GOPS2017
Shenzhen

GOPS

2017 全球运维大会

深圳站

指导单位： 数据中心联盟 Data Center Alliance
主办单位： 高效运维社区 GreatOPS Community   开放运维联盟 OOPSA Open OPS Alliance

Apache Eagle

Architecture Evolvement and New Features

Hao Chen, Lead PMC and Committer of Apache Eagle

个人简介

Hao Chen / 陈浩

Apache Eagle 联合发起人（PMC 以及Committer）

eBay基础架构部 资深工程师 (Staff Engineer, Member of Technical Staff)

QCon, Hadoop Summit (中国/北美/日本), GOPS 等国内外知名会议讲师

Agenda

Introduction

New Features and Use Cases

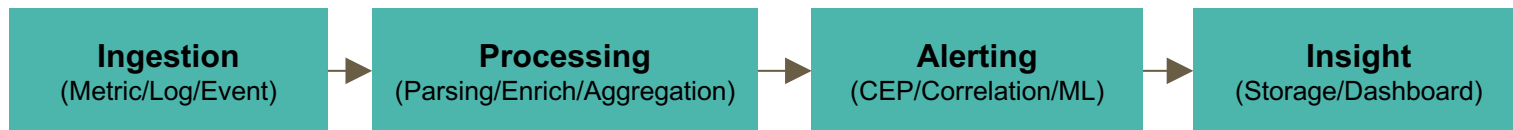Architecture Evolvement

What's Next

Q&A

Apache Eagle - Introduction

Apache® Eagle™ analyzes data activities, yarn applications, JMX metrics, and daemon logs etc., provides state-of-the-art alert engine to identify security breach, performance issues and shows insights.

	Apache Eagle v0.3 Release	Apache Eagle v0.4 Release	Apache Eagle v0.5 Release

Oct, 2015

Jan, 2017

Apache Incubation

Apr, 2016

Jul, 2016

★ **Apache Top Level Project**

March, 2017

Apache Eagle - Introduction

Apache® Eagle™ analyzes data activities, yarn applications, JMX metrics, and daemon logs etc., provides state-of-the-art alert engine to identify security breach, performance issues and shows insights.

Ingestion (Metric/Log/Event)	→	Processing (Parsing/Enrich/Aggregation)	→	Alerting (CEP/Correlation/ML)	→	Insight (Storage/Dashboard)

Global Marketplace

http://www.ebay.com

291M
MOBILE APP DOWNLOADS GLOBALLY

162M
ACTIVE BUYERS
Q3 2016

800M
ACTIVE LISTINGS
Q3 2016

1.3B
LISTINGS CREATED VIA MOBILE EVERY WEEK

25M
ACTIVE SELLERS

8.8M
NEW LISTINGS EVERY WEEK

42%
GMV VIA MOBILE

Trustable Ecommerce Platform

APPLICATIONS / DATABASE

BIG DATA PLATFORM

CLOUD PLATFORM

Log

Metric

Event

Critical Event

Vision
- Availability
- Security

Capability
- Monitoring
- Alerting
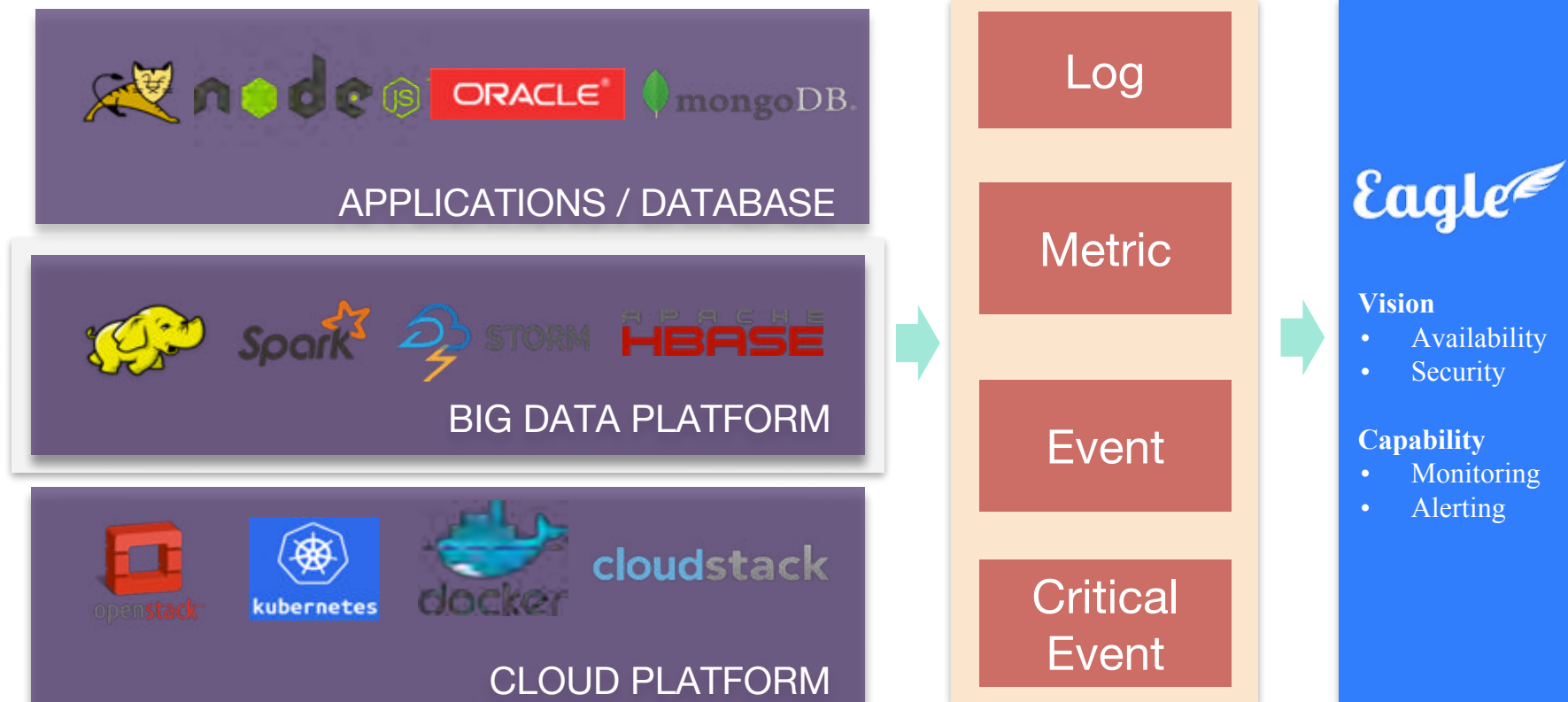
Big Data in eBay

Eagle was initialized by end of 2013 for hadoop ecosystem monitoring as any existing tool like zabbix, ganglia can not handle the huge volume of metrics/logs generated by hadoop system in eBay.

