NJSD

中国（南京）软件开发者大会

China (Nanjing) Software Developers Conference

2016

当计算资源抽象为原子单位后

SPARK之上的结构

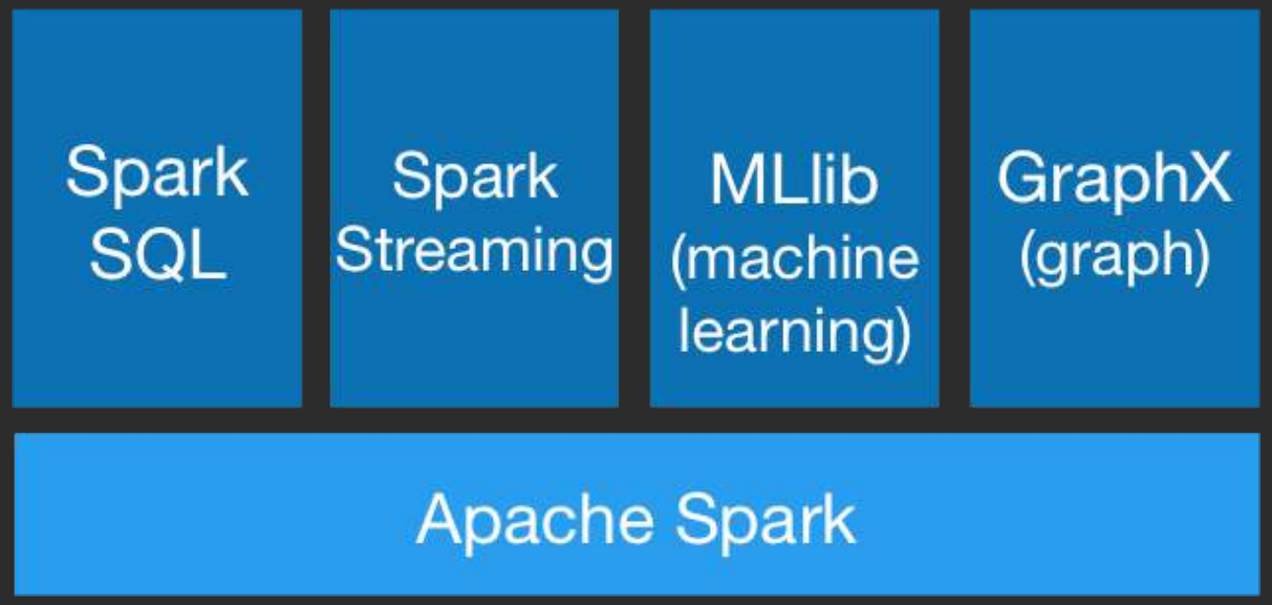
@时金魁 2016.4 chizhu@wacai.com

AGENDA

- I. Spark之上
- II. Spark云产品的实现
- III. spark-notebook/hue/zepplin
- IV. 数据产品的结构

关于SPARK

- 统一的RDD计算模型
- 高效
- all in one program platform
- 容器支持:mesos,yarn,standalone
- 数据源支持到位
- 已成势，潮流



