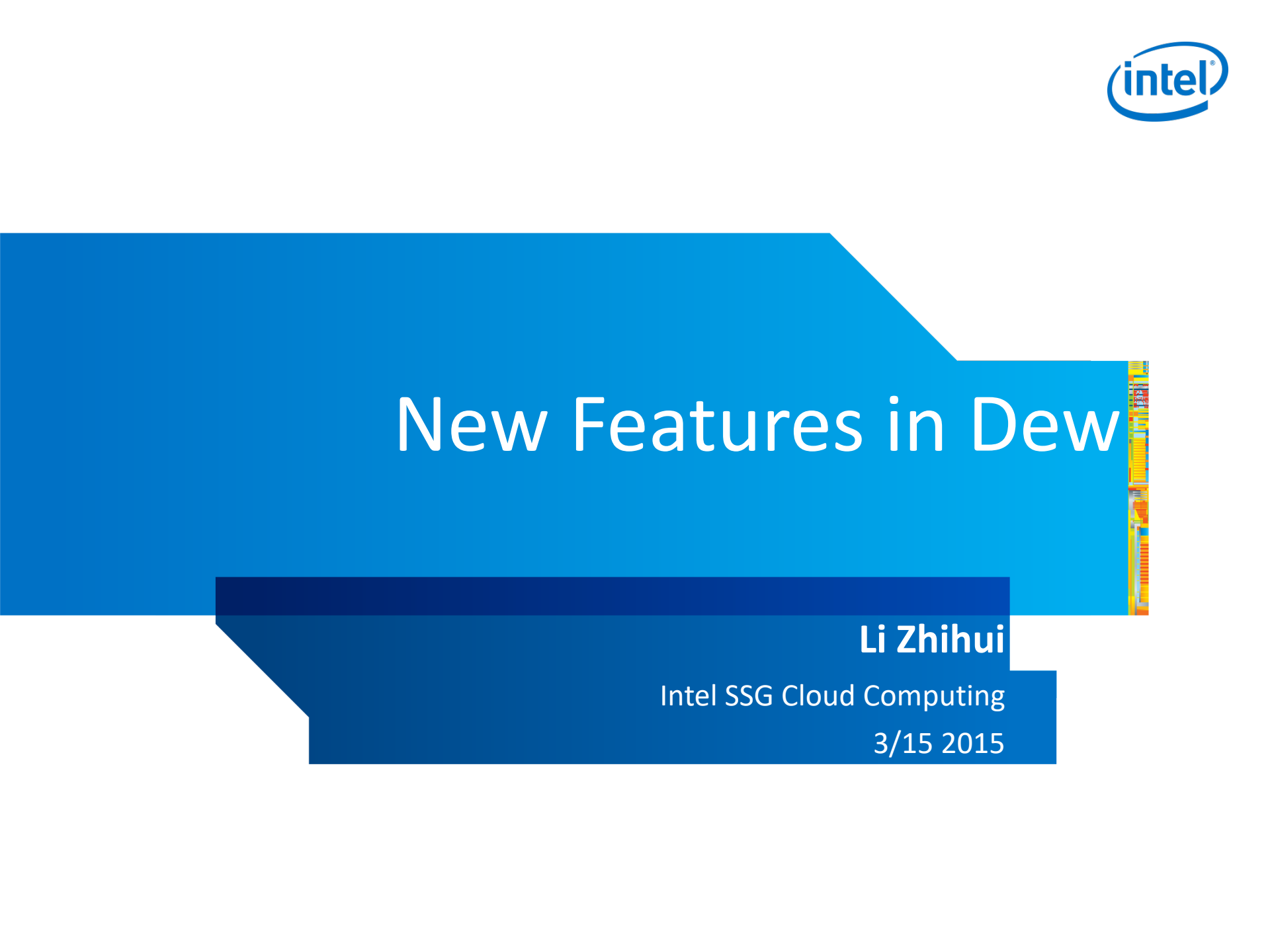


New Features in Dew

Li Zhihui

Intel SSG Cloud Computing

3/15 2015

Outlines

- Before
- New Features
- User Guide
- Architecture

Before

The Dew is a light-weight distributed Spark performance analysis framework.