Hadoop @ eBay Inc

3000+ nodes
10,000+ cores
50+ PB
2012

100+ nodes
1000 + cores
1 PB
2010

2011
1000+ nodes
10,000+ cores
10+ PB

2015/2016
10,000+ nodes
150,000+ cores
250 PB
2000+ user

2009
50+ nodes

2007
1-10 nodes
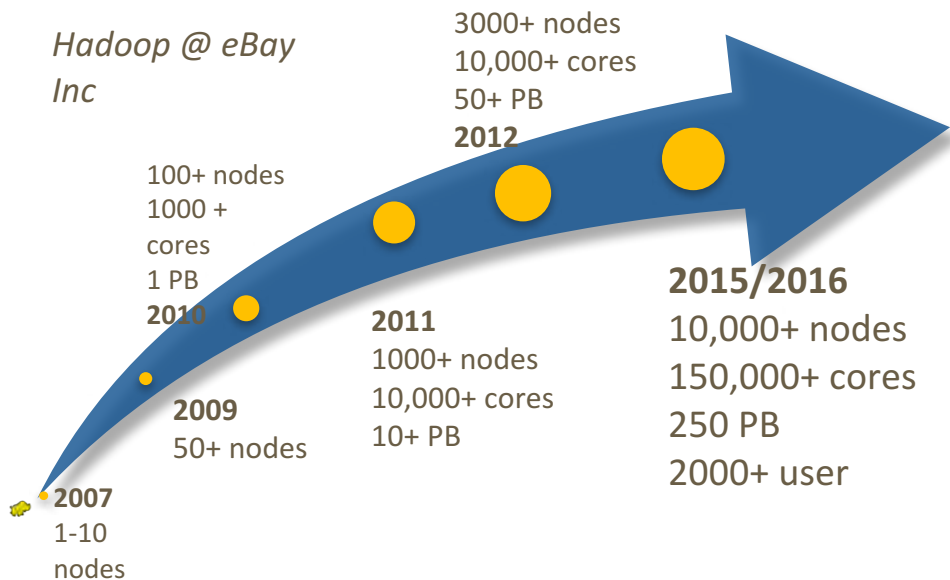
Hadoop Data
- *Security*
- *Activity*

Hadoop Platform
- *Heath*
- *Availability*
- *Performance*

7+		
	CLUSTERS	
10000+		*NODES*
250+ PB		*DATA*
10 B+ DAY		*EVENTS /*
500+		
	METRIC TYPES	
50,000+		*JOBS /*

Apache Eagle - Typical Use Cases

1 **Service Health Check**
Service and process aliveness, JMX status as well as JVM GC

2 **Bad Node Detection**
Detect soft failure issues, Linux filesystem ACL, disk full
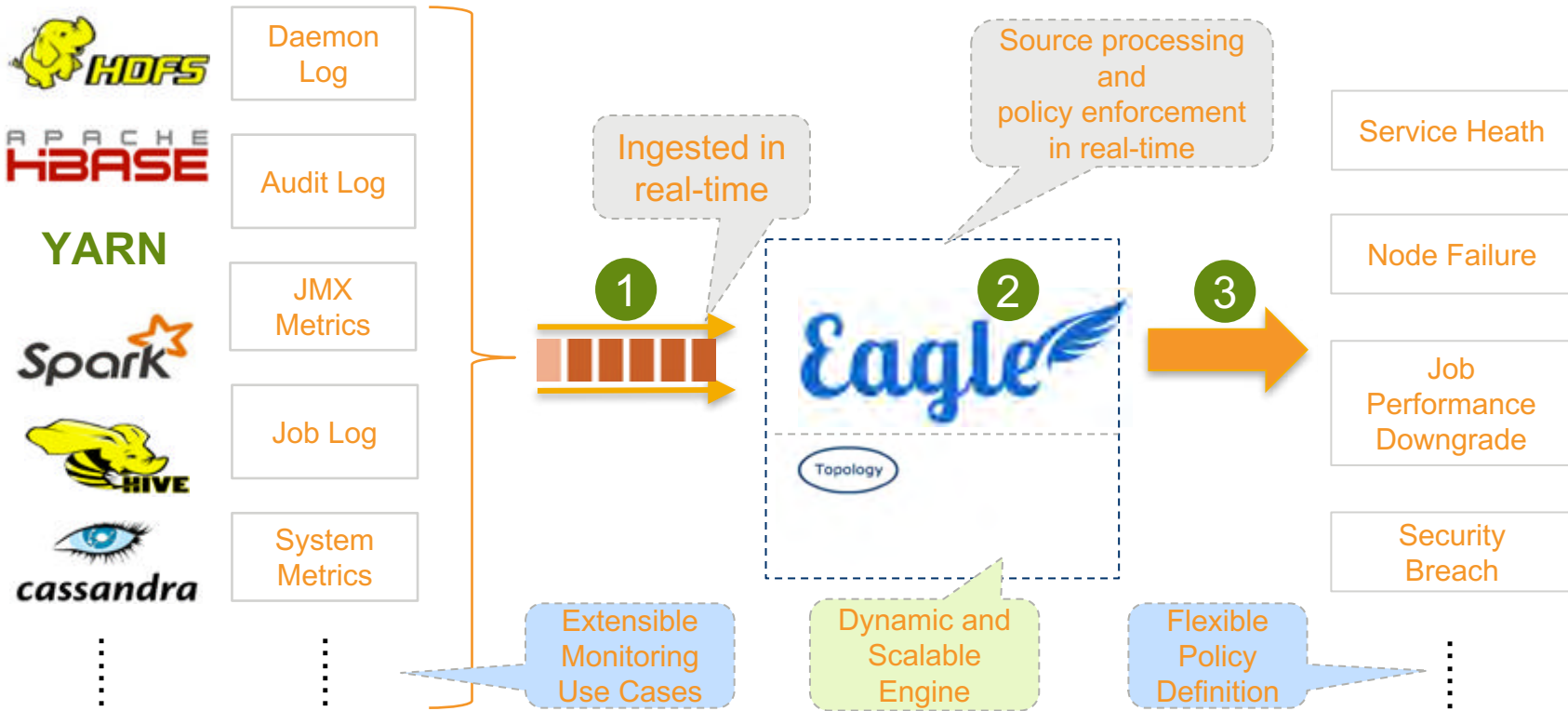
3 **Security Monitoring**
Instantly identify sensitive data access and malicious operations
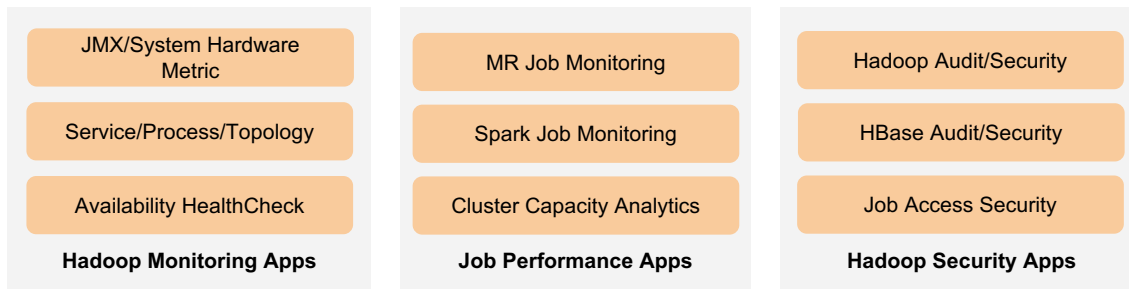
4 **Job Performance Monitoring**
Hadoop, Spark job profiling and performance analysis

