



GOPS 2020  
Shanghai

# GOPS

2020 全球运维大会  
- AIOps 风向标



指导单位：



主办单位：



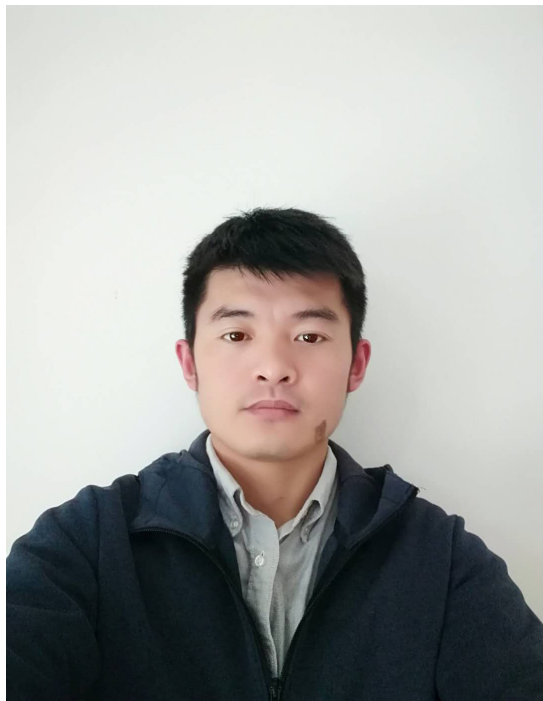
大会时间：2020年11月27日-28日

大会地点：上海中庚聚龙酒店



# 容量管理体系建设思考及实践

叶传伟 运维开发专家



# 叶传伟

## 运维开发专家

2015年加入京东，主要从事京东前台相关工具平台和监控体系的建设；目前主要负责CMDB、智效平台、基础监控、业务监控等相关工具平台和监控系统的持续运营。

CONTENTS

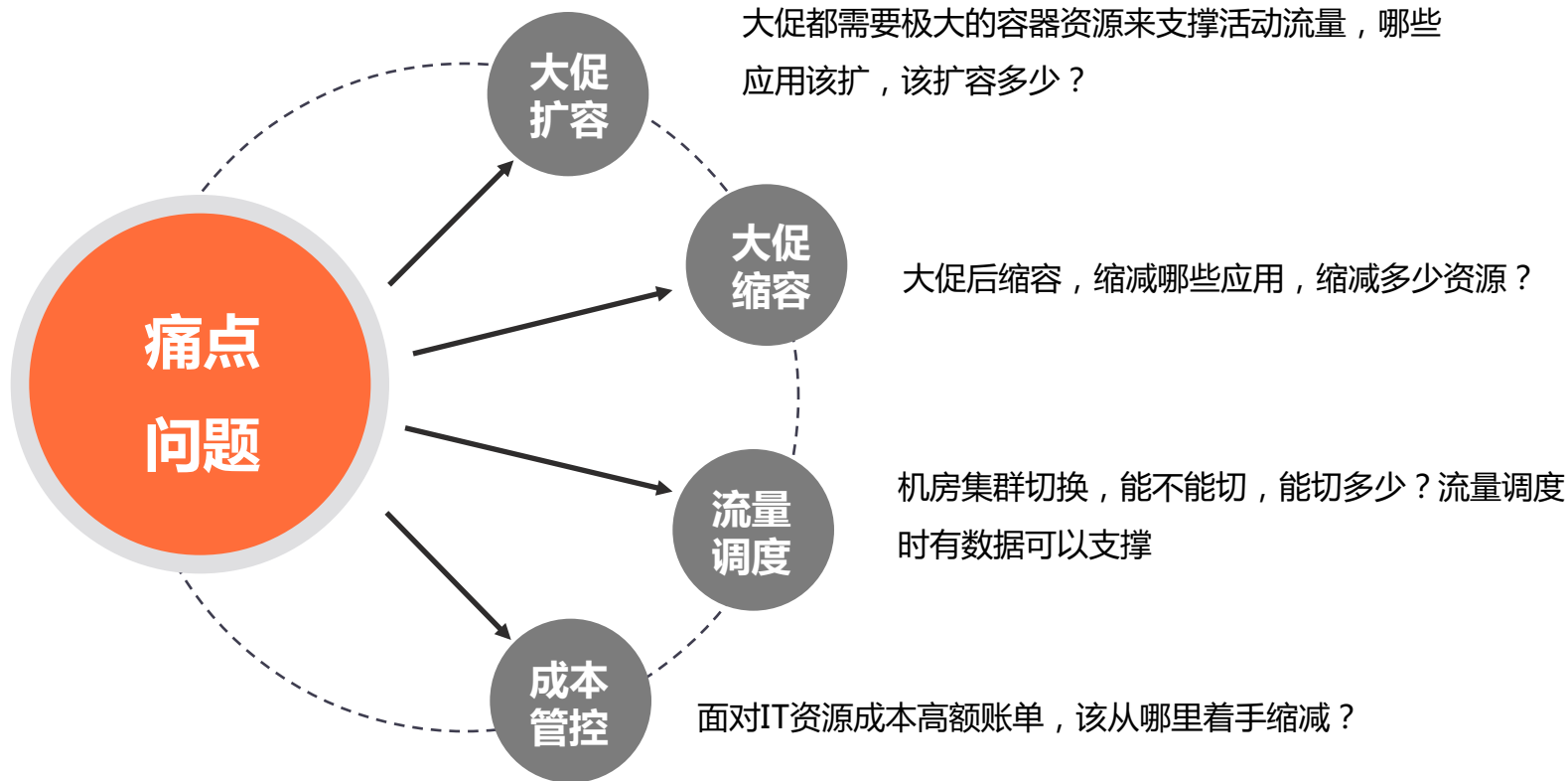
# 目录

- ① 运维痛点&解决之道
- ② 容量管理体系建设思路
- ③ 精细化运营
- ④ 总结与展望



# 运维痛点&解决之道

# 运维痛点问题 – 为什么需要容量管理





# 容量管理体系建设思路

# 容量管理 - 运维立体化监控



容量监控是运维立体化监控体系中一环，并且处于核心位置



