

BDTC

2014 中国大数据技术大会

BIG DATA TECHNOLOGY CONFERENCE

暨第二届CCF大数据学术会议

OpenStack在大数据方面的思考： 技术、应用和生态系统

杜玉杰(Ben)

OpenStack基金会独立董事

中国企业级云计算联盟（ECA）副秘书长



@ben_杜玉杰



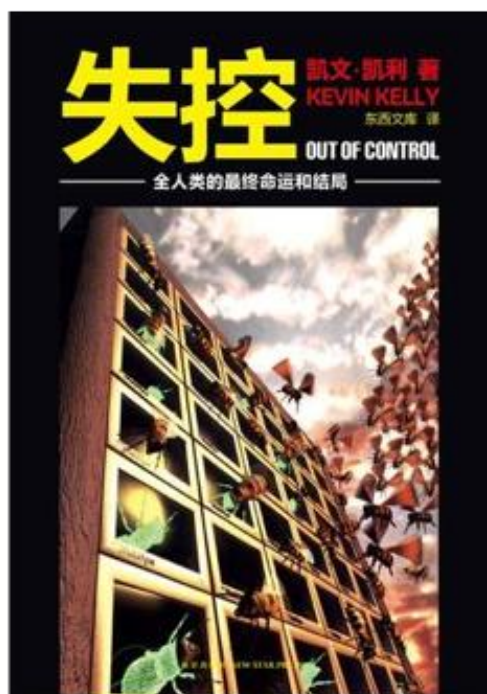
duyujie.dyj@gmail.com



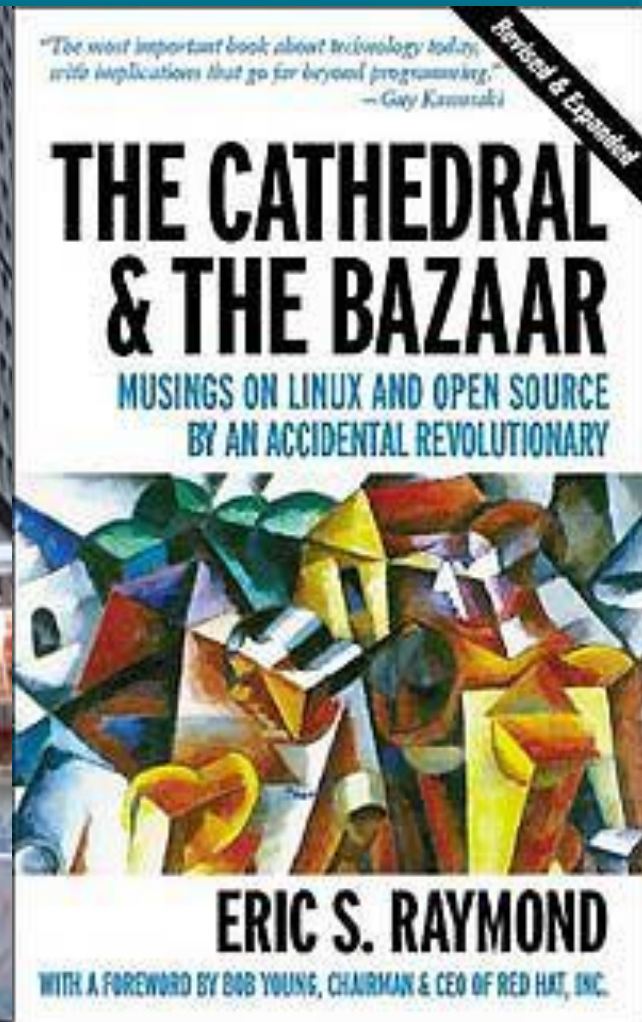
<http://www.linkedin.com/in/duyujie/>



KK，互联网思维，《失控》，大数据



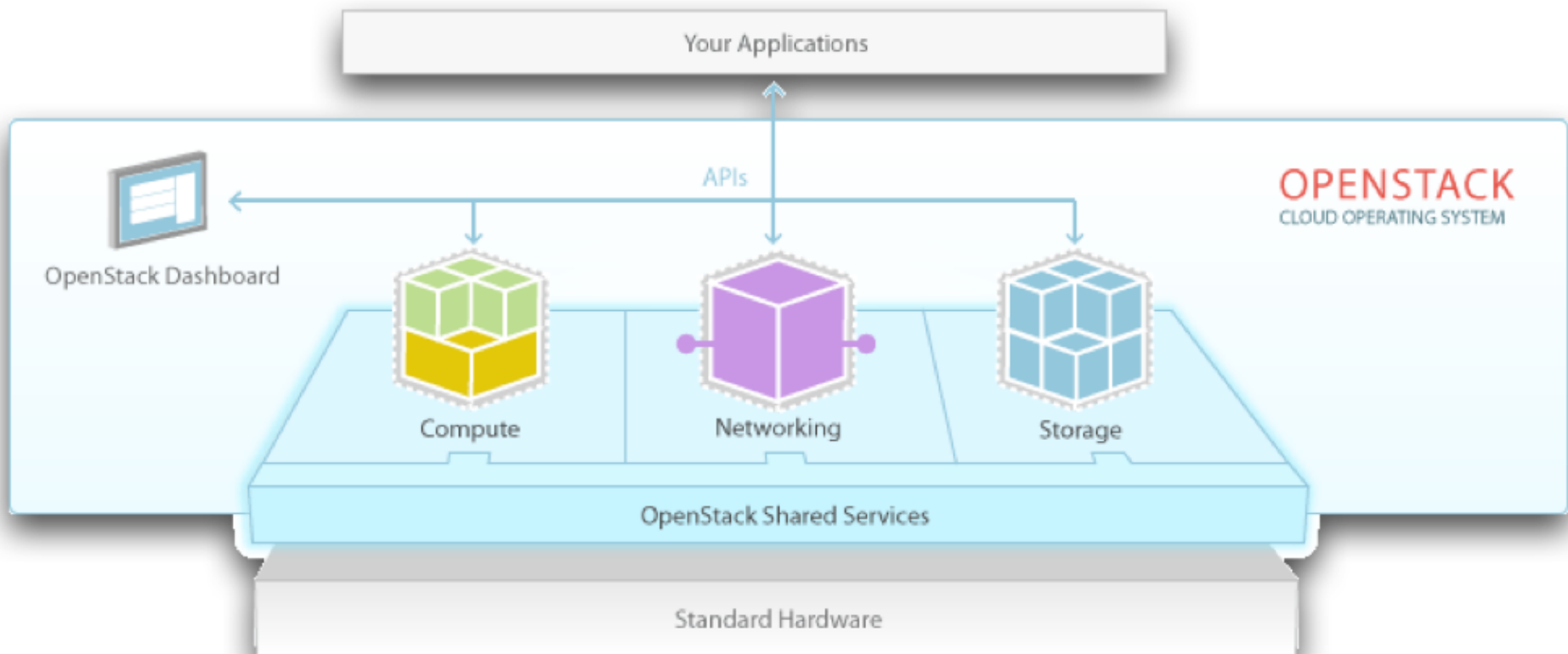
书中提到，互联网的进化经历了三个阶段：从文件（Files）通过文件夹（Folders）的组织形式存储在台式机（Desktop）中，到网页（Pages）以链接（Links）的方式形成网络（Web），再到数据流（Streams）借助标签（Tags）构成云（Cloud）。



