XDF XILINX DEVELOPER FORUM

智捷分析 洞鉴先机：SumUp Nucleus实时文本分析平台

Presented By

SUMUP ANALYTICS

吴永俊（Jim Wu）

10月16日

XILINX

智捷分析 洞鉴先机

SUMUP ANALYTICS

THE PAIN POINT

Too Much To Read

Time Is Scarce

WHAT WE DO

High-granularity real-time analysis on Unstructured data

- x Time-consuming
- x Voluminous
- x Unstructured
- x Overwhelming

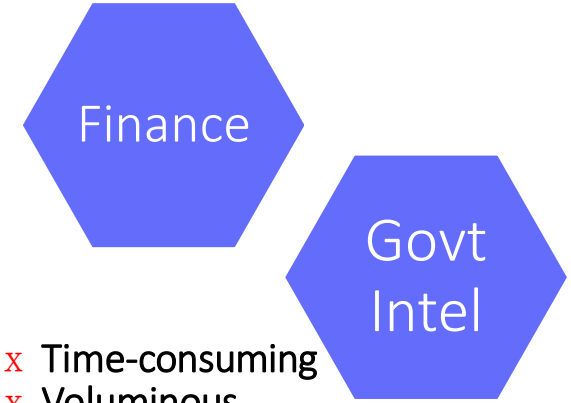
WHAT WE DO

High-granularity real-time analysis on Unstructured data

- x Time-consuming
- x Voluminous
- x Unstructured
- x Overwhelming

WHAT WE DO

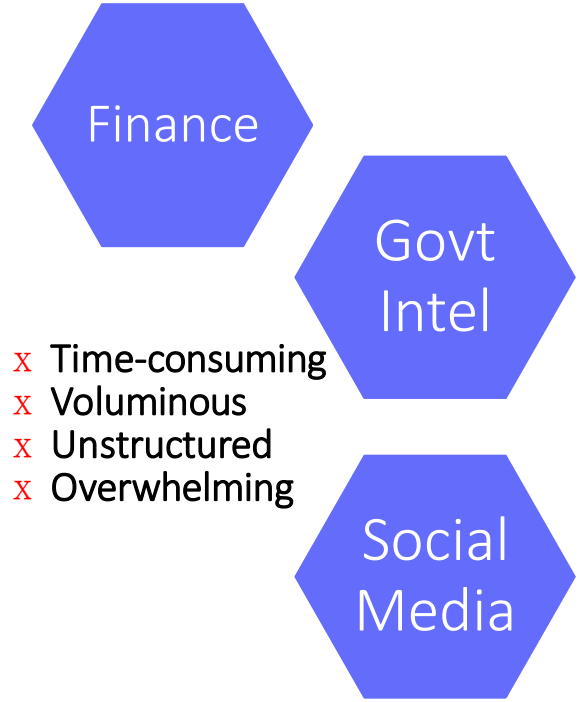
High-granularity real-time analysis on Unstructured data



