



Spark Service on Kubernetes

Motivation, current status and objectives



Spark Service on Kubernetes

- Current state of the art for Spark at CERN
- Future demand outlook – LHC experiments and big data
- Spark on Kubernetes
- State of the art in industry
- Current progress of Spark on Kubernetes
- What next?

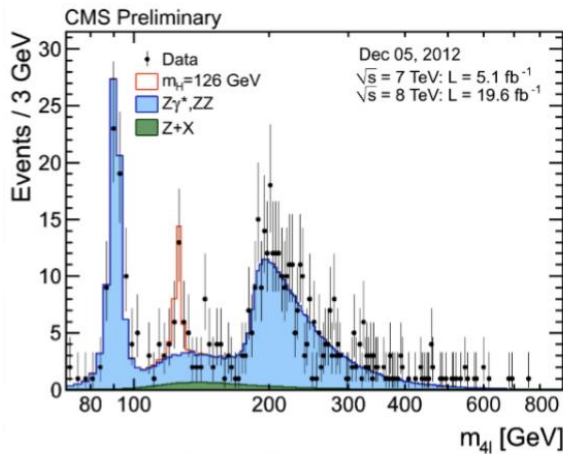
Current state of the art for Apache Spark at CERN

- Spark running on top of Hadoop/YARN (distributed filesystem for big data and cluster resource manager).
- Physical machines allocated – means no elasticity, no isolation, not cloud-ready model (Openstack)
- Stable workloads from monitoring, security, plus some other smaller communities
- Sometimes more busy due to high load from physics analysis

Spark Service on Kubernetes

- Current state of the art for Spark at CERN
- LHC Experiments and big data – future demand outlook
- Spark on Kubernetes
- State of the art in industry
- Current progress of Spark on Kubernetes
- What next?

LHC Experiments and big data – future outlook

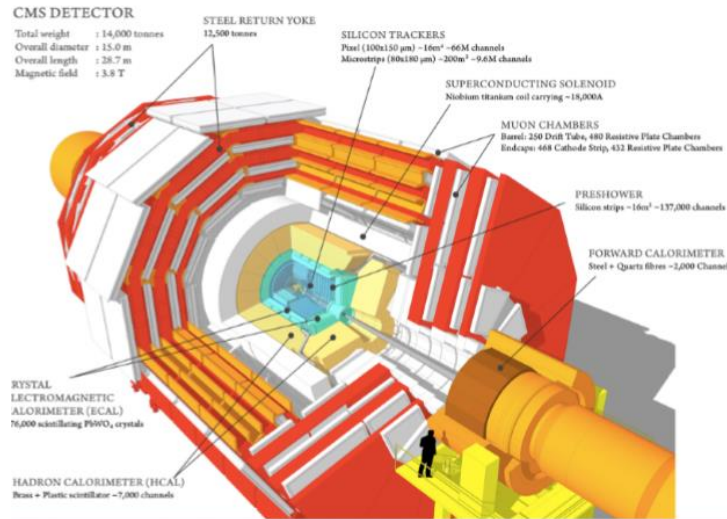


Detect particle interactions (data), compare with theory predictions (simulation)



Intel Big Data Analytics
CMS Data Analysis with Apache Spark

Particle detection analysis



Large scale data reduction facility

Experiments and big data – future outlook

Towards interactive data analysis for TOTEM experiment using Apache Spark



📅 29 Jan 2018, 11:30

🕒 30m

📍 AGH Computer Science Building D-17

Presentation

User Voice: Novel Applic...

Speakers

- 👤 Mr. Grzegorz Bogdał (AGH University of Sc...)
- 👤 Mr. Piotr Gawryś (AGH University of Sc...)
- 👤 Mr. Paweł Nowak (AGH University of Sc...)
- 👤 Mr. Łukasz Plewnia (AGH University of Sc...)
- 👤 Leszek Grzanka (AGH University of Sc...)
- 👤 Maciej Malawski (AGH University of Sc...)

