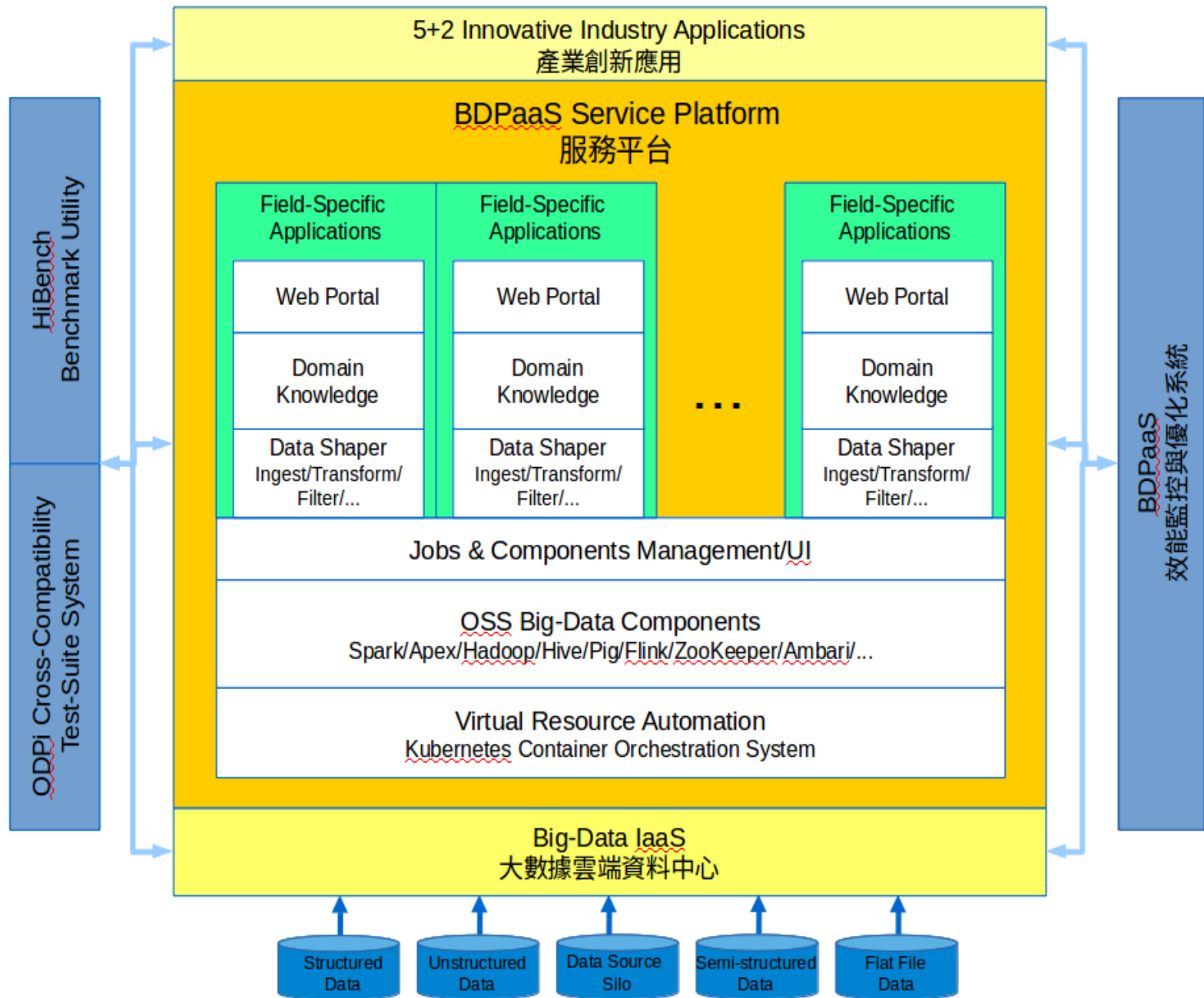


BDPaaS Big-Data Platform as a Service

實現 BDPaaS 服務平台之完整架構：



BDPaaS 實現三個模組:

- Jobs & Components Management/UI
 - 提供使用者整合的 Web 界面, 使用者可以建立與刪除 Spark 與 Apex project
 - 提供使用者統一的工作運行與操作的 Workspace
 - Spark workspace 由 Zeppelin(Apache 開源計畫)支援
<https://zeppelin.apache.org/>
 - Apex workspace 由 DataTorrent RTS 支援
<https://www.datatorrent.com/products-services/datatorrent-rts/>
- OSS Big-Data Components
 - 提供 Spark 與 Apex 兩個大數據開源模組
- Virtual Resource Automation
 - Kubernetes Container Orchestration System
 - Kubernetes 容器管理平台的建置