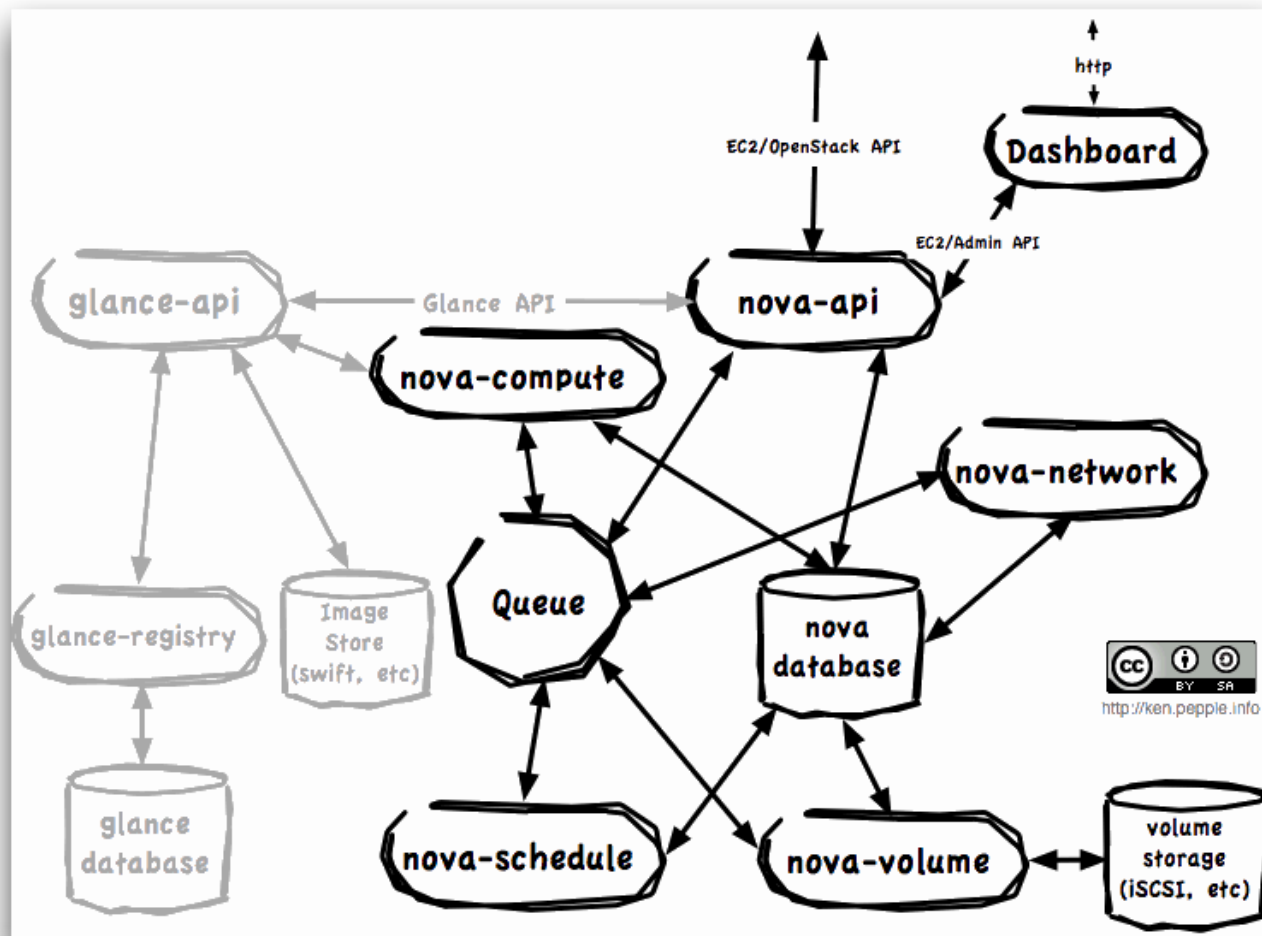
我们想造出一个大教堂，可是眼下只有一个集市，怎么办？

https://www.slideshare.net/ben_duyujie/community-leader-open-dayopenstack-community