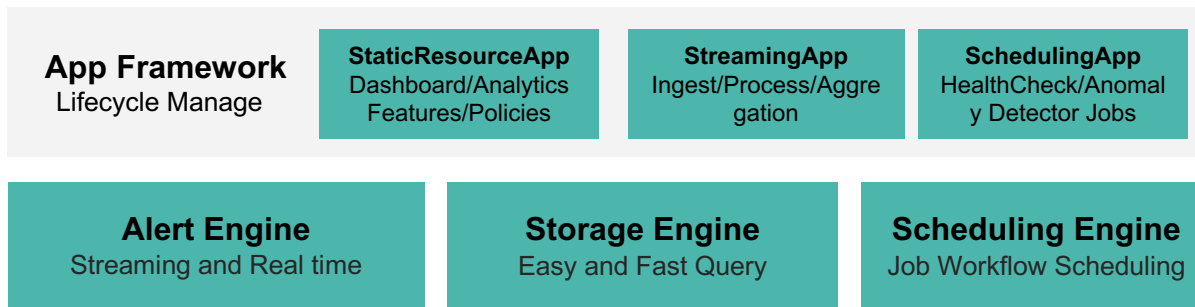# Apache Eagle - Overview

Apache Eagle - Components

Eagle Apps

Hadoop Monitoring Apps
- JMX/System Hardware Metric
- Service/Process/Topology
- Availability HealthCheck

Job Performance Apps
- MR Job Monitoring
- Spark Job Monitoring
- Cluster Capacity Analytics

Hadoop Security Apps
- Hadoop Audit/Security
- HBase Audit/Security
- Job Access Security