PoC 工作共分為五個項目:

1. Kubernetes Cluster 平台建置

- 工作機台
 - master node x 1 + minion node x 2
- 機台作業系統: Ubuntu 16.04
- 佈建 pods
 - Spark pods
 - master x 1 + worker x 2, Zeppelin x 1
 - docker image:
 - master/worker : gcr.io/google_containers/spark:1.5.2_v1
 - zeppelin : gcr.io/google_containers/zeppelin:v0.5.6_v1
 - Apex pods
 - sandbox x 1, DataTorrent x 1
 - docker image:
 - sandbox : apacheapex/sandbox:latest
 - DataTorrent binary :
<https://www.datatorrent.com/downloads/datatorrent-rts.bin>

2. Job Executer & Monitor

- Spark job script
- Apex job script

3. Job Builder & Controller

- Python codes

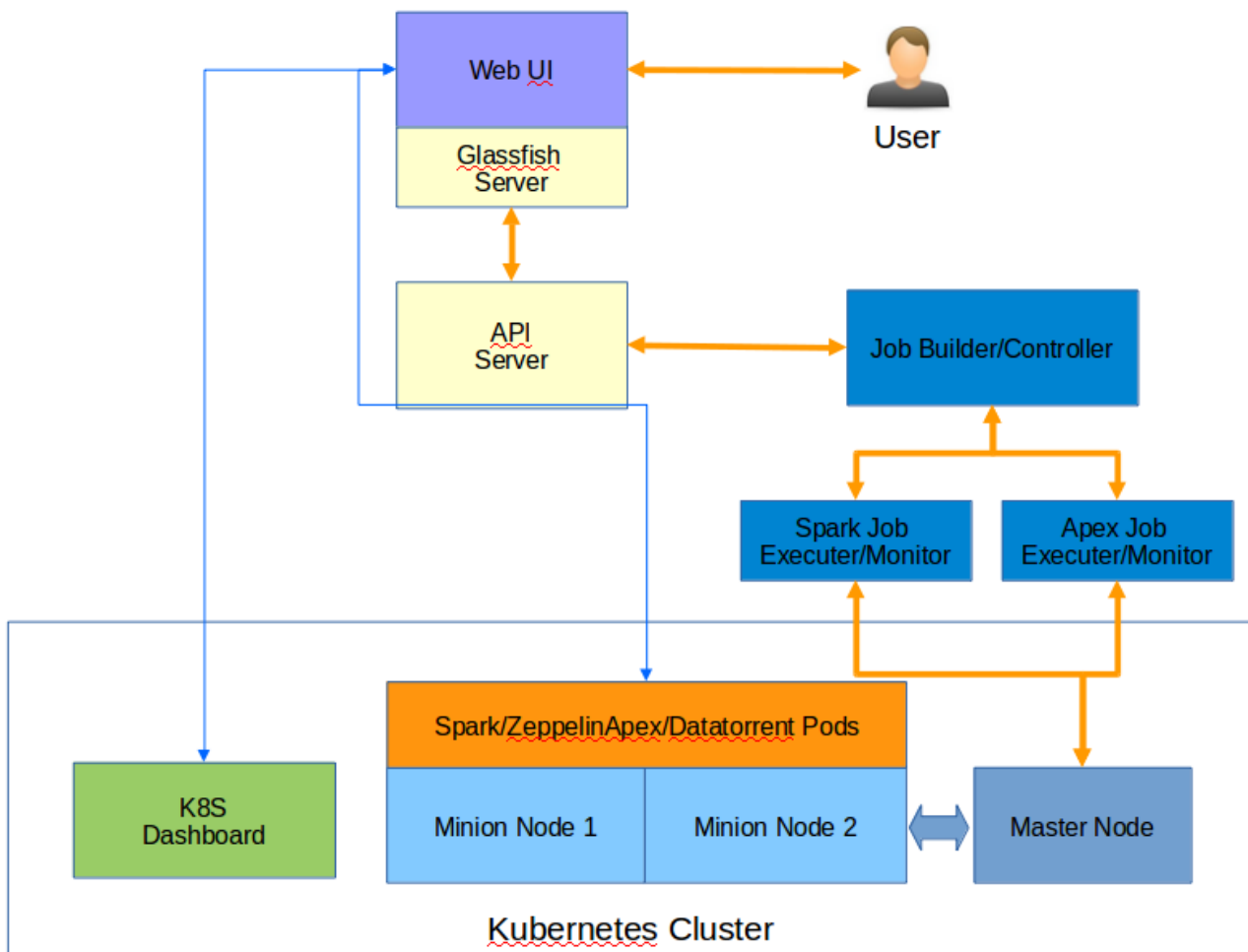
4. API Server

- provide RESTful APIs to interface with Glassfish server
- Python codes