# 容量管理 – 建设思路



梳理资源对象



明确度量指标



建立数据模型



自动化工具



数据校对检验

1

梳理运维资源，  
哪些资源需要纳  
入容量管理体系

2

针对不同类型资  
源，采集不同数  
据项，建立度量  
指标

3

针对不同类型资  
源，建立不同数  
据模型，基于实  
时数据和历史数  
据建立模型

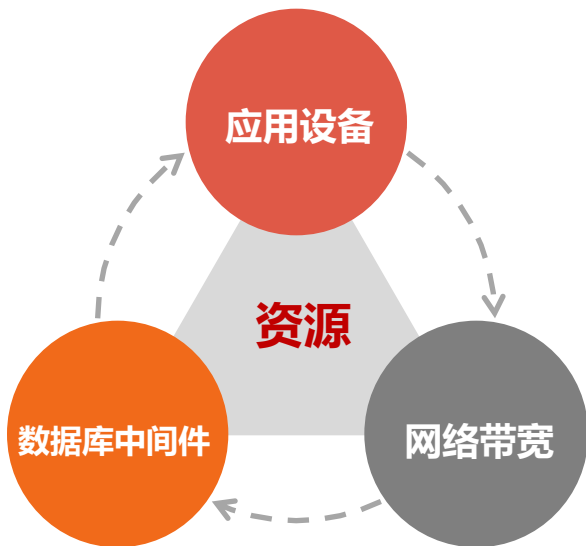
4

及时响应业务问  
题，积累场景，  
沉淀解决方案，  
提升工作效率

5

通过全链路压测  
及大促数据，验  
证数据模型预测  
准确性，优化模  
型，提升预测准  
确率

# 容量管理 - 资源对象和度量指标



## 资源类型

- 一、应用设备：容器、物理机
- 二、数据库：MySQL
- 三、中间件：Jimdb、Elasticsearch、Hbase、Hadoop
- 四、网络流量：域名流量、机房流量