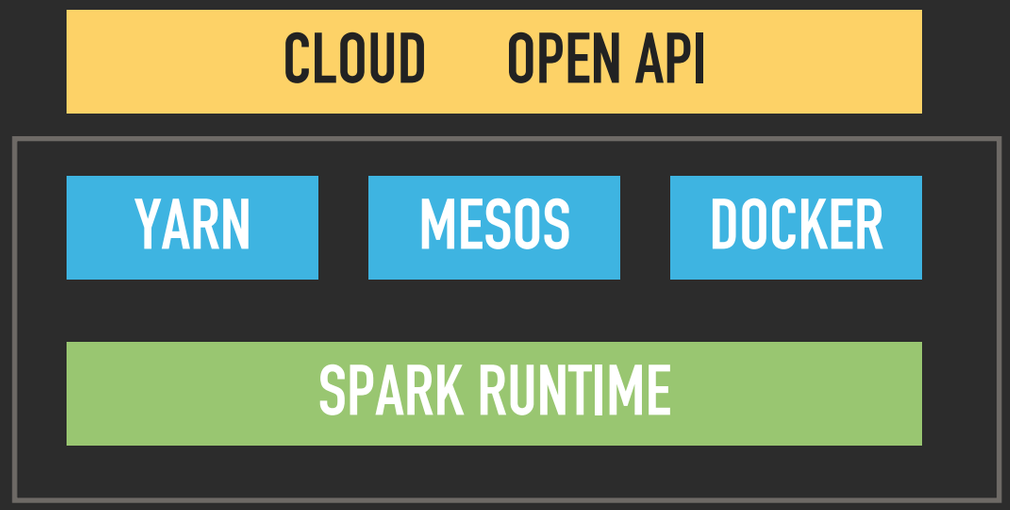
图片来源:spark.apache.org

SPARK之上

- yarn
- mesos
- standalone
- docker
- aws/aliyun/baidu cloud/qing cloud ...

支持SAPRK的云彩会下雨

- aws
- baidu cloud(bce)
- aliyun
- qing cloud



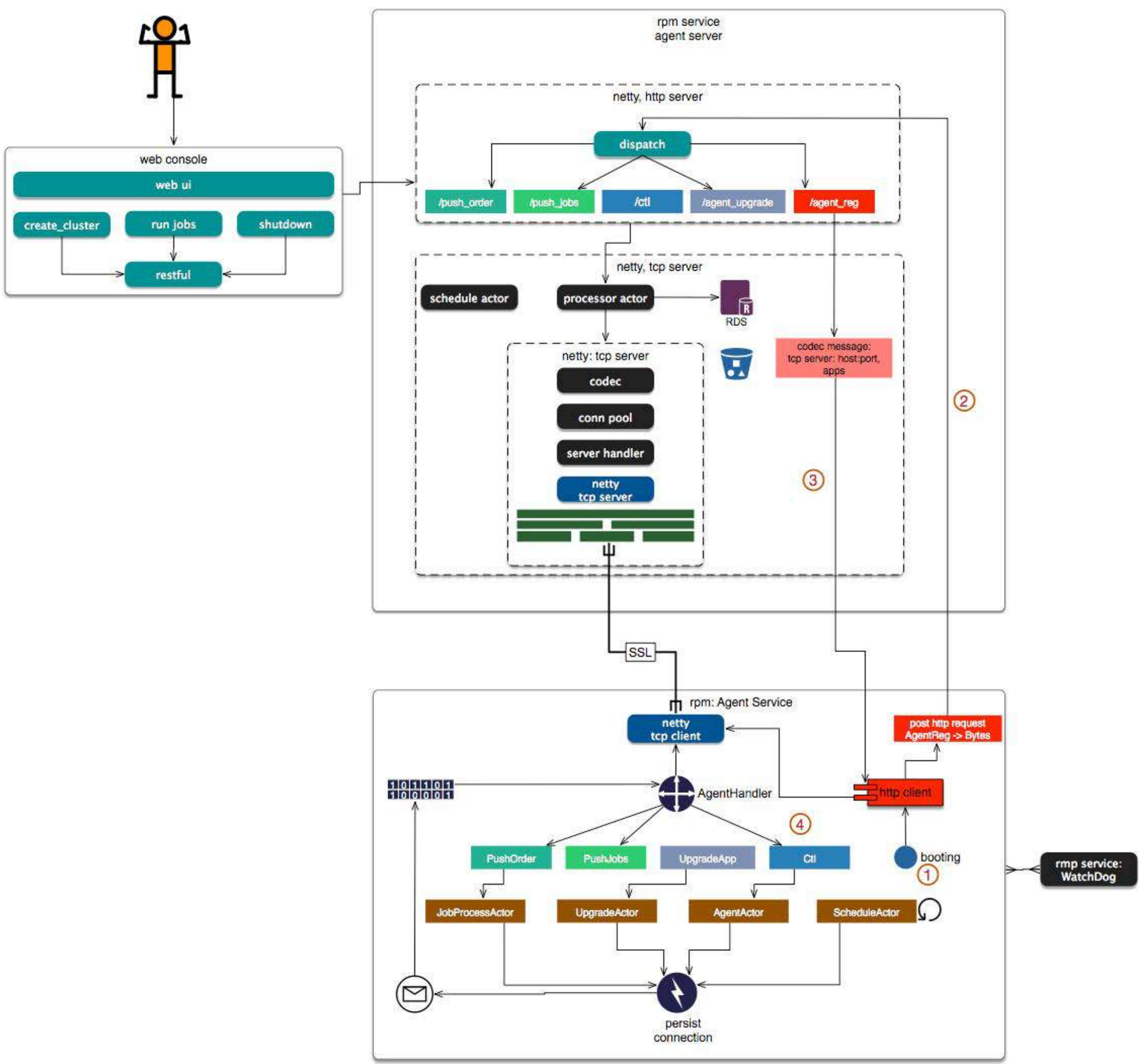
1. 三层结构
2. 边界和传输格式
3. netty: 传输、tcp长连接
4. akka: 异步消息处理和系统边界
5. sbt-native-packager: package rpm
6. 升级agent和application