Eagle Core

App Framework
Lifecycle Manage

StaticResourceApp
Dashboard/Analytics Features/Policies

StreamingApp
Ingest/Process/Aggregation

SchedulingApp
HealthCheck/Anomaly Detector Jobs

Alert Engine
Streaming and Real time

Storage Engine
Easy and Fast Query

Scheduling Engine
Job Workflow Scheduling

Eagle Interface

UI
Dashboard
API

Integration

kafka APACHE STORM APACHE HBASE Apache Ambari docker
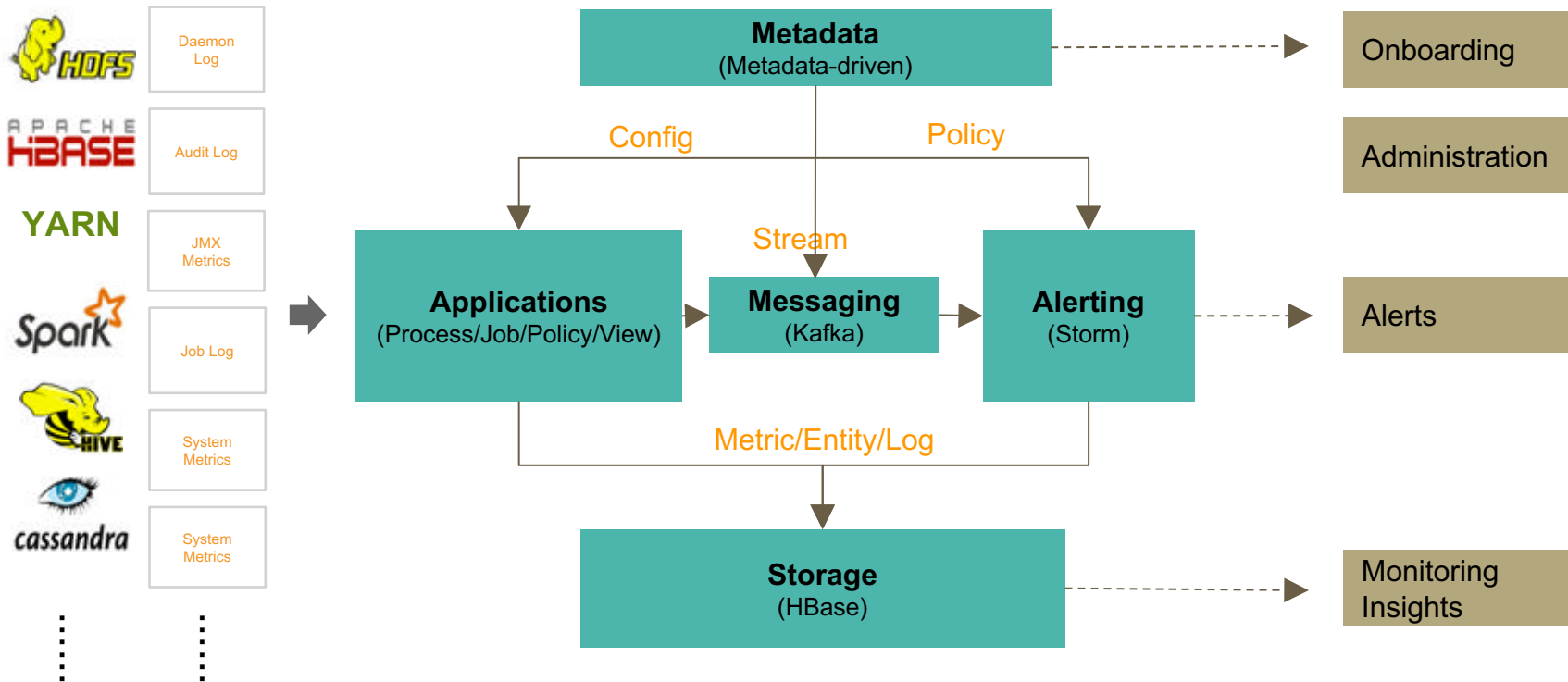
Apache Eagle - Case Onboarding

1. Register a new Monitored Site

2. Choose and Install Application

3. Configure Application

4. Administrate Application

5. Define Alert Policies and Explore Alerts

6. Analyze with Dashboards and Insights
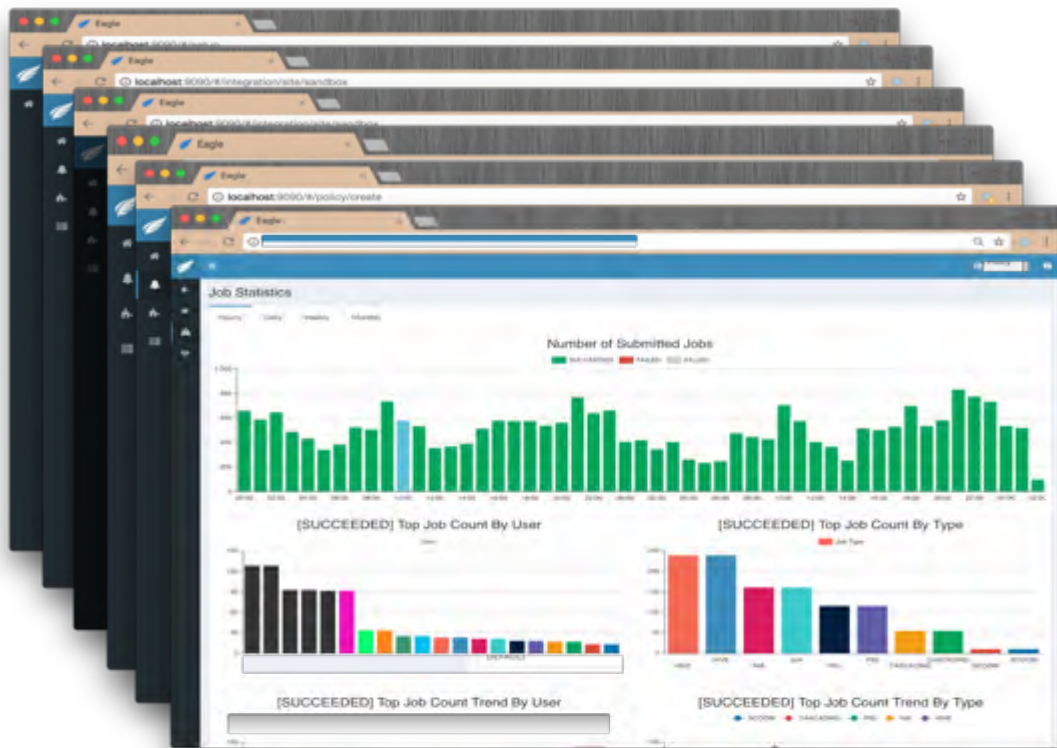
Apache Eagle - Architecture

Apache Eagle - Application Framework

An **"Application"** is case-oriented solution package

Installation: Application user guide, configuration, management

Ingestion: Provide data ingestion/collection approaches to integrate any kinds of monitor data sources