不同资源类型有不同的**度量指标**

# 容量管理 - 基于应用的度量指标



应用度量指标



应用特征分析



数据模型

# 容量管理 - 应用CPU峰值利用率算法

## 根据机器数加权平均

- 1、先平均，再峰值
- 2、应用容器规格不统一，会影响数据准确性
- 3、排除预发布设备
- 4、采样率问题

## 根据CPU核数加权平均

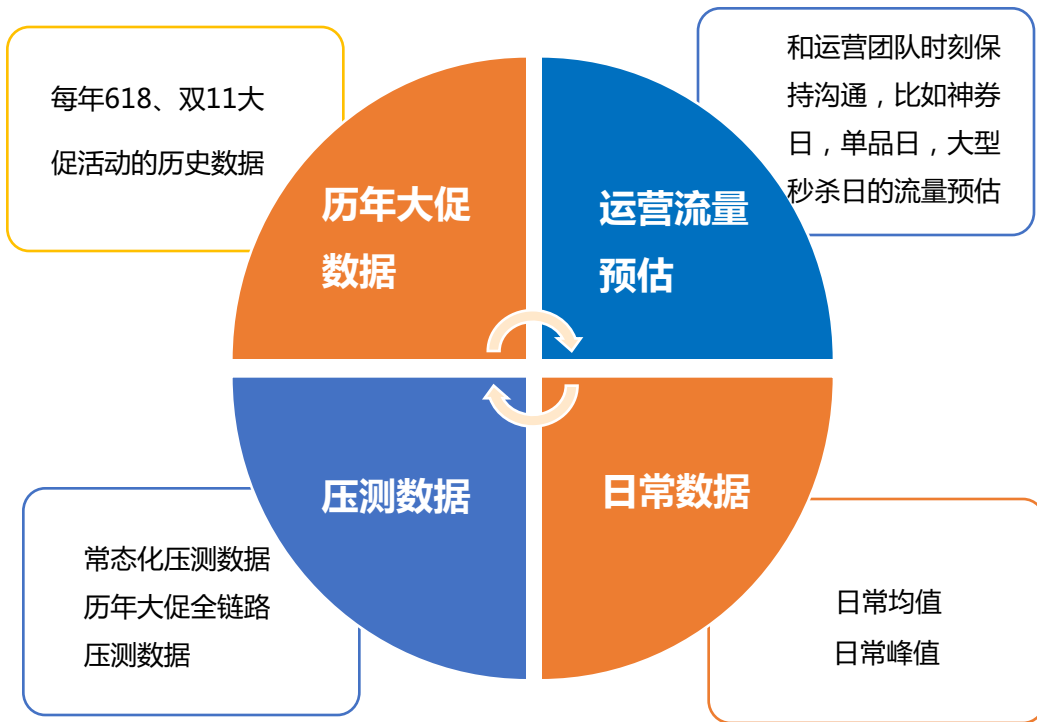
- 1、先平均，再峰值
- 2、应用容器规格不统一，不会影响数据准确性
- 3、排除预发布设备