- Analyze Spark performance with data flow.
- No coupling with Spark application.
- No affecting Spark application performance.
- No cluster size limitation (powerful scalability)

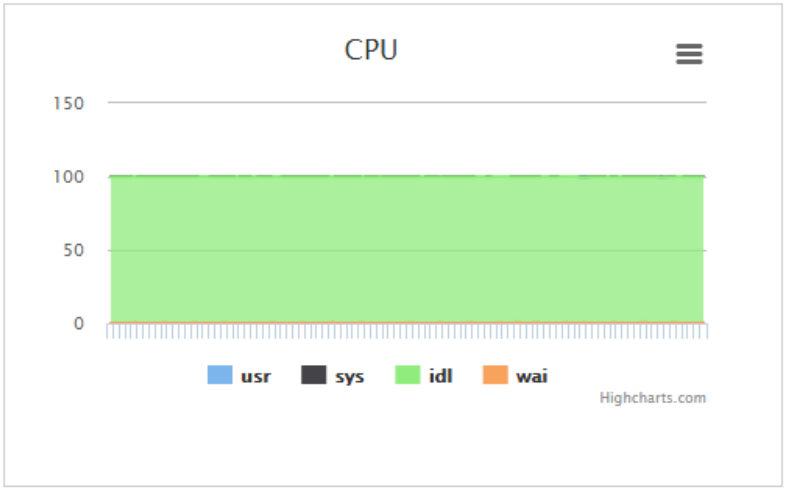
New Features

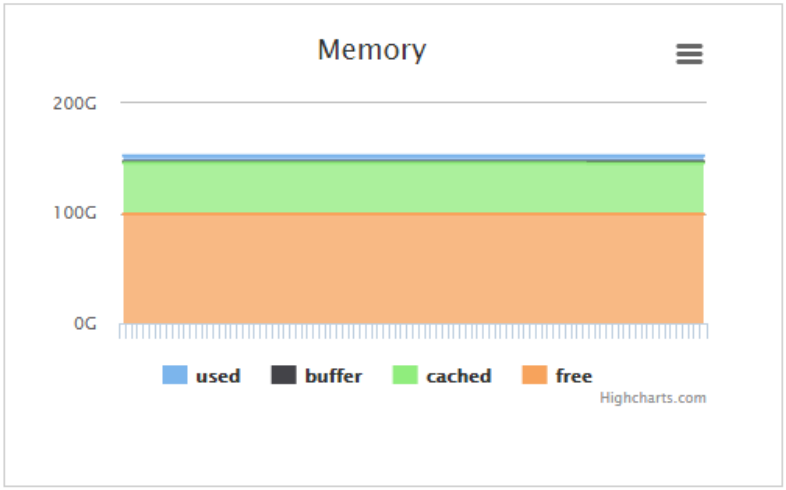
Now Dew is a big data application management and analysis system.

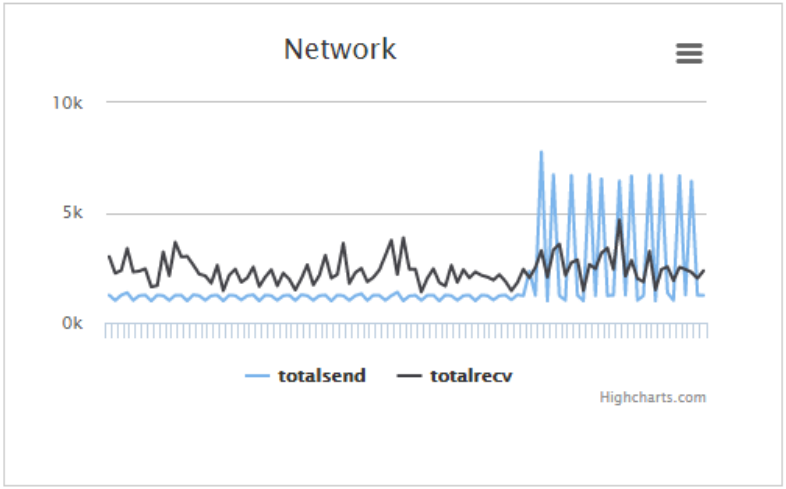
- **Distributed log collection and query**
- **Distributed command execution**
- **Spark performance diagnosis**
- **Spark application management and report**

