Design and Evaluation of an HPC-based Expert System to speed-up Retail Data Analysis using Residual Networks Combined with Parallel Association Rule Mining and Scalable Recommenders

C. Barakat, M. Riedel, S. Brynjólfsson, G. Cavallaro, J. Busch, R. Sedona

2021-10-01

Introduction

- Everyday life becoming more and more digital:
  - Boom in online marketplaces.
  - Further escalated by the Covid-19 pandemic and the ensuing lockdowns.
- Suppliers and retailers have growing databases of transaction data.
- This data is helpful in many ways:
  - Highlight product popularity and seasonal/regional trends[1].
  - Improve product recommendation systems[2,3].
  - Improve search functionality[4].

Main Hurdles

- Retail staff can not be expected to remember product recommendations for every product available.
- Increase in customers reduces staff ability to make more personalised recommendations.
- Turning to online sales and not being properly equipped for it.
- Limited or no access to the computing power necessary for data processing.
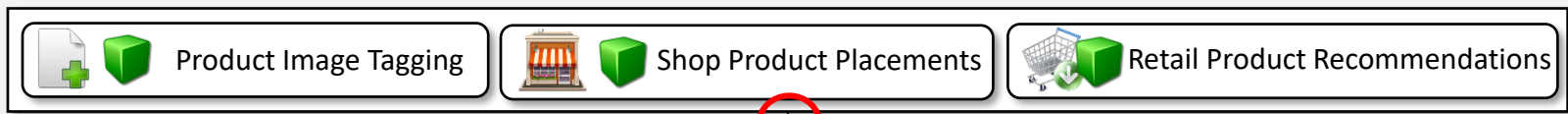- Big data analytics is not a human skill.

Solution

- Put available High-Performance Computing (HPC) resources[6,7] to use (ON4OFF):
  - Use transaction data to extract association rules
  - Improve personalised recommendation systems with machine learning approaches.
  - Automate product tagging for search using image recognition.
  - Make the online resources available to smaller retailers.
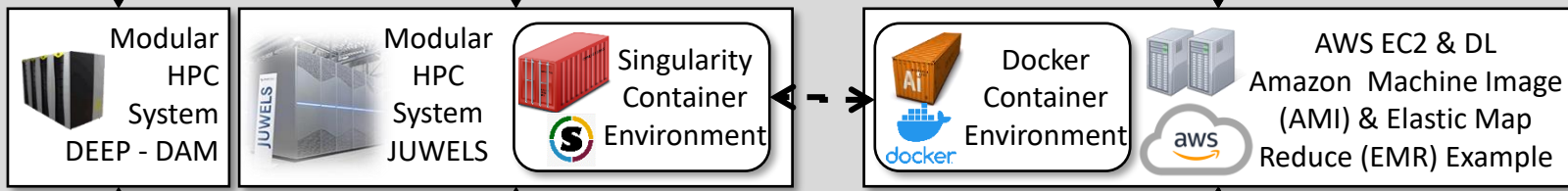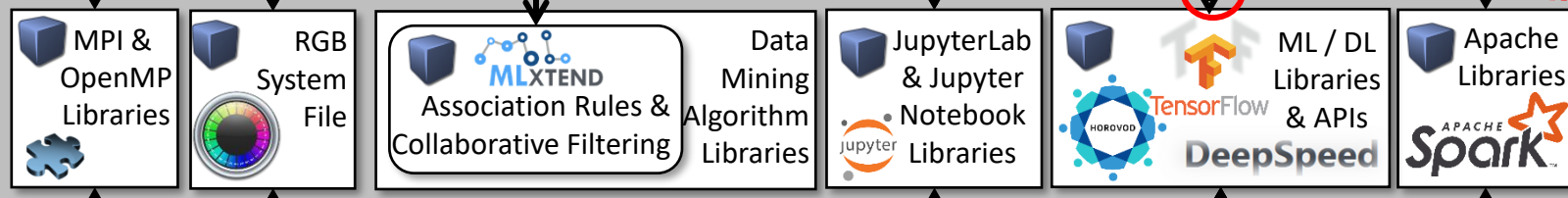
HPC-enabled Expert System
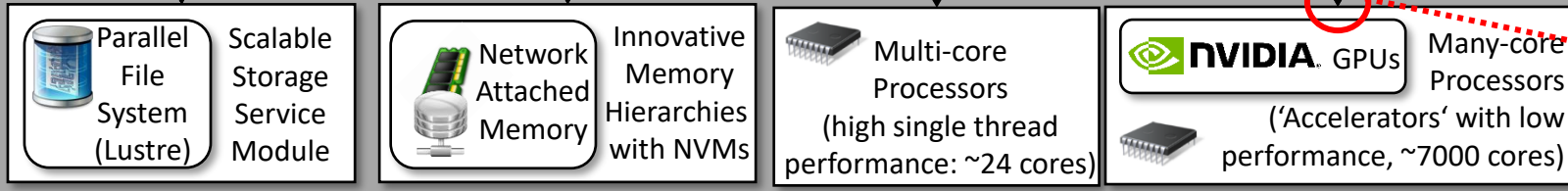
System Blueprint

