
Apache Eagle

— Architecture Evolvment and New Features —

Hao Chen, PMC and Committer of Apache Eagle

OSC - Shanghai - 2017/05/13

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 Evolvment

What's Next

Q&A

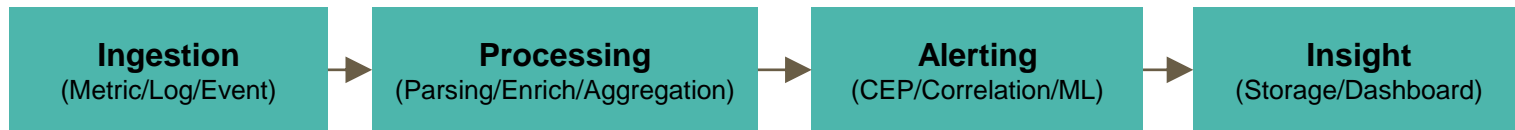
Apache Eagle - Introduction

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 - 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.



ebay Global Marketplace

<http://www.ebay.com>

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

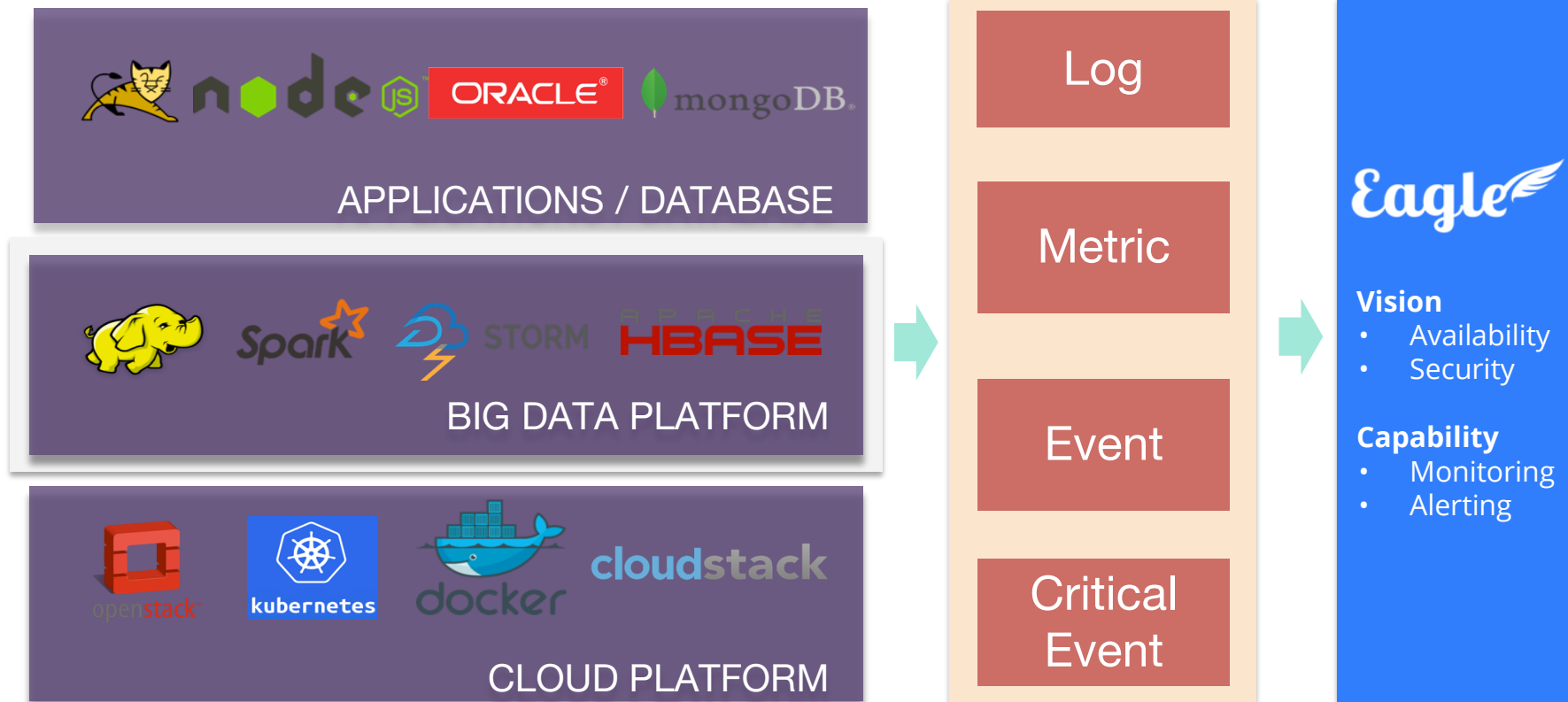