II. SPARK云产品的实现

透明和信任

1. 三层结构

- web console
- agent server
- agent



2. 系统边界和传输格式

- json: web console <–> agent server
- custom payload:agent server <–> agent

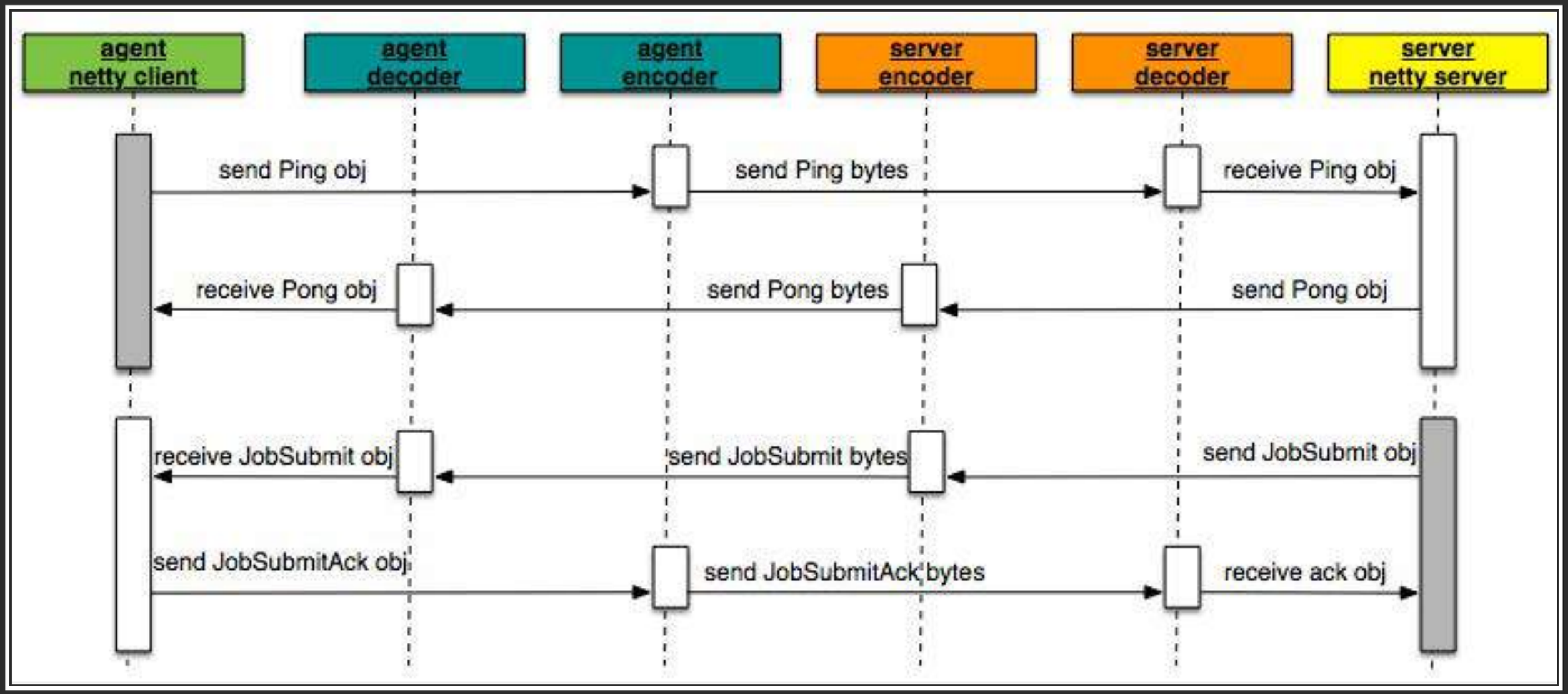
2.1 JSON格式和RESTFUL

```json
{
  "content": {
    "clusterId": 2323232,
    "ct": 1436179875098,
    "uid": 123413423432,
    "deploy": "standalone",
    "instanceIds": [
      "ECSInstanceId1",
      "ECSInstanceId2"
    ],
    "master": [
      "ECSInstanceId1"
    ],
    "isOpenOss": true,
    "jobs": [],
    "apps": [
      {
        "name": "HDFS",
        "ver": "2.6.0"
      }
    ],
    "nodeCount": 10,
    "userId": 1111
  },
  "sign": "f"
}
```