- x Time-consuming
- x Voluminous
- x Unstructured
- x Overwhelming

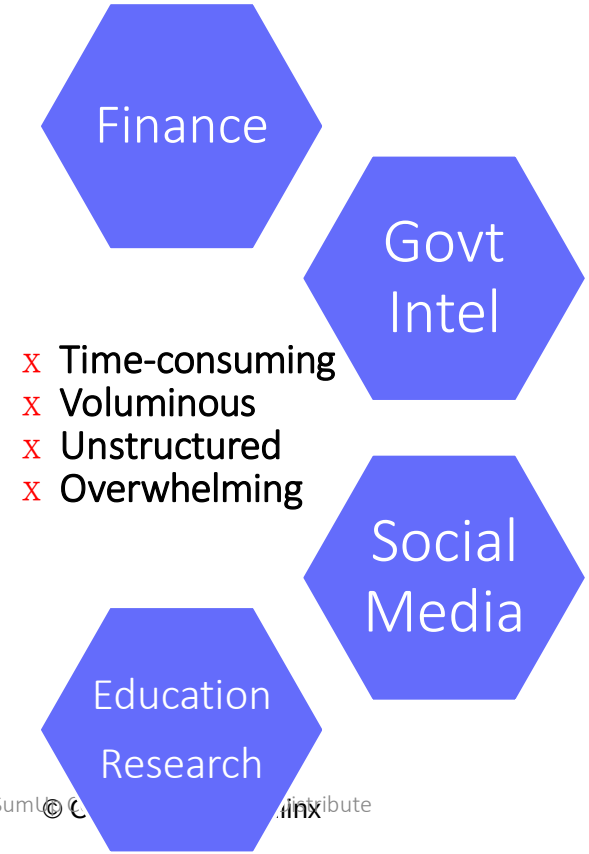
WHAT WE DO

High-granularity real-time analysis on Unstructured data



WHAT WE DO

High-granularity real-time analysis on Unstructured data



WHAT WE DO

High-granularity real-time analysis on Unstructured data

WHAT WE DO

High-granularity real-time analysis on Unstructured data

WHAT WE DO

High-granularity real-time analysis on Unstructured data

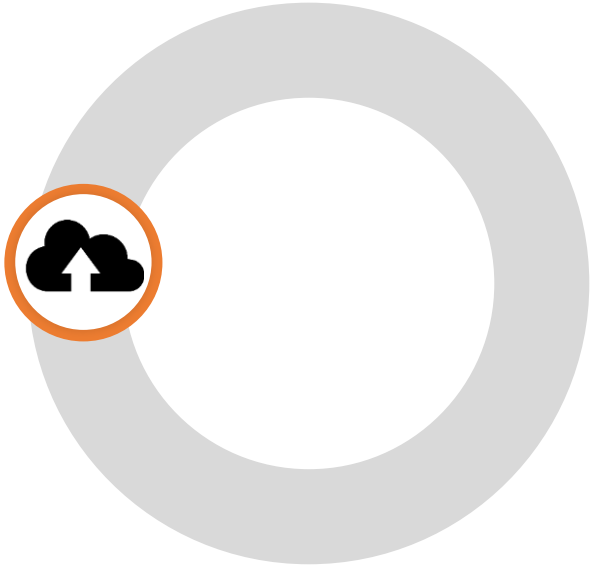
> 100x Speed-Up

> 100x Speed-Up

OUR SOLUTIONS

Nucleus SaaS | On-Demand Analytics APIs | Streaming Analytics APIs

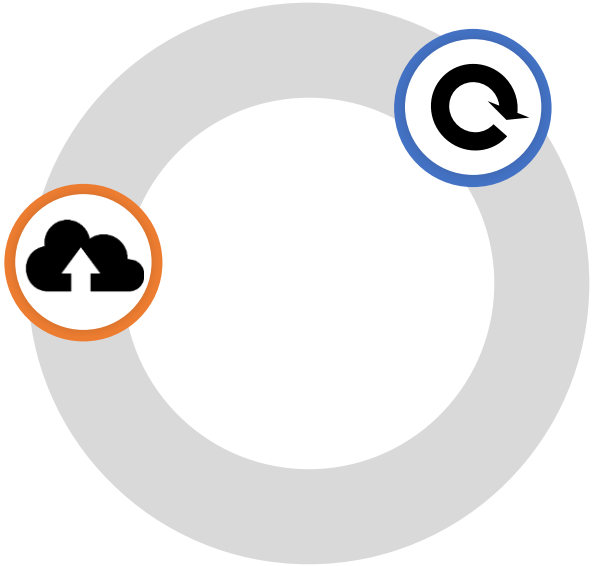
Step 1: Upload | DataFeeds



OUR SOLUTIONS

Nucleus SaaS | Streaming Analytics APIs | On-Demand Analytics APIs

Step 1: Upload | DataFeeds



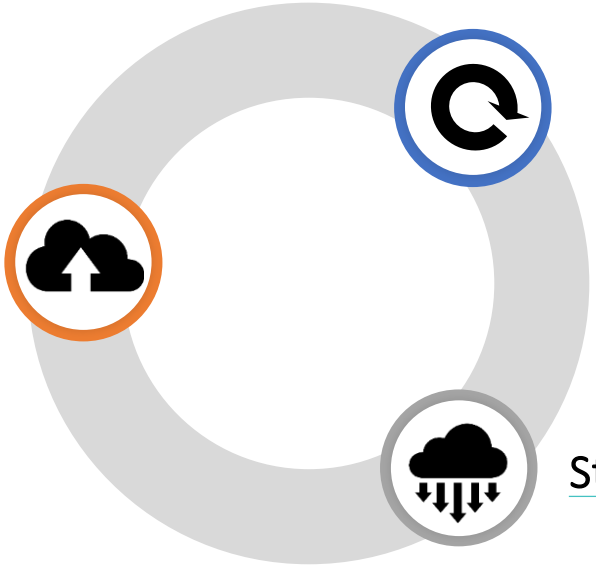
Step 2: Iterative Analysis

- Identify Topics & Summarize
- Analyze Sentiment & Consensus
- Get document recommendations
- Deep-dive on-demand