42%

GMV VIA MOBILE

291M

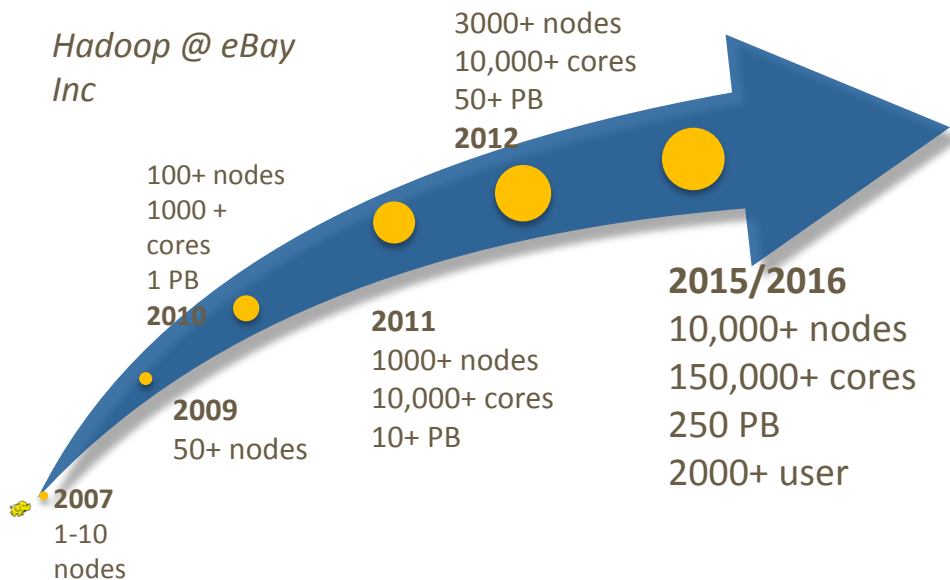
MOBILE APP DOWNLOADS
GLOBALLY

Trustable Ecommerce Platform



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 Data*
- Security
 - Activity
- Hadoop Platform*
- Health
 - Availability
 - Performance

7+	CLUSTERS	
10000+		NODES
250+ PB		DATA
10 B+ DAY		EVENTS /
500+	METRIC TYPES	

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

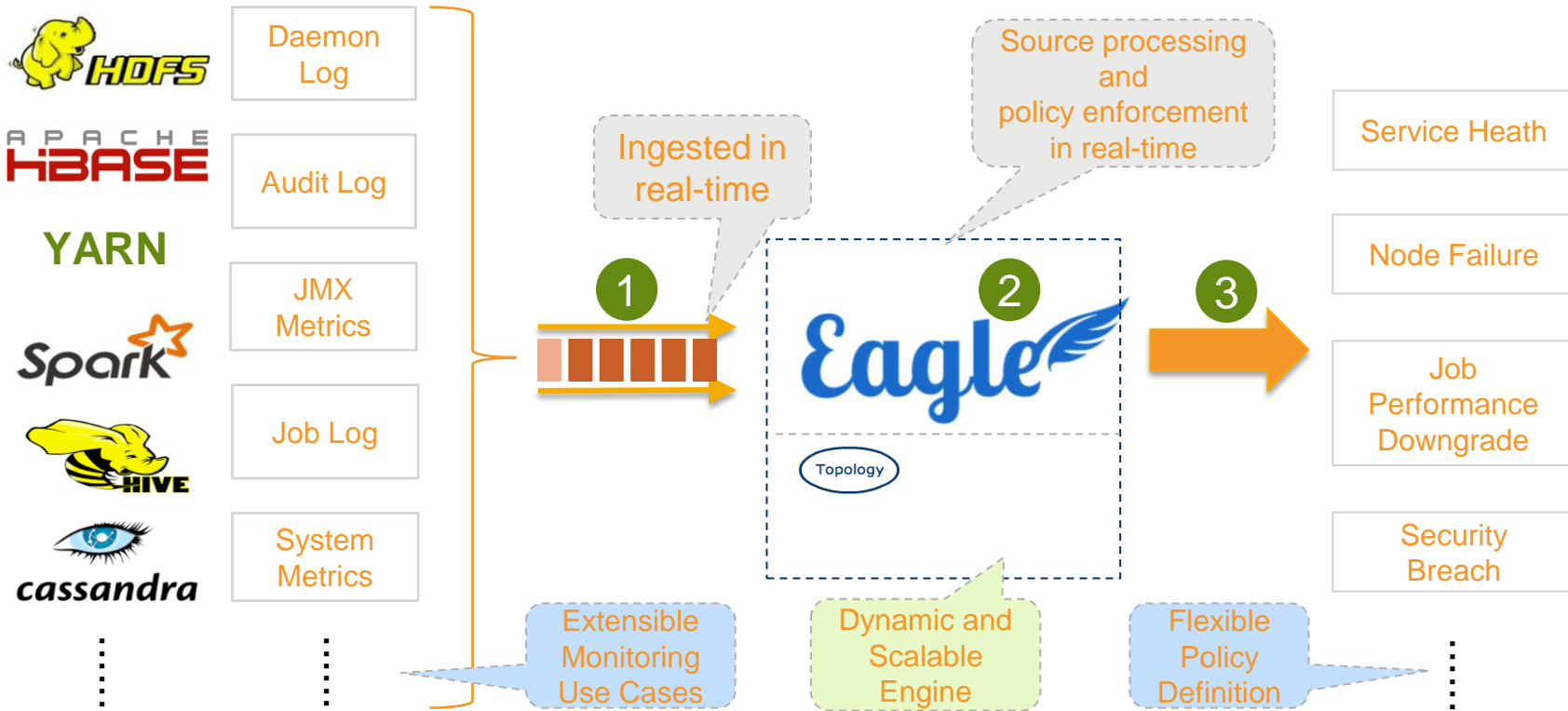
Instantly identify sensitive data access and malicious operations