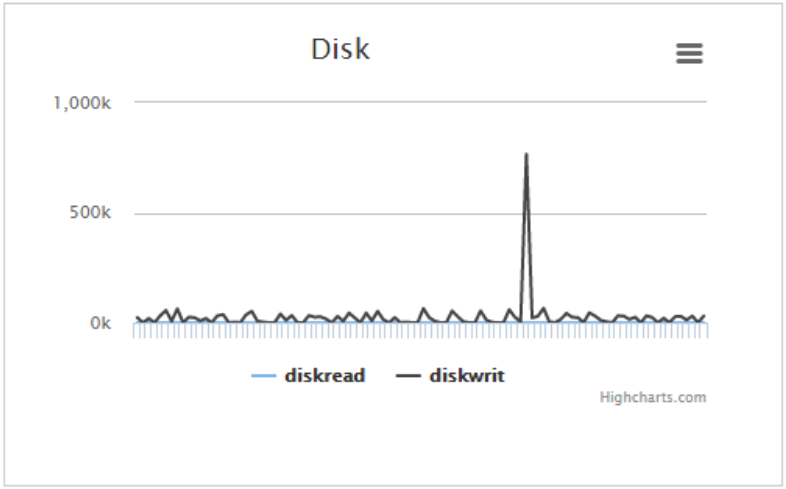
- **WebCenter: Big Data App Management**
  - Application registration and execute
  - Application execution result analysis and report
  - Cluster performance monitor
  - Dew registered services monitor

Demo - Cluster Status









Demo – Agents Status

Dew Agents Status

IP	HostName	URL	Type	Services
10.1.0.45	sr145	akka.tcp://Agent@sr145:56312/user/dew/agent	branch	[logcollection, shell, dstat]
10.1.2.53	sr453	akka.tcp://Agent@sr453:46684/user/dew/agent	branch	[logcollection, shell, dstat]
10.1.2.53	sr453	akka.tcp://Agent@sr453:40501/user/dew/agent	branch	[logcollection, shell, dstat]
10.1.2.53	sr453	akka.tcp://Agent@sr453:59272/user/dew/agent	leaf	[logaggregation]
10.1.2.54	sr454	akka.tcp://Agent@sr454:47323/user/dew/agent	branch	[logcollection, shell, dstat]
10.1.2.54	sr454	akka.tcp://Agent@sr454:43381/user/dew/agent	branch	[logcollection, shell, dstat]

Demo - Application & Job Registration

Add New Application

Name

kmeans

Host

sr145

Path

/home/username/workload/kmeans

Executable

./run.sh

Strategy

reExecute

Type

spark

Submit

Add New Job

Name

daily

Defination

nweight,wordcount

Cycle

0 0 2

Submit

Demo - Execution Result Report

Application Record List

AppName	StartTime	EndTime	Result	Operation
test1	3/5/15 12:56:00 PM.512	3/5/15 12:57:09 PM.565	success	Analysis LogQuery Diagnosis DriverLog
test1	3/4/15 12:56:00 PM.077	3/4/15 12:57:06 PM.458	success	Analysis LogQuery Diagnosis DriverLog
test1	3/3/15 12:56:00 PM.122	3/3/15 12:57:06 PM.241	success	Analysis LogQuery Diagnosis DriverLog
test1	2/11/15 11:14:18 AM.916	2/11/15 11:15:26 AM.452	success	Analysis LogQuery Diagnosis DriverLog
test1	2/11/15 9:28:59 AM.589	2/11/15 9:30:10 AM.583	success	Analysis LogQuery Diagnosis DriverLog
test1	2/6/15 3:06:55 PM.842	2/6/15 3:08:01 PM.985	failure	Analysis LogQuery Diagnosis DriverLog
test1	2/6/15 3:01:15 PM.239	2/6/15 3:02:11 PM.310	failure	Analysis LogQuery Diagnosis DriverLog

Job Record List

JobName	StartTime	EndTime	Result
app1	3/5/15 12:56:00 PM.004	3/5/15 12:56:00 PM.004	success
app1	3/4/15 12:56:00 PM.020	3/4/15 12:57:06 PM.458	success
app1	3/3/15 12:56:00 PM.042	3/3/15 12:57:06 PM.241	success
app1	2/11/15 11:14:18 AM.839	2/11/15 11:15:26 AM.452	success
app1	2/11/15 9:28:59 AM.513	2/11/15 9:30:10 AM.583	success
app1	2/6/15 3:06:55 PM.724	2/6/15 3:08:01 PM.985	failure
app1	2/6/15 3:01:15 PM.209	2/6/15 3:02:11 PM.310	failure
app1	2/6/15 2:58:23 PM.768	2/6/15 2:59:19 PM.838	failure
app1	2/6/15 2:55:13 PM.727	2/6/15 2:56:19 PM.657	success

Demo – Diagnosis

Show DiagnosisResult

hostName	diagnosisName	level	describe	advice
sr453	waste-CPU	middle	Cpu resources waste percent is 69.76%. More time on non-computation task.	Improve node's disk and network performance.
sr454	waste-CPU	middle	Cpu resources waste percent is 69.09%. More time on non-computation task.	Improve node's disk and network performance.
sr453	load-Net-Send	high	load-Net-Send is lower than cluster average by 81.74%	Check the node or your application algorism.