## 分钟级CPU核数加权平均

- 1、先分钟取峰值，再根据核数取平均，再全天取峰值
- 2、峰值不一定是在同一秒

在基于**应用**的容量管理实践中，我们通过引入**应用峰值利用率**来实现应用的负载度量。

# 容量管理 - 数据



建立容量数据是容量管理中最为基础、重要的一个环节

基于准确的容量数据有利于更有效的做容量预测和规划

数据标准化

数据统一化

# 容量管理 - 容量预测方法

VS

## 1. 凭经验预估容量

- 基于历史数据，做出预测

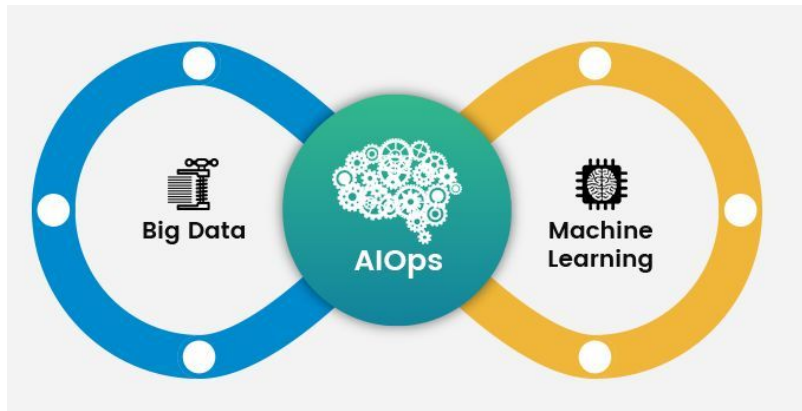
## 2. 按比例扩充

- 基于当前值和目标值的按比例换算

## 3. 压测

- 单机压测
- 全链路压测

常规方式