OUR SOLUTIONS

Nucleus SaaS | Streaming Analytics APIs | On-Demand Analytics APIs

Step 1: Upload | DataFeeds

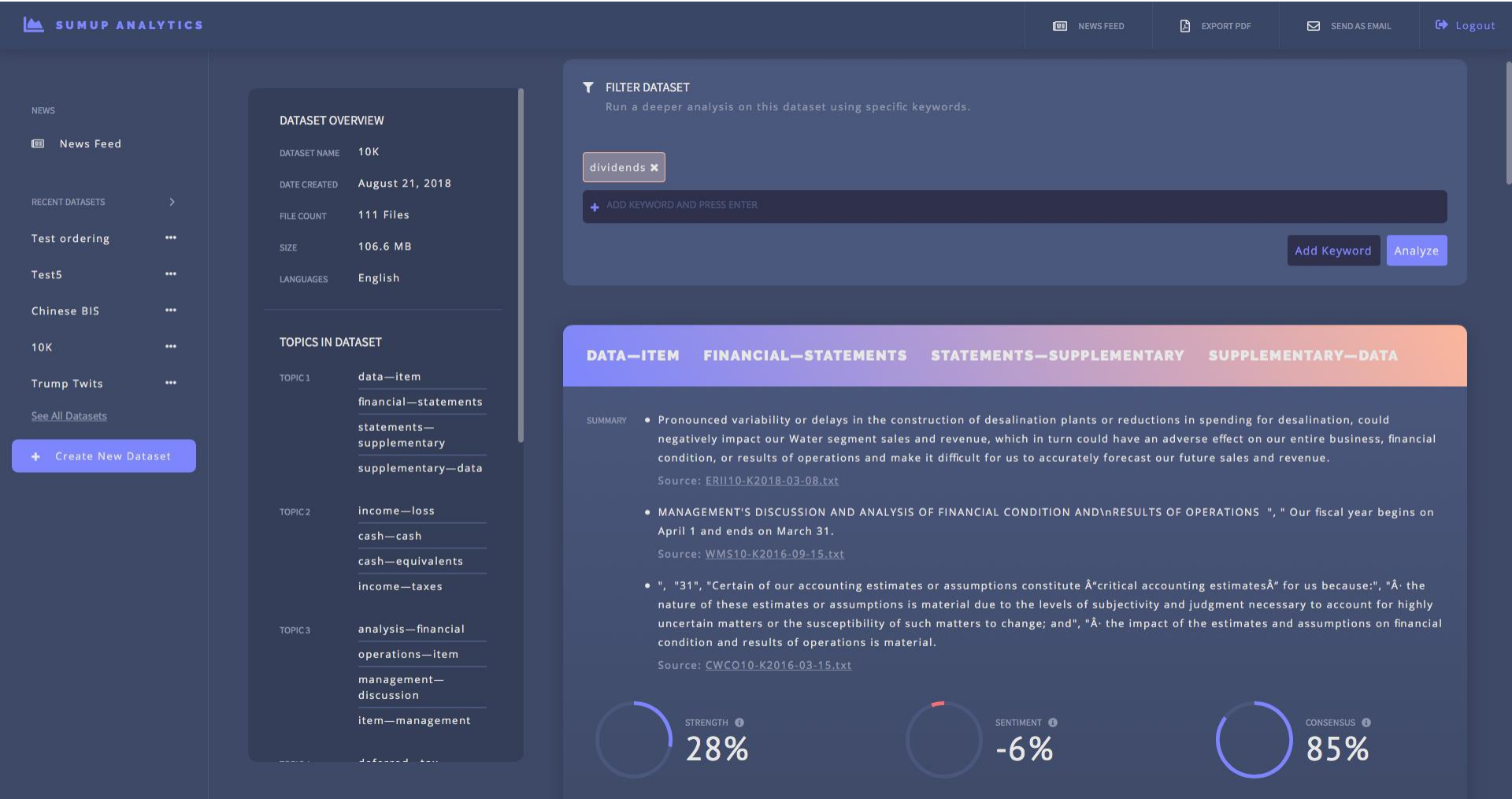


Step 2: Iterative Analysis

- Identify Topics & Summarize
- Analyze Sentiment & Consensus
- Get document recommendations
- Deep Dive on Demand

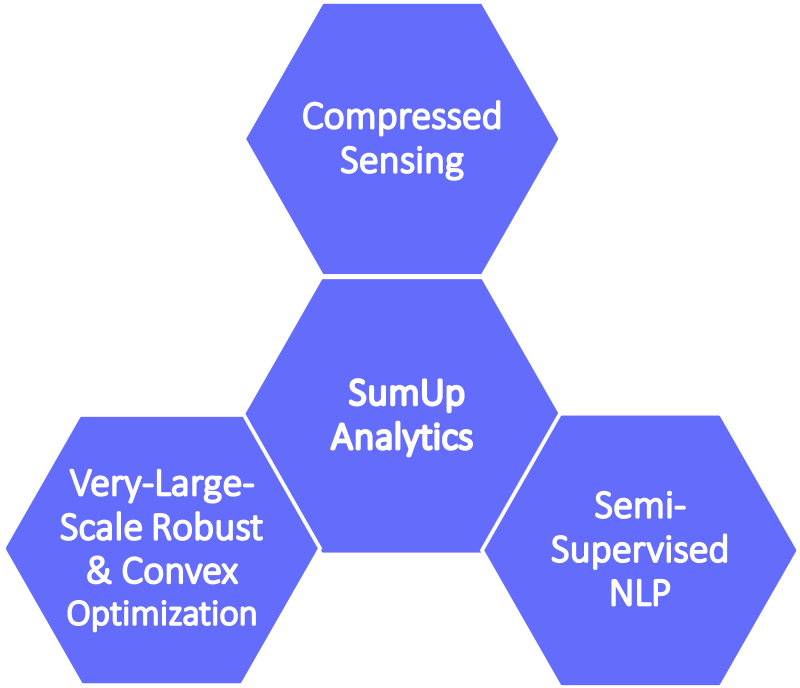
Step 3: Export & Share

NUCLEUS BY SUMUP



DIFFERENTIATED INTELLECTUAL PROPERTIES

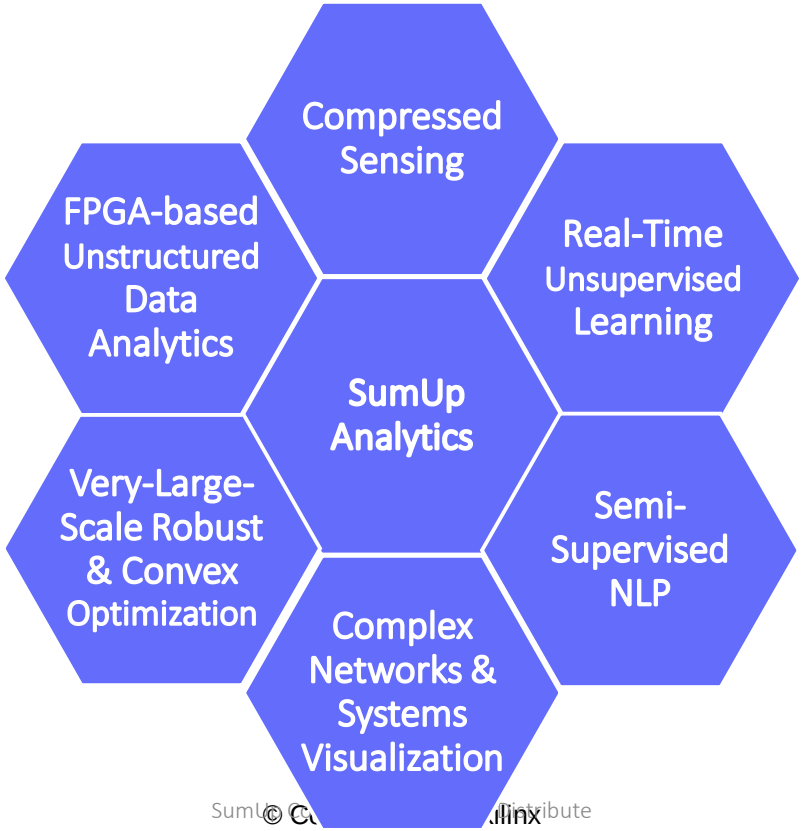
R&D in Analytics & Computational Systems