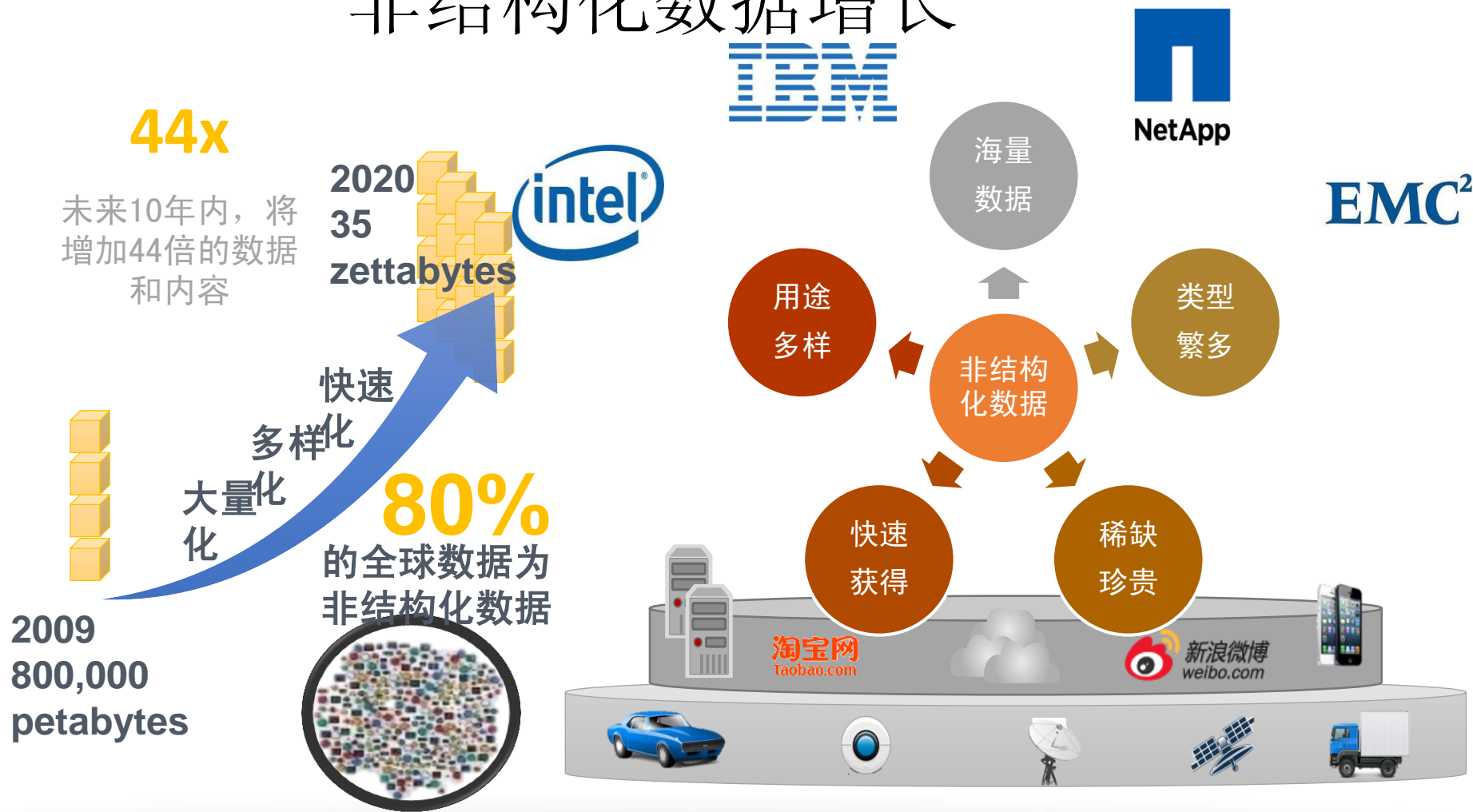
什么是OpenStack

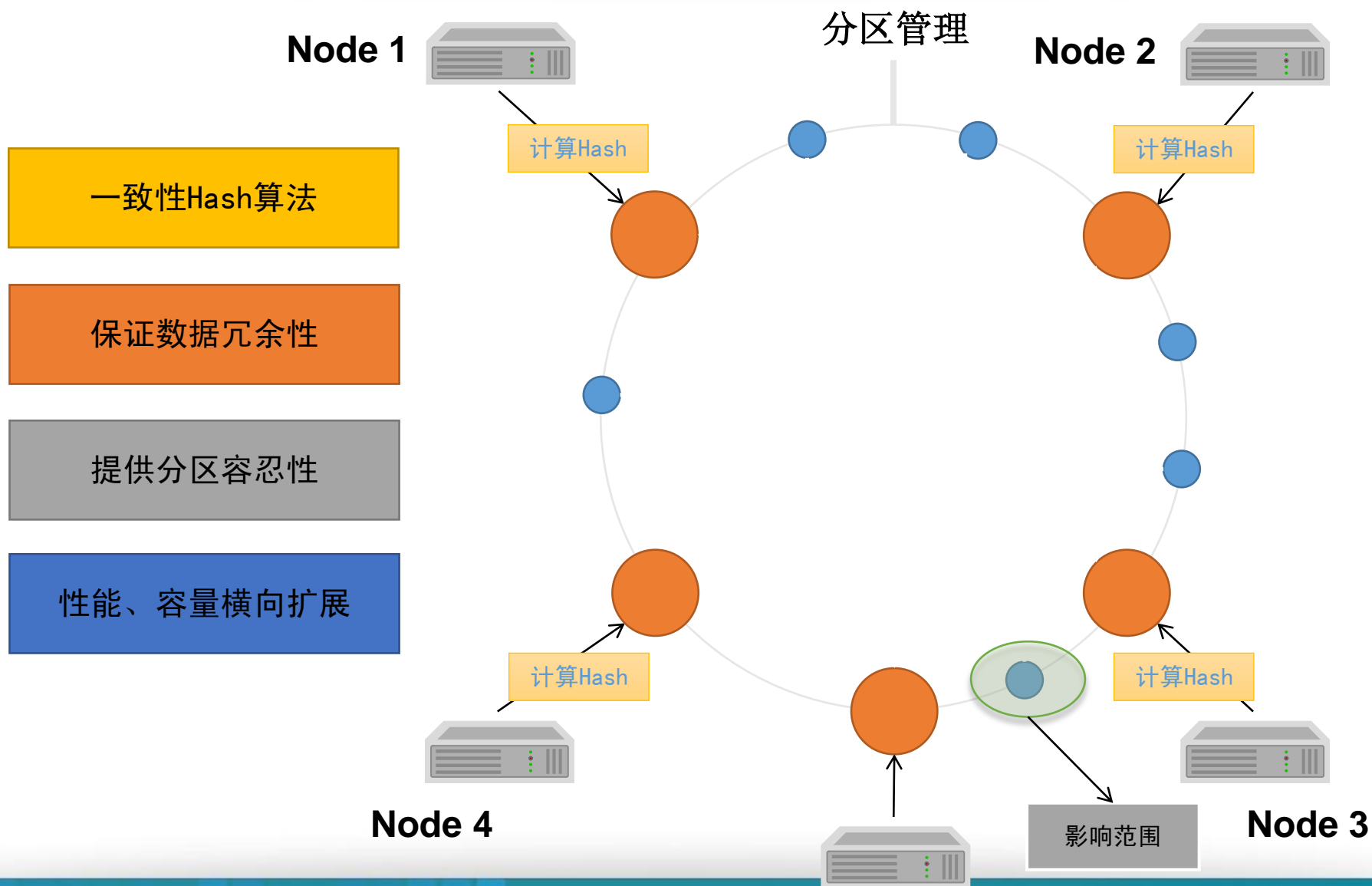


早期的OpenStack

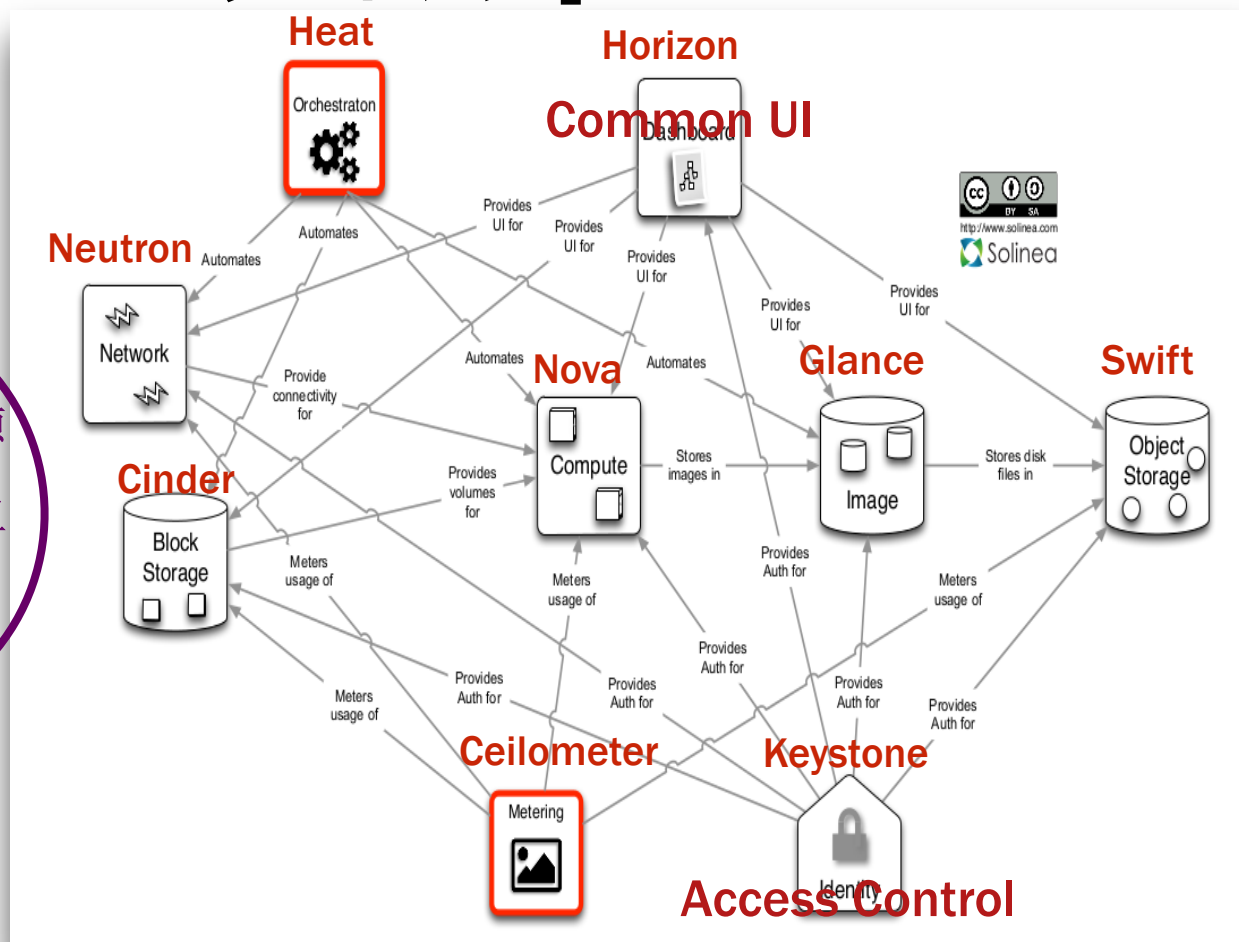


非结构化数据增长





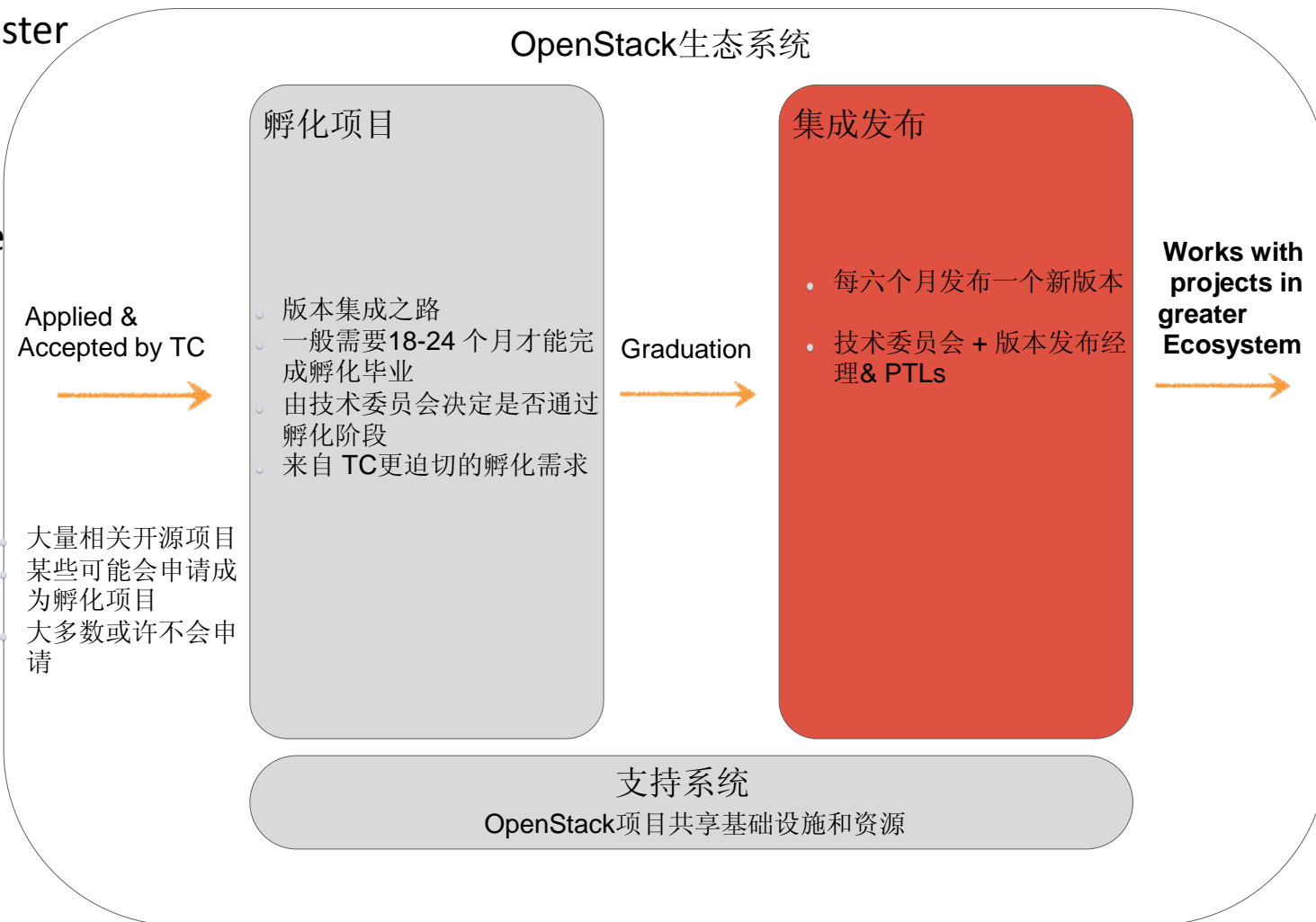
如今的OpenStack



不是OpenStack项目，但是有助于理解OpenStack组件间的关系

RabbitMQ
Open Source Enterprise Messaging

Staccato
Libra
Turbo-hipster
Mistral
Climate
Entropy
Designate
Pecan
Kite
Milk
Graffiti
Rubick