4

Job Performance Monitoring

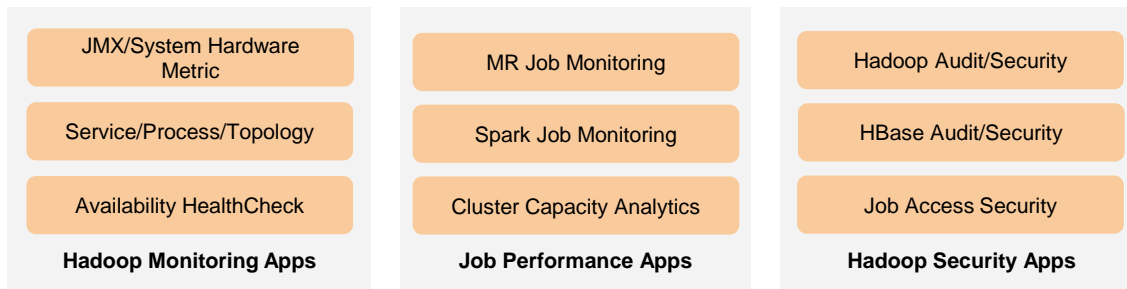
Hadoop, Spark job profiling and performance analysis

Apache Eagle - Overview



Apache Eagle - Components

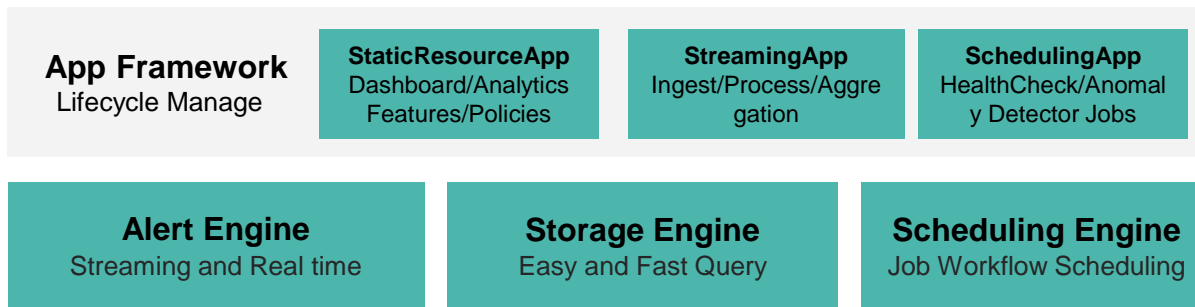
Eagle Apps



Eagle Interface

UI
Dashboard