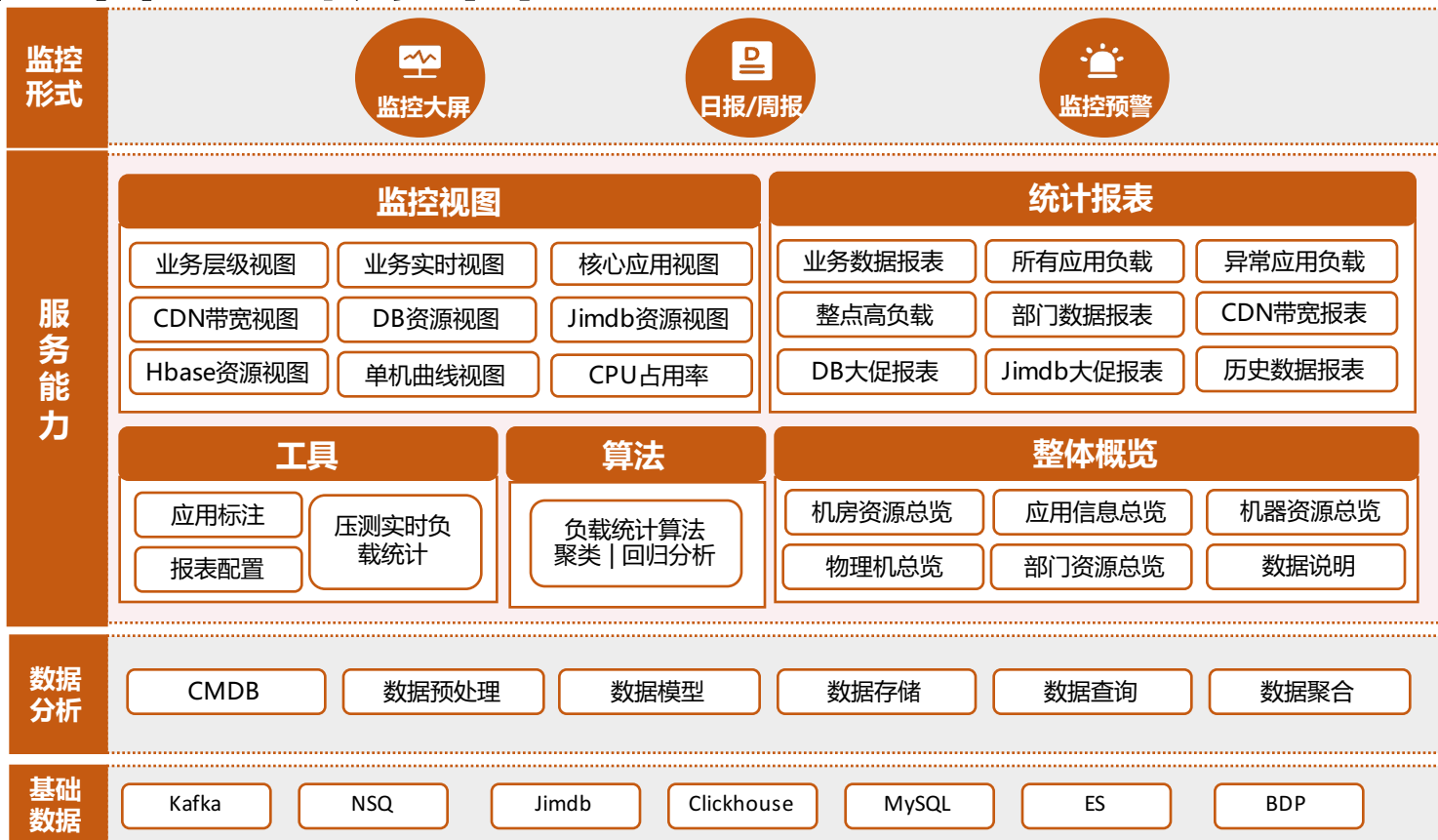


智能预测

# 容量管理 - 智能预测模型



# 容量管理 - 概览图



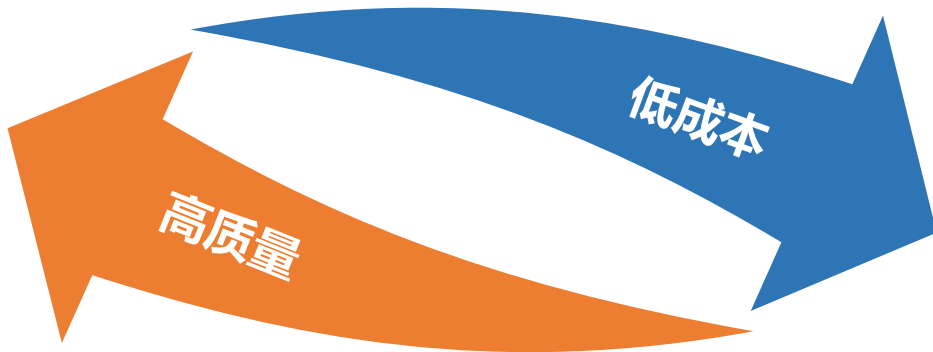




# 精细化运营

# 容量管理核心 - 质量和成本

应用稳定性压倒一切，各方需要确保服务稳定性和高可用，为业务迅速发展保驾护航



业务迅速发展，带来运营成本飞速增加，如何确保不陷入成本失控的困境，一直是困扰各方的难题

## ● 质量和效率

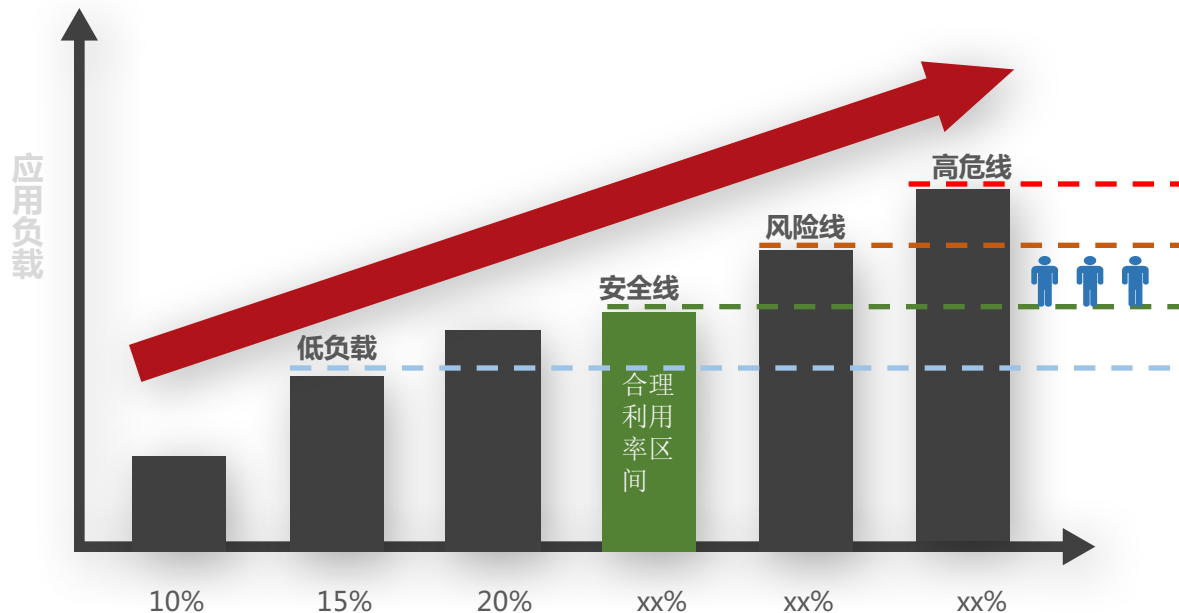
容量管理体系为运维同学跟进应用运营提供数据支持和决策依据，提升应用可用性的同时，提高工作效率。