Packstack
Murano
Manila
Docker
Solum
Rally
Barbican
Sitori
Fuel



Havana Release (Oct 2013)

孵化项目

Relational DB(Trove)

Hadoop(Savannah)

Bare Metal(Ironic)

Messaging(Marconi)

Deployment(TripleO)

集成发布

Compute(Nova)

Object Storage(Swift)

Block Storage(Cinder)

Networking(Neutron)

Dashboard(Horizon)

Image Service(Glance)

Identity(Keystone)

Telemetry(Ceilometer) New

Orchestration(Heat) New

支持系统

Infrastructure | Docs | Common Libraries | QA | Release Mgmt

Icehouse Release (April 2014)

孵化项目

Bare Metal(Ironic)

Queue Service(Marconi)

Data Processing(Sahara)

集成发布

Compute(Nova)

Object Storage(Swift)

Block Storage(Cinder)

Networking(Neutron)

Dashboard(Horizon)

Image Service(Glance)

Identity(Keystone)

Telemetry(Ceilometer)

Orchestration(Heat)

Database Service(Trove) New

支持系统

Infrastructure | Docs | Common Libraries | QA | Release Mgmt

Juno Release (Oct 2014)

孵化项目

Zaqar (queue service)

Designate (DNS service)

Manila (shared file system)

Barbican (key management)

集成发布

Compute(Nova)

Object Storage(Swift)

Block Storage(Cinder)

Networking(Neutron)

Dashboard(Horizon)

Image Service(Glance)

Identity(Keystone)

Telemetry(Ceilometer)

Orchestration(Heat)

Database Service(Trove)