API

Eagle Core

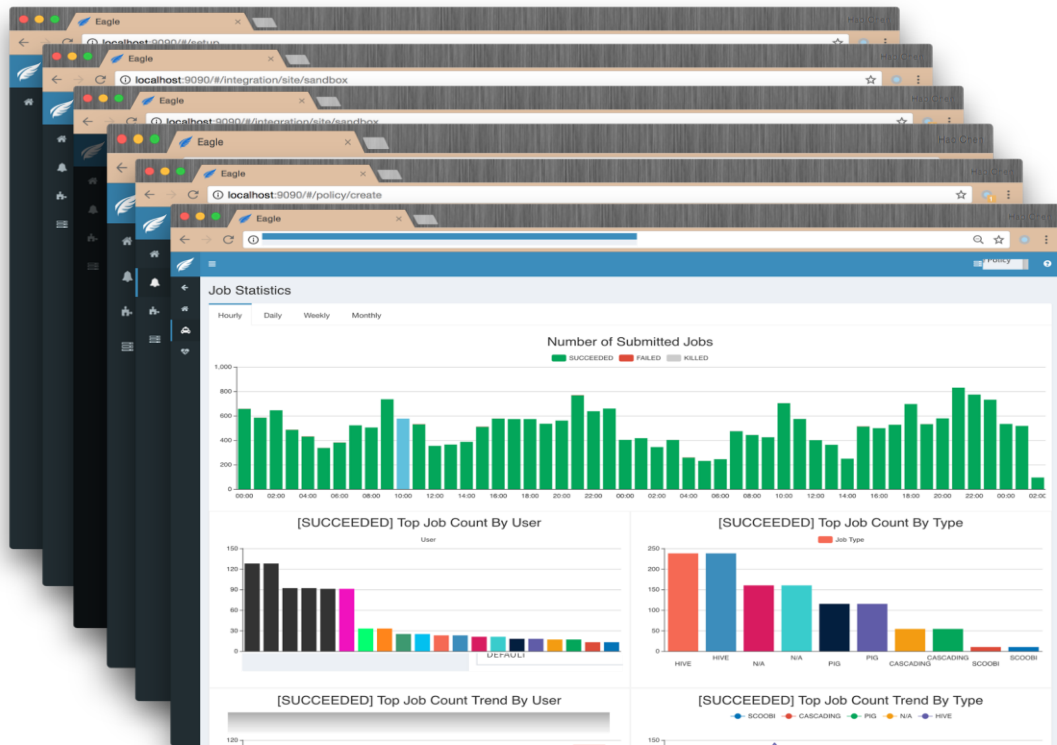


Integration

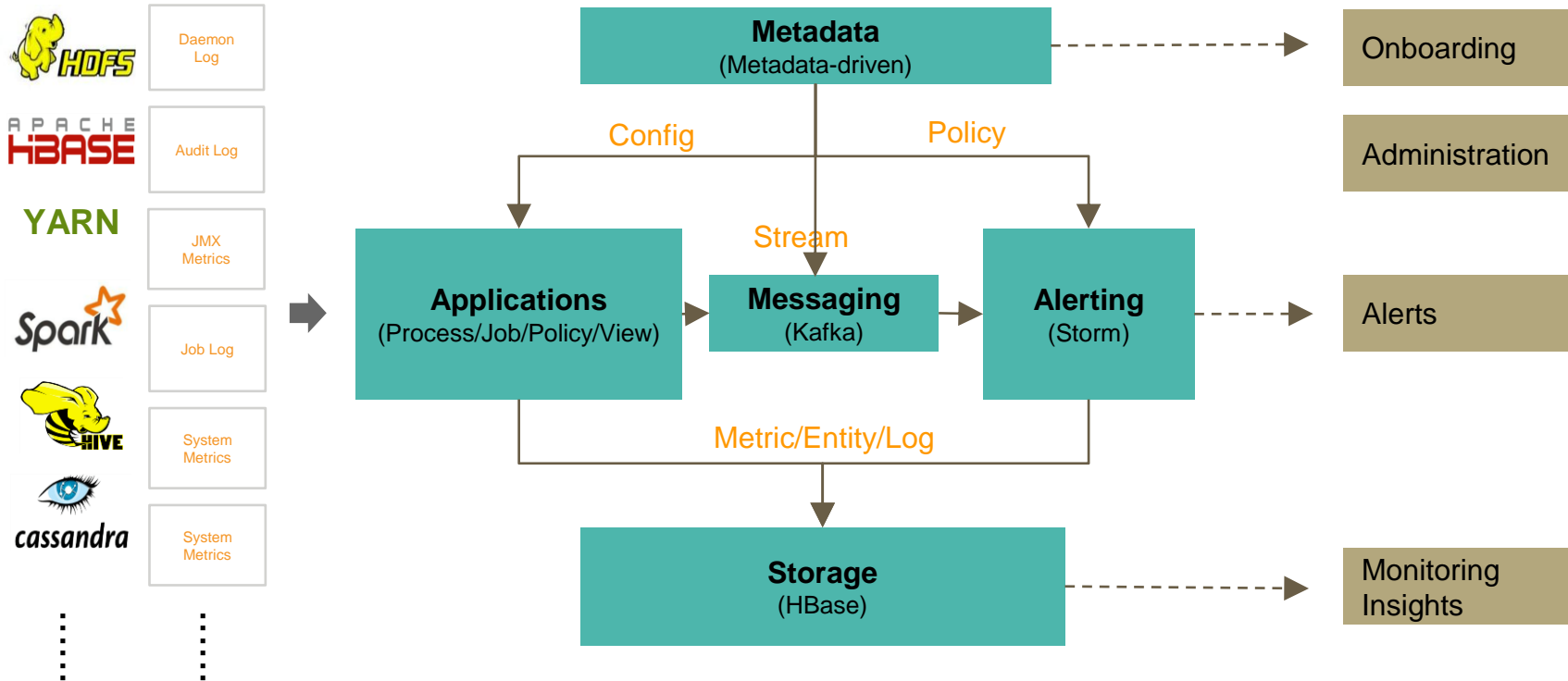


Apache Eagle - Case Onboarding

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

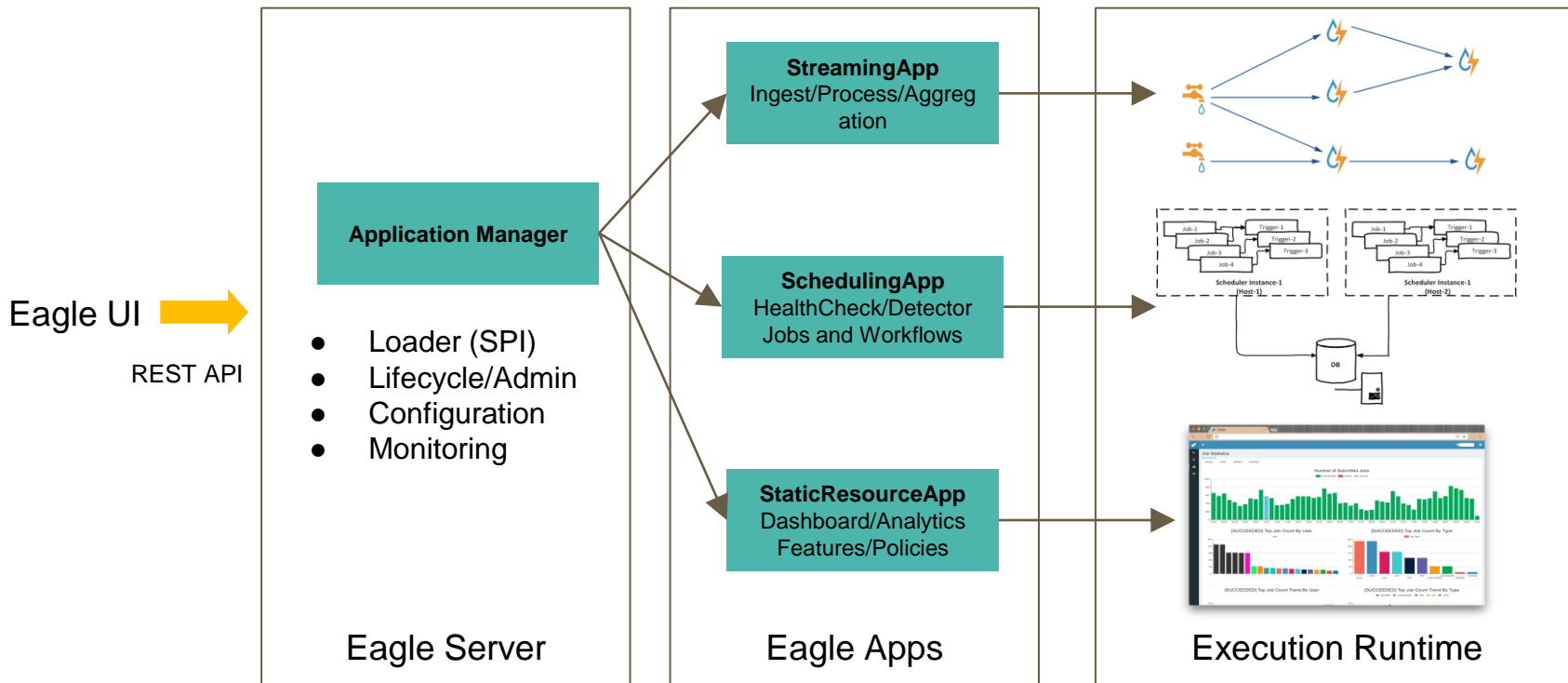
