Turbo-Charge Your Insight

Presented By

SUMUP ANALYTICS

Tanya Roosta, PhD
Lead AI Researcher
October 2nd 2018

Jim Wu
Lead Acceleration Engineer
TURBO CHARGE YOUR INSIGHT

SUMUP ANALYTICS
THE PAIN POINT

Too Much To Read

Time Is Scarce
WHAT WE DO

High-granularity real-time analysis on Unstructured data

- Time-consuming
- Voluminous
- Unstructured
- Overwhelming
WHAT WE DO

High-granularity real-time analysis on Unstructured data

- Time-consuming
- Voluminous
- Unstructured
- Overwhelming
WHAT WE DO

High-granularity real-time analysis on Unstructured data

- Finance
- Govt Intel

- Time-consuming
- Voluminous
- Unstructured
- Overwhelming
WHAT WE DO

High granularity real-time analysis on Unstructured data

- Finance
- Govt Intel
- Social Media

× Time-consuming
× Voluminous
× Unstructured
× Overwhelming
WHAT WE DO

High-granularity real-time analysis on Unstructured data

Finance

Govt Intel

x Time-consuming
x Voluminous
x Unstructured
x Overwhelming

Social Media

Education
WHAT WE DO

High-granularity real-time analysis on Unstructured data

- Finance
- Govt Intel
- Social Media
- Education
- Law

x Time-consuming
x Voluminous
x Unstructured
x Overwhelming
WHAT WE DO

High-granularity real-time analysis on Unstructured data

- Finance
- Govt Intel
- Health care
- Law
- Social Media
- Education

- Time-consuming
- Voluminous
- Unstructured
- Overwhelming
WHAT WE DO

High-granularity real-time analysis on Unstructured data

Finance
Govt Intel
Social Media
Education
Health care
Law

> 100x Speed-Up

SumUp Analytics

> 100x Speed-Up
<table>
<thead>
<tr>
<th>OUR SOLUTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleus SaaS</td>
</tr>
</tbody>
</table>

Step 1: Upload | DataFeeds
OUR SOLUTIONS

Step 1: Upload | DataFeeds

Step 2: Iterative Analysis
- Identify Topics & Summarize
- Analyze Sentiment & Consensus
- Get document recommendations
- Deep-dive on-demand

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

OUR SOLUTIONS

Nucleus SaaS  Streaming Analytics APIs  On-Demand Analytics APIs
DIFFERENTIATED INTELLECTUAL PROPERTIES

R&D in Analytics & Computational Systems

- Compressed Sensing
- SumUp Analytics
- Very-Large-Scale Robust & Convex Optimization
- Semi-Supervised NLP
DIFFERENTIATED INTELLECTUAL PROPERTIES

R&D in Analytics & Computational Systems

Leverages our Xilinx partnership

- FPGA-based Unstructured Data Analytics
- Very-Large-Scale Robust & Convex Optimization
- Complex Networks & Systems Visualization
- Compressed Sensing
- Real-Time Unsupervised Learning
- Semi-Supervised NLP
FPGA: NATURAL FIT FOR UNSTRUCTURED DATA

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+PCIe
- FPGA+SSD boards
- Vectorization, Tokenization, Regexp IPs
- Text AI algos on FPGA
- Text AI algos on FPGA
FPGA: POWERS NUCLEUS END TO END

**Workflow**
- Raw Text Upload
- Raw Text Preprocessing
- Core AI-based Analytics
- Peripheral AI-based Analytics

**FPGA Solution**
- FPGA+PCIe
- FPGA+SSD boards
- Vectorization, Tokenization, Regexp IPs
- Text AI algos on FPGA
- Text AI algos on FPGA

**Fits in single FPGA chip, min. data transfer**
FPGA: POWERS NUCLEUS END TO END

Workflow

Raw Text Upload
Raw Text Preprocessing
Core AI-based Analytics
Peripheral AI-based Analytics

FPGA Solution

FPGA+PCIe FPGA+SSD boards
Vectorization, Tokenization, Regexp IPs
Text AI algos on FPGA
Text AI algos on FPGA

Fits in single FPGA chip, min. data transfer
Takes out CPU, max I/O
## ACCELERATED DEV & DEPLOY W/ XILINX

<table>
<thead>
<tr>
<th>Xilinx Public IPs</th>
<th>SumUp Python/SDAccel Wrapper</th>
<th>AWS F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prototyping</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>08 / 2018</th>
<th>05 / 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python library</td>
<td>Core algo</td>
</tr>
<tr>
<td>Xilinx IPs</td>
<td>Standalone FPGA-based, v1</td>
</tr>
<tr>
<td>SDAccel</td>
<td></td>
</tr>
</tbody>
</table>
# ACCELERATED DEV & DEPLOY W/ XILINX