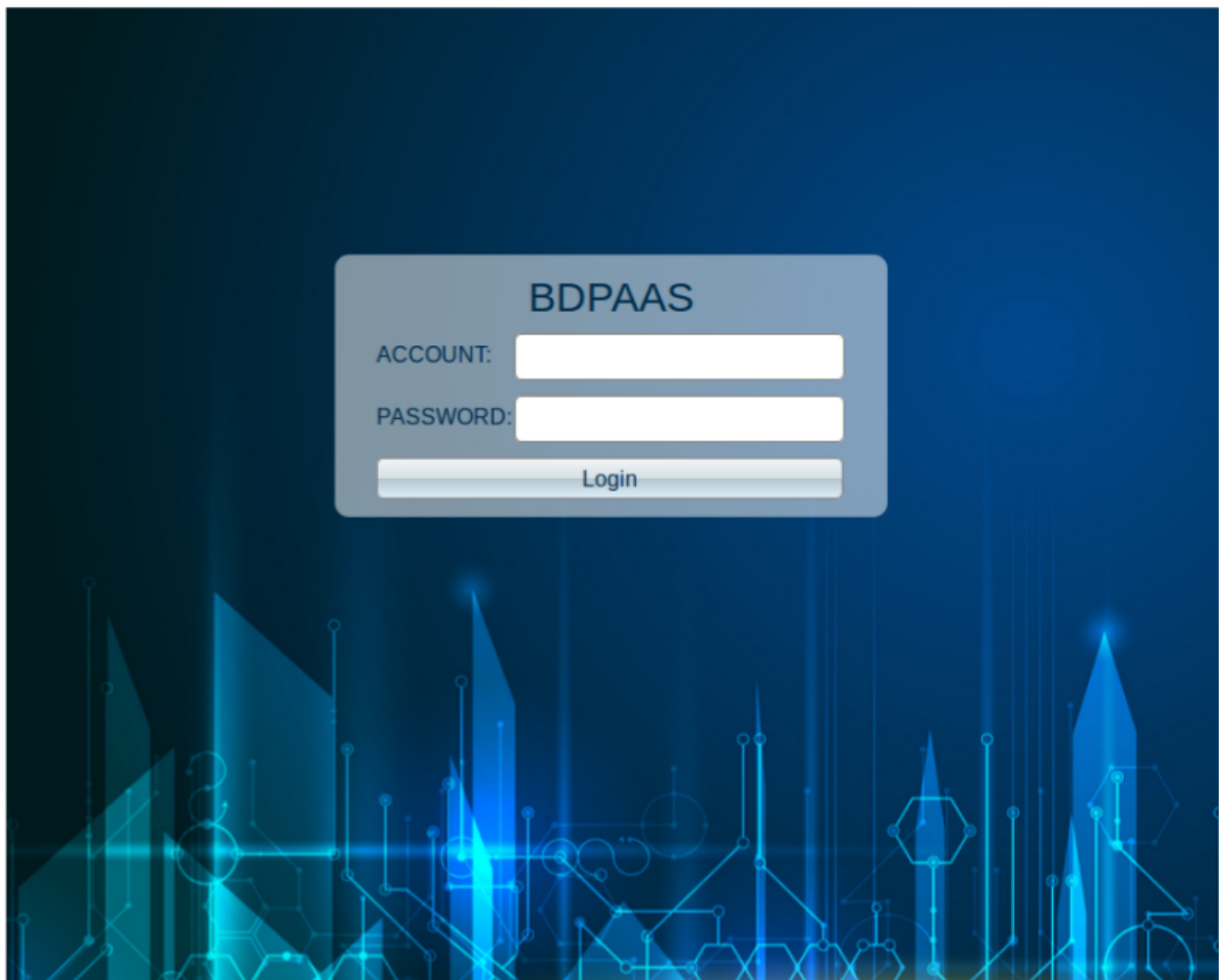
5. Glassfish Server

- Web UI

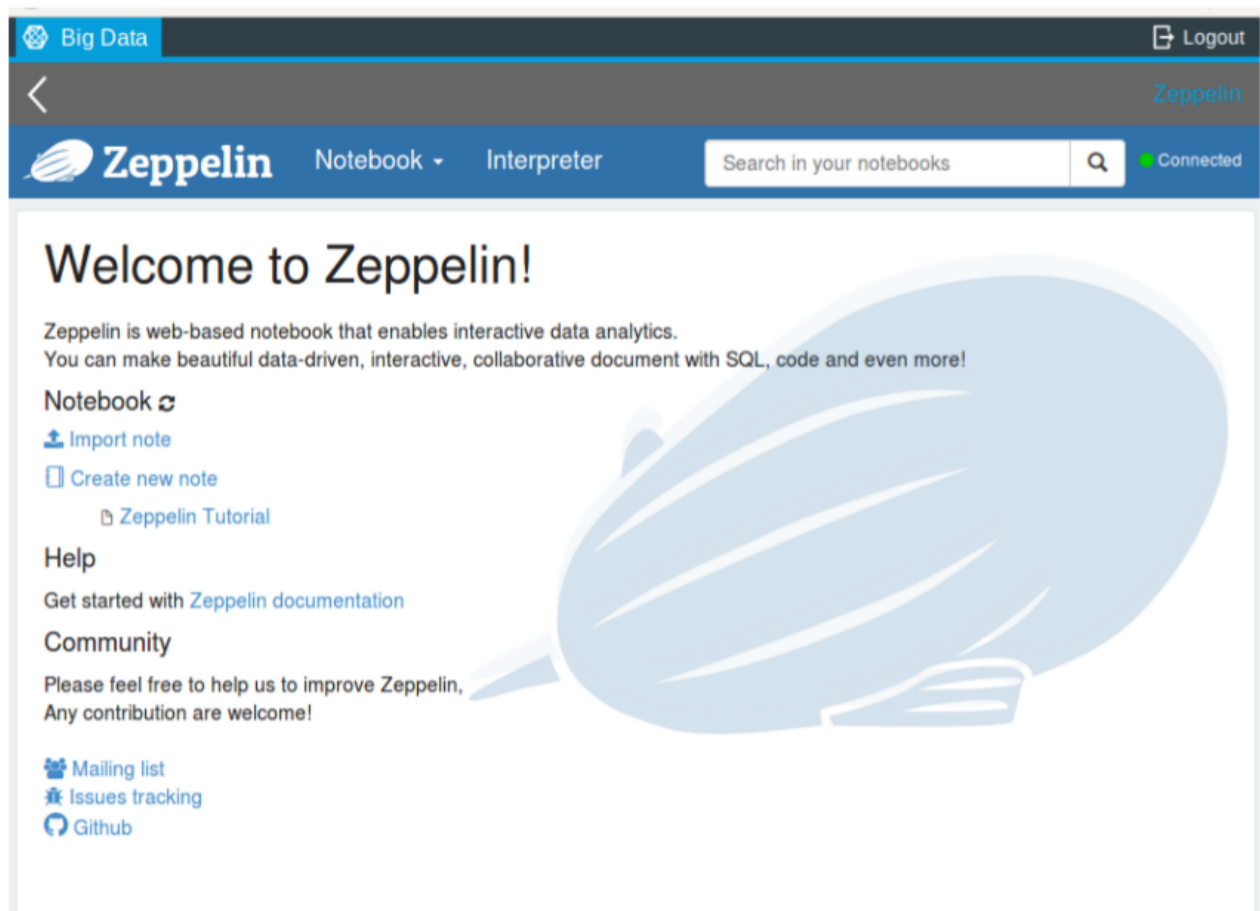
BDPaaS PoC Architecture

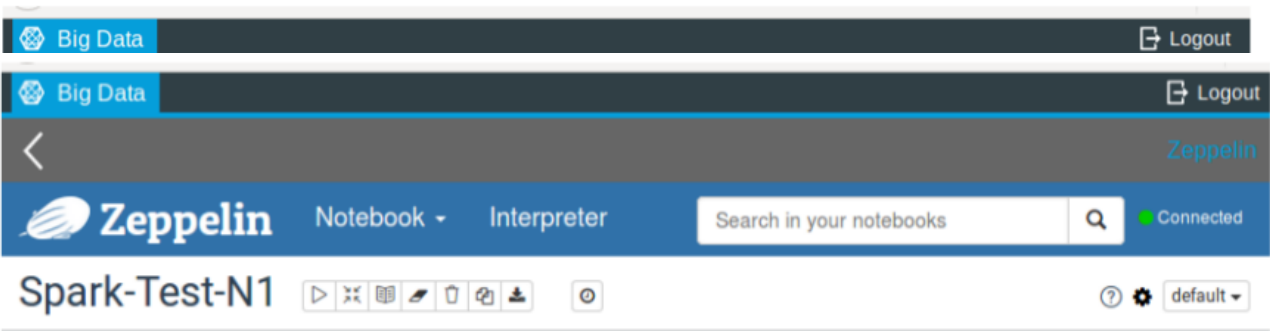
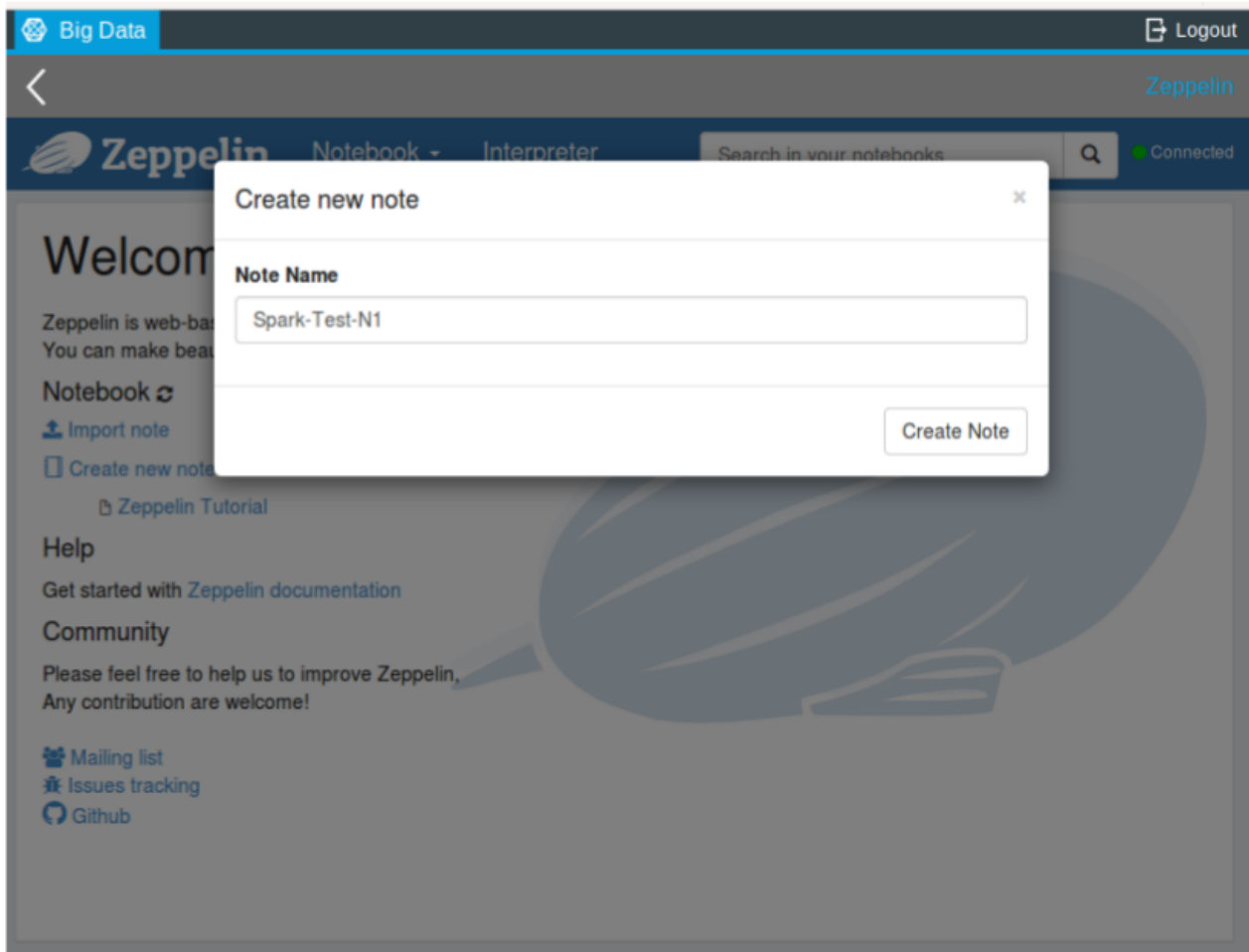


BDPaaS 操作畫面:

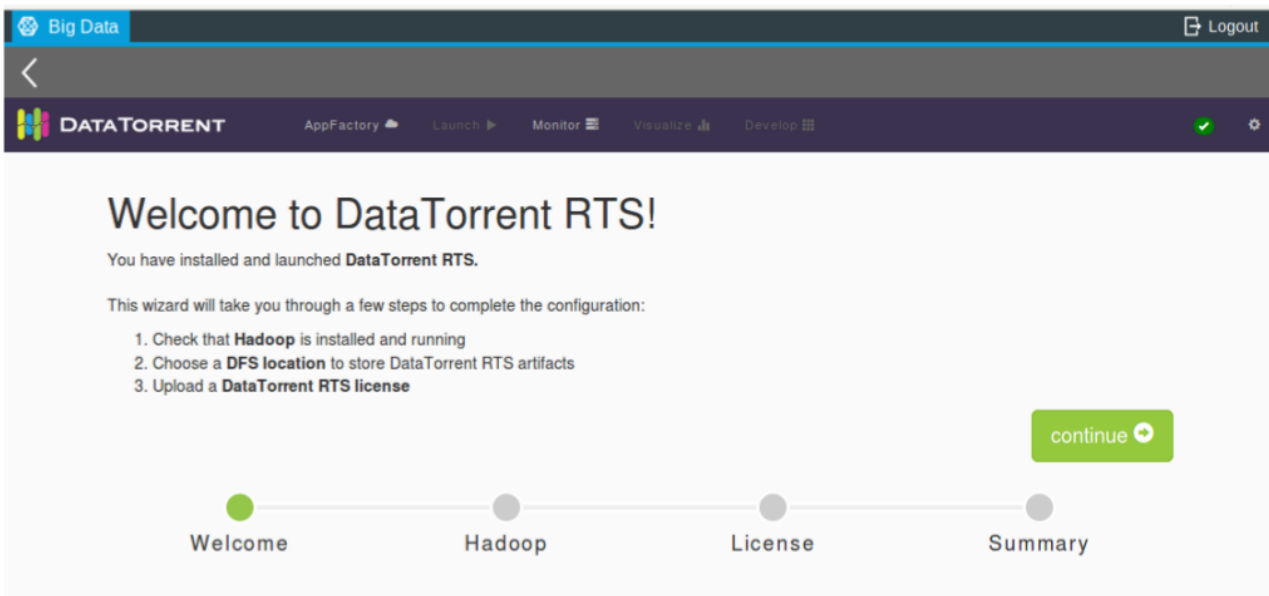


User Spark Workspace





User Apex Workspace



Big Data

Logout

DATATORRENT

AppFactory

Launch

Monitor


Visualize

Develop

Welcome to DataTorrent RTS!


DataTorrent RTS, built on Apache Apex, provides a high-performing, fault-tolerant, scalable, easy to use data processing platform for both batch and streaming workloads. DataTorrent RTS includes advanced management, monitoring, development, visualization, and data ingestion and distribution features.

Getting Started




Run Demos

Experience the power of RTS quickly. Import, launch, manage and visualize applications in minutes.



Discover RTS

Learn about the architecture and the rich feature set of the DataTorrent RTS platform and applications.



Create Apps

Tutorials and code samples to rapidly create applications on the RTS platform using Java or dtAssemble.

Using Console

AppFactory

Application packages, including solutions, application templates, demos, and examples can be imported from AppFactory.

Launch

View and launch all available applications and configurations, connected with their active instances.

Monitor

Monitor, manage, and troubleshoot the DataTorrent RTS platform and applications.

Visualize

Visualize real-time and historical application data using widgets and dashboards.

Develop

Add packages, manage schemas, and create new applications visually.

Configure

Configure settings, manage licenses, administer users and roles, and view system status information.

Big Data

Logout

DATATORRENT

AppFactory

Launch

Monitor

Visualize

Develop

Launch

Launch Applications

Applications and Configurations available for launch are listed below, along with their active instances. Applications launched with custom settings can be saved as Application Configurations to quickly launch with same settings next time.

upload package

manage packages

owner	pkg ty	source package	app ty	application	active instances	actions
<input type="text" value="string search"/>	<input type="text" value="string s"/>	<input type="text" value="string search"/>	<input type="text" value="string s"/>	<input type="text" value="string search"/>	<input type="text"/>	
dtadmin	app	pi-demo @ 3.5.0	java	PiDemo	● PiDemo	<button>▶ launch</button>
dtadmin	app	pi-demo @ 3.5.0	java	PiDemoAppData		<button>▶ launch</button>
dtadmin	app	pi-demo @ 3.5.0	java	PiJavaScriptDemo		<button>▶ launch</button>
dtadmin	app	pi-demo @ 3.5.0	json	PIJsonDemo		<button>▶ launch</button>
dtadmin	app	pi-demo @ 3.5.0	java	PILibraryDemo		<button>▶ launch</button>
dtadmin	app	wordcount-demo @ 3.4.0	java	TopNWordsWithQueries		<button>▶ launch</button>
dtadmin	app	wordcount-demo @ 3.4.0	java	WordCountDemo	● WordCountDemo	<button>▶ launch</button>

Big Data
Logout

DATA TORRENT
AppFactory
Launch
Monitor
Visualize
Develop

Operations / apps : application_1505890600256_0001

logical
physical
physical-dag-view
metric-view
attempts

Application Overview

WordCountDemo

shutdown
kill
set logging level
AM logs
visualize

id: 0001 user: dtadmin source: WordCountDemo uptime: 00:01:37 queue: default

State

● **RUNNING**
187
179

state
current w/d
recovery w/d

Performance

15
118
118
11,456
11,456

latency (ms)
processed/s
emitted/s
total processed
total emitted

Physical

3
3 / 3
6.0 GB

operators
planned/alloc. cntns
allocated mem.