Description

Data analysis in High Energy Physics experiments requires processing of large amounts of data. As the main objective is to find interesting events from among those recorded by detectors, the typical operations involve data filtering by applying cuts

100s of Root files for offline analysis

LHC Experiments and big data – future outlook

Data Intensive – lots of data to be processed and reduced

Data stored in external place – EOS with around 250 PB

Sporadically analyzed

How to achieve elasticity?

How to easily deploy applications?

How to make use of data stored on EOS and not HDFS ?

How to ensure that service can self-heal and recover from failures easier?

Spark Service on Kubernetes

- Current state of the art for Spark at CERN
- Future demand outlook – LHC experiments and big data
- Spark on Kubernetes
- State of the art in industry
- Current progress of Spark on Kubernetes
- What next?

Spark on Kubernetes – why for this use case?



Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications.

15 years of virtualization experience from Google

Scales horizontally. Self-healing system - restarts containers that fail, replaces and reschedules containers when nodes die or container not respond

Large and industry-grade community

Kubernetes already gets adopted at CERN.

Spark on Kubernetes – why for this use case?

In development by:



Bloomberg
Google
Haiwen
Hyperpilot
Intel
Palantir
Pepperdata
Red Hat



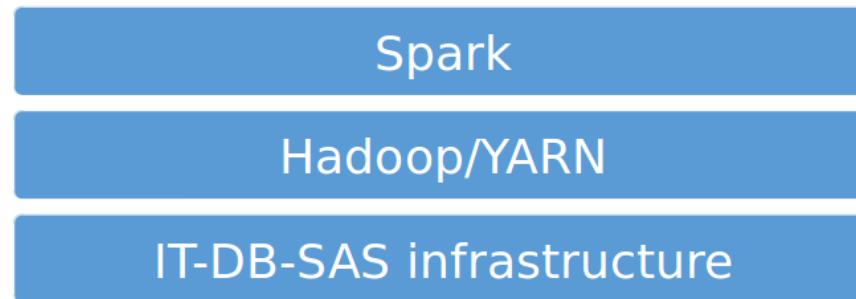
Spark runs as any other application in Kubernetes cluster

Spark on Kubernetes released this month in Spark 2.3 and in active development currently.

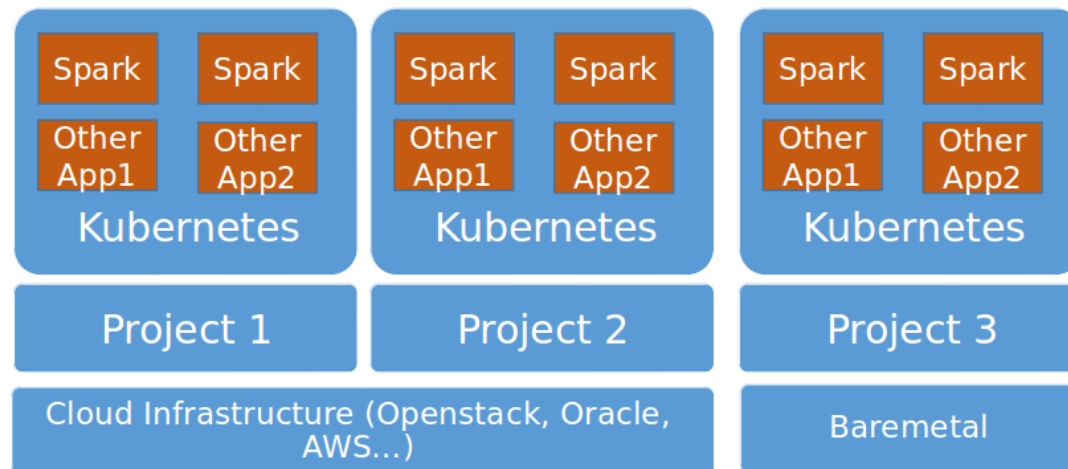
Get better understanding of the system for big data applications at CERN

