How we built a highly scalable Machine Learning platform using Apache Mesos

Daniel Sârbe
Development Manager, BigData and Cloud Machine Translation @ SDL
Co-founder of BigData/DataScience Meetup Cluj, Romania

October 2017
Agenda

1. Some Machine Translation context
2. Why we needed a new platform?
3. Architecture overview of our current SaaS platform
4. New microservices platform
5. Lesson Learned
6. Demo – Scaling micro-services to handle traffic increase
7. Q&A
• Software development background (13+ years)
• Passionate about people and technology
• Interest in anything that is related to Scalability, BigData, Machine Learning
• Currently leading the BigData and Cloud Machine Translation group at SDL Cluj, Romania
• Co-founder of BigData/DataScience Meetup Cluj
  – 1200+ members
  – 25+ events organized
    ▪ meetups with more 100+ participants
    ▪ workshops with 30 people
Machine Translation context

Technological Progress

2000  2005  2015
Machine Translation quality improvements

Technological Progress

2000 2005 2015

Customization

Generic  Industry Vertical  Custom Trained
Machine Translation quality improvements

On average, customized engines handled industry terminology 24% better than standard MT baselines.
Adaptive Machine Translation idea

Machine translation

Machine learns during the statistical training process

Machine does not learn or improve during the translation process

Machine translation + AdaptiveMT

Machine learns during the statistical training process

Machine learns from user feedback during the translation process
Learning from Post-Edits

Update operation

Adaptive MT Model

Source
MT Output
Post-Edit
An update of one of the statistical MT models, the translation model

Source: No further requirements are needed
MT: Pas d'autres exigences sont requises
PE: Aucune exigence supplémentaire n’est nécessaire
### Good New Translations
- needed -> *requises* - nécessaire
- further -> *autres* - supplémentaire
- no further requirements are -> aucune exigence supplémentaire n’est

### Bad New Translations
- are -> n’est
- requirements -> exigence
- no -> aucune

- Statistical features help choose good rules, and decide when to use them
AdaptiveMT progressive impact

To PE 300 segments with AdaptiveMT, we need 250 fewer edits than without AdaptiveMT.
Second use-case – Neural MT

• Rule-based
  – define and build the model by hand

• Traditional SMT
  – define the model by hand then statistically learn it from data

• Neural MT
  – design architecture to automatically discover, define, and learn the models from data
Neural MT

• Uses a deep learning architecture capable of learning the meaning of the text
  – fluent and naturally sounding translation output

• Neural MT shows significant translation quality improvement over SMT
  – captures both local and global dependencies and can handle long-range word reordering
  – e.g. we observe an impressive 30% improvement on English-German
Neural MT in the Cloud

To accommodate NMT in the Cloud we need:

– new hardware: GPUs
– flexible infrastructure (new&old engines)
– break the old implementation (independent services)
– new modern API (for new clients on-bording)
How can we do this?
What we had before?

Our Legacy SaaS platform

Mature, Iteratively-developed platform

- >15 billion words translated in average month
- >200 million translation request/month

No P1/P2 Bugs in last 24 months
Availability: 99.9x

The only large-scale, commercial-grade MT solution other than Google and Microsoft
Translation engines are not modular and is difficult to add new functionality.

Redundant flows and services based on outdated requirements.

Scaling up-down requires manual intervention and allocation of new VMs.

Overall, monolithic design that is hard to adopt for new use-cases.
A new platform
What do we want to achieve? – Key Concepts

• Scalability
• Latency
• Independent (micro-)services
• Elasticity (auto-scaling)
• Fault-tolerance & robustness
• Infrastructure automation
• Reliable monitoring and alerts
Architecture evolution

• eBay
  • 5th generation today
  • Monolithic Perl -> Monolithic C++ -> Java -> microservices
• Twitter
  • 3rd generation today
  • Monolithic Rails -> JS / Rails / Scala -> microservices
• Amazon
  • Nth generation today
  • Monolithic C++ -> Java / Scala -> microservices
• SDL MT
  • 3rd generation today
  • Monolithic Rails -> Monolithic Java -> microservices
# Our Technology Stack – New microservices platform

<table>
<thead>
<tr>
<th><strong>Mesos</strong></th>
<th>Cluster manager</th>
<th><strong>Scaling</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Marathon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBase</td>
<td>NoSQL, column-oriented db</td>
<td></td>
</tr>
<tr>
<td>Hadoop</td>
<td>Storage</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Kafka</strong></th>
<th>Messaging system</th>
<th><strong>Latency, Fault-tolerance</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Zookeeper</td>
<td>Centralized coordination service</td>
<td></td>
</tr>
<tr>
<td>Protocol Buffers</td>
<td>Serializing structured data – developed by Google</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Ansible</strong></th>
<th>IT automation</th>
<th><strong>Infrastructure automation, Elasticity</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS</td>
<td>Cloud Computing Services</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>ELK stack</strong></th>
<th>Log aggregation, search server(indexing &amp; querying logs)</th>
<th><strong>Monitoring and alerts</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grafana</strong></td>
<td>Beautiful metric &amp; analytic dashboards</td>
<td></td>
</tr>
<tr>
<td><strong>OpenTSDB</strong></td>
<td>Real-time metrics</td>
<td></td>
</tr>
<tr>
<td><strong>Zabbix</strong></td>
<td>Monitoring solution</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Docker</strong></th>
<th>Containerization platform</th>
<th><strong>Microservices</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DropWizard</strong></td>
<td>REST application bootstrap framework</td>
<td></td>
</tr>
<tr>
<td><strong>SpringBoot</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Java8</strong></td>
<td>Programming language</td>
<td></td>
</tr>
</tbody>
</table>