DIFFERENTIATED INTELLECTUAL PROPERTIES

R&D in Analytics & Computational Systems

Leverages our Xilinx partnership



FPGA: NATURAL FIT FOR UNSTRUCTURED DATA

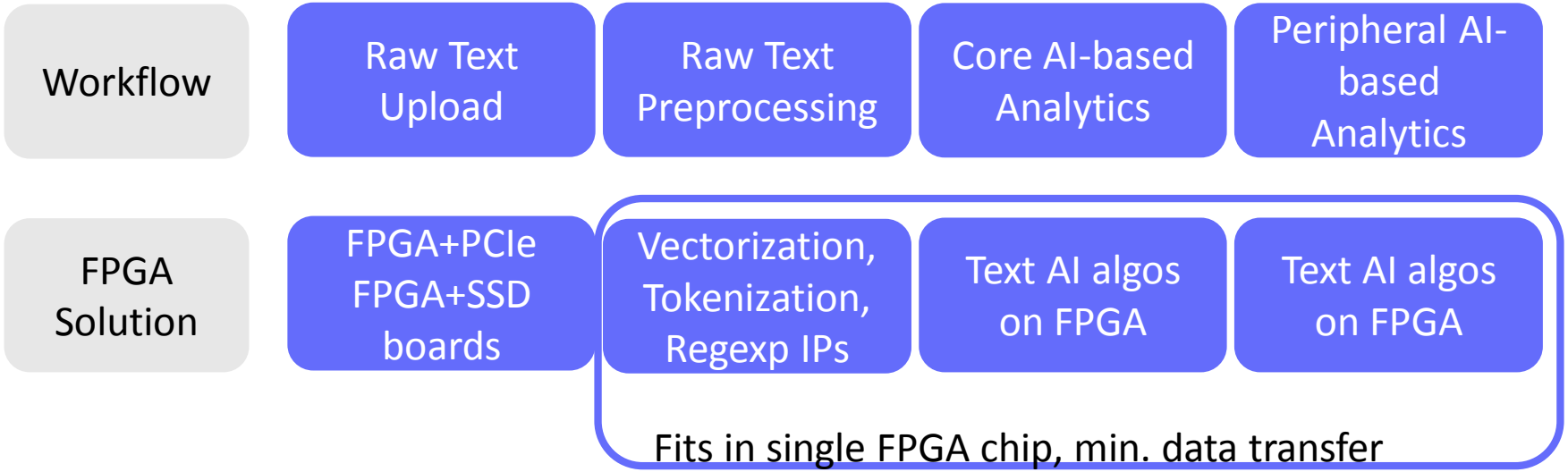
Data Challenge	Extremely Sparse	Extremely Large	Frequently Changing
FPGA Solution	Extensive Decision Logic	Very Deep Pipeline	Direct Data Streaming

FPGA very appealing to implement AI algos for Unstructured Data

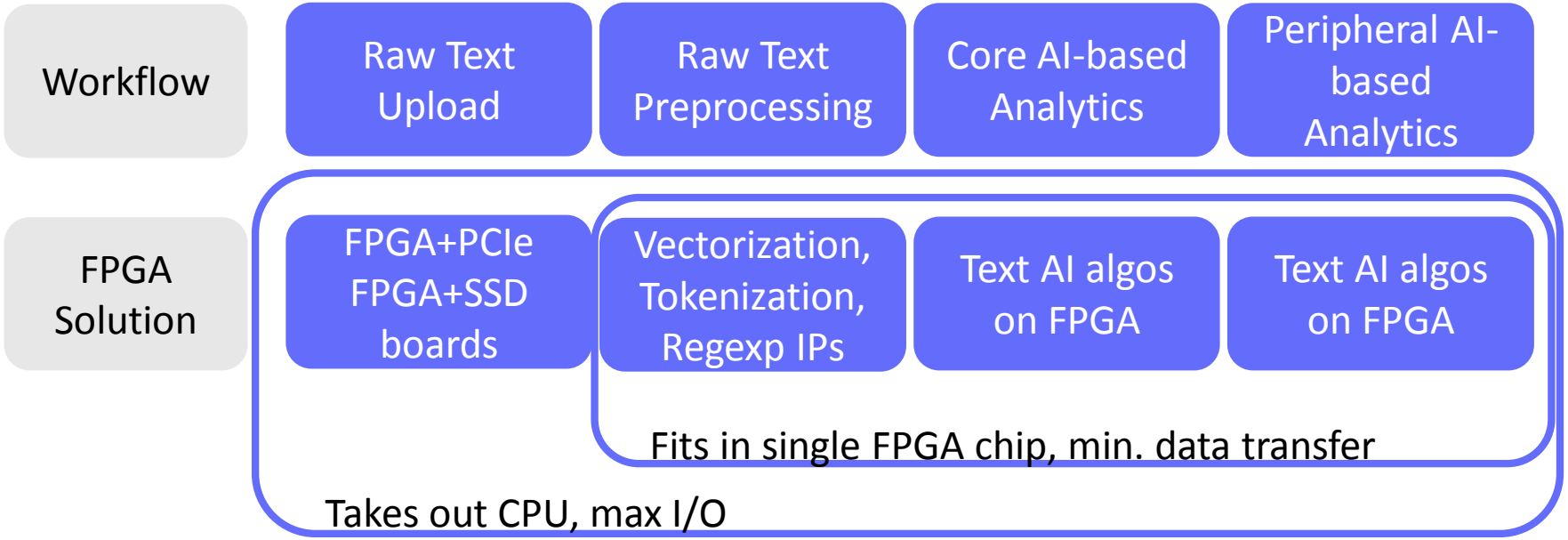
FPGA: POWERS NUCLEUS END TO END

Workflow	Raw Text Upload	Raw Text Preprocessing	Core AI-based Analytics	Peripheral AI-based Analytics
FPGA Solution	FPGA+NIC FPGA+NVME boards	Vectorization, Tokenization, Regexp IPs	Text AI algos on FPGA	Text AI algos on FPGA