processing-intensive applications

- Product Image Tagging
- Shop Product Placements
- Retail Product Recommendations

computing platforms

- HPC-based Expert System for Retail
- Service Landscape hosted by Adesso
- Microsoft Azure, aws, Google Cloud — Commercial Cloud Vendors

innovative computing resources

- Modular HPC System DEEP - DAM
- Modular HPC System JUWELS
- Singularity Container Environment
- Docker Container Environment
- AWS EC2 & DL Amazon Machine Image (AMI) & Elastic Map Reduce (EMR) Example

technology libraries & packages

- MPI & OpenMP Libraries
- RGB System File
- MLXTEND Association Rules & Collaborative Filtering
- Data Mining Algorithm Libraries
- JupyterLab & Jupyter Notebook Libraries
- HOROVOD TensorFlow DeepSpeed — ML / DL Libraries & APIs
- Apache Spark Libraries

key hardware technologies

- Parallel File System (Lustre) — Scalable Storage Service Module
- Network Attached Memory — Innovative Memory Hierarchies with NVMs
- Multi-core Processors (high single thread performance: ~24 cores)
- NVIDIA GPUs — Many-core Processors ('Accelerators' with low performance, ~7000 cores)
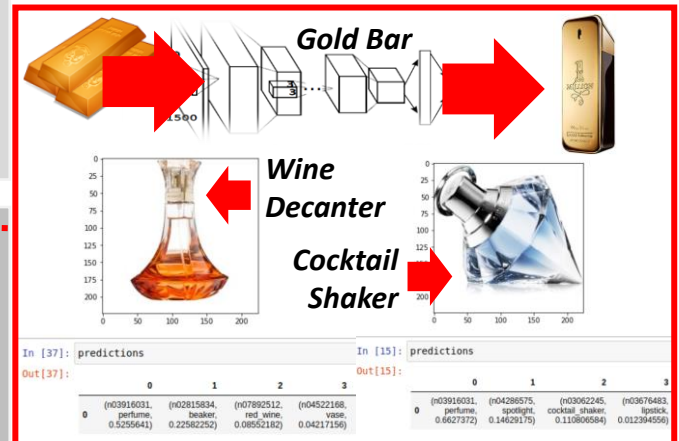
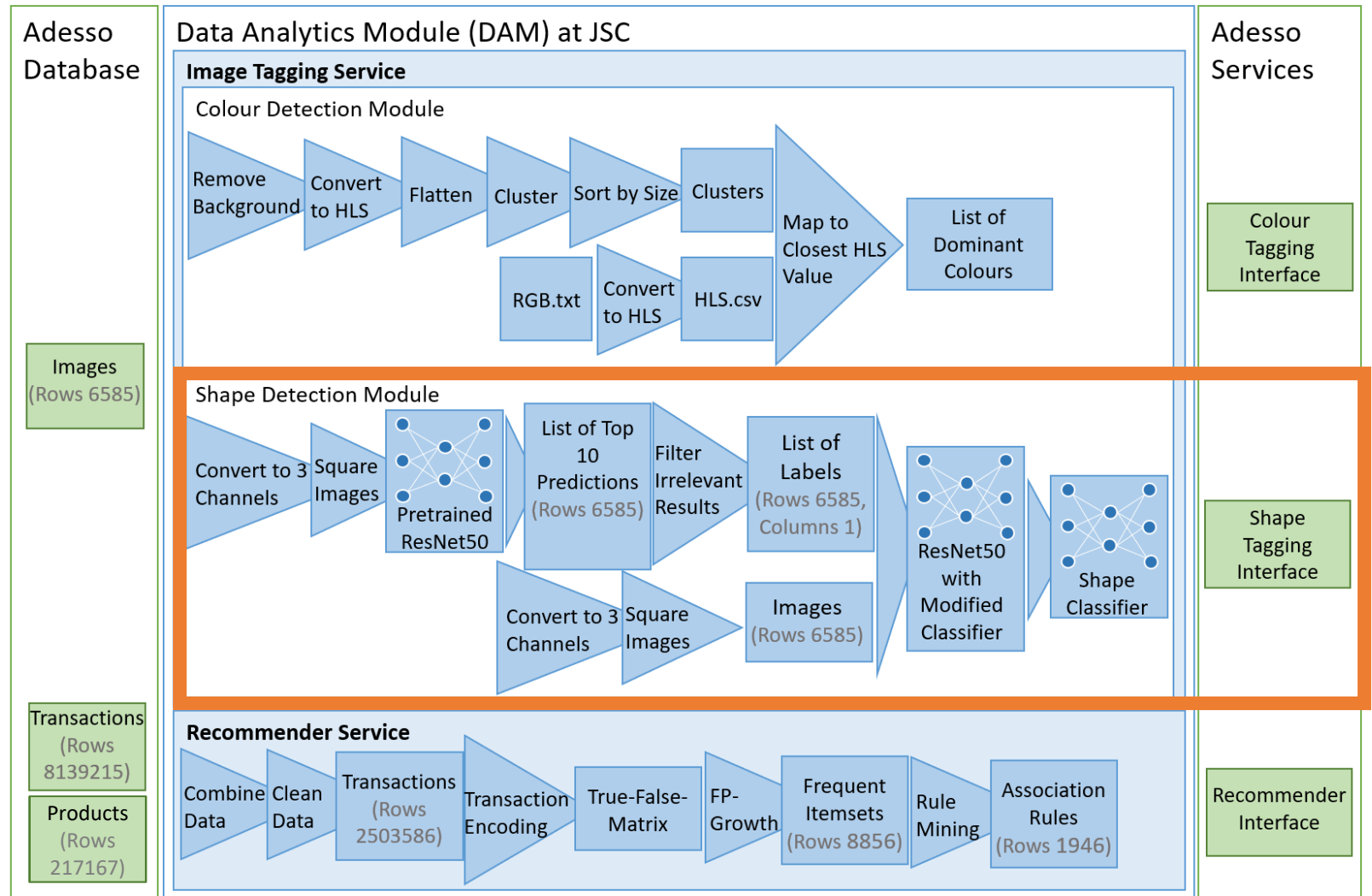| #GPUs	images/s	speedup	Performance per GPU [images/s]
1	55	1.0	55
4	178	3.2	44.5
8	357	6.5	44.63
16	689	12.5	43.06
32	1230	22.4	38.44
64	2276	41.4	35.56
128	5562	101.1	43.45

