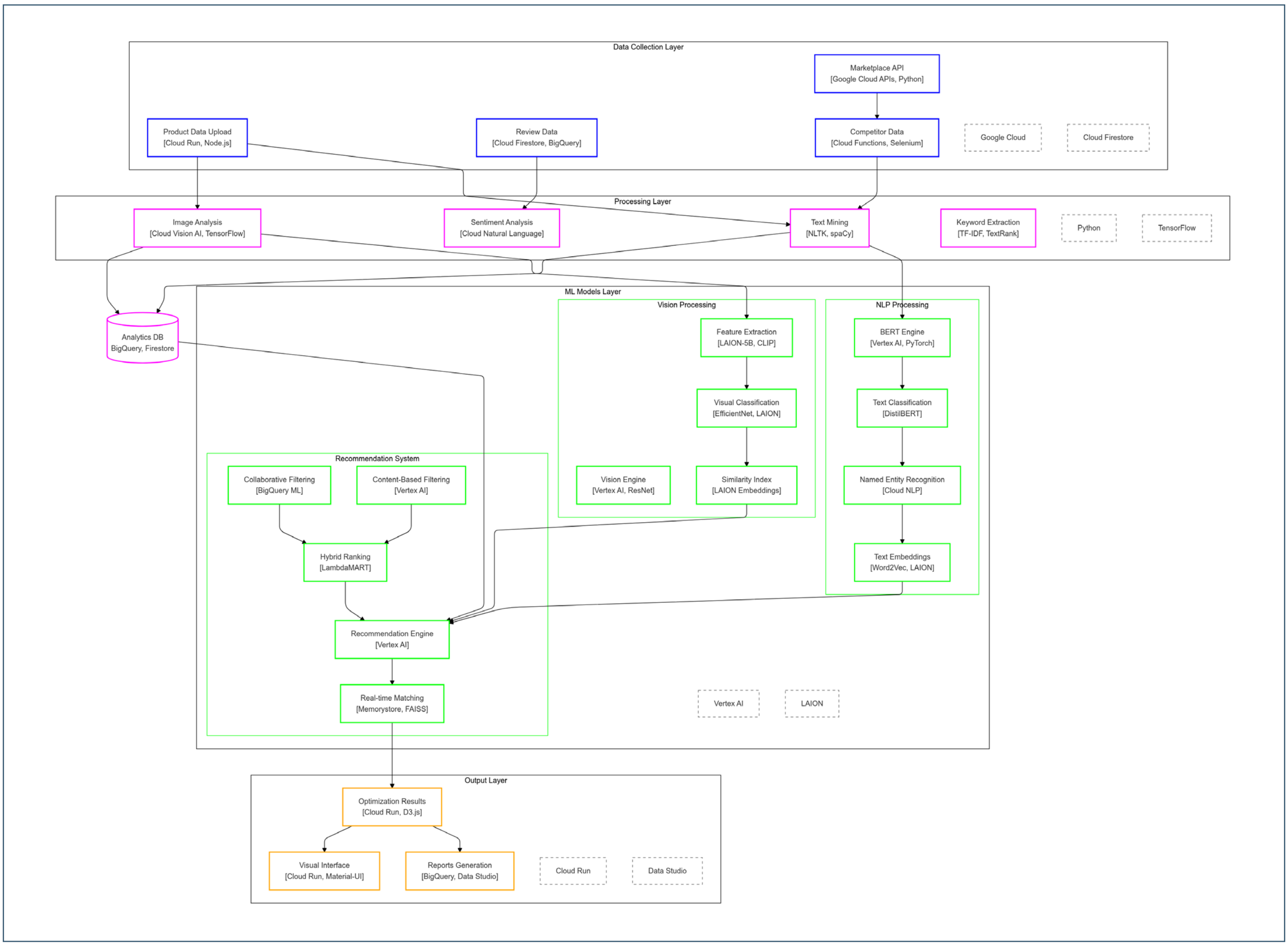
In the case of the PoC, light models like tiny-GPT or tiny-BERT were used for a more limited number of categories.

Following the successful proof of concept, we collaborated with the customer to outline the next development phases. A key focus was implementing an advanced image-to-text taxonomy encoding system. This required developing a sophisticated search engine that could identify media based on semantic meaning rather than just visual characteristics.