FPGA: POWERS NUCLEUS END TO END



FPGA: POWERS NUCLEUS END TO END



ACCELERATED DEV & DEPLOY W/ XILINX

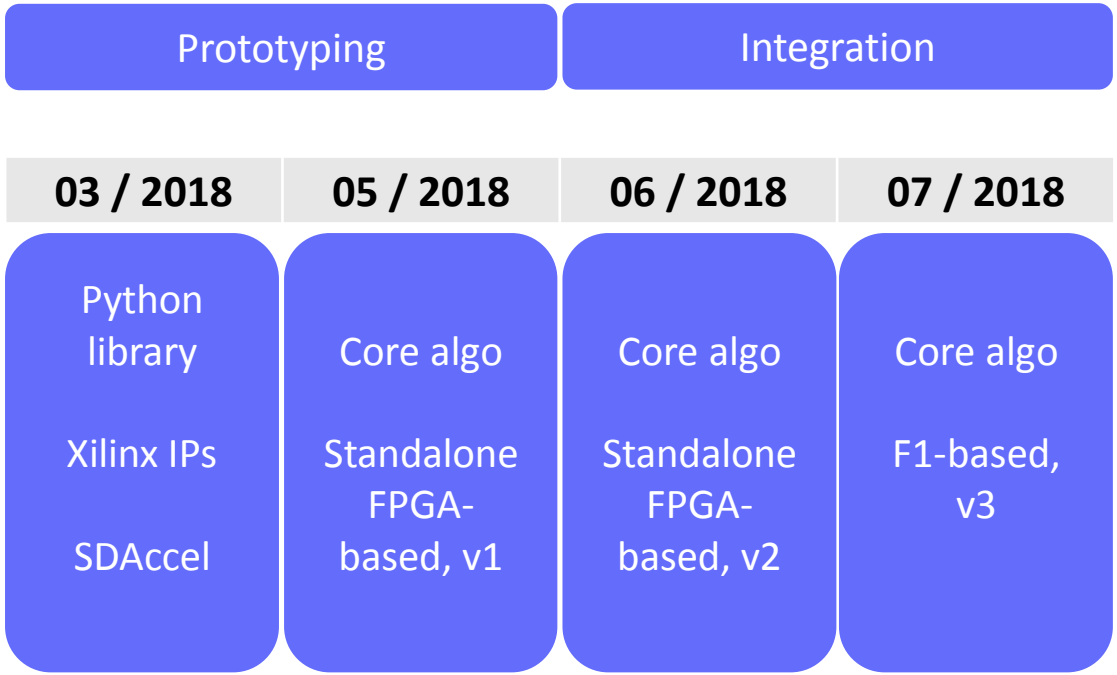
Xilinx Public IPs | SumUp Python/SDAccel Wrapper | AWS F1

Prototyping

03 / 2018	05 / 2018
Python library Xilinx IPs SDAccel	Core algo Standalone FPGA-based, v1