path	desc	method	from
/push_order	下推job列表和订单信息	POST	web console
/push_job	下推job列表	POST	web console
/ctl	下发控制命令	POST	web console
/ping	心跳	GET	agent/web console

```json
{
  "RESULT": {
    "clusterId": 11111,
    "ct": 1437449452246,
    "ip": "1.1.1.1"
  },
  "SIGN": "3e1273ad2bbbd0a930f541f9d0bc36b0",
  "SUCCESS": true
}
```

2.2 自定义传输格式


agent
netty client
agent
decoder
agent
encoder
server
encoder
server
decoder
server
netty server
send Ping obj
send Ping bytes
receive Ping obj
receive Pong obj
send Pong bytes
send Pong obj
receive JobSubmit obj
send JobSubmit bytes
send JobSubmit obj
send JobSubmitAck obj
send JobSubmitAck bytes
receive ack obj

2.3 自定义传输格式: CASE CLASS

```
case class JobSubmit(
  jobs: Array[Job],
  sessionId: Long
) extends BaseMsg(MsgType.job_submit, sessionId)

case class JobSubmitAck(
  clusterId: Long,
  jobs: Array[Long],
  sessionId: Long
) extends BaseMsg(MsgType.job_submit_ack, sessionId)
```

```
private def decodeJobSubmit(buf: ByteBuf, sid: Long): JobSubmit
  val joblen = buf.readInt()
  val jobs = ArrayBuffer[Job]()
  var i = 0
  while (i < joblen && buf.isReadable) {
    i = i + 1
    jobs += new Job(
      buf.readLong(),
      buf.readLong(),
      ClusterMode(buf.readInt()),
      JobType(buf.readInt()),
      JobFailAct(buf.readInt()),
      decodeString(buf)
    )
  }

  new JobSubmit(jobs.toArray, sid)
}
```

```
private def encodeJobSubmitAck(ack: JobSubmitAck
  out.writeLong(ack.clusterId)
  out.writeInt(ack.jobs.length)
  ack.jobs.foreach(out.writeLong)
}
```

3. NETTY: 传输、TCP长连接

- codec, msg <–> bytes
- tcp transfer bytes payload
- tcp server: persist connect map, push
- http server: only for agent register and validate

4. AKKA: 异步消息处理和系统边界

- message-driven
- asynchronous and parallel
- module boundary is message
- module entrance is actor ref
- business logic inside is private

5. SBT-NATIVE-PACKAGER: PACKAGE RPM

- rpm:packageBin
- sbt-release: version auto increase
- 服务管理归于OS
- 监控服务状态 { from “Linux Standard Base Core Specification” }
  - 0: OK,
  - 1: dead, pid exist
  - 2: dead, lock file exist
  - 3: not running
  - 4: unknown

