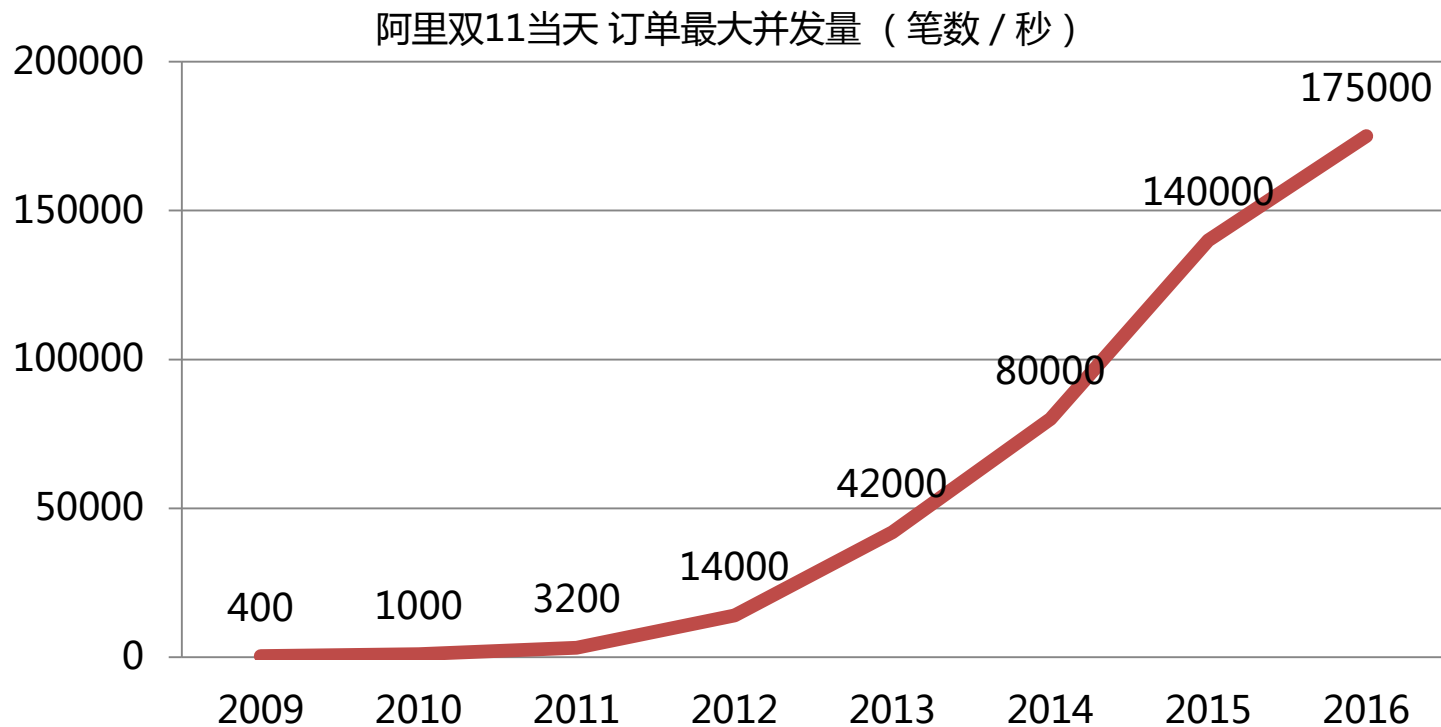


分布式架构与Aliware平台

中间件技术部架构师

平安

阿里技术确保支撑电商的海量访问和处理



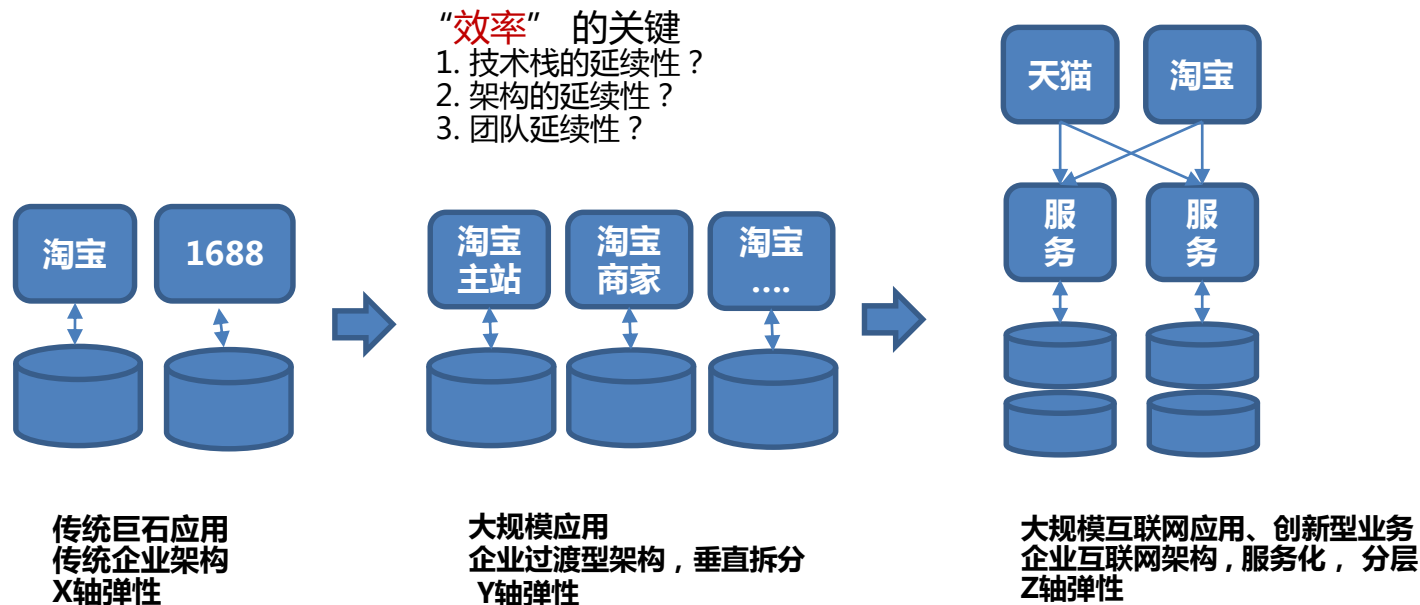
百亿级
商品

1207亿
成交额

3万亿
GMV

系统压力，7年加大了400+倍，系统越来越稳定！

系统架构演进路径 - - - “Scale Cube” 适应于不同业务阶段



分布式架构下挑战

技术团队的奇怪现象

- 线上故障，击鼓传花找背锅侠
- 救火队规模快速扩充，包含了所有业务的核心成员
- 技术创新百花齐放，造出了各种轮子求晋升
- 应对大促营销活动，技术团队长期通宵达旦，老大一直提心吊胆
- 业务没单点了，但是团队有单点
- 人手永远不够, boss和team各说各话
-

分布式架构下常见的技术问题

- 问题排查，快速定位并解决？
- 业务大促和营销活动的怎么做容量规划和准备系统资源？
- 服务化后分布式事务问题？
- 关系数据库如何也具备可运营的扩缩能力？
- 业务监控告警怎么能真正有效？
- 大量开源框架，怎么选没有错？
- 服务治理、集群管理、灰度发布、无感知升级、等等？

解决问题要对症下药，构筑有力的PAAS层能力

■ 治理能力是分布式架构下核心问题

- 别被框架、协议、性能等遮住眼
- 整体综合效率提升避免木桶效应
- 请关注每个团队都在投入资源解决的非业务问题

■ 服务运营能力团队价值最直接体现

- 每个服务状态owner应该最清楚, 不要等别人来报故障
- 营销活动支持应该工具化, 配置化, 常态化
- 肯定会犯错, 肯定会有bug, 你的系统做好准备没有?