Data Processing(Sahara) New

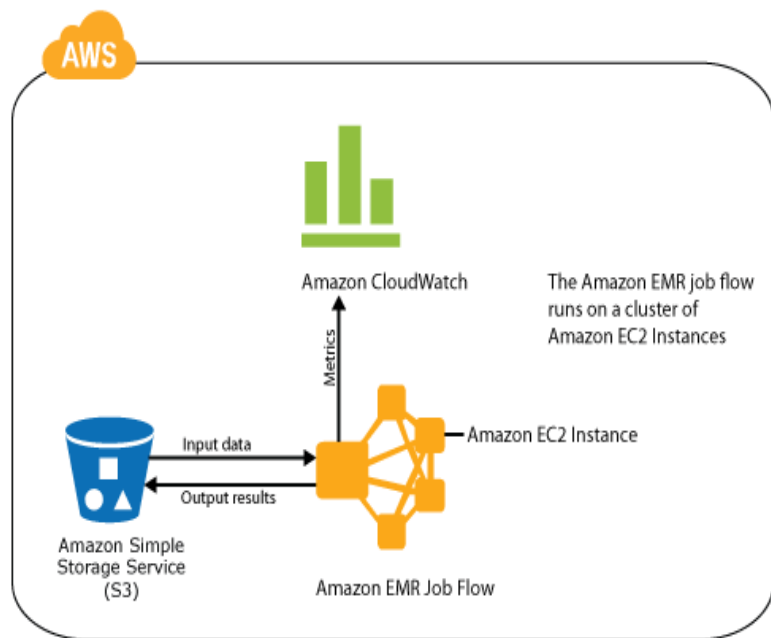
Bare Metal(Ironic) New

支持系统

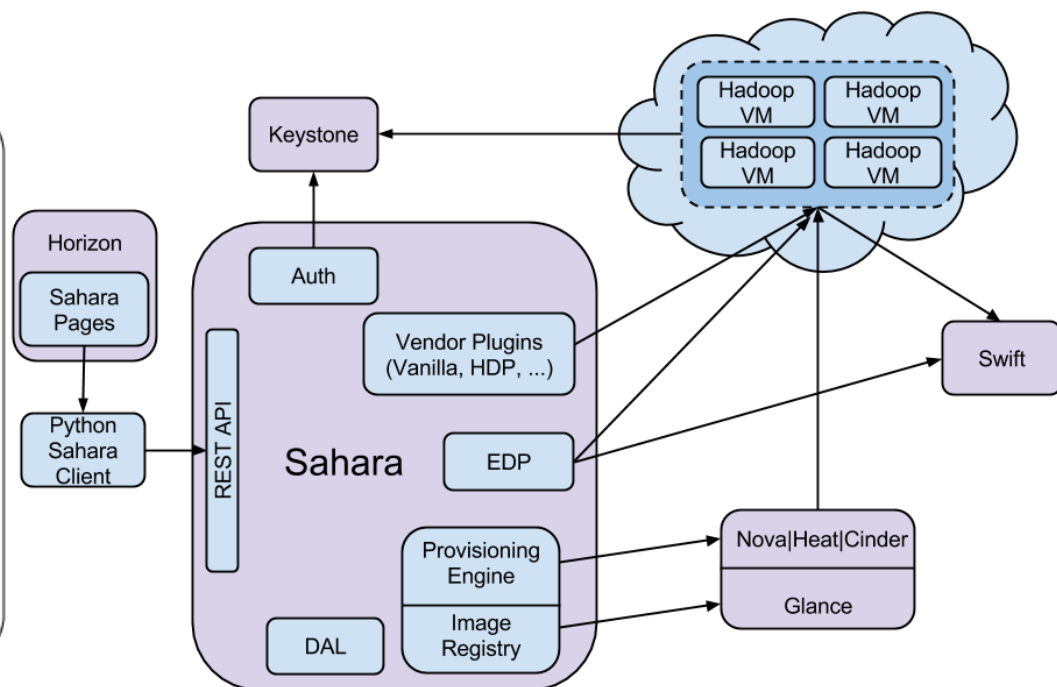
Infrastructure | Docs | Common Libraries | QA | Release Mgmt

关于Hadoop云化

AWS EMR



OpenStack Sahara



OpenStack Sahara简介

- 之前名称为Savanna
- 在OpenStack上快速创建Hadoop集群
- 利用IaaS上空闲的计算能力
- Analytics as a Service



OpenStack Sahara

- Sahara的使用模式

- 模式一：IaaS模式、管理模式
- 模式二：PaaS/DAaaS模式、用户模式、EDP模式

- 节点 (node)

- 节点组(node group)

- 节点组模板

(node group template)

- 集群 (cluster)

- 集群模板

(cluster template)

- 任务 (job)