Process: Analyze data source based on Storm Topology or Spark Streaming App
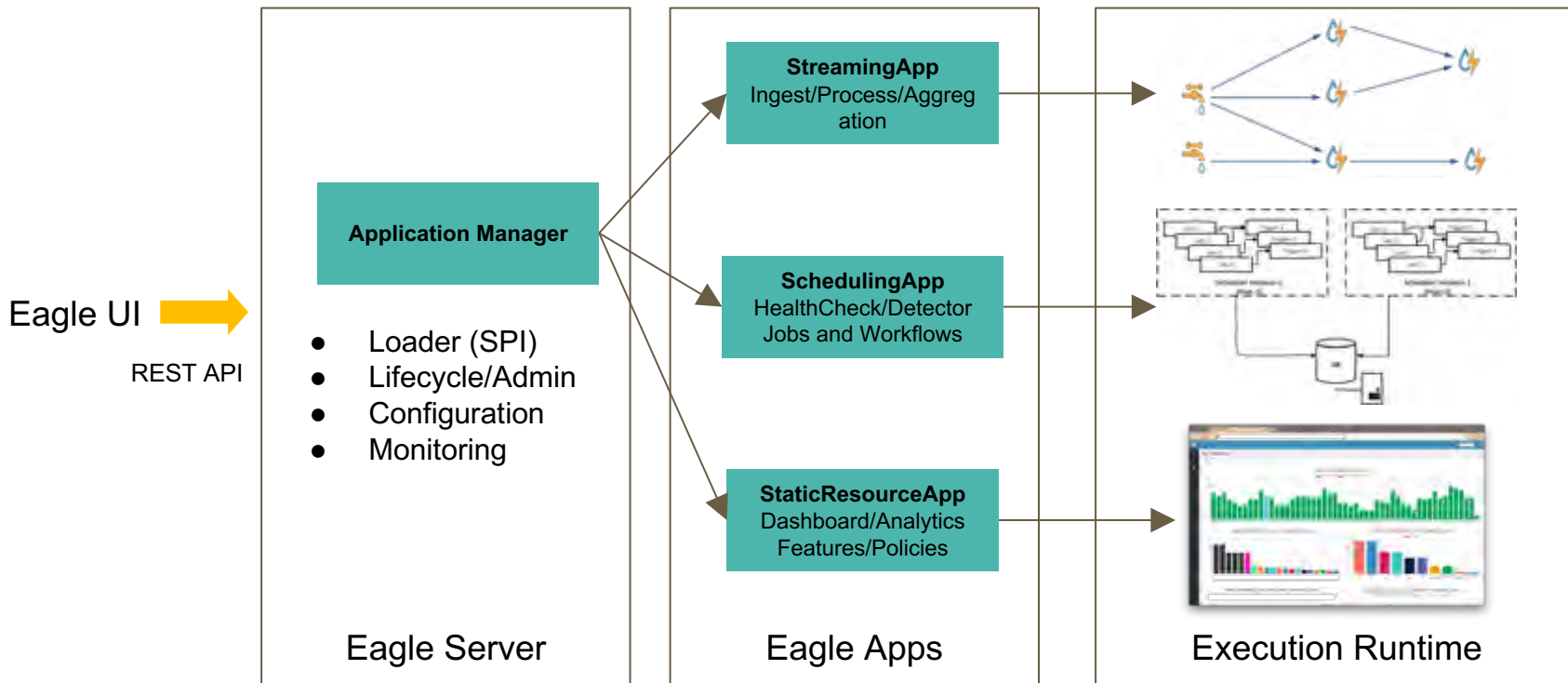
Alerting:

    Stream: Structured stream exported for alerting with eagle alerting engine or persistence in eagle storage

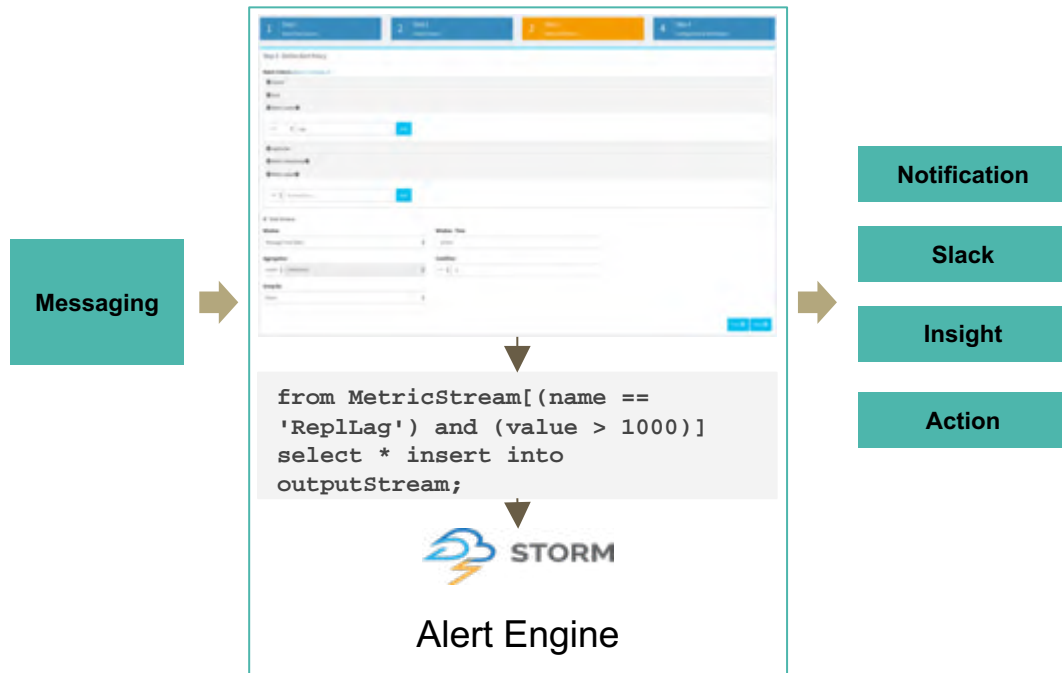   **Model:** Complex built-in policies or policy templates defined in SQL/Java code/ML model, etc.

Insight: Monitoring Analytics UI or Dashboard

Apache Eagle - Application Execution

Apache Eagle - Distributed Alert Engine

- **Real-time Streaming**: Apache Storm (Execution Engine) + Kafka (Messaging)

- **Declarative Policy**: CEP and Extensible Alert Model in streaming way

- **Dynamical Onboarding & Correlation**: Connect to new stream and change Stream Grouping in Runtime

- **Hot Deploy & No Downtime**: Metadata-driven and lightweight alert logic assignment

Messaging

```
from MetricStream[(name ==
'ReplLag') and (value > 1000)]
select * insert into
outputStream;
```