Process: Analyze data source based on Storm Topology or Spark Streaming App

Alerting:

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

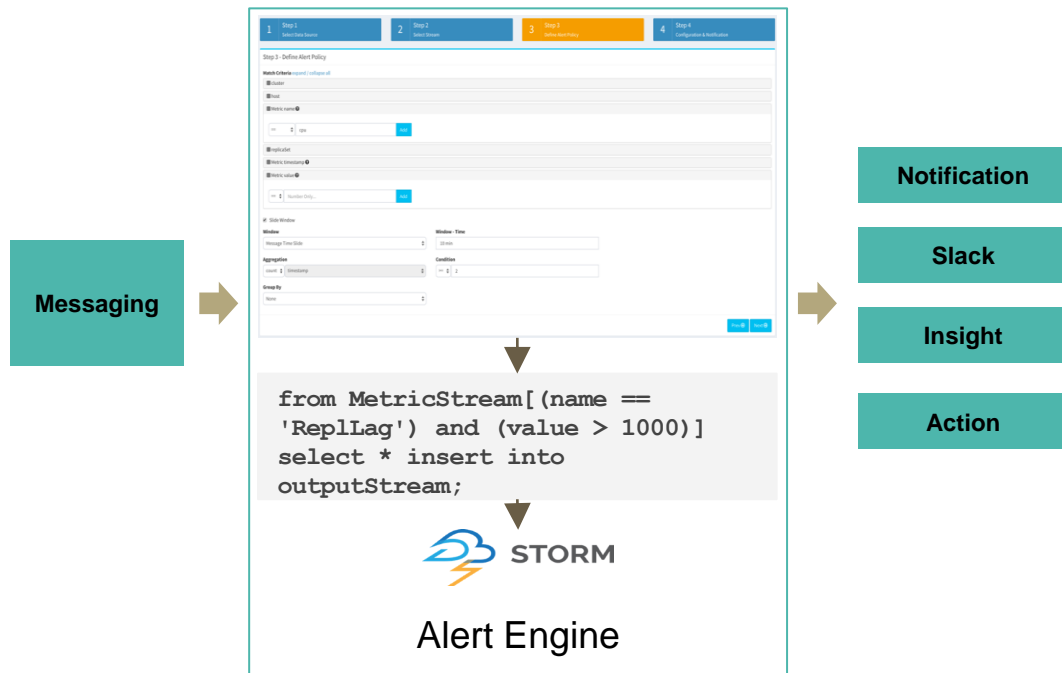
Model: Complex built-in policies or policy templates defined in SQL/Java code/ML model, etc.