Spark on Kubernetes – Comparison to YARN

Spark on Hadoop/YARN



Spark on Kubernetes

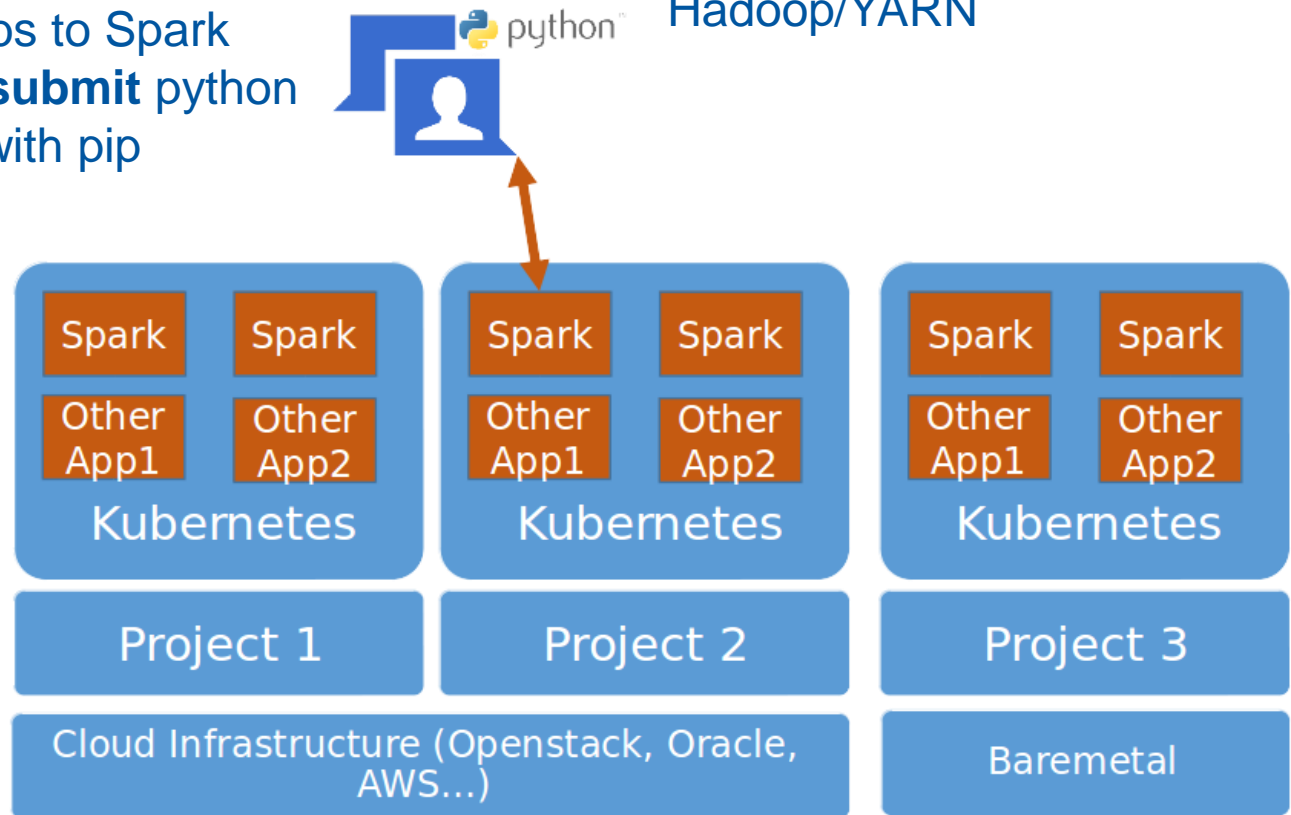


Spark on Kubernetes – How it works?