#128 GPUs in parallel

Colour Detection

- Uses K-means Clustering to detect most dominant colours.

- Compares to standard RGB colours available in most Linux distributions.

- Outputs the nearest recognized colours as list of tags.

- Module is the most portable element of the service platform.
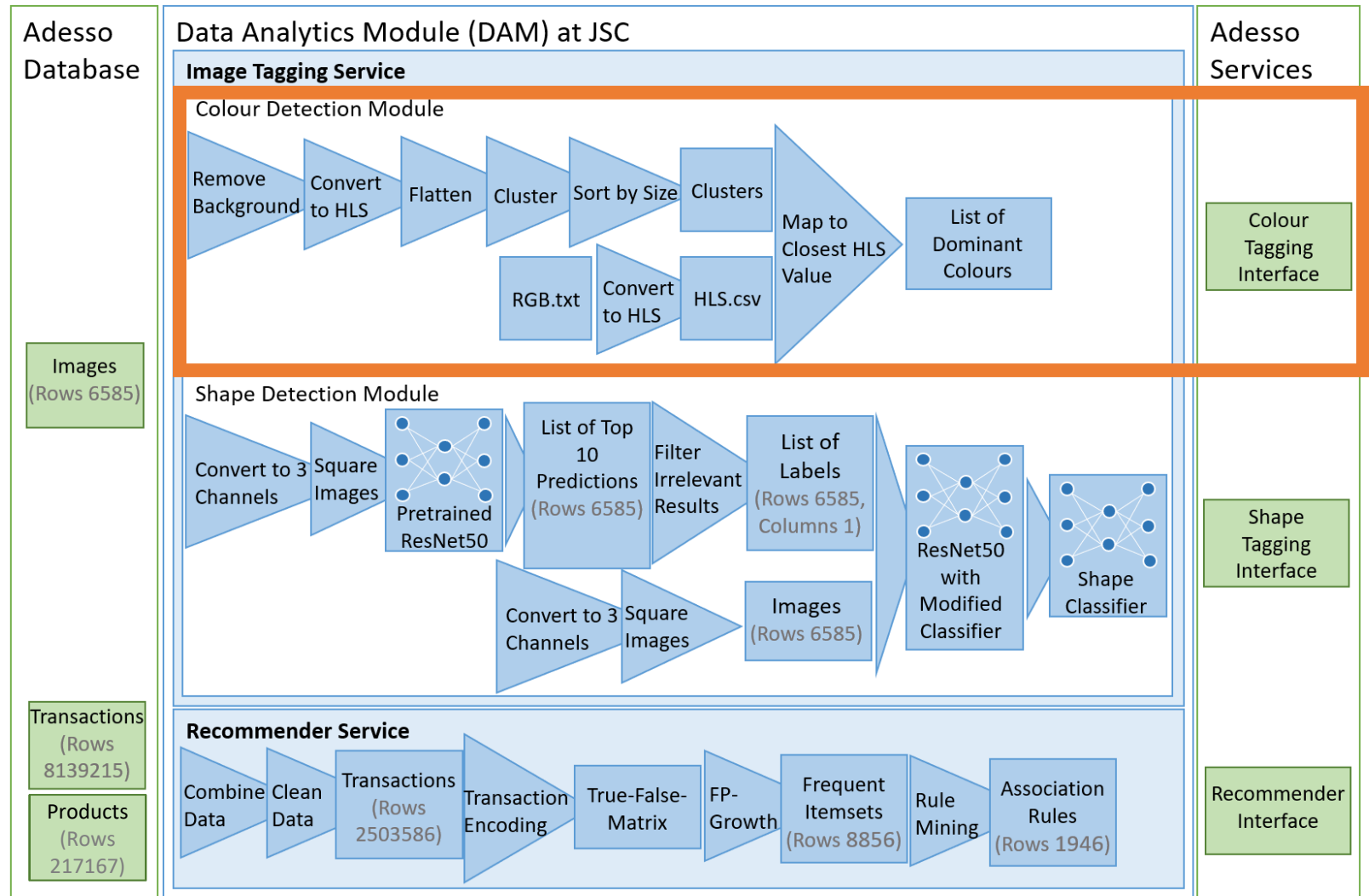
Shape Detection