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



Apache Eagle - Distributed Alert Engine

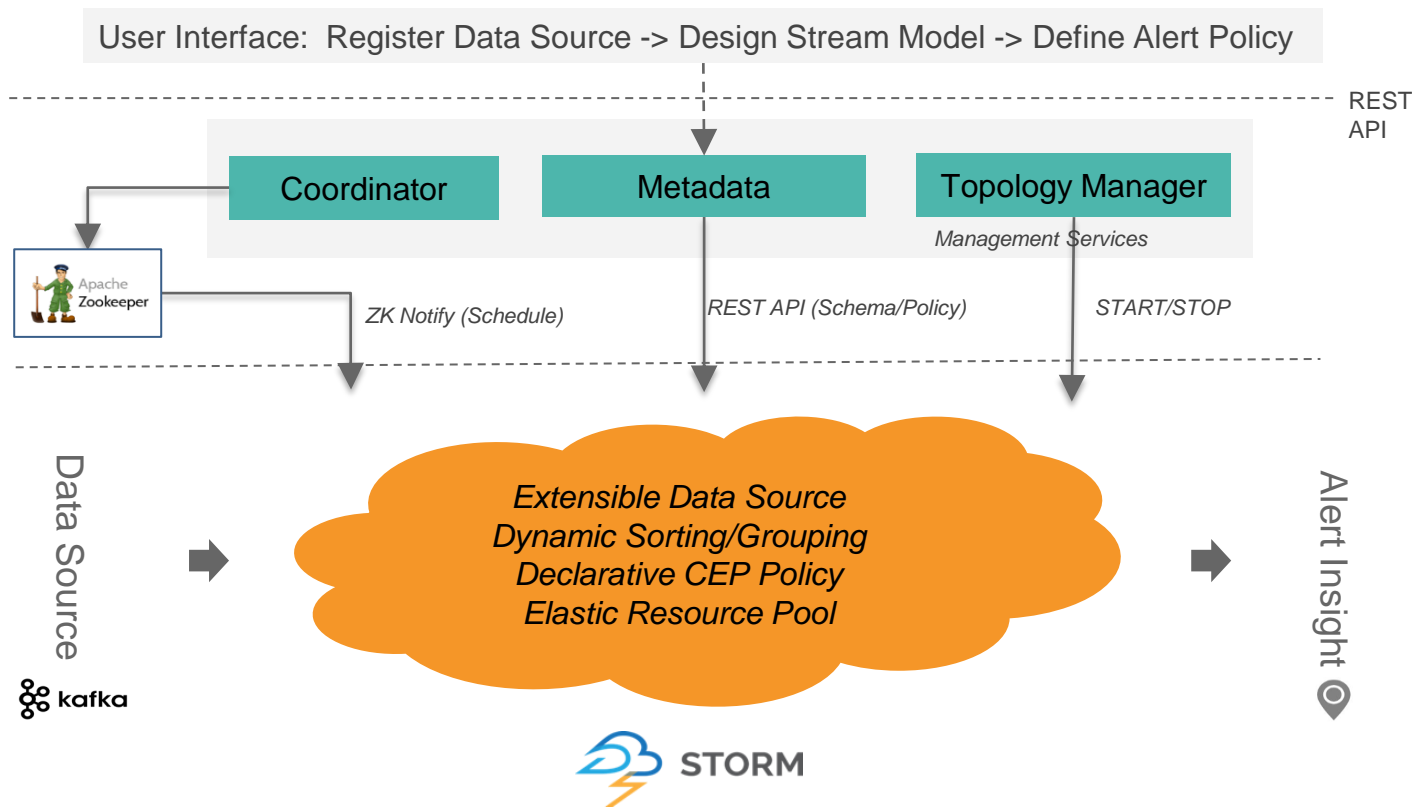
Example 1: Alert if hadoop namenode capacity usage exceed 90 percentages