1. Create Kubernetes cluster and initialize its dependencies

2. Submit Spark jobs to Spark using **cern-spark-submit** python package installed with pip

3. Our docker images will deploy and run your Spark application over Kubernetes as it was on Hadoop/YARN



```
→ ~ cern-spark-service kube-create --openstack --only-spark-deps
Kubernetes cluster exists, fetch only cluster configuration and initialize spark dependencies
```

```
-- MENU: Select Openstack project in which to configure cluster --
```

- (0) IT Hadoop
- (1) IT Hadoop service development
- (2) Spark as a Service
- (3) IT Hadoop HDP1 - Ironic
- (4) IT Data Analytics Development
- (5) IT Data Analytics - Container Pilot
- (6) Personal pmrowczy
- (7) IT Hadoop on VMs with local storage
- (8) IT Database Cloud 02 Test
- (9) IT Spark as a Service
- (10) Config Training
- (11) IT Application Server Infrastructure

```
[?] Enter menu selection: 9
```

```
-- MENU: Select cluster to configure --
```

- (0) spark-service-cluster

```
[?] Enter menu selection: 0
```

```
Kubernetes cluster [name: spark-service-cluster] configuration init in progress..
```

```
Kubernetes client configuration..
```

```
spark-service-cluster
```

```
https://137.138.150.224:6443
```

```
/home/mrow4a/.cern-spark-service/spark-service-cluster/ca.pem
```

```
/home/mrow4a/.cern-spark-service/spark-service-cluster/cert.pem
```

```
/home/mrow4a/.cern-spark-service/spark-service-cluster/key.pem
```

```
Init Spark on Kubernetes Resource Staging Server..
```

```
spark-resource-staging-server-7f49df6fc-26hv9 default Running
```

```
2018-03-21 09:10:32 INFO log:192 - Logging initialized @19794ms
```

```
2018-03-21 09:10:36 WARN ContextHandler:1444 - o.s.j.s.ServletContextHandler@301eda63{/,null,null} contextPath end
```

```
2018-03-21 09:10:42 WARN NativeCodeLoader:62 - Unable to load native-hadoop library for your platform... using built-in
```

```
2018-03-21 09:10:44 INFO SecurityManager:54 - Changing view acls to: root
```

```
2018-03-21 09:10:44 INFO SecurityManager:54 - Changing modify acls to: root
```

```
2018-03-21 09:10:44 INFO SecurityManager:54 - Changing view acls groups to:
```

```
2018-03-21 09:10:44 INFO SecurityManager:54 - Changing modify acls groups to:
```

```
2018-03-21 09:10:44 INFO SecurityManager:54 - SecurityManager: authentication disabled; ui acls disabled; users with
```

```
permissions: Set(root); groups with modify permissions: Set()
```

```
2018-03-21 09:10:45 INFO Server:345 - jetty-9.3.z-SNAPSHOT
```

```
2018-03-21 09:10:57 INFO ContextHandler:781 - Started o.s.j.s.ServletContextHandler@301eda63{/api,null,AVAILABLE}
```

```
2018-03-21 09:10:57 INFO AbstractConnector:270 - Started ServerConnector@2546d2b8{HTTP/1.1,[http/1.1]}{0.0.0.0:10000}
```

```
2018-03-21 09:10:57 INFO Server:403 - Started @44995ms
```

```
2018-03-21 09:10:57 INFO ResourceStagingServer:54 - Resource staging server started on port 10000.
```

```
Finished kubernetes cluster creation!
```