Demo - Driver Log

Driver Log

```
========== running Scala WordCount bench ==========
HADOOP_HOME=/home/liyezhan/work/hadoop/hadoop-2.2.0
HADOOP_EXECUTABLE=/home/liyezhan/work/hadoop/hadoop-2.2.0/bin/hadoop
HADOOP_CONF_DIR=/home/liyezhan/work/hadoop/hadoop-2.2.0/etc/hadoop
HADOOP_EXAMPLES_JAR=/home/liyezhan/work/hadoop/hadoop-2.2.0/hadoop-examples*.jar
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

rmr: DEPRECATED: Please use 'rm -r' instead.
15/03/05 14:10:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
15/03/05 14:10:36 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted hdfs://sr145:8020/SparkBench/Wordcount/Output
dus: DEPRECATED: Please use 'du -s' instead.
15/03/05 14:10:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
/home/liyezhan/frank/spark/bin/spark-submit --properties-file /home/liyezhan/frank/Sparkbench/wordcount/scala/../conf/._prop.conf --class
```

Demo - Log Collection

Contents of directory /dewlog/application_1422431846398_0127

Goto : /dewlog/application_1422431 go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
driver.log	file	191.54 KB	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup
sr453.container_1422431846398_0127_01_000002.stderr	file	15.37 KB	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup
sr453.container_1422431846398_0127_01_000002.stdout	file	0 B	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup
sr453.container_1422431846398_0127_01_000004.stderr	file	14.86 KB	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup
sr453.container_1422431846398_0127_01_000004.stdout	file	0 B	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup
sr453.container_1422431846398_0127_01_000006.stderr	file	16.80 KB	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup
sr453.container_1422431846398_0127_01_000006.stdout	file	0 B	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup
sr453.container_1422431846398_0127_01_000008.stderr	file	18.38 KB	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup
sr453.container_1422431846398_0127_01_000008.stdout	file	0 B	3	128 MB	2015-03-05 14:11	rw-r--r--	liyezhan	supergroup

Demo - Log query

Master

Search

Query Result

driver.log 15/03/05 14:10:39 INFO spark.SparkEnv: Registering BlockManagerMaster
driver.log ApplicationMaster host: N/A
driver.log ApplicationMaster RPC port: 0
driver.log 15/03/05 14:10:45 INFO cluster.YarnClientSchedulerBackend: ApplicationMaster registered as Actor[akka.tcp://sparkYarnAM@sr454:39826/user/YarnAM#-1427037721]
driver.log ApplicationMaster host: sr454
driver.log ApplicationMaster RPC port: 0
driver.log 15/03/05 14:10:45 INFO storage.BlockManagerMaster: Trying to register BlockManager
driver.log 15/03/05 14:10:45 INFO storage.BlockManagerMasterActor: Registering block manager sr145:50518 with 265.0 MB RAM, BlockManagerId(<driver>, sr145, 50518)
driver.log 15/03/05 14:10:45 INFO storage.BlockManagerMaster: Registered BlockManager
driver.log 15/03/05 14:10:53 INFO storage.BlockManagerMasterActor: Registering block manager sr454:53267 with 2.1 GB RAM, BlockManagerId(6, sr454, 53267)
driver.log 15/03/05 14:10:53 INFO storage.BlockManagerMasterActor: Registering block manager sr454:60428 with 2.1 GB RAM, BlockManagerId(8, sr454, 60428)
driver.log 15/03/05 14:10:54 INFO storage.BlockManagerMasterActor: Registering block manager sr453:37544 with 2.1 GB RAM, BlockManagerId(3, sr453, 37544)