StreamEvents

filter

- 09/20/2017 08:25:55 UTC **StartContainer** id: 01_000002 node: apex-sandbox-deployment-2-2543423778-xsmtt:36387
- 09/20/2017 08:25:57 UTC **StartContainer** id: 01_000003 node: apex-sandbox-deployment-2-2543423778-xsmtt:36387
- 09/20/2017 08:25:57 UTC **StartContainer** id: 01_000004 node: apex-sandbox-deployment-2-2543423778-xsmtt:36387
- 09/20/2017 08:26:03 UTC **StartOperator** id: 2 logical: count
- 09/20/2017 08:26:03 UTC **StartOperator** id: 1 logical: wordinput
- 09/20/2017 08:26:04 UTC **StartOperator** id: 3 logical: console

Big Data
Logout

DATA TORRENT
AppFactory
Launch
Monitor
Visualize
Develop

Operations

Cluster Overview

CPU/Memory

0.01 6.0 GB 6.0 GB

cores current alloc mem peak alloc mem

Applications

●1
●1
●0
●0
●0
●2

running
pending
failed
finished
killed
submitted

Performance

3 3 138 138

containers operators tuples/s emitted/s

Issues

1 0

warnings errors

DataTorrent Applications

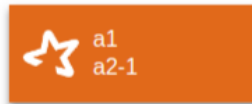
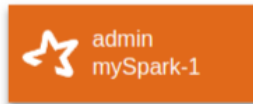
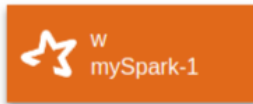
ended apps 0 selected

id	name	state	user	queue	started	lifetime	memory
0002	PiDemo	ACCEPTED	dtadmin	default	2 minutes, 32 secon...	00:02:32	0 B
0001	WordCountDemo	RUNNING	dtadmin	default	2 minutes, 56 secon...	00:02:56	6.0 GB

Admin Console

- Spark -

Refresh



Spark Master at spark://spark-master:7077

URL: spark://spark-master:7077
REST URL: spark://spark-master:6066 (cluster mode)
Alive Workers: 2
Cores in use: 4 Total, 4 Used
Memory in use: 17.5 GB Total, 1024.0 MB Used
Applications: 1 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State	Cores	Memory
worker-20170920021205-10.244.1.218-45016	10.244.1.218:45016	ALIVE	2 (2 Used)	8.8 GB (512.0 MB Used)
worker-20170920021222-10.244.2.178-45993	10.244.2.178:45993	ALIVE	2 (2 Used)	8.8 GB (512.0 MB Used)

Running Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20170920032728-0000	(kill) Zeppelin	4	512.0 MB	2017/09/20 03:27:28	root	RUNNING	30 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
----------------	------	-------	-----------------	----------------	------	-------	----------

Admin Console

- Apex -



All Applications

Logged in as: dt.who

- Cluster
 - About
 - Nodes
 - Node Labels
 - Applications
 - NEW
 - NEW SAVING
 - SUBMITTED
 - ACCEPTED
 - RUNNING
 - FINISHED
 - FAILED
 - KILLED
 - Scheduler
- Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
2	1	1	0	4	6 GB	8 GB	0 B	4	8	0	1	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:8>

Show 20 entries Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
application_1505890600256_0002	dtadmin	PIDemo	ApacheApex	default	Wed Sep 20 16:26:12 +0800 2017	N/A	ACCEPTED	UNDEFINED		ApplicationMaster
application_1505890600256_0001	dtadmin	WordCountDemo	ApacheApex	default	Wed Sep 20 16:25:48 +0800 2017	N/A	RUNNING	UNDEFINED		ApplicationMaster