- Feeds sample images into pretrained ResNet50[9].

- Produced tags are filtered (remove irrelevant tags).

- Used to retrain a new ResNet50 model with a reduced classifier layer.

- Training on GPU takes about 10 minutes to complete.

- Retrained model (~105 MB) is exported to the service platform.

Adesso Database

Images (Rows 6585)

Transactions (Rows 8139215)

Products (Rows 217167)

Data Analytics Module (DAM) at JSC

Image Tagging Service

Colour Detection Module

Remove Background | Convert to HLS | Flatten | Cluster | Sort by Size | Clusters | Map to Closest HLS Value | List of Dominant Colours

RGB.txt | Convert to HLS | HLS.csv

Shape Detection Module

Convert to 3 Channels | Square Images | Pretrained ResNet50 | List of Top 10 Predictions (Rows 6585) | Filter Irrelevant Results | List of Labels (Rows 6585, Columns 1) | ResNet50 with Modified Classifier | Shape Classifier

Convert to 3 Channels | Square Images | Images (Rows 6585)

Recommender Service

Combine Data | Clean Data | Transactions (Rows 2503586) | Transaction Encoding | True-False-Matrix | FP-Growth | Frequent Itemsets (Rows 8856) | Rule Mining | Association Rules (Rows 1946)

Adesso Services

Colour Tagging Interface

Shape Tagging Interface

Recommender Interface

Recommendation

- Transaction data is cleaned, reformatted, and encoded.

- This step is impossible without HPC -> 500GB stored in Memory.

- The association rules are loaded into the service platform and used for product recommendation.