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

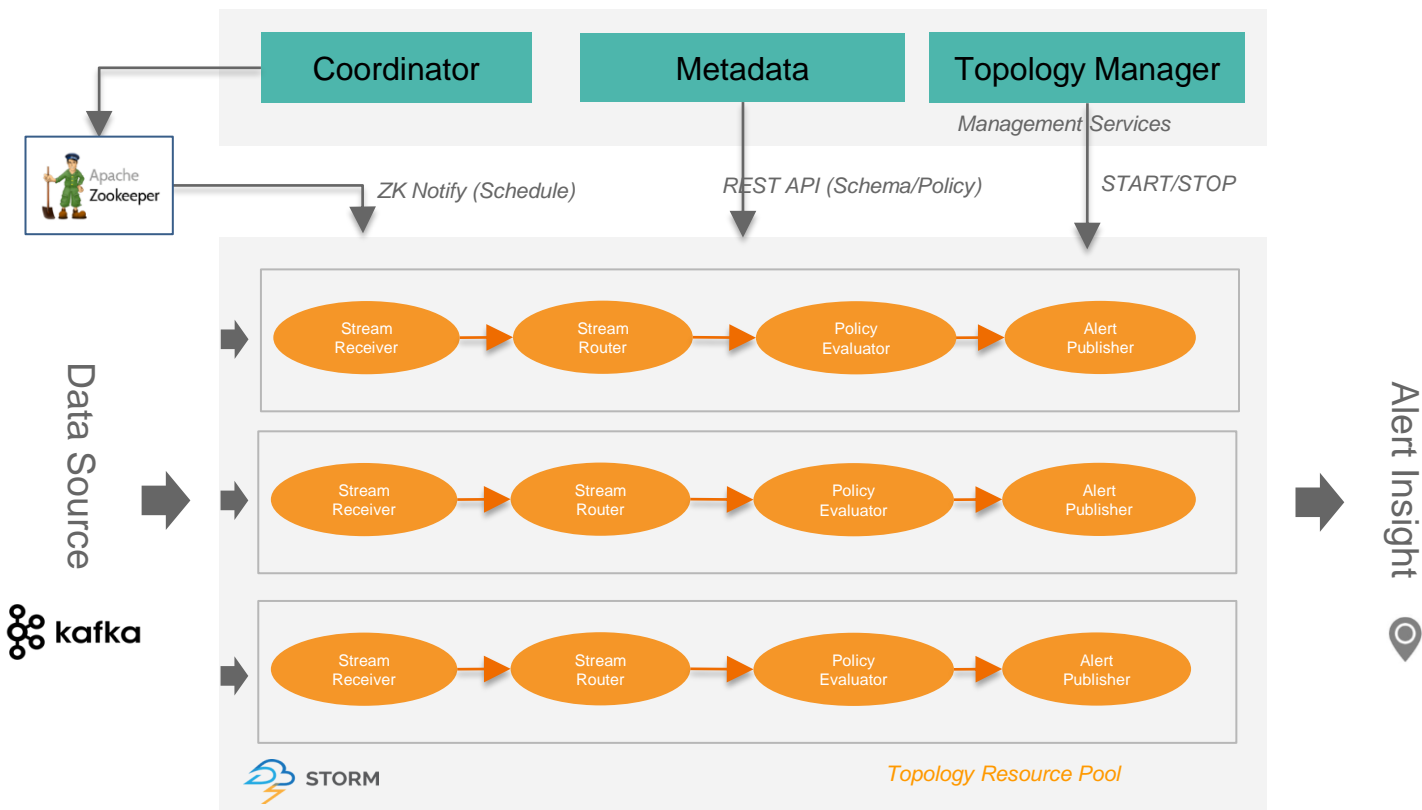
Example 2: Alert if hadoop namenode HA switches

```
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

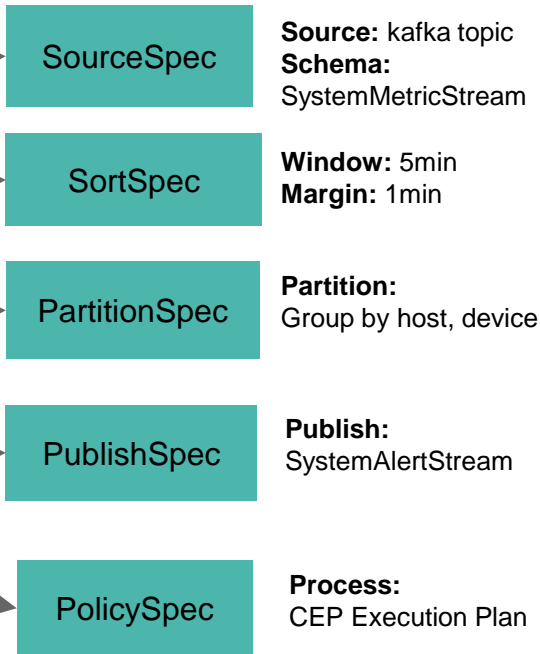


Apache Eagle - Distributed Alert Engine

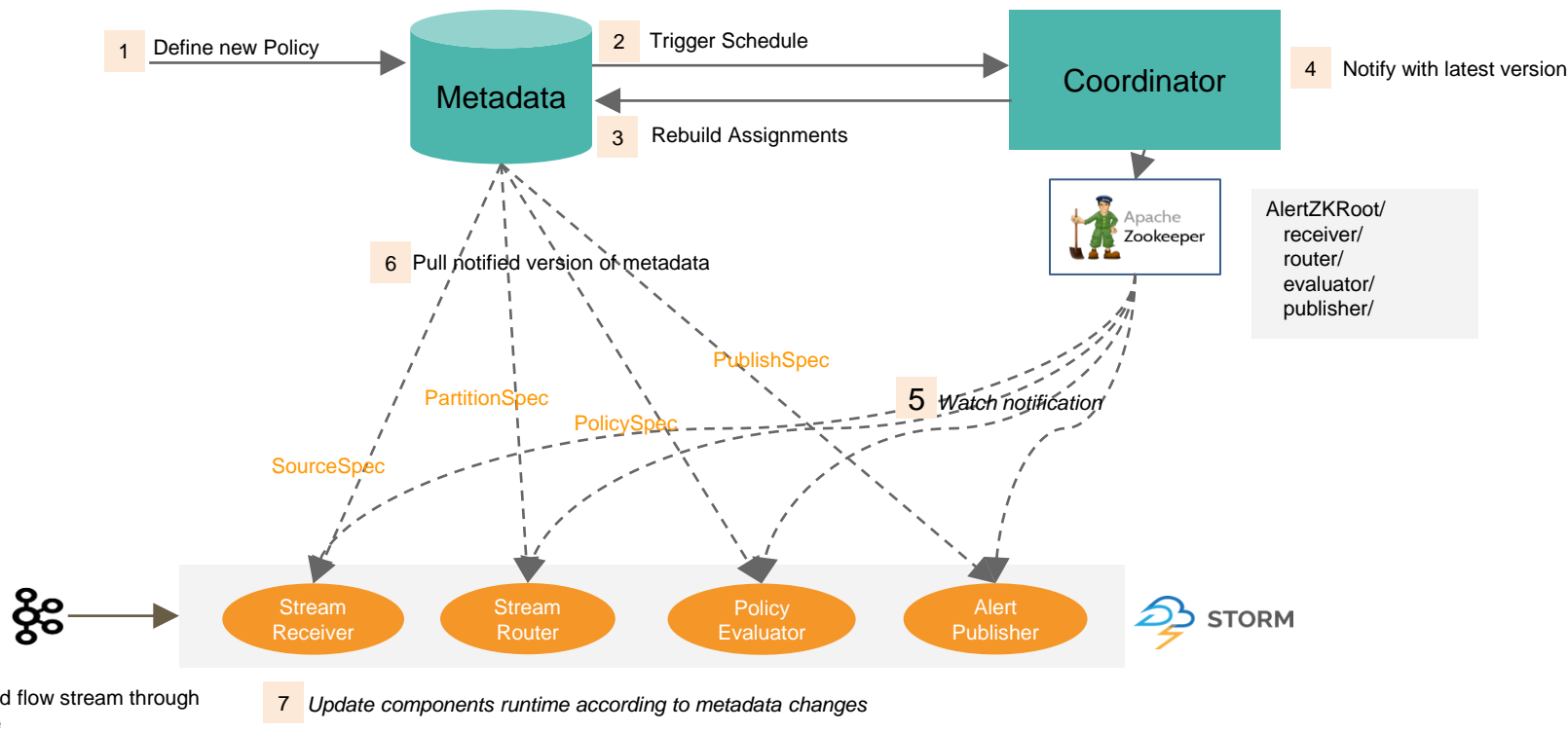
From Policy Definition in User View Schedule Assignment → to Engine View