Execute command

Choose project on openstack

Choose cluster in the project

Script will initialize configuration on the client and initialize required spark dependencies on the cluster (resource staging, shuffle service)

```
→ spark-service-tools git:(master) X cern-spark-service spark-submit \  
--conf spark.executor.instances=3 \  
--class org.sparkservice.sparkrootapplications.examples.DimuonReductionAOD \  
--jars \  
http://central.maven.org/maven2/org/diana-hep/spark-root_2.11/0.1.16/spark-root_2.11-0.1.16.jar,\  
http://central.maven.org/maven2/org/diana-hep/histogrammar-sparksql_2.11/1.0.3/histogrammar-sparksql_2.11-1.0.3.jar, \  
http://central.maven.org/maven2/org/diana-hep/root4j/0.1.6/root4j-0.1.6.jar,\  
http://central.maven.org/maven2/org/diana-hep/histogrammar_2.11/1.0.3/histogrammar_2.11-1.0.3.jar,\  
http://central.maven.org/maven2/org/apache/bcel/bcel/5.2/bcel-5.2.jar,\  
http://central.maven.org/maven2/org/tukaani/xz/1.2/xz-1.2.jar,\  
http://central.maven.org/maven2/jakarta-regexp/jakarta-regexp/1.4/jakarta-regexp-1.4.jar \  
/home/mrow4a/Projects/spark-service-examples/target/scala-2.11/spark-service-examples_2.11-0.0.1.jar \  
root://eospublic.cern.ch//eos/opendata/cms/MonteCarlo2012/Summer12_DR53X/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/AODSIM/PU_RD1_ST
```



```
Running spark-submit with cluster config...
/home/mrow4a/.cern-spark-service/spark-service-cluster
spark-service-cluster
https://137.138.150.224:6443
org.sparkservice.sparkrootapplications.examples.dimuonreductionaod.1521621773
137.138.122.195

2018-03-21 09:42:57 WARN NativeCodeLoader:62 - Unable to load native-hadoop library for your platform... us
2018-03-21 09:42:57 INFO SecurityManager:54 - Changing view acls to: root
2018-03-21 09:42:57 INFO SecurityManager:54 - Changing modify acls to: root
2018-03-21 09:42:57 INFO SecurityManager:54 - Changing view acls groups to:
2018-03-21 09:42:57 INFO SecurityManager:54 - Changing modify acls groups to:
2018-03-21 09:42:57 INFO SecurityManager:54 - SecurityManager: authentication disabled; ui acls disabled; u
permissions: Set(root); groups with modify permissions: Set()
2018-03-21 09:42:57 WARN DriverServiceBootstrapStep:66 - Driver's hostname would preferably be org-sparkser
long (must be <= 63 characters). Falling back to use spark-1521621777548-driver-svc as the driver service's
2018-03-21 09:42:58 INFO LoggingPodStatusWatcherImpl:54 - State changed, new state:
  pod name: org.sparkservice.sparkrootapplications.examples.dimuonreductionaod.1521621773
  namespace: default
  labels: spark-app-selector -> spark-d354b4dcc7844a509b687adf59b4e237, spark-role -> driver
  pod uid: f9e9ac44-2ce3-11e8-abe6-fa163e4a4036
  creation time: 2018-03-21T08:43:49Z
  service account name: default
  volumes: kubernetes-credentials, spark-local-dir-0-spark-296413b5-5c28-451a-8ca7-cbad9e0ade1e, spar
  node name: N/A
  start time: N/A
  container images: N/A
  phase: Pending
  status: []
2018-03-21 09:42:58 INFO LoggingPodStatusWatcherImpl:54 - State changed, new state:
  pod name: org.sparkservice.sparkrootapplications.examples.dimuonreductionaod.1521621773
  namespace: default
  labels: spark-app-selector -> spark-d354b4dcc7844a509b687adf59b4e237, spark-role -> driver
  pod uid: f9e9ac44-2ce3-11e8-abe6-fa163e4a4036
  creation time: 2018-03-21T08:43:49Z
  service account name: default
  volumes: kubernetes-credentials, spark-local-dir-0-spark-296413b5-5c28-451a-8ca7-cbad9e0ade1e, spar
  node name: spark-service-cluster-avzbbdfg6m3d-minion-11
  start time: N/A
  container images: N/A
  phase: Pending
  status: []
2018-03-21 09:42:58 INFO LoggingPodStatusWatcherImpl:54 - State changed, new state:
  pod name: org.sparkservice.sparkrootapplications.examples.dimuonreductionaod.1521621773
  namespace: default
  labels: spark-app-selector -> spark-d354b4dcc7844a509b687adf59b4e237, spark-role -> driver
  pod uid: f9e9ac44-2ce3-11e8-abe6-fa163e4a4036
  creation time: 2018-03-21T08:43:49Z
  service account name: default
  volumes: kubernetes-credentials, spark-local-dir-0-spark-296413b5-5c28-451a-8ca7-cbad9e0ade1e, spar
  node name: spark-service-cluster-avzbbdfg6m3d-minion-11
  start time: 2018-03-21T08:43:49Z
  container images: gitlab-registry.cern.ch/db/spark-service/docker-registry:spark-driver
```

```

-- sumChargedParticlePt: float (nullable = true)
-- sumNeutralHadronEt: float (nullable = true)
-- sumPhotonEt: float (nullable = true)
-- sumNeutralHadronEtHighThreshold: float (nullable = true)
-- sumPhotonEtHighThreshold: float (nullable = true)
-- sumPUpt: float (nullable = true)
-- type_: integer (nullable = true)
-- pfP4_: struct (nullable = true)
  |-- fCoordinates: struct (nullable = true)
    |-- fx: double (nullable = true)
    |-- fy: double (nullable = true)
    |-- fz: double (nullable = true)
    |-- ft: double (nullable = true)