- Also implemented collaborative filtering for personalised recommendations
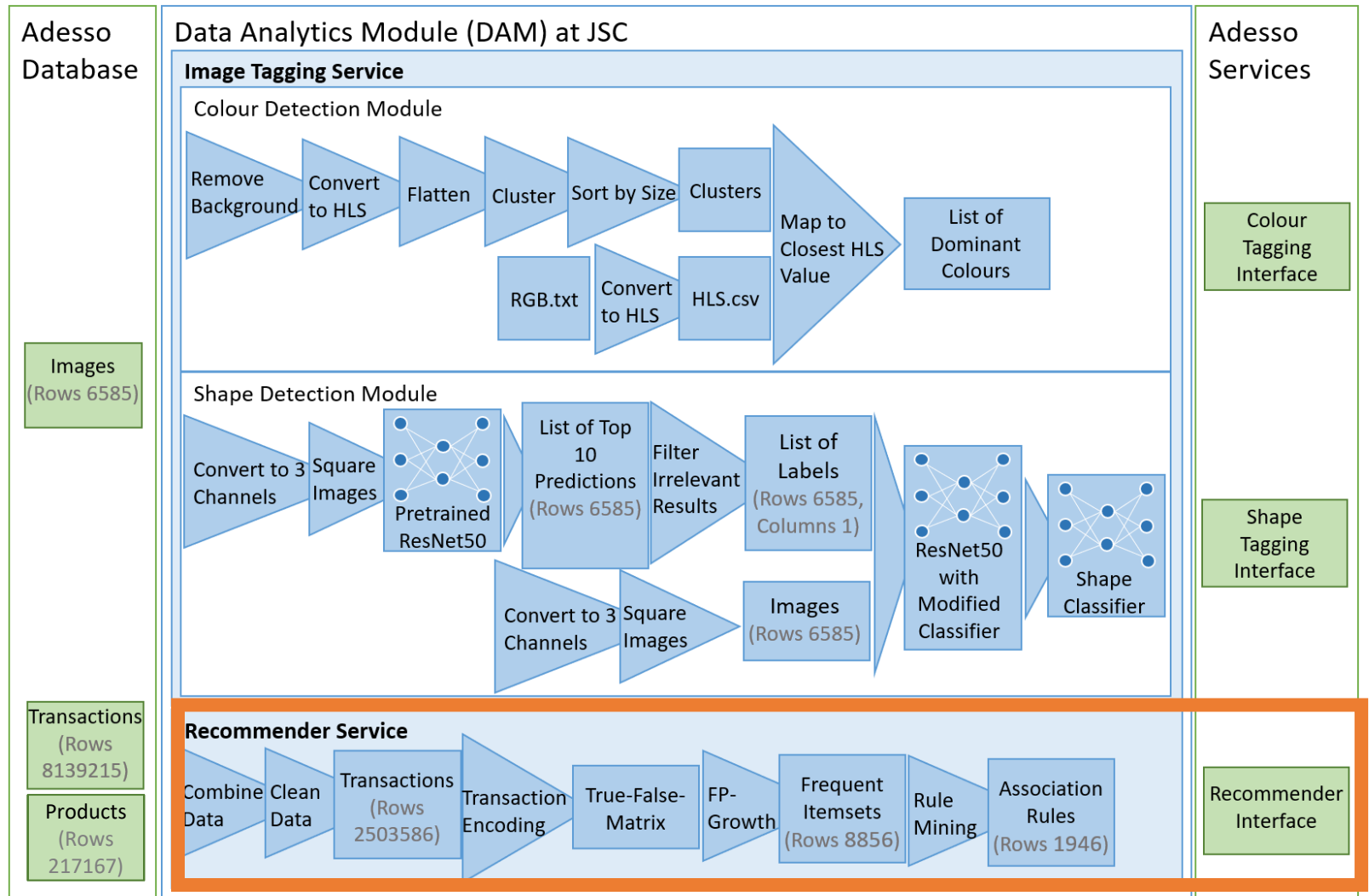
Conclusions

- Developed a comprehensive approach to implement HPC resources in the service of the retail sector.

- It is meant to improve the shopping experience for customers, improve recommendations, and therefore sales, and simplify product tagging on websites.

- The service platform also integrates with resources available from cloud computing vendors.

- Currently in the process of being applied for new projects as part of the ON4OFF project.

Bibliography

1. A. L. D. Loureiro et al., "Exploring the use of deep neural networks for sales forecasting in fashion retail," *Decision Support Systems*, vol. 114, pp. 81–93, 2018.
2. A. N. Sagın et al., "Determination of association rules with market basket analysis: An application in the retail sector," *Southeast Europe Journal of Soft Computing*, vol. 7, no. 1, pp. 10–19, 2018.
3. S. Kumar et al., "A review on big data based parallel and distributed approaches of pattern mining," *Journal of King Saud University – Computer and Information Sciences*, 2019.
4. S. Advani et al., "Visual co-occurrence network: Using context for large-scale object recognition in retail," in 2015 13th IEEE Symposium on Embedded Systems For Real-time Multimedia (ESTIMedia), pp. 1–10.
5. [11] K. Fuchs et al., "Towards identification of packaged products via computer vision: Convolutional neural networks for object detection and image classification in retail environments," in 9th International Conference on the Internet of Things (IoT 2019), 2019, pp. 1–8.
6. T. Lippert et al., Modular Supercomputing Architecture: From Idea to Production, pp. 223–255, Imprint CRC Press, 1st ed. edition, 2019, Contemporary High Performance Computing.
7. https://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUWELS
8. https://www.deep-projects.eu/
9. K. He et al., "Deep residual learning for image recognition," December 2015.