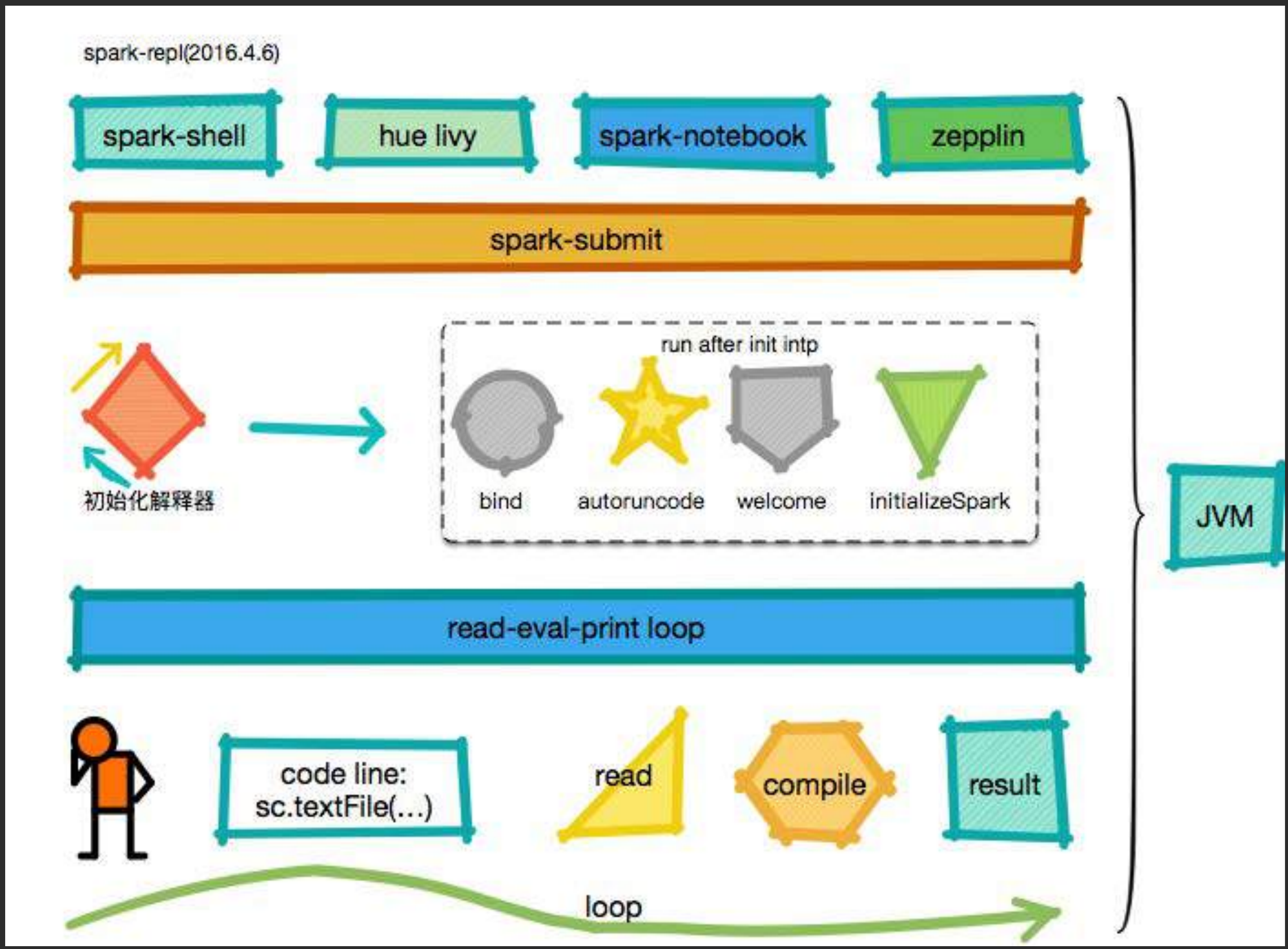
打包发布

```
{
  "app_info": {
    "name": "server",
    "owner": "玄畅"
  },
  "package": {
    "type": "scala",
    "script": "sbt",
    "args": "-Dsbt.override.build.repos=true clean dist",
    "conf": "",
    "dist_file": "server/target/universal/server-*.zip"
  },
  "deploy": {
    "type": "standalone",
    "script": "bin/server",
    "args": "",
    "ping_url": "http://127.0.0.1:9999/health",
    "ping_living": "success",
    "pid_file": "run/RUNNING_PID"
  }
}
```

6. 升级AGENT和APP

- schedule check service status
- upgrade type: zip, rpm, tgz, conf
- upgrade trait:
  - download
  - validate
  - upgrade
  - restart
  - callback msg