Alert Engine

Notification

Slack

Insight

Action

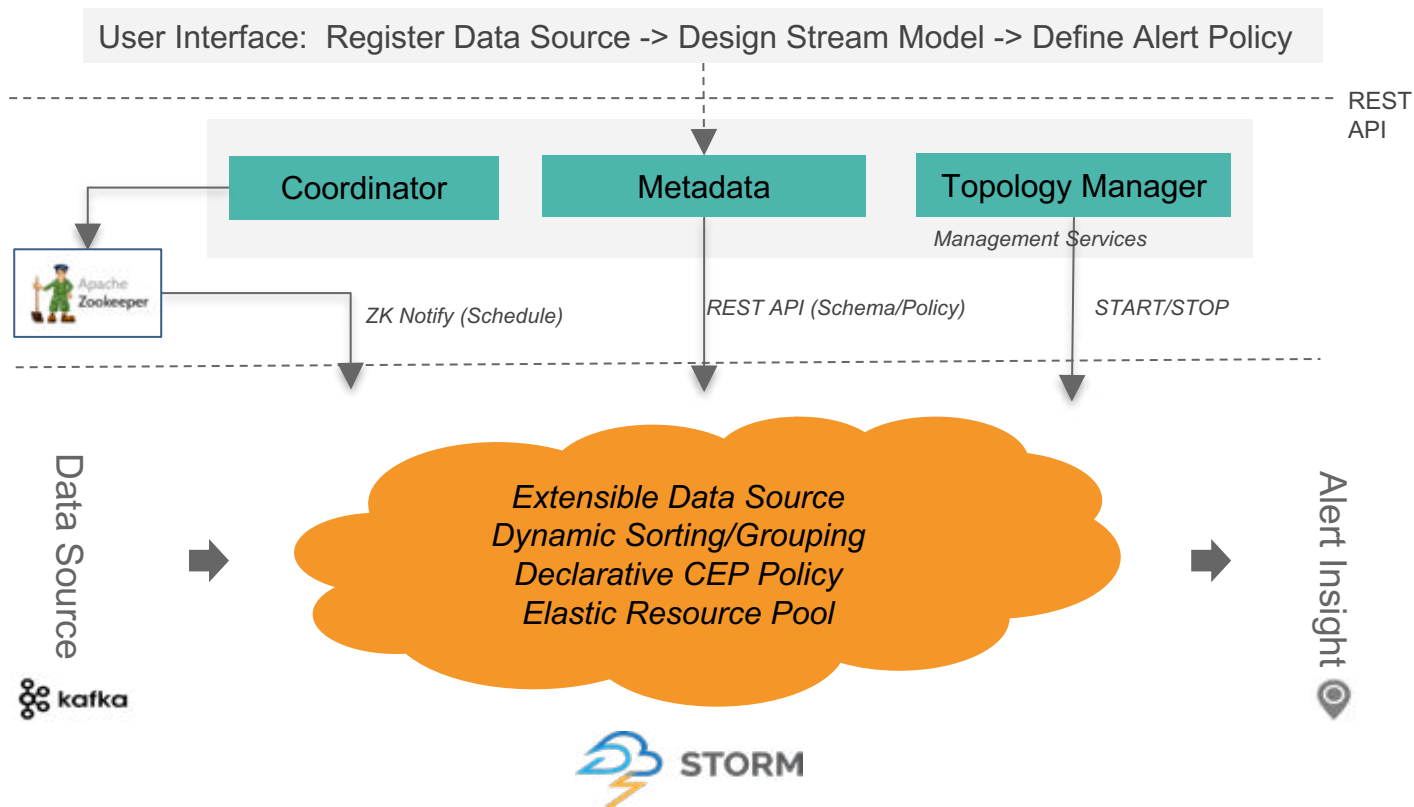# Apache Eagle - Distributed Alert Engine

Example 1: Alert if hadoop namenode capacity usage exceed 90 percentages

```
from hadoopJmxMetricEventStream
[metric == "hadoop.namenode.fsnamesystemstate.capacityused" and value > 0.9] select
metric, host, value, timestamp, component, site insert into alertStream;
```

Example 2: Alert if hadoop namenode HA switches

```
from every
a = hadoopJmxMetricEventStream[metric=="hadoop.namenode.fsnamesystem.hastate"]
->
b = hadoopJmxMetricEventStream[metric==a.metric and b.host == a.host and a.value !=
value)]
within 10 min
select a.host, a.value as oldHaState, b.value as newHaState, b.timestamp as timestamp,
b.metric as metric, b.component as component, b.site as site insert into alertStream;
```