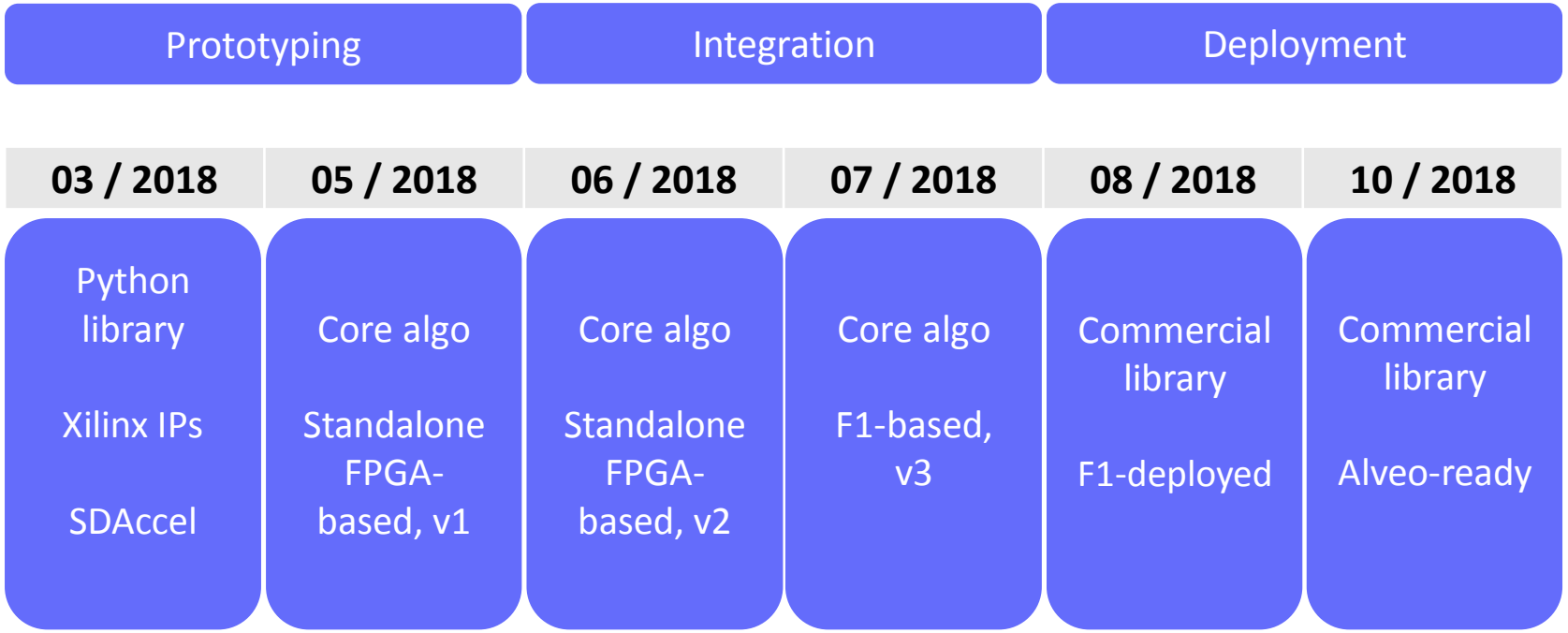
ACCELERATED DEV & DEPLOY W/ XILINX

Xilinx Public IPs | SumUp Python/SDAccel Wrapper | AWS F1

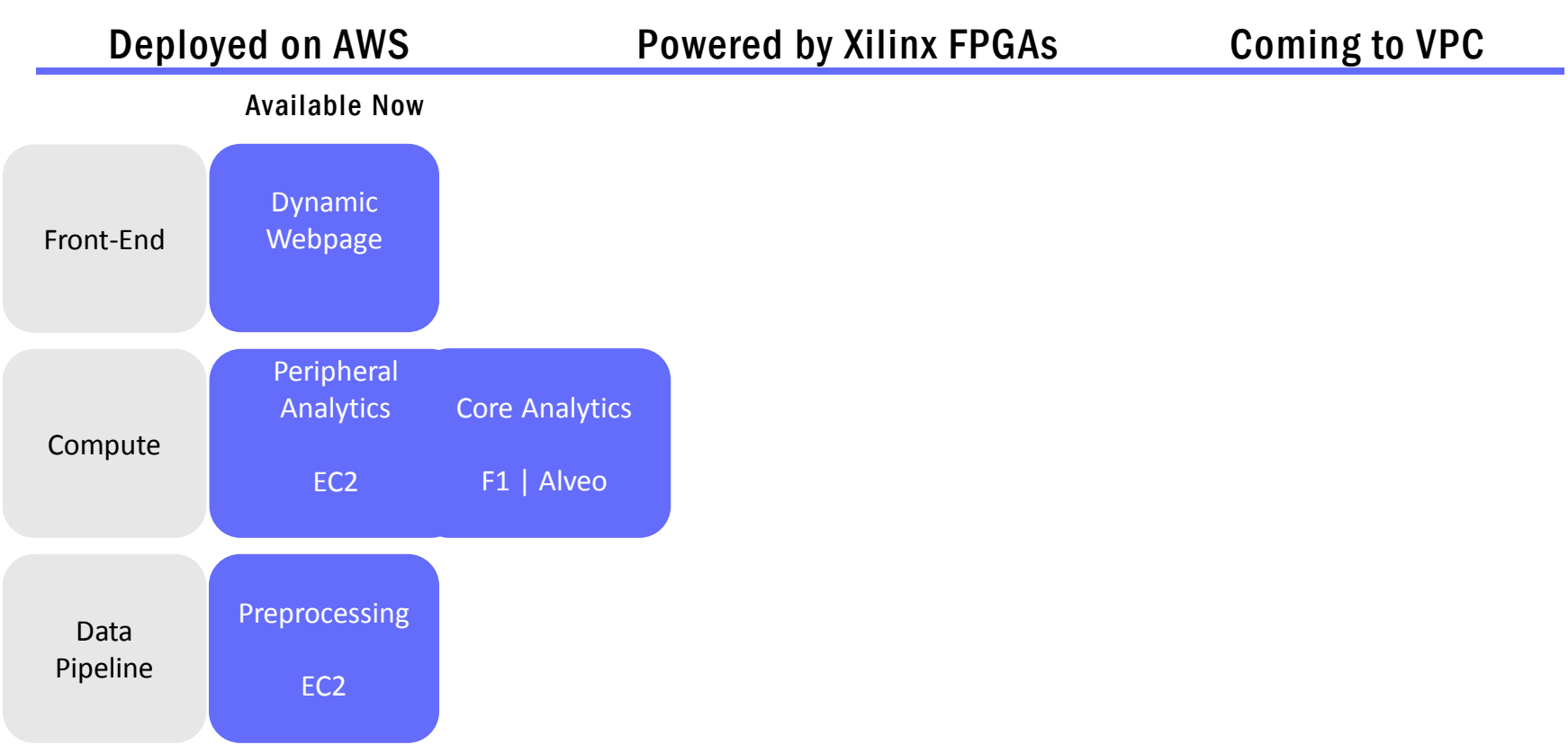