**AWS** – Amazon Web Services

**Protocol Buffers** – Serializing structured data – developed by Google

**Mesos** – Cluster manager

**Marathon** – IT automation

**HBase** – NoSQL, column-oriented db

**Hadoop** – Storage

**Kafka** – Messaging system

**ZooKeeper** – Centralized coordination service

**Protocol Buffers** – Serializing structured data

**Ansible** – IT automation

**AWS** – Cloud Computing Services

**ELK stack** – Log aggregation, search server(indexing & querying logs)

**Grafana** – Beautiful metric & analytic dashboards

**OpenTSDB** – Real-time metrics

**Zabbix** – Monitoring solution

**DropWizard** – REST application bootstrap framework

**SpringBoot** – Framework

**Java8** – Development language

**Docker** – Containerization platform

**SDS** – Service Delivery System

**Ansible** – IT automation

**AWS** – Cloud Computing Services

**DropWizard** – REST application bootstrap framework

**SpringBoot** – Framework

**Java8** – Development language

**Docker** – Containerization platform

**SDS** – Service Delivery System
Lessons Learned

“No regrets in life. Just lesson learned.”
1. Cost efficient

- Dev/QA/Clone clusters – ~40% cost
  - issues found only in aws-qa
  - prod clone has the same IPs/confs as prod

- AWS
  - periodical cleanup
  - email alerts on no of instances running
  - r3.4xlarge -> r4.4xlarge ($1.33/h -> $1.06/h)
  - ElasticBlockStore(EBS) vs ElasticFileSystem(EFS)
  - reserved instances
2. Security

• ssh via a single aws bastion machine
• gpg encryption of confs
  – no clear passwords in git
  – restrict access to specific envs
• secure Marathon/Kibana/HAPProxyUI
• AWS termination protection
3. Platform high availability

• Infrastructure allocation
  – +1 node/cluster
  – one instance decommissioned/stopped (AWS EC2/human error)

• Microservices
  – 2 instances/micoservice
  – unique constraints should be set

• Test with 5x-10x more traffic

• Early monitoring on all fronts
  – infrastructure - Zabbix
  – app metrics - OpenTSDB
  – usage stats - ELK
  – external - Pingdom
4. Resource allocation

- Memory limitations for containers
  - Marathon memory settings were not enforced on container level
  - reported container memory = host memory
  - enforce Xmx, Xms (OOM)
  - crash dumps (mount partitions to have a crash dump)

- Under high load
  - hard to monitor resource usage (CPU/IOWAIT/Network/IOPS)

- CPU weights (Marathon)
  - reduced 1 to 0.1 – overprovision
5. Releases are not as easy as expected

• **No downtime releases**
  - simulate 2-4 times the releases in prod-clone
  - scripts to monitor downtime during deployment
  - connection draining (killed by default)
  - messages compatibility (using protoBuff)

• **Ansible-ize the manual steps**
  - prod-clone commands run on prod (gpg to fix)
  - 0 to cluster (in x min)
6. Investigations become more complex

- **Logs**
  - file based logs vs centralized logs
  - aggregated logs into ELK(requestId)
  - using stdout -> no disk space on mesos-slaves (disable)
  - under high load the gelf appender caused slowdowns
    - move to log4j-kafka

- **Metrics in OpenTSDB**
  - application-specific-metrics

- **Correlation between various sources**
7. Independent microservices

• Keep microservices as independent and as small as possible
  – 30+ microservices
  – a challenge, especially for legacy code (unit tests)
  – continuous refactoring even for all microservices
8. Periodically reevaluate assumptions

- Follow user behavior over time
  - Users behavior is different from what we expected
  - API flow changes (v2/v3 for some APIs)
  - Speed is more important on some flows (sync)

- Scale the microservices based on usage
Future improvements
Future improvements

• Maintenance and evolution of the platform
  – Periodical upgrades of the stack
  – Improve monitoring

• Auto-scaling
  – based on usage patterns
  – use of spot instances

• Move all components in Mesos(DCOS)
  – HBase
  – ElasticSearch
  – Kafka
  – HDFS
Demo Time!

- Scaling micro-services to handle traffic increase
Questions?
MesosCon
EUROPE