## 平衡

## ● 成本管控

容量管理体系更进一步的作用在于成本管控，如何在低成本运营前提下，确保服务稳定性，确保服务质量不降低，是运维同学面临的一大难题。

# 容量管理 - 水位线机制



## 追求平衡

显而易见，提高使用率也**不是越高越好**。

应用的稳定性和高可用性为**前提**

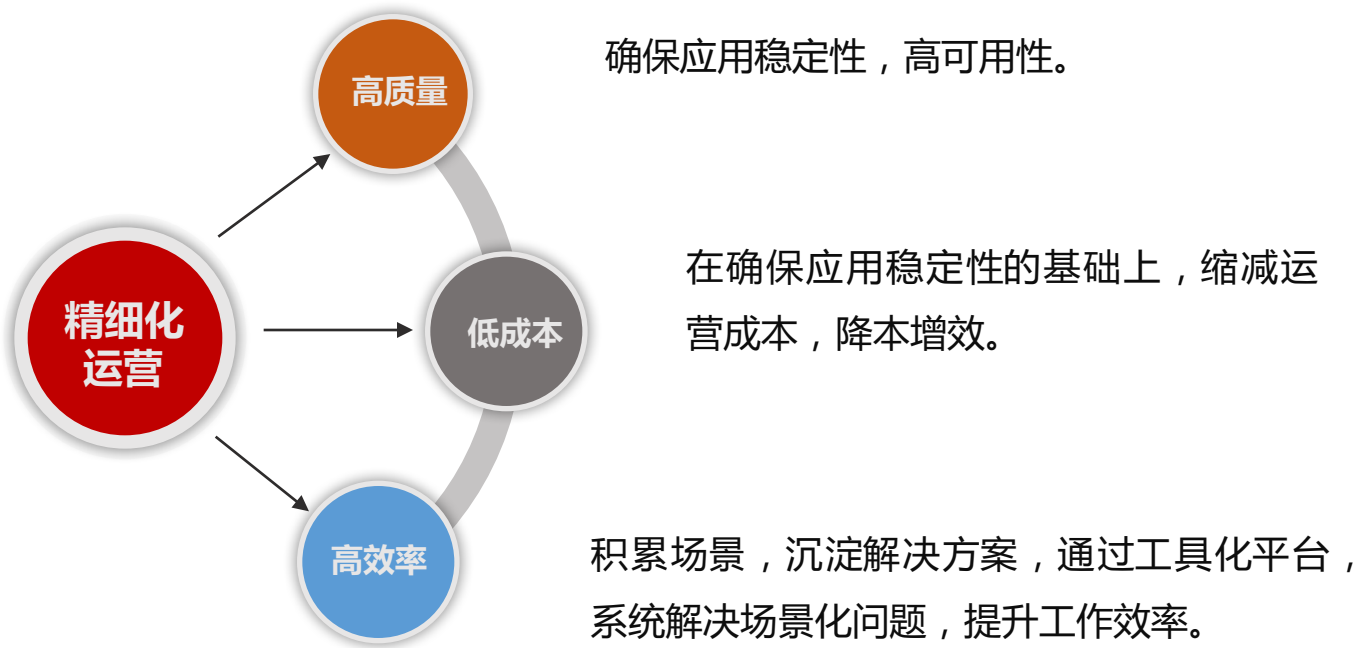
### 水位线

安全线

风险线：紧急扩容

高危线：限流降级

# 容量管理 – 精细化运营

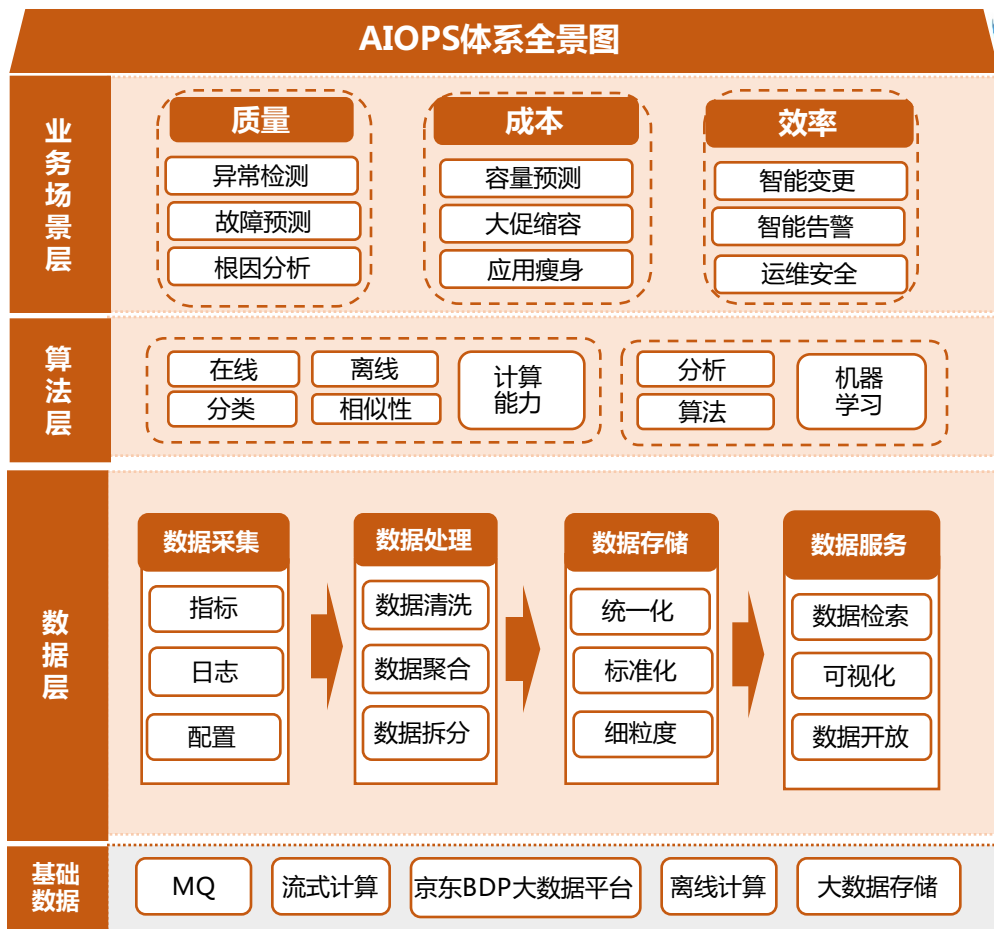
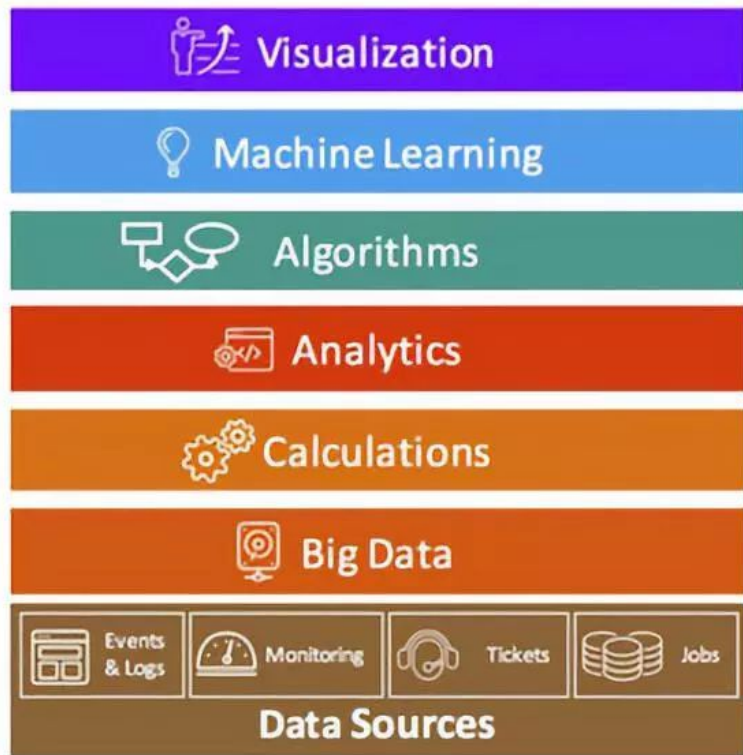