ACCELERATED DEV & DEPLOY W/ XILINX

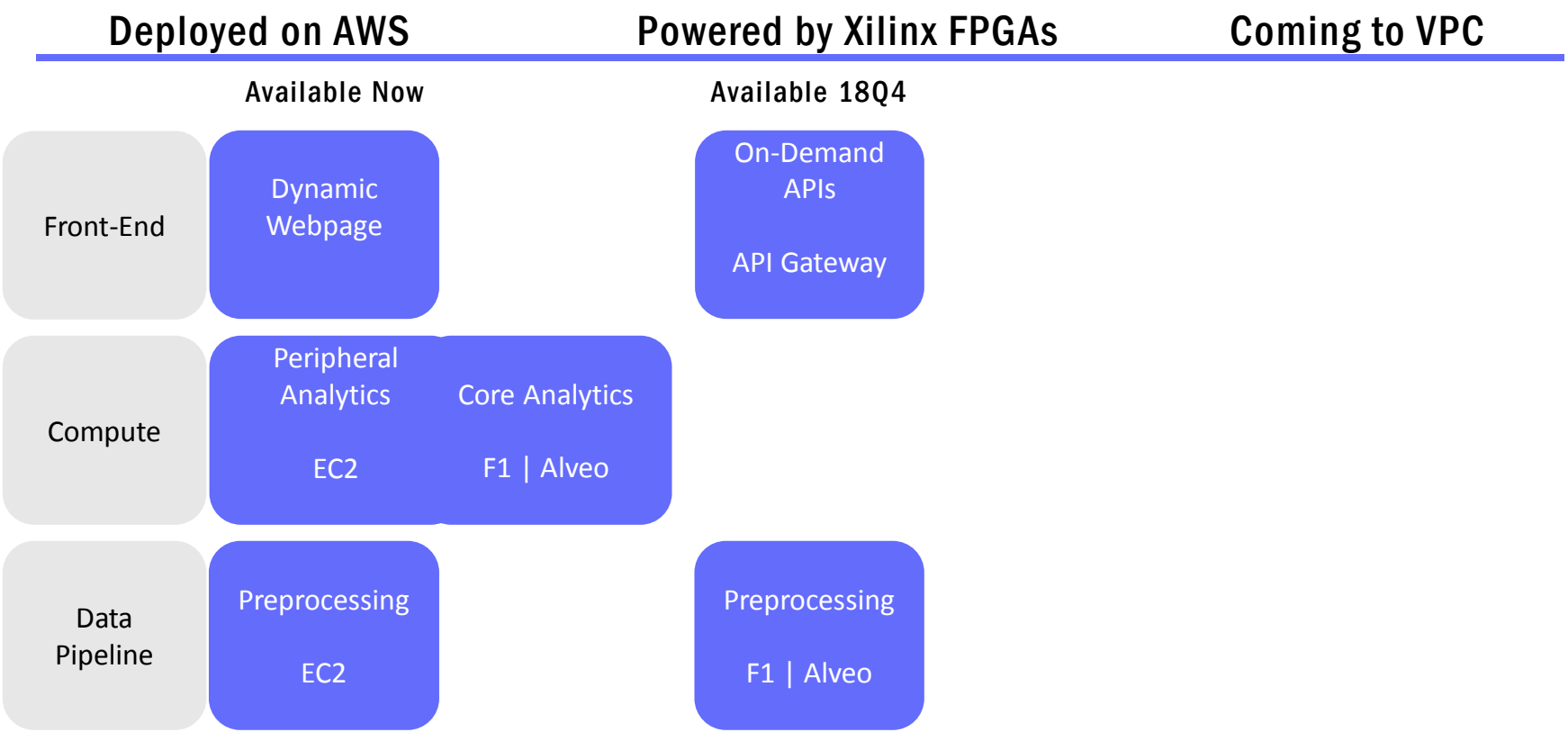
Xilinx Public IPs | SumUp Python/SDAccel Wrapper | AWS F1