Quick Start

Link:https://github.com/Intel-bigdata/Dew

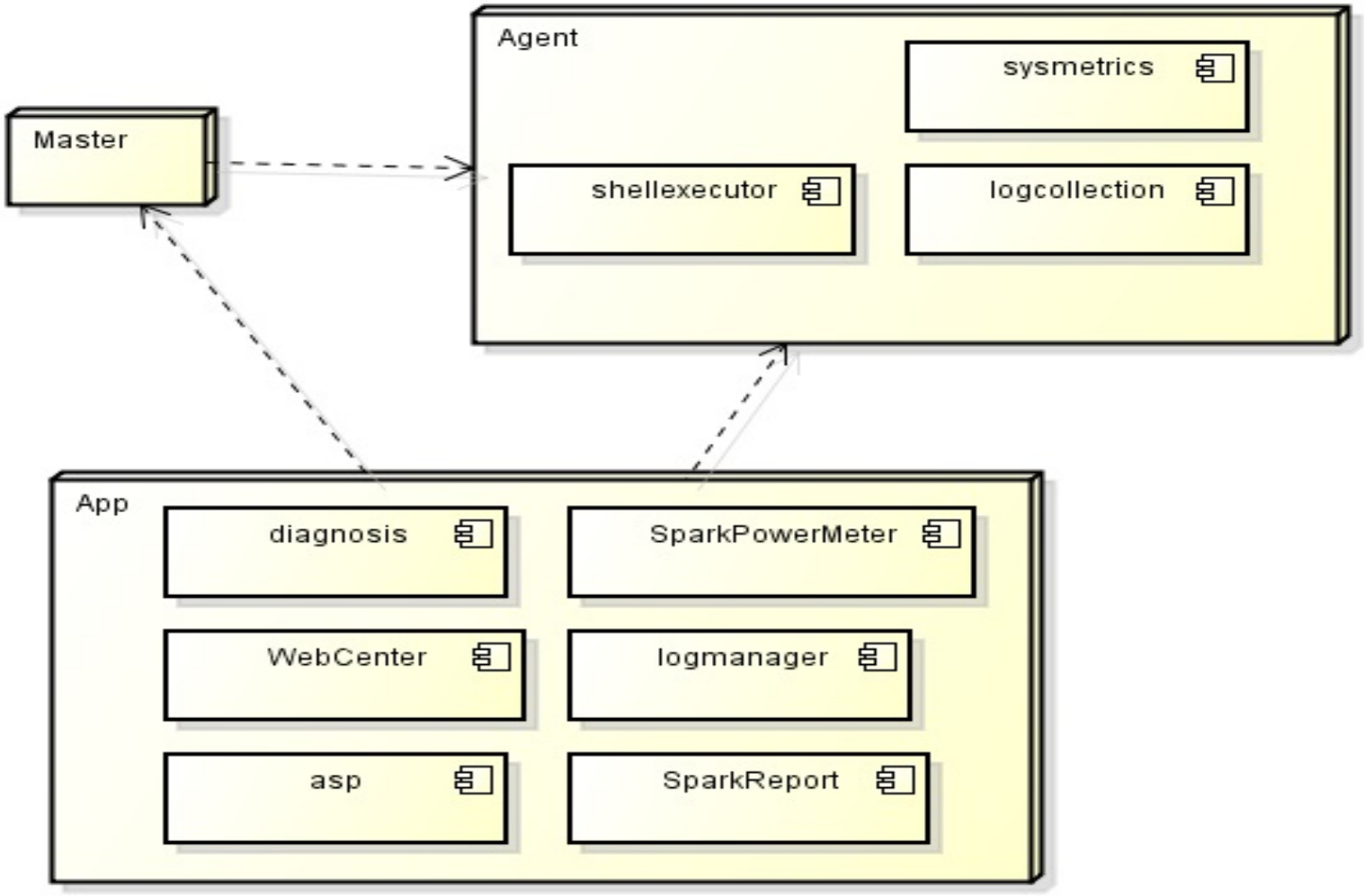
- Configuration File
  app.sparkpowermeter/conf.properties.template
  conf/dew.conf.template
  conf/slaves.template