4

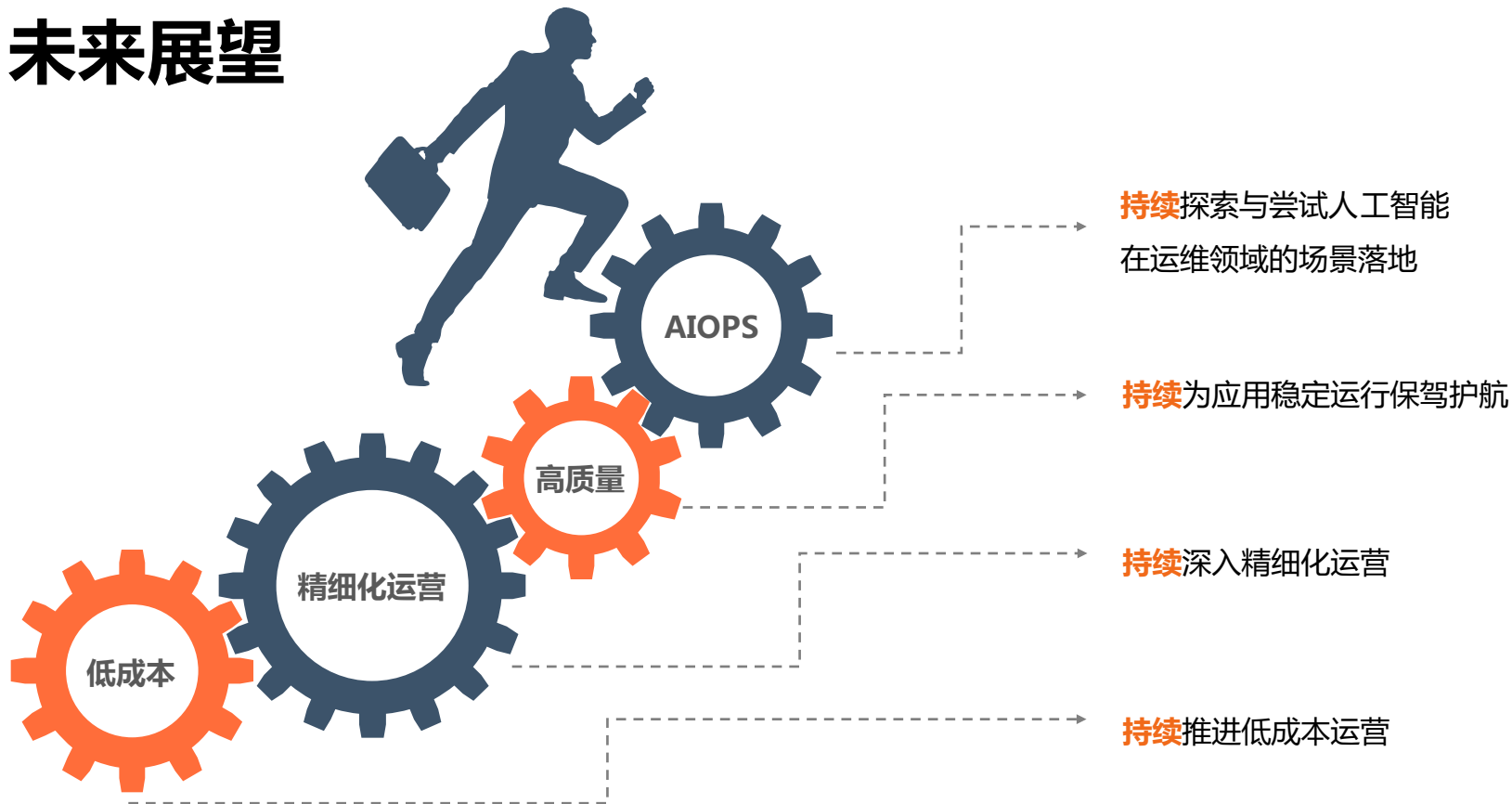


# 总结与展望

# 总结 - 主动运维



# 未来展望



# Thanks