The system is designed to leverage an existing database of annotated images. When new images are introduced, the engine analyzes them against this database, identifying semantically similar media. Based on the similarity scores, the system aggregates taxonomy tags from these matches, enabling automatic taxonomy generation for newly uploaded images.

Product deployment and further development

Through iterative discussions and development cycles following the successful POC, we moved forward with MVP and FVP development phases, establishing and measuring specific success criteria at each stage. We systematically designed and integrated various components, refining our approach based on performance metrics and stakeholder feedback. This methodical process led to the development of a scalable and efficient model that optimally serves the system's requirements. Working in close collaboration with stakeholders, we carefully evaluated each component's performance and impact before incorporation, ensuring the final architecture could effectively handle growing data volumes while maintaining high performance standards.



The PoC and further architecture patterns were architected as a modular microservices system deployed in Google Cloud Platform, utilizing containerized applications for maximum flexibility and scalability. The architecture consists of four distinct layers.

Data Collection Layer

- Marketplace API integration using Cloud Functions for efficient data harvesting
- Product data upload service running on Cloud Run for scalable processing
- Review data collection utilizing Cloud Firestore and BigQuery for storage optimization
- Competitor data analysis through serverless functions with Selenium

Processing Layer

- Image Analysis pipeline powered by Cloud Vision AI and TensorFlow
- Sentiment Analysis using Cloud Natural Language services
- Text Mining implementation with NLTK and spaCy
- Keyword Extraction utilizing TF-IDF and TextRank algorithms

ML Models Layer

- Vision Processing with LAION-CLIP for feature extraction
- NLP Processing using Vertex AI with PyTorch integration
- Advanced recommendation system combining:
  - Collaborative Filtering using BigQuery ML
  - Content-Based Filtering through Vertex AI
  - Hybrid Ranking system with LambdaMART
  - Real-time Matching using Memorystore and FAISS

Output Layer

- Optimization results delivery through Cloud Run
- Visual Interface built on Material UI
- Reports Generation using BigQuery and Data Studio

PoC RESULTS ACHIEVED

- ☑ **Successfully implemented the core recommendation engine, processing 100,000 products daily.**
- ☑ **Validated custom computer vision models for product image analysis and optimization.**
- ☑ **Proved the concept of an adaptive taxonomy system adaptable across product categories and industries.**

Expected MVP/FVP Business Impact

Projected Operational Improvements

MVP and FVP phases are designed to deliver significant operational enhancements:

- Manual optimization time reduction by 70%
- Market trend identification acceleration by 85%
- Product update cycle time improvement by 60%

Anticipated Revenue Impact

The fully developed system is projected to deliver:

- 15–20% increase in product conversion rates
- 25–30% improvement in product visibility
- 40–45% reduction in optimization-related costs

Planned Scalability Features

The target architecture is designed to provide:

- Capability to handle 10x current product volume
- 95% automation in recommendation generation
- Support for multiple marketplace integrations

WHAT HAPPENS NOW

Our active preparation for production development is currently in progress. The PoC has validated core technical approaches including our ML model selection, cloud infrastructure setup, and data processing pipeline. We project a 6–9 month implementation timeline across three phases.

Phase 1 (Months 1–3):

- Infrastructure scaling and reliability improvements
- Multi-region deployment on Google Cloud
- Enhanced monitoring and alerting system
- Advanced error handling implementation
- Data validation and consistency mechanisms

Phase 2 (Months 3–6):

- Core system enhancement
- Scaled ML model deployment
- Real-time processing pipeline optimization
- Advanced security implementation
- Performance optimization

Phase 3 (Months 6–9):

- Production readiness
- Full high-availability setup
- Comprehensive testing
- Documentation and training
- Production deployment

We estimate maintaining our current core team of 8 specialists with a potential addition of 2–3 specialists during peak development phases, with projected ROI within 7–10 months post-implementation.

We're looking forward to starting the main development phase as soon as the customer approves technical requirements and implementation timeline.

Estimated Business Impact:

- 65% reduction in manual optimization work
- 35% improvement in product performance
- Processing capacity increase to 5,000+ products/hour
- System response time under 1s

SCALABILITY & ADAPTABILITY

While initially developed for e-commerce optimization, the system's architecture can be adapted for various data-intensive industries requiring pattern recognition and recommendation generation. The core technologies – computer vision, natural language processing, and predictive analytics – can be repurposed for applications ranging from retail to agriculture.

Ready to take the next step?

Reach us today at info@brightgrove.com