- Run Dew
  sbin/start-all.sh

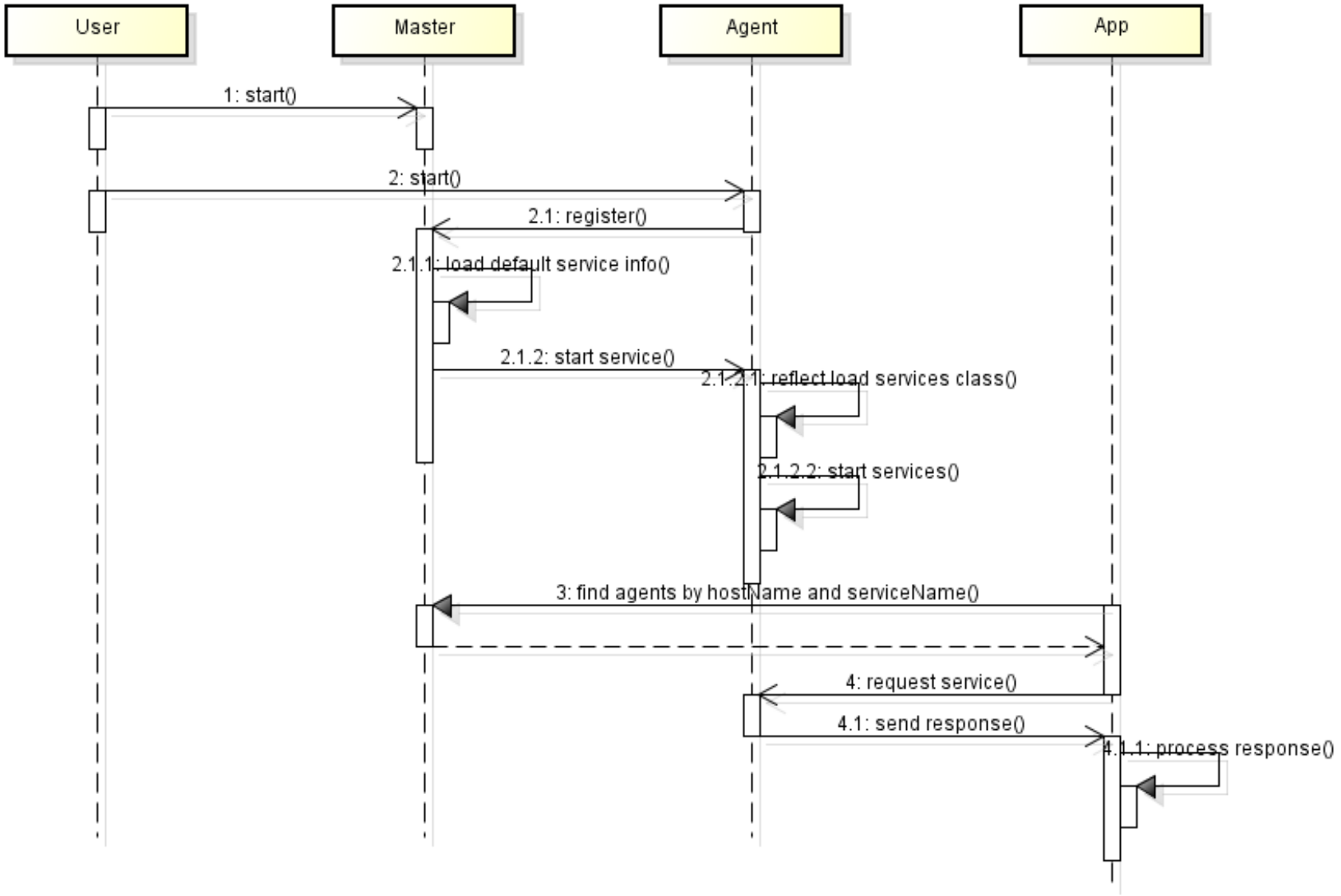
- Start webcenter
  app.webcenter → ./create-db.sh
  ./start-web.sh

- Login
  IP:6077 admin:admin

Dew Architecture



Dew Architecture



Advantages

- Friendly user interface
  - Easy to build, easy to use
  - Do anything with web console
- Flexible architecture
  - Easy to build large scale distributed computation cluster
  - Easy to implement new distributed service and application
- No couple but tightly integrate big data engine(Spark, Hadoop)
  - With plugin distributed service and application

TODO List

High available when some servers crashed.

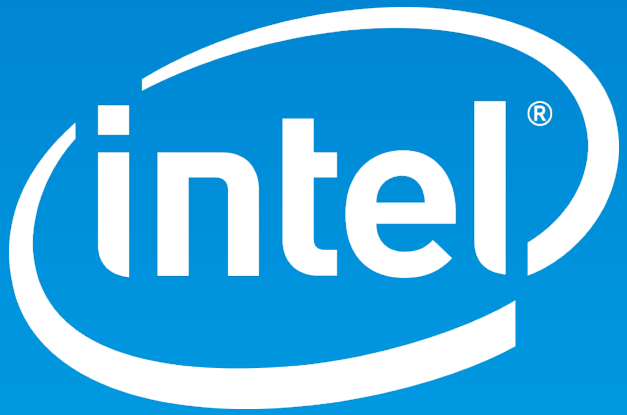
System metrics archive to hdfs.

Better user and developer documentation.

Better quality codes.

More dew service and application.

https://github.com/zhihuili/Dew


intel