```

```

2018-03-21 09:44:53 INFO spark-root:166 - Map(tree -> Events, path -> root://eospublic.cern.ch//eos/opensdata/cms/MonteCarlo2012/Summer12_DR53X/DYJetsToLL_M-50_TuneZ2Star_8
D1_START53_V7N-v1/20000/DCF94DC3-42CE-E211-867A-001E67398011.root)
2018-03-21 09:44:53 INFO FileSourceScanExec:54 - Planning scan with bin packing, max size: 134217728 bytes, open cost is considered as scanning 4194304 bytes.
2018-03-21 09:44:53 INFO SparkContext:54 - Starting job: show at DimuonReductionAOD.scala:300
2018-03-21 09:44:53 INFO DAGScheduler:54 - Got job 1 (show at DimuonReductionAOD.scala:300) with 1 output partitions
2018-03-21 09:44:53 INFO DAGScheduler:54 - Final stage: ResultStage 2 (show at DimuonReductionAOD.scala:300)
2018-03-21 09:44:53 INFO DAGScheduler:54 - Parents of final stage: List()
2018-03-21 09:44:53 INFO DAGScheduler:54 - Missing parents: List()
2018-03-21 09:44:53 INFO DAGScheduler:54 - Submitting ResultStage 2 (MapPartitionsRDD[11] at show at DimuonReductionAOD.scala:300), which has no missing parents
2018-03-21 09:44:53 INFO MemoryStore:54 - Block broadcast_2 stored as values in memory (estimated size 135.9 KB, free 408.8 MB)
2018-03-21 09:44:53 INFO MemoryStore:54 - Block broadcast_2_piece0 stored as bytes in memory (estimated size 21.4 KB, free 408.7 MB)
2018-03-21 09:44:53 INFO BlockManagerInfo:54 - Added broadcast_2_piece0 in memory on spark-1521621777548-driver-svc.default.svc.cluster.local:7079 (size: 21.4 KB, free: 408.7 MB)
2018-03-21 09:44:53 INFO SparkContext:54 - Created broadcast 2 from broadcast at DAGScheduler.scala:1006