III. SPARK-NOTEBOOK/HUE/ZEPPLIN与SPARK-REPL的关系

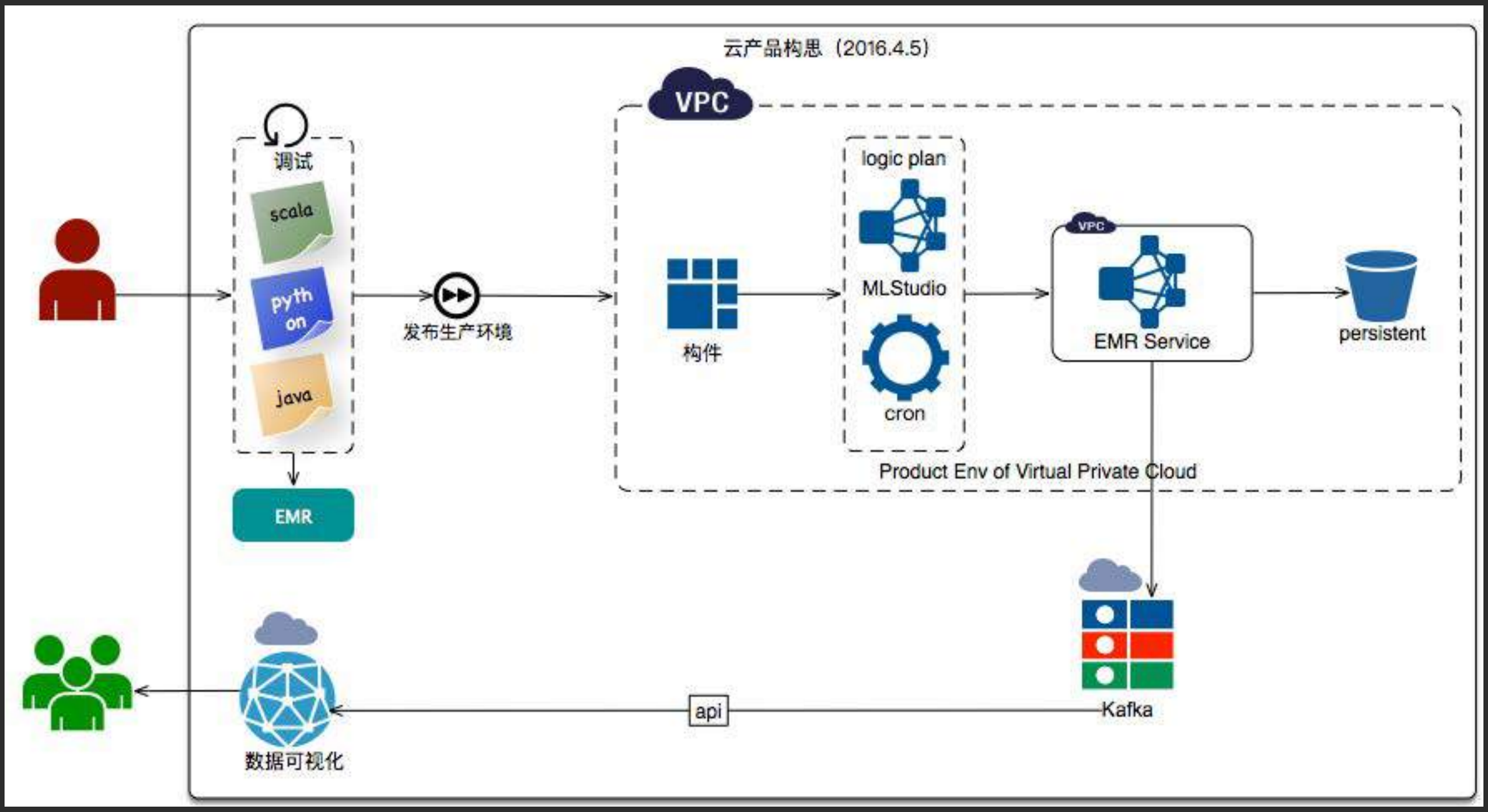


III. SPARK-NOTEBOOK/HUE/ZEPPLIN

- hue　大杂烩
- spark-notebook 仅支持spark
- incubator-zeppelin

- 强依赖spark-repl

构思


云产品构思（2016.4.5）
VPC
调试
scala
pyth on
java
EMR
发布生产环境
构件
logic plan
MLStudio
cron
VPC
EMR Service
persistent
Product Env of Virtual Private Cloud
Kafka
api
数据可视化