Apache Eagle - Distributed Alert Engine

Apache Eagle - Distributed Alert Engine
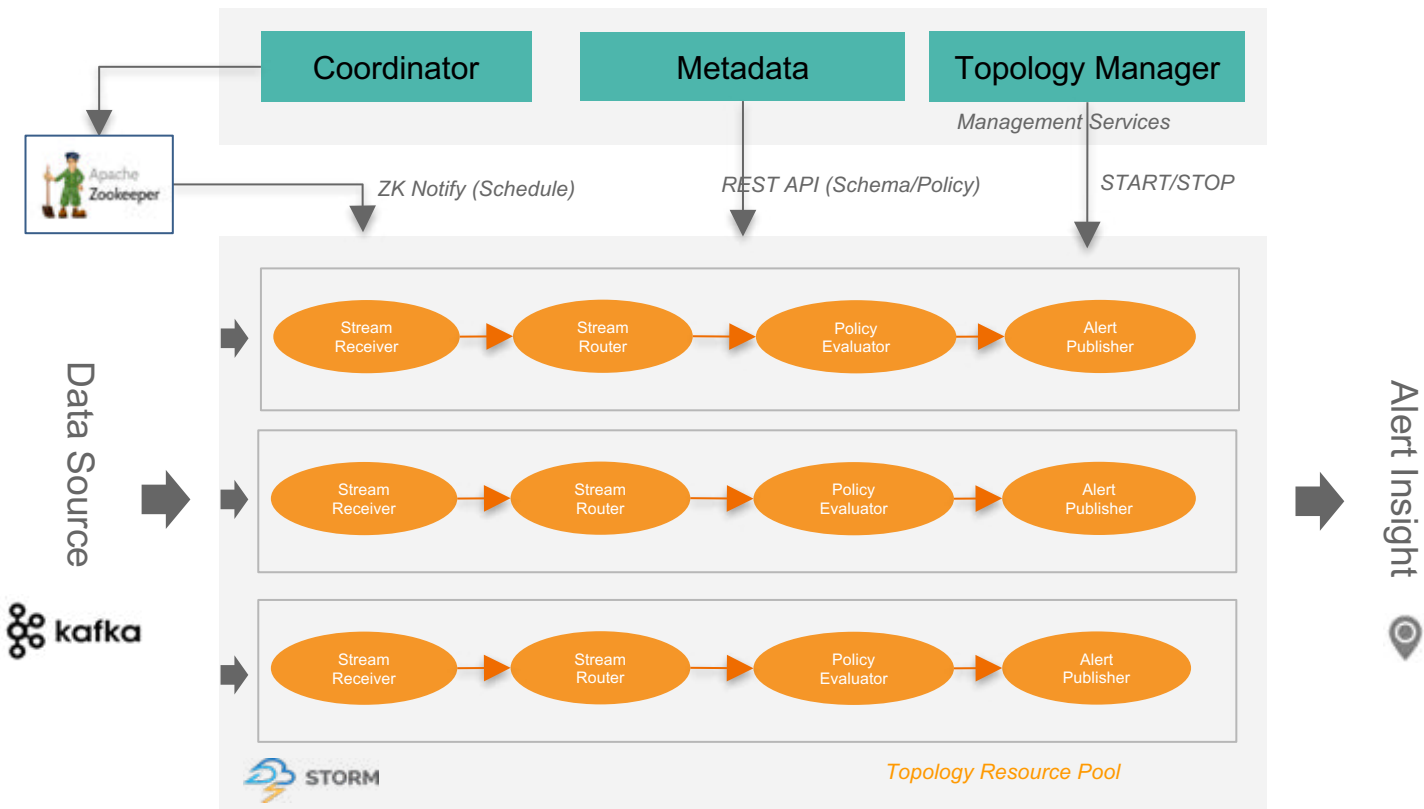
Apache Eagle - Distributed Alert Engine

From Policy Definition in User View ----→ Schedule Assignment ----→ to Engine View

```
define stream SystemMetricStream (
    metric string,
    host string,
    device string
    value double);

from  SystemMetricStream
[name = "disk.usage.metric" and value > 0.99 ]
#window.time(30 min)
group by host, device
insert into SystemAlertStream
```

SourceSpec

Source: kafka topic
Schema: SystemMetricStream

SortSpec

Window: 5min
Margin: 1min

PartitionSpec

Partition:
Group by host, device

PublishSpec

Publish:
SystemAlertStream

PolicySpec

Process:
CEP Execution Plan

Apache Eagle - Distributed Alert Engine
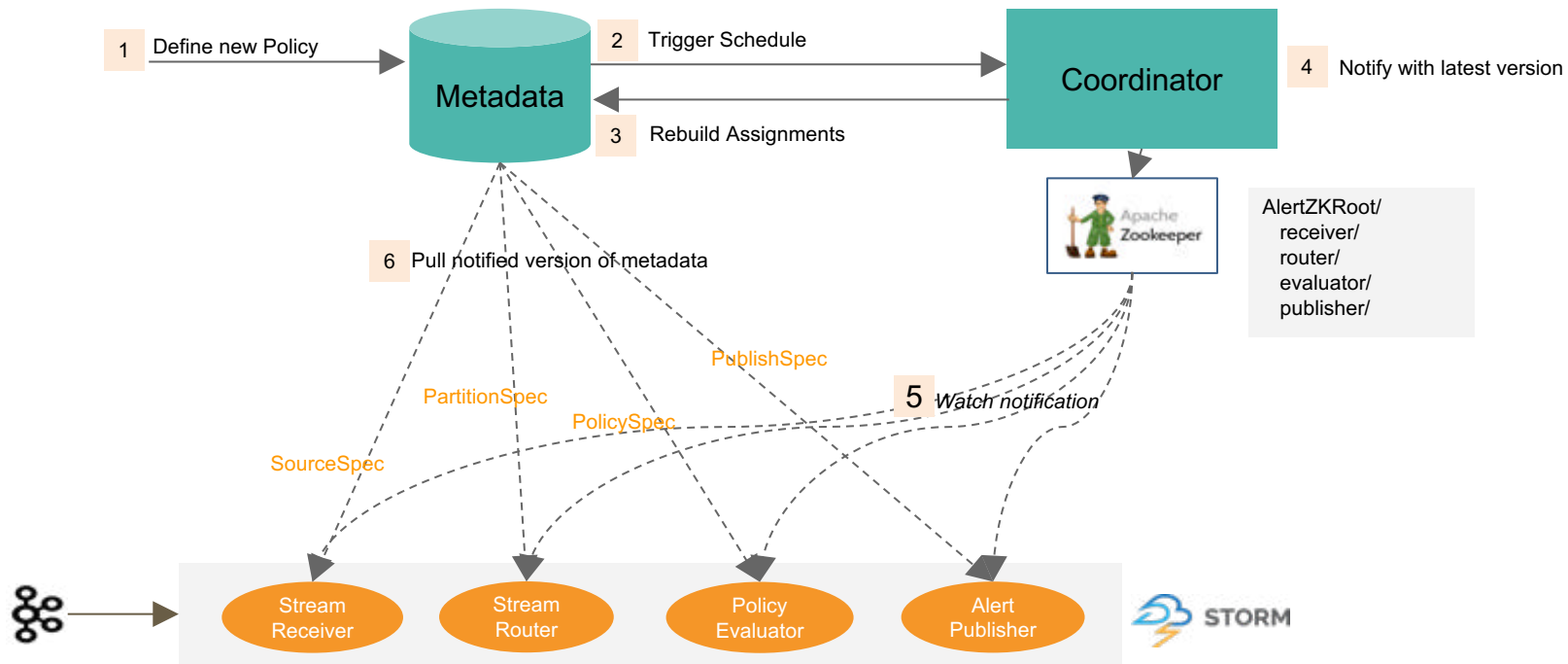
Apache Eagle - TSDB Storage Engine