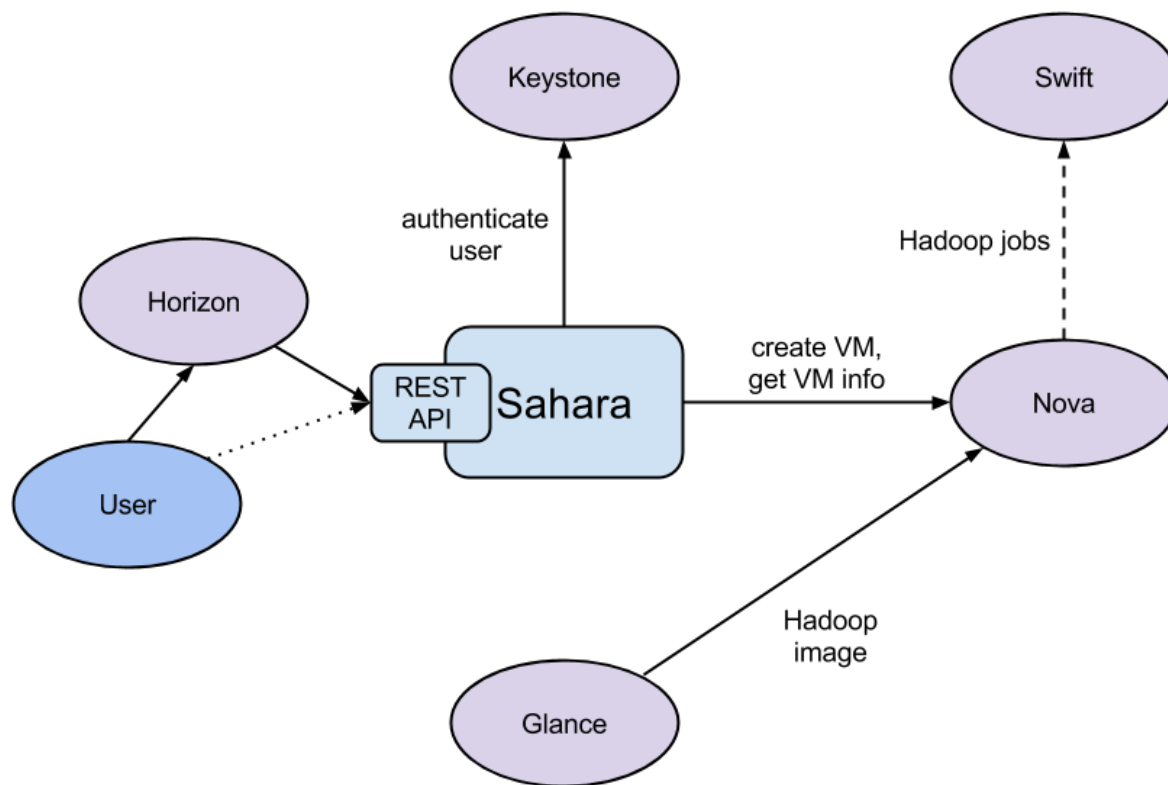
```
{  
  "name": "test-master-tmpl",  
  "flavor_id": "2",  
  "plugin_name": "vanilla",  
  "hadoop_version": "1.2.1",  
  "node_processes": ["jobtracker", "namenode"]  
}
```

节点组模板

OpenStack Sahara

```
{
  "name": "demo-cluster-template",
  "plugin_name": "vanilla",
  "hadoop_version": "1.2.1" ,
  "node_groups": [
    {
      "name": "master",
      "node_group_template_id": "b1ac3f04-c67f-445f-
b06c-fb722736ccc6",
      "count": 1
    },
    {
      "name": "workers",
      "node_group_template_id": "dbc6147e-4020-4695-
8b5d-04f2efa978c5",
      "count": 2
    }
  ]
}
```

集群模板



目前已经能够支持Spark
能够和Heat集成

Chrome File Edit View History Bookmarks Window Help

172.18.79.141/project/data_processing/data_plugins/

openstack demo admin Sign Out

Data Processing Plugins

Title	Supported Versions	Description
Vanilla Apache Hadoop	1.2.1 2.4.1	This plugin provides an ability to launch vanilla Apache Hadoop cluster without any management consoles. Also it can deploy Oozie and Hive
Hortonworks Data Platform	1.3.2 2.0.6	The Hortonworks OpenStack plugin works with project Sahara to automate the deployment of the Hortonworks data platform on OpenStack based public & private clouds
Cloudera Plugin	5	This plugin provides an ability to launch CDH clusters with Cloudera Manager management console.
Apache Spark	1.0.0 0.9.1	This plugin provides an ability to launch Spark on Hadoop CDH cluster without any management consoles.

Displaying 4 items

Project

- Compute
- Network
- Object Store
- Orchestration
- Data Processing
 - Clusters
 - Cluster Templates
 - Node Group Templates
 - Job Executions
 - Jobs
 - Job Binaries
 - Data Sources

我认为，未来会更多地属于那些告别大教堂、拥抱集市的人们。

这不是说个人的远见和才华不再重要；而是在在我看来，未来的成功者只是从自己的远见和才华开始工作，然后通过有效的社区合作，将其不断地放大。

开放式的文化会最终胜利，这或许不是因为"开放"在道德上正确，或者"封闭"在道德上错误，而只是因为开放式合作可以在一个问题上投入多几个数量级的技术工时，封闭的世界无法赢得这样的竞争。

——Eric Raymond 《大教堂和集市》

BDTC

2014 中国大数据技术大会

BIG DATA TECHNOLOGY CONFERENCE

暨第二届CCF大数据学术会议

谢谢！

开放你的人生，世界一定因此不同～