思考

- 数据仓库、模型、计算资源各自独立，横向扩展性
- 闭环: 用户无感知, 降低数据分析门槛
- streaming: 实时当为标配
- 可视化: 最后一公里
- Open API: 黑盒服务入口
- 好的数据产品是可轻度付费交易的

IV. 数据产品的结构

延续思考之二:

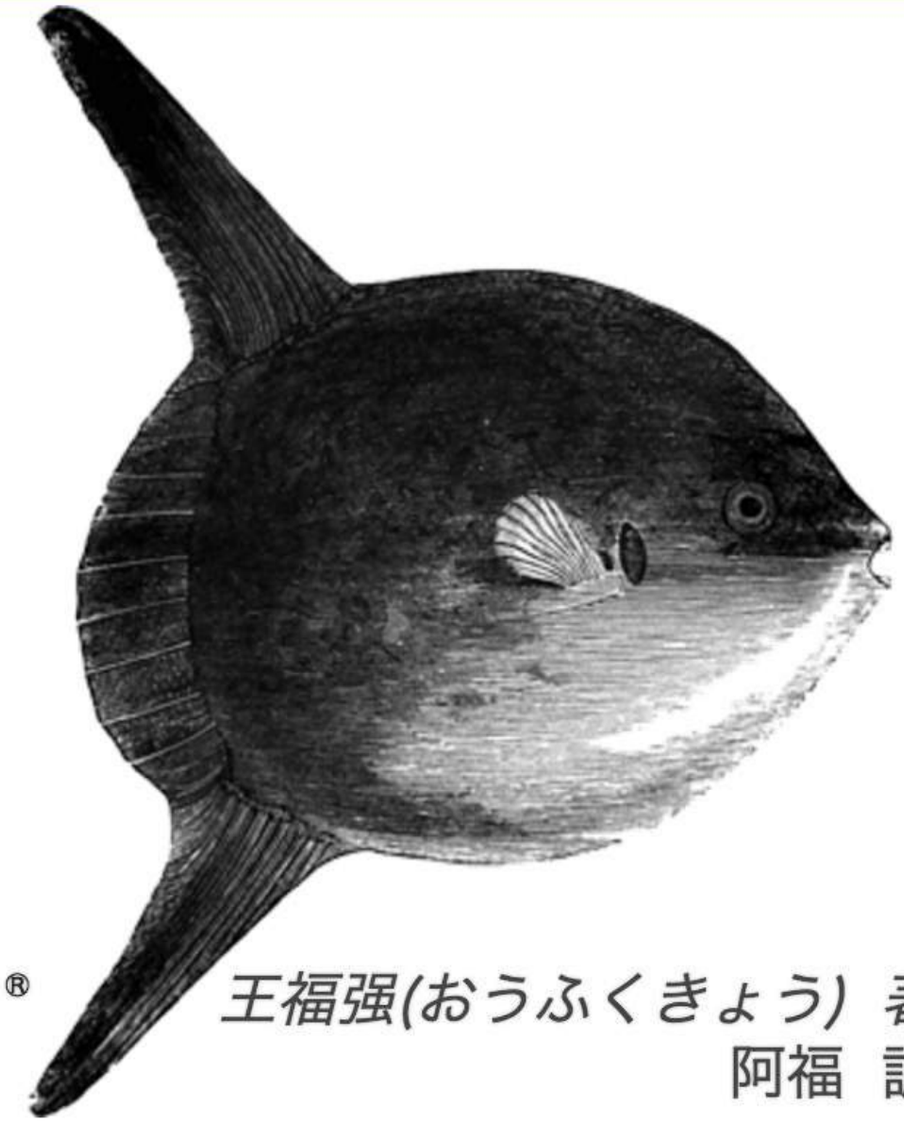
安全和云厂商的可信值

延续思考

- spark-repl是否会提供restful接口
- notebook是否会合并或新的选择
- 数据交易市场是否会出现
- 各大云厂商多久能把自己开源
- 搭VPC平民化

阿里与平安集团技术高层倾心倾情推荐，
互联网与互联网金融行业各大技术掌门一致好评

快速构建微服务体系
SpringBoot揭秘



KEEVOL®

王福强(おうふくきょう) 著
阿福 訳

华章出版社

谢谢

挖财招人
数据挖掘工程师和SCALA程序员

chizhu@wacai.com