2018-03-21 09:44:53 INFO DAGScheduler:54 - Submitting 1 missing tasks from ResultStage 2 (MapPartitionsRDD[11] at show at DimuonReductionAOD.scala:300) (first 15 tasks are
2018-03-21 09:44:53 INFO KubernetesTaskSchedulerImpl:54 - Adding task set 2.0 with 1 tasks
2018-03-21 09:44:53 INFO KubernetesTaskSetManager:54 - Starting task 0.0 in stage 2.0 (TID 2, 10.100.23.3, executor 2, partition 0, PROCESS_LOCAL, 5453 bytes)
2018-03-21 09:44:54 INFO BlockManagerInfo:54 - Added broadcast_2_piece0 in memory on 10.100.23.3:39015 (size: 21.4 KB, free: 408.9 MB)
2018-03-21 09:45:02 INFO SparkContext:54 - Created broadcast 2 from broadcast at DAGScheduler.scala:1006
2018-03-21 09:45:02 INFO KubernetesTaskSetManager:54 - Finished task 0.0 in stage 2.0 (TID 2) in 9324 ms on 10.100.23.3 (executor 2) (1/1)
2018-03-21 09:45:02 INFO KubernetesTaskSchedulerImpl:54 - Removed TaskSet 2.0, whose tasks have all completed, from pool
2018-03-21 09:45:02 INFO DAGScheduler:54 - ResultStage 2 (show at DimuonReductionAOD.scala:300) finished in 9.327 s
2018-03-21 09:45:02 INFO DAGScheduler:54 - Job 1 finished: show at DimuonReductionAOD.scala:300, took 9.483278 s
2018-03-21 09:45:02 INFO CodeGenerator:54 - Code generated in 25.914246 ms

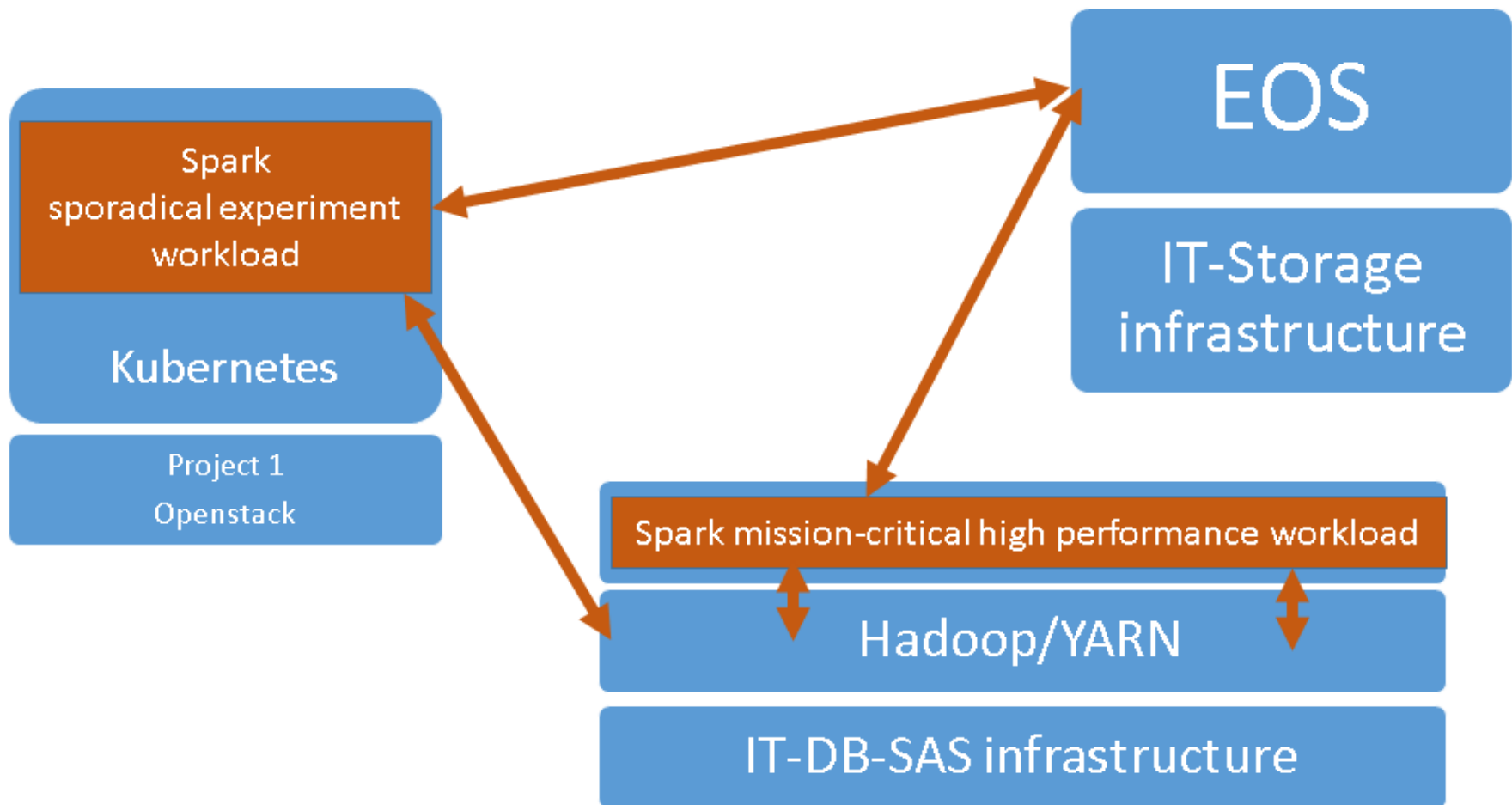
```

```

+-----+
|ml1
+-----+
|88.332664
|87.45533
|92.8502
|95.65151
|89.15023
|91.41383
|75.830734
|90.908035
|94.0488
|88.95186
|89.58696
|90.14875
|93.84803
|91.53471
|89.955284
|92.35516
|64.48392
|96.913216
|101.840485