- *Light-weight ORM Framework for HBase/RDMBS*
- *Full-function SQL-Like REST Query*
- *Optimized Rowkey design for time-series data*
- *Native HBase Coprocessor*
- *Secondary Index Support*

```
@Table("alertdef")
@ColumnFamily("f")
@Prefix("alertdef")
@Service(AlertConstants.ALERT_DEFINITION_SERVICE_ENDPOINT_NAME)
@JsonIgnoreProperties(ignoreUnknown = true)
@TimeSeries(false)
@Tags({"site", "dataSource", "alertExecutorId", "policyId", "policyType"})
@Indexes({
    @Index(name="Index_1_alertExecutorId", columns = { "alertExecutorID" },
unique = true),
})
public class AlertDefinitionAPIEntity extends TaggedLogAPIEntity{
    @Column("a")
    private String desc;
    @Column("b")
    private String policyDef;
    @Column("c")
    private String dedupeDef;
```


Eagle Query Framework
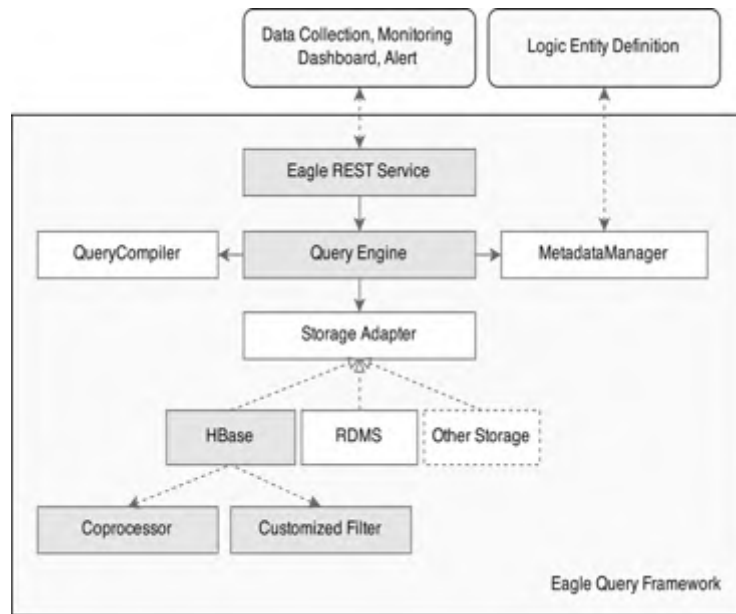
```
Query=AlertDefinitionService[@dataSource="hiveQueryLog"]{@policyDef}
```

Apache Eagle - What's Next

Eagle Alert Engine on Apache Beam

- Unified streaming on Spark/Flink

Eagle Integration with Ambari/Cloudera Manager

- Seamless connect monitoring data source

Eagle on Cloud

- Support deployment and monitor service on AWS

Unified Monitoring Applications

- Monitor real-time/online platform like Storm/Kafka/Database, etc.

Apache Eagle - Learn more

Community

- Website: http://eagle.apache.org

- Github: http://github.com/apache/eagle

- Mailing list: dev@eagle.incubator.apache.org

Publications

- *EAGLE: USER PROFILE-BASED ANOMALY DETECTION IN HADOOP CLUSTER (IEEE)*

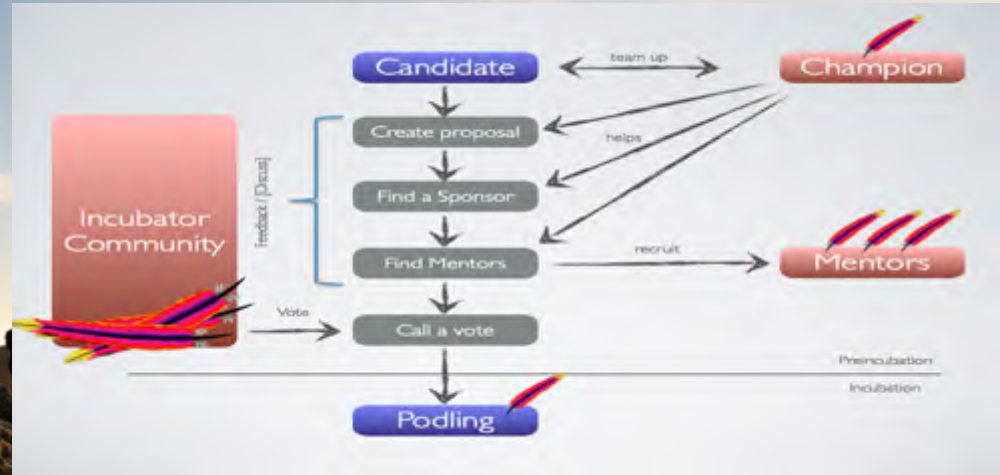- *EAGLE: DISTRIBUTED REAL-TIME MONITORING FRAMEWORK FOR HADOOP CLUSTER*

Apache Eagle - Community

Open Source

If you want to go fast, go alone.
If you want to go far, go together.

-- African Proverb

Open Sourced By **ebay**

Thanks and We are Hiring!

http://eagle.apache.org

✉ dev@eagle.incubator.apache.org

⊙ apache/incubator-eagle

🐦 @TheApacheEagle

高效运维社区
GreatOPS Community

会议

培训

咨询

- 3月18日 DevOpsDays 北京
- 8月18日 DevOpsDays 上海
- 全年 DevOps China 巡回沙龙
- 4月21日 GOPS深圳
- 11月17日 DevOps金融上海

- EXIN DevOps Master 认证培训
- DevOps 企业内训
- DevOps 公开课
- 互联网运维培训

- 企业DevOps 实践咨询
- 企业运维咨询

商务经理：刘静女士
电话 / 微信：13021082989
邮箱：liujing@greatops.com

Thanks

高效运维社区

开放运维联盟

荣誉出品

GOPS2017
Shenzhen

想第一时间看到
高效运维社区公众号
的好文章吗？

请打开高效运维社区公众号，点击右上角小人，如右侧所示设置就好