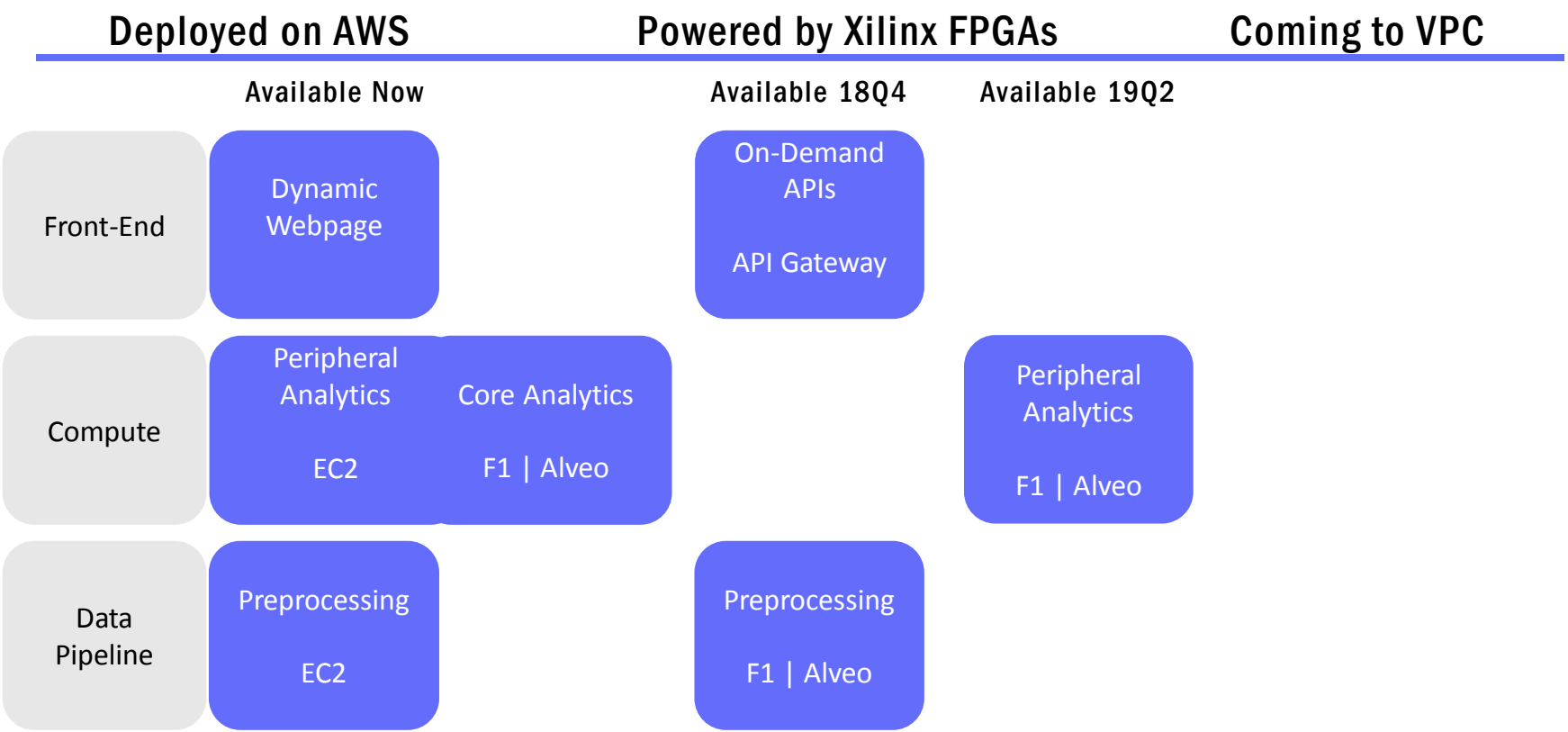
PRODUCT ROADMAP



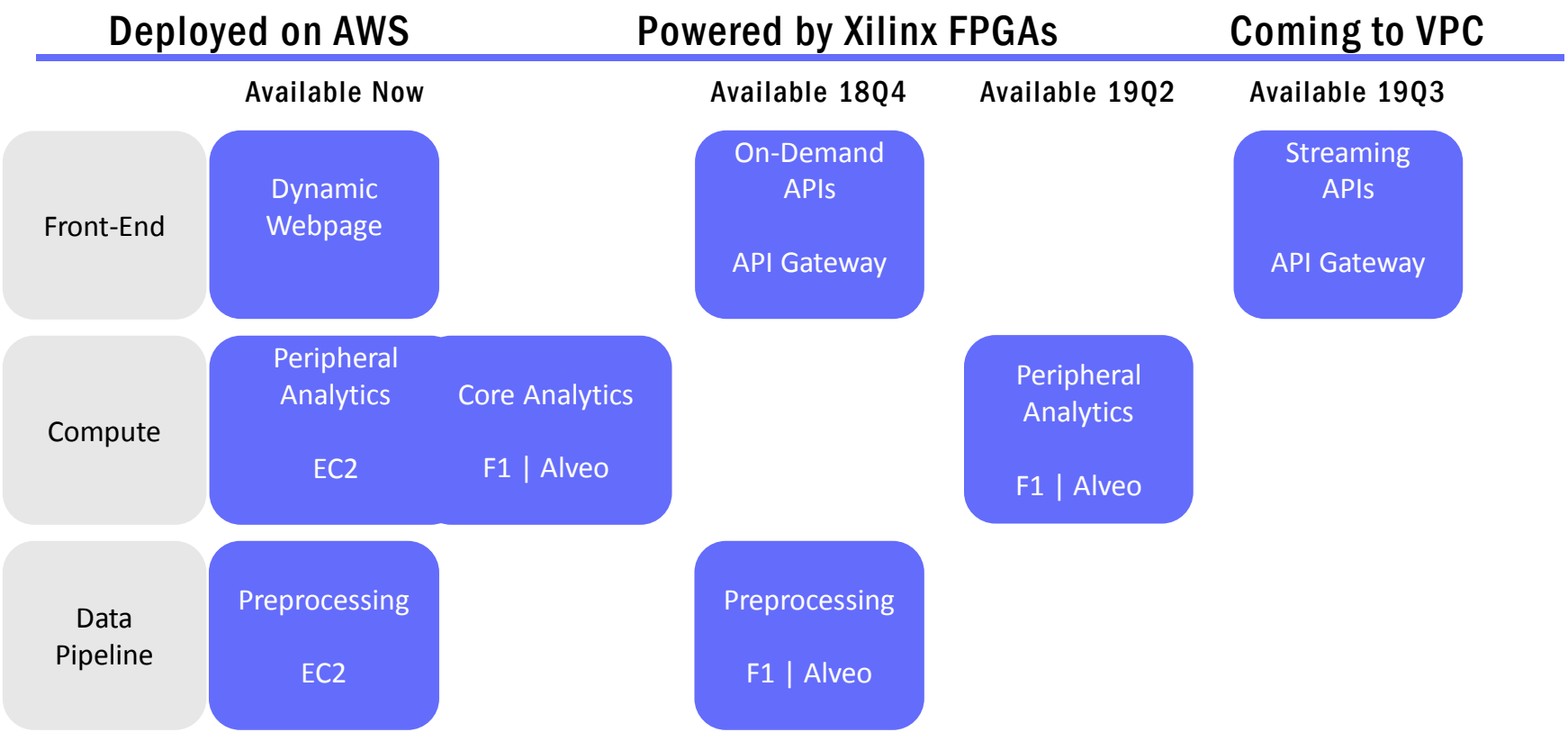
PRODUCT ROADMAP



PRODUCT ROADMAP



PRODUCT ROADMAP



BENCHMARKS

> 100x Faster

Much higher topic quality

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	10	-9.7	-19.9	-19.9
		4	0.3	648	32	-9.4	-14.7	-15.5
		8	0.6	1157	51	-11.9	-11.8	-16.0
29,287	78,063	1	0.1	220	10	-4.9	-5.8	-18.2
		4	0.2	588	36	-9.8	-14.9	-16.9
		8	0.7	1023	57	-8.3	-15.8	-16.3
158,348	336,532	1	0.4	2817	98	-4.8	-14.0	-18.0
		4	1.0	11282	371	-6.8	-12.7	-17.6
		8	1.7	N/A	685	-6.9	N/A	-16.0
2,620,008	3,253,518	1	9	39600	8135	-10.0	-15.5	-15.4
		4	21	N/A	34850	-9.0	N/A	-16.8
		8	34		N/A	-6.3		N/A
7,539,661	7,817,625	1	20	N/A	N/A	-2.5	N/A	N/A
		4	54			-3.5		
		8	113			-5.8		

U_Mass: less negative is better

AWS c4.8xlarge, f1.2xlarge

N/A indicates run not completed within 36 hours

DEMO

Alpha Generation in Equities

1day 14h -> 3min 40s

- Portfolio managers, financial analysts, quants

See for yourself

Social Media Monitoring

6days -> 55min

- Intelligence analysts, compliance officers, data scientists

IN SUMMARY

- Nucleus by SumUp offers the fast, flexible and transparent text analytics solution needed by professionals
- Nucleus typically is > 100x faster than alternative solutions
- Further acceleration is achievable through complete integration of data pipeline and algos on FPGA board

THANK YOU

NUCLEUS
by SumUp


XDF
XILINX
DEVELOPER
FORUM