```

Spark on Kubernetes – why for this use case?



State of the art in the industry - Kubernetes



"Kubernetes basically solved most of our problems. Before, the time of deployment took about a week, now it only takes minutes."

[Read about Huawei](#)



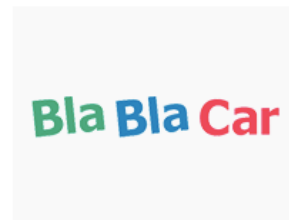
"Over the next couple of years, people won't even think that much about it when they want to run containers. Kubernetes is going to be the go-to solution."

[Read about Haufe Group](#)



"My message to other enterprises like us is you can actually integrate Kubernetes into an existing, well-orchestrated machinery."

[Read about BlackRock](#)



"[With Kubernetes] our infrastructure is much more resilient and we have better availability than before."

[Read about BlaBlaCar](#)



Using Kubernetes to reinvent the world's largest educational company



Kubernetes at Box: Microservices at Maximum Velocity



Inside eBay's shift to Kubernetes and containers atop OpenStack



Migrating from a homegrown 'cluster' to Kubernetes

State of the art in the industry – Spark Service



Hops

Hops - Apache Spark and Tensorflow as a Service on YARN (Zeppelin and Notebooks integrated)



Apache Spark for Azure HDInsight

Apache Spark in the cloud for mission critical deployments



PRODUCTS ▾

ELEMENTAL TO BIG DATA

Apache Spark
as a Service

IBM Analytics for Apache Spark

The Databricks Unified Analytics Platform

Harness the power of AI through a truly unified approach to data analytics from the team that created Apache Spark™.

Spark on Kubernetes

- Current state of the art for Spark at CERN
- Future demand outlook – experiments and big data
- Spark on Kubernetes
- State of the art in industry
- Current progress of Spark on Kubernetes
- What next?

Current progress

- Successfully deployed Spark on Kubernetes on OpenStack and built spark images and tooling
- We prototyped and made a proof of concept. Able to run root file analysis on EOS
- Work on the **cern-spark-service** package - <https://pypi.python.org/pypi/cern-spark-service> (installation with pip **install --upgrade cern-spark-service**)

Next Steps

April 2018 – Allow creation of spark-on-kub cluster on OpenStack and run spark workloads accessing EOS and HDFS (including necessary auth). Further improve usability of tooling.

May 2018 – Make spark-submission compatible with Kubernetes on baremetal / Helix Nebula test.

June 2018 – Benchmarks, fixes and adjustments to run large scale workloads

December 2018 – Multi tenancy on Spark-on-kub cluster

December 2018 – Integration of Spark as a Service with SWAN

Thank you!

Questions?

- Current state of the art for Spark at CERN
- Future demand outlook – LHC experiments and big data
- Spark on Kubernetes
- State of the art in industry
- Current progress of Spark on Kubernetes
- What next?

Spark on Kubernetes – Conclusions

Spark on Hadoop/YARN:

- good for production, high-availability workloads
- infrastructure, service and software stack maintained by IT-DB-SAS
- adding/removing physical machine is not that trivial

Spark on Kubernetes:

- allows extreme scale and sporadical workloads on your own project resources (Openstack, Cloud, Baremetal).
- shutdown and create on demand via Kubernetes over Openstack / Cloud
- lacks in performance and reliability (assumption), not suitable for continuous workloads