<table>
<thead>
<tr>
<th>Xilinx Public IPs</th>
<th>SumUp Python/SDAccel Wrapper</th>
<th>AWS F1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prototyping</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Integration</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>08 / 2018</th>
<th>05 / 2018</th>
<th>06 / 2018</th>
<th>07 / 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python library</td>
<td>Core algo</td>
<td>Core algo</td>
<td>Core algo</td>
</tr>
<tr>
<td>Xilinx IPs</td>
<td>Standalone FPGA-based, v1</td>
<td>Standalone FPGA-based, v2</td>
<td>F1-based, v3</td>
</tr>
<tr>
<td>SDAccel</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*SumUp Confidential - Do Not Distribute*
<table>
<thead>
<tr>
<th>Xilinx Public IPs</th>
<th>SumUp Python/SDAccel Wrapper</th>
<th>AWS F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prototyping</td>
<td>Integration</td>
<td>Deployment</td>
</tr>
<tr>
<td>08 / 2018</td>
<td>05 / 2018</td>
<td>08 / 2018</td>
</tr>
<tr>
<td>Python library</td>
<td>Core algo</td>
<td>Commercial library</td>
</tr>
<tr>
<td>Xilinx IPs</td>
<td>Core algo</td>
<td>F1-deployed</td>
</tr>
<tr>
<td>SDAccel</td>
<td>Core algo</td>
<td>Alveo-ready</td>
</tr>
<tr>
<td></td>
<td>Standalone FPGA-based, v1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standalone FPGA-based, v2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Core algo F1-based, v3</td>
<td></td>
</tr>
</tbody>
</table>
# PRODUCT ROADMAP

<table>
<thead>
<tr>
<th>Deployed on AWS</th>
<th>Powered by Xilinx FPGAs</th>
<th>Coming to VPC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Available Now</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Front-End</strong></td>
<td><strong>Dynamic Webpage</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Compute</strong></td>
<td><strong>Peripheral Analytics</strong></td>
<td>Core Analytics</td>
</tr>
<tr>
<td></td>
<td><strong>EC2</strong></td>
<td>**F1</td>
</tr>
<tr>
<td><strong>Data Pipeline</strong></td>
<td><strong>Preprocessing</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>EC2</strong></td>
<td></td>
</tr>
</tbody>
</table>

SumUp Confidential - Do Not Distribute
<table>
<thead>
<tr>
<th>Deployed on AWS</th>
<th>Powered by Xilinx FPGAs</th>
<th>Coming to VPC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Front-End</strong></td>
<td><strong>Available Now</strong></td>
<td><strong>Available 18Q4</strong></td>
</tr>
<tr>
<td>Dynamic Webpage</td>
<td></td>
<td>On-Demand APIs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>API Gateway</td>
</tr>
<tr>
<td><strong>Compute</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peripheral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analytics</td>
<td>Core Analytics</td>
<td></td>
</tr>
<tr>
<td>EC2</td>
<td>F1</td>
<td>Alveo</td>
</tr>
<tr>
<td><strong>Data Pipeline</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preprocessing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*SumUp Confidential - Do Not Distribute*
<table>
<thead>
<tr>
<th>Deployed on AWS</th>
<th>Powered by Xilinx FPGAs</th>
<th>Coming to VPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front-End</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Webpage</td>
<td>On-Demand APIs</td>
<td></td>
</tr>
<tr>
<td>Peripheral</td>
<td>Core Analytics</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Analytics</td>
<td></td>
<td>Analytics</td>
</tr>
<tr>
<td>EC2</td>
<td></td>
<td>F1</td>
</tr>
<tr>
<td>Preprocessing</td>
<td></td>
<td>F1</td>
</tr>
<tr>
<td>Data Pipeline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Available Now

Available 18Q4

Available 19Q2

On-Demand APIs

API Gateway

Coming to VPC

SumUp Confidential - Do Not Distribute
PRODUCT ROADMAP

Deployed on AWS | Powered by Xilinx FPGAs | Coming to VPC

Front-End
- Dynamic Webpage

Compute
- Peripheral Analytics
  - EC2
- Core Analytics
  - F1 | Alveo
- Preprocessing
  - EC2
- Preprocessing
  - F1 | Alveo
- On-Demand APIs
- API Gateway
- Peripheral Analytics
  - F1 | Alveo
- Streaming APIs
- API Gateway

Available Now
- On-Demand APIs
- API Gateway

Available 18Q4
- Peripheral Analytics
  - F1 | Alveo

Available 19Q2
- Peripheral Analytics
  - F1 | Alveo

Available 19Q3
- Streaming APIs
- API Gateway

SumUp Confidential - Do Not Distribute
### Benchmarks

- **> 100x Faster**

<table>
<thead>
<tr>
<th>Nbr docs</th>
<th>Dictionary length</th>
<th>Nbr topics</th>
<th>Runtime in seconds</th>
<th>U_Mass Coherence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nucleus</td>
<td>Scikit</td>
</tr>
<tr>
<td>31,496</td>
<td>81,801</td>
<td>1</td>
<td>0.1</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.3</td>
<td>648</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>0.6</td>
<td>1157</td>
</tr>
<tr>
<td>29,287</td>
<td>78,063</td>
<td>1</td>
<td>0.1</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.2</td>
<td>588</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>0.7</td>
<td>1023</td>
</tr>
<tr>
<td>158,348</td>
<td>336,532</td>
<td>1</td>
<td>0.4</td>
<td>2817</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>1.0</td>
<td>11282</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>1.7</td>
<td>N/A</td>
</tr>
<tr>
<td>2,620,008</td>
<td>3,253,518</td>
<td>1</td>
<td>9</td>
<td>39600</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>21</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>34</td>
<td>N/A</td>
</tr>
<tr>
<td>7,539,661</td>
<td>7,817,625</td>
<td>1</td>
<td>20</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>54</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>113</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**U_Mass:** less negative is better

---

**AWS c4.8xlarge, f1.2xlarge**

N/A indicates run not completed within 36 hours

---

SumUp Confidential - Do Not Distribute
**Alpha Generation in Equities**

1 day 14h -> 3min 40s

- Portfolio managers, financial analysts, quants

**Social Media Monitoring**

6 days -> 55min

- Intelligence analysts, compliance officers, data scientists

See for yourself
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