Showing 1 to 2 of 2 entries

First Previous 1 Next Last



Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities -

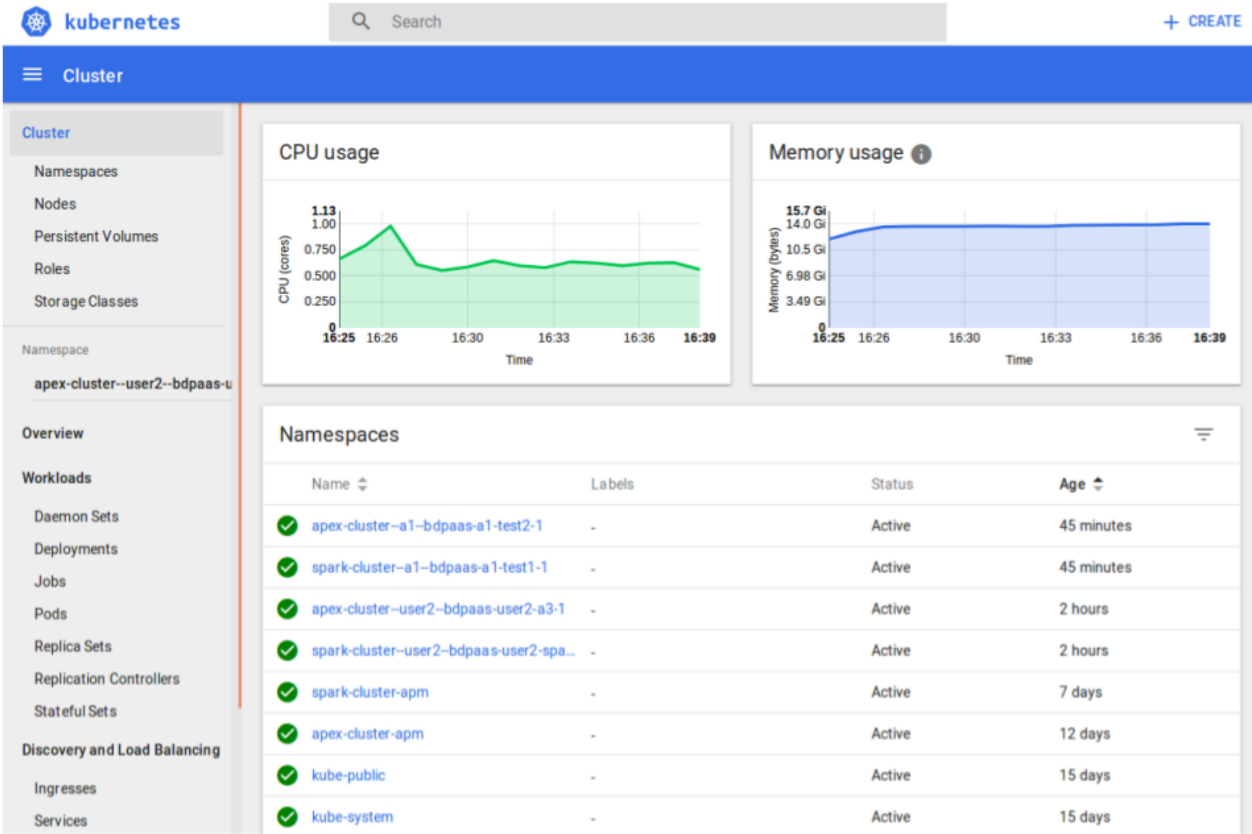
Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
apex-sandbox-deployment-2-2543423778-xsmitt:50010 (127.0.0.1:50010)	2	In Service	16.61 GB	90.97 MB	12.52 GB	4 GB	96	90.97 MB (0.53%)	0	2.7.1

Admin K8S Dashboard

CONFIDENTIAL



Workloads

Cluster

- Namespaces
- Nodes
- Persistent Volumes
- Roles
- Storage Classes

Namespace

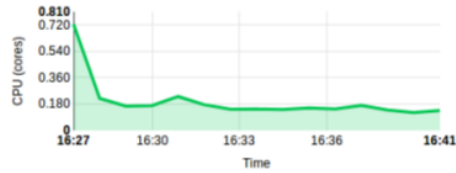
apex-cluster--user2--bdpaas-u

Overview

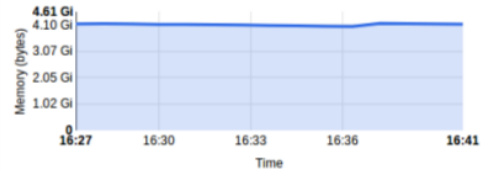
Workloads

- Daemon Sets
- Deployments
- Jobs
- Pods
- Replica Sets
- Replication Controllers
- Stateful Sets

CPU usage



Memory usage



Deployments

Name	Labels	Pods	Age	Images
✓ apex-sandbox-deployme...	app: apex-sandbox2	1 / 1	2 hours	apacheapex/sandbox2.v...

Pods

Name	Node	Status	Restarts	Age	CPU (cores)	Memory (bytes)
✓ apex-sandbox-depl...	k8ss2	Running	1	2 hours	0.136	4.134 Gi

Workloads > Pods

Cluster

- Namespaces
- Nodes
- Persistent Volumes
- Roles
- Storage Classes

Namespace

apex-cluster--user2--bdpaas-u

Overview

Workloads

- Daemon Sets
- Deployments
- Jobs
- Pods
- Replica Sets
- Replication Controllers
- Stateful Sets

CPU usage



Memory usage



Pods

Name	Node	Status	Restarts	Age	CPU (cores)	Memory (bytes)
✓ apex-sandbox-depl...	k8ss2	Running	1	2 hours	0.132	4.131 Gi