```
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;
```



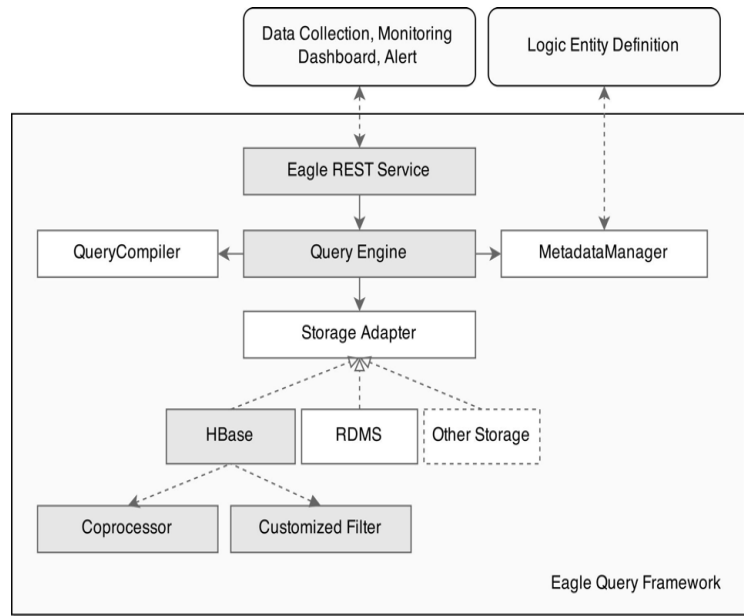
Apache Eagle - Distributed Alert Engine



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;
```



```
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 Agent

Unified Log/Metric/Event Shipper

Eagle on Cloud

Support deployment and monitor service on AWS/Aliyun

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*

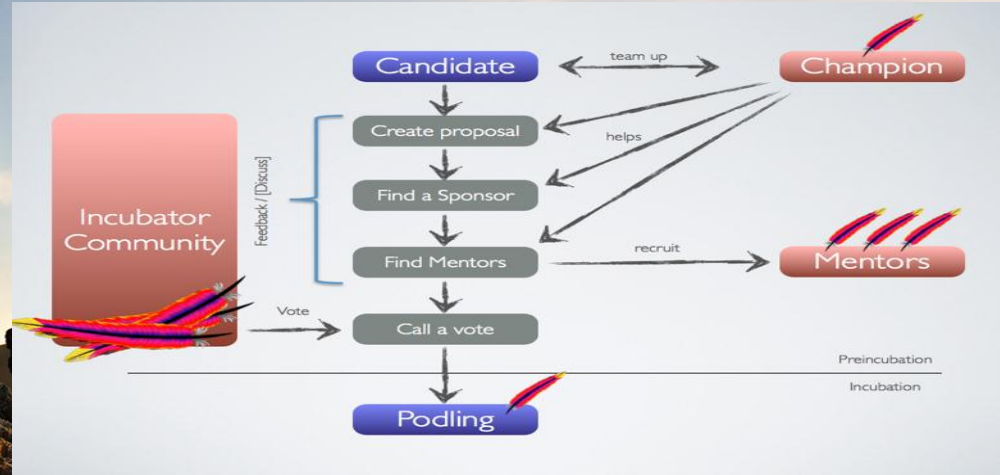
Apache Eagle - Community



Open Source

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

-- African Proverb



Open Sourced By 



Thanks and We are Hiring!