

Real-Time Text Analytics on AWS F1™ and Xilinx Alveo™ Accelerator Cards

Advancing the way businesses leverages large-scale, text data

- Speeds up analysis from hours to seconds
- Simplifies source modification & streamlines decision workflows
- Provides actionable results across languages and formats

INTRODUCTION

Artificial Intelligence & Text

80% of world data is in the form of text and 45% of that text data is not in English. Artificial intelligence is rapidly evolving particularly in algorithms, computing capability, and storage infrastructure. However, most of the focus has been on numerical data or English content in well-defined and controlled situations.

Challenges with Current Text Analytics

Professionals in information-focused roles such as investors, intelligence analysts, lawyers, and researchers spend at least a third of their day reading information. An overwhelming majority express frustration towards the cost, bias, and incompleteness of that process. These groups of users demand a transparent, real-time, and flexible software that allows them to spend less time searching for insights and more on making recommendations, decision-making, and attending to their clients. They want and need to work with their own text data under a very fast loopback to direct the analysis towards specific points of relevance. Available market solutions are either providing pre-packaged answers to pre-defined questions on controlled data, or are too slow to compel professionals to use them in any realistic setting.

PRODUCT OVERVIEW

Working in partnership with Xilinx, SumUp Analytics' Nucleus platform operates seamlessly on local appliances with Xilinx Alveo™ Data Center Accelerator Cards or on the public cloud. The platform is currently available on AWS EC2 F1 instances.

SOLUTION OVERVIEW

SumUp Analytics' Nucleus is a Xilinx FPGA-powered web-based SaaS and API-based SaaS platform for identifying, extracting, and analyzing critical information from unstructured text. Nucleus delivers topic extraction, summarization, sentiment and consensus analysis, historical analysis, named-entity tagging, transfer learning, author connectivity analysis, and content recommendation across 7 major file formats (.doc, .docx, .pdf, .txt, .csv, .html, .rtf) and 13 languages (English, Traditional and Simplified Chinese, Japanese, Russian, German, Spanish, Portuguese, Italian, French, Arabic, Farsi and Hindi).

Nucleus is comprised of a Python/SDAccel hybrid library running on Xilinx FPGAs for core analytics and distributed-CPU for peripheral analytics. Data on the back-end is stored in MySQL and a React/Node front-end is overlaid for user interaction through the SaaS product. APIs are available through a Python SDK end-point on a REST server.

Key Benefits of Proprietary Algorithms

Very low-latency at scale

Nucleus algorithms are tailored to unstructured data and extensively leverage unsupervised learning. This results in highly responsive and highly adaptive analytics derived from just a few specialized publications up to millions of social media posts, delivering insight 100x faster than major competitors.

High Information Capture

Nucleus algorithms have been designed to quickly identify differentiated discussions within a dataset. Furthermore, these algorithms characterize each discussion in a clear, legible, and concise manner. This results in a markedly higher capture of important information within a dataset compared to major competitors.

Accelerated by Xilinx FPGAs

Nucleus' core algorithm exploits the sparsity in the machine learning model for unstructured data to reduce the computation complexity. It also takes advantage of abundant decision logic in the FPGA to manage irregular memory access introduced by the sparsity that allows creation of deep pipelined and parallel computation architecture.

Nbr docs	Dictionary length	Nbr topics	Runtime in seconds			U_Mass Coherence		
			Nucleus	Scikit	Gensim	Nucleus	Scikit	Gensim
31,496	81,801	1	0.1	230	3	-9.7	-19.9	-19.0
		4	0.3	648	4	-9.4	-14.7	-15.9
		8	0.6	1157	4	-11.9	-11.8	-15.9
29,287	78,063	1	0.1	220	3	-4.9	-5.8	-18.8
		4	0.2	588	3	-9.8	-14.9	-18.6
		8	0.7	1023	4	-8.3	-15.8	-17.9
211,687	434,336	1	0.2	2817	27	-4.8	-14.0	-18.4
		4	0.5	11282	33	-6.8	-12.7	-16.4
		8	1.0	N/A	38	-6.9	N/A	-15.7
2,995,495	3,609,526	1	4	39600	1877	-10.3	-15.5	-16.8
		4	23	N/A	2063	-10.3	N/A	-17.2
		8	34	N/A	2274	-7.1	N/A	-16.7
9,328,305	9,245,680	1	11	N/A	N/A	-2.8	N/A	N/A
		4	37	N/A	N/A	-3.5	N/A	N/A
		8	94	N/A	N/A	-7.6	N/A	N/A

U_Mass: less negative is better

LDA on Scikit
LDAMultiCore on Gensim

AWS c4.8xlarge, f1.2xlarge
N/A indicates Scikit run not completed within 36 hours or
Gensim peak memory issue (2GB dataset, 60GB available)

Web-based SaaS Turn-key solution

Nucleus is web-hosted, provides embedded data-feeds, and accepts uploads from any of the user local folders. Minimum configuration is required, thus enabling teams to rapidly deploy and ramp-up.

Flexible API-based SaaS solution

Our APIs offer companies easy integration into current workflow and flexibility to specify customized parameters. They are available for on-premise deployment, and on-cloud deployment via Amazon AWS and Alibaba Cloud.

Company-wide scalability

Guided onboarding and support ensures rapid adoption across both the technical and non-technical workforce.

CONCLUSION

Fast, precise, simple. SumUp Analytics' Nucleus is the real-time text analytics solution to mission-critical professionals, across languages and formats.

TAKE THE NEXT STEP

To request a trial and for on-premise solutions: sales@sumup.ai

Explore SumUp Analytics' Real-Time Text Solution running on Xilinx Alveo™: www.xilinx.com/alveo

For more information: www.sumup.ai



329 Bryant Street
Suite 3D
San Francisco, CA, 94107

Email: info@sumup.ai