■ 平台化服务化是提高整体效率的有效途径

- 专事专干, 不要把有限的团队资源用到无限的造轮子上
- 业务域和技术域隔离, 不要一团麻
- 业务服务化、技术服务化、运营服务化

从共享业务事业部到大中台的顶层架构



飞天平台：不仅仅是云计算，是云计算，大数据，云安全的统一平台。

企业级互联网架构平台：实现业务能力云化的基础，支撑业务微服务化共享。

共享服务层：对集团业务的服务化抽象共享运营，实现业务能力共享。

业务层面特征：

- 创新快，敏捷
- 数据实时打通无孤岛
- 高并发，线性扩展，
- 所有业务同架构大平台
- 高可靠，无单点

看看淘宝的大促备战技术演进



全链路压测挑战

➤ 业务范围很复杂、业务量级很大

海量数据！1000w/s请求,剁手党们疯狂的热情！每秒100GB的网络带宽！
每一条业务链路的压测可行性需要保障（系统改造）

➤ 测试流量与正常流量隔离，不影响正常业务和报表，预防脏数据

不能影响bi报表、推荐算法等等

不能影响正常的业务监控和报表

不产生脏数据

➤ 贴近大促业务场景，保证压测结果的准确性和可参考性

业务数据量、业务模型、转化率、流量模型等需与大促场景类似

流量的发起地需要真实

协助众多的政府和大中型企业进行互联网转型

在政府，能源，制造，金融，商贸，媒体，电信等行业，阿里企业级互联网架构及中间件已成为关键支撑

政府部
委：



广东省人民政府



广东省气象局
Guangdong Meteorological Service



央企：



国家电网
STATE GRID
国网浙江省电力公司
STATE GRID ZHEJIANG ELECTRIC POWER COMPANY



CHEMCHINA
中国化工集团公司
China National Chemical Corporation



中国航空工业集团公司
AVIC AVIATION INDUSTRY CORPORATION OF CHINA



工业制
造：



金融：



中信集团



物流行
业：



大型分布式架构应该具备的PaaS能力

• 开发

持续开发
持续测试
持续集成
持续交付

• 运营

集群管理

- > 自动弹性
- > 无感知发布

数据化能力

- > 链路分析
- > 依赖可视
- > 能力数字化

实时监控告警

- > 系统、容器、业务

稳定性保障

- > 限流降级
- > 故障演练

容量评估

- > 容量评估
- > 线上实时压测

系统实时扩缩

- > 业务扩缩
- > 数据库扩缩

• 迭代

灰度发布
AB测试
故障预案

融入了阿里架构精华的互联网中间件---Aliware

阿里分布式
应用服务器
EDAS



分布式服务框架
数据化运营
自动化运维
线性扩展

阿里分布式
数据库服务
DRDS



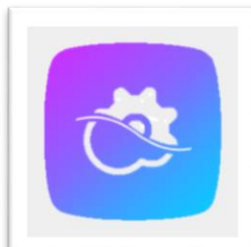
高可用
自动化
线性扩展

阿里分布式
消息服务
MQ



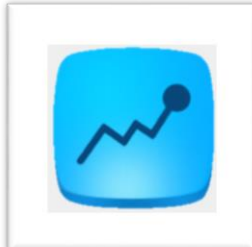
异步化
最终一致
线性扩展

阿里能力
开放平台
CSB



能力开放
服务管控与治理
异构系统整合

阿里实时
监控大盘
ARMS



流式计算引擎
实时业务看盘

阿里全局
事务服务
GTS



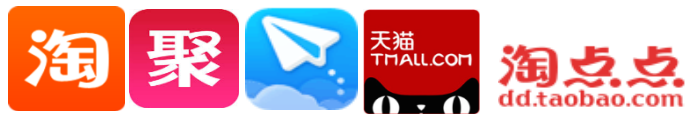
高性能分布式
事务支持

阿里云效
开发平台
OPS

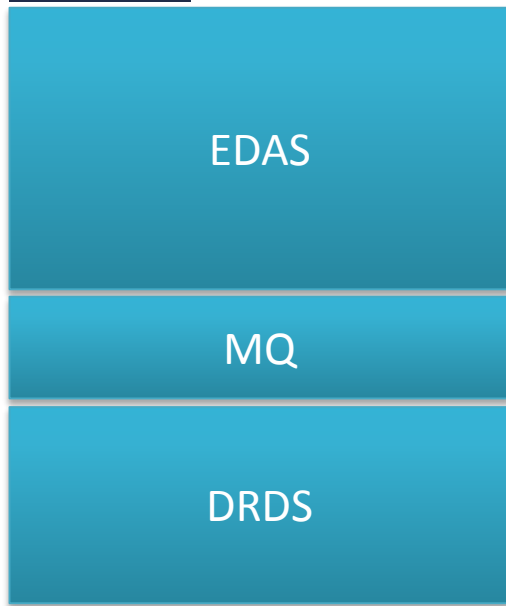


项目管理
自动化测试
持续集成
持续交付

通过阿里云输出产品与能力, 架构统一



弹内（内部集群）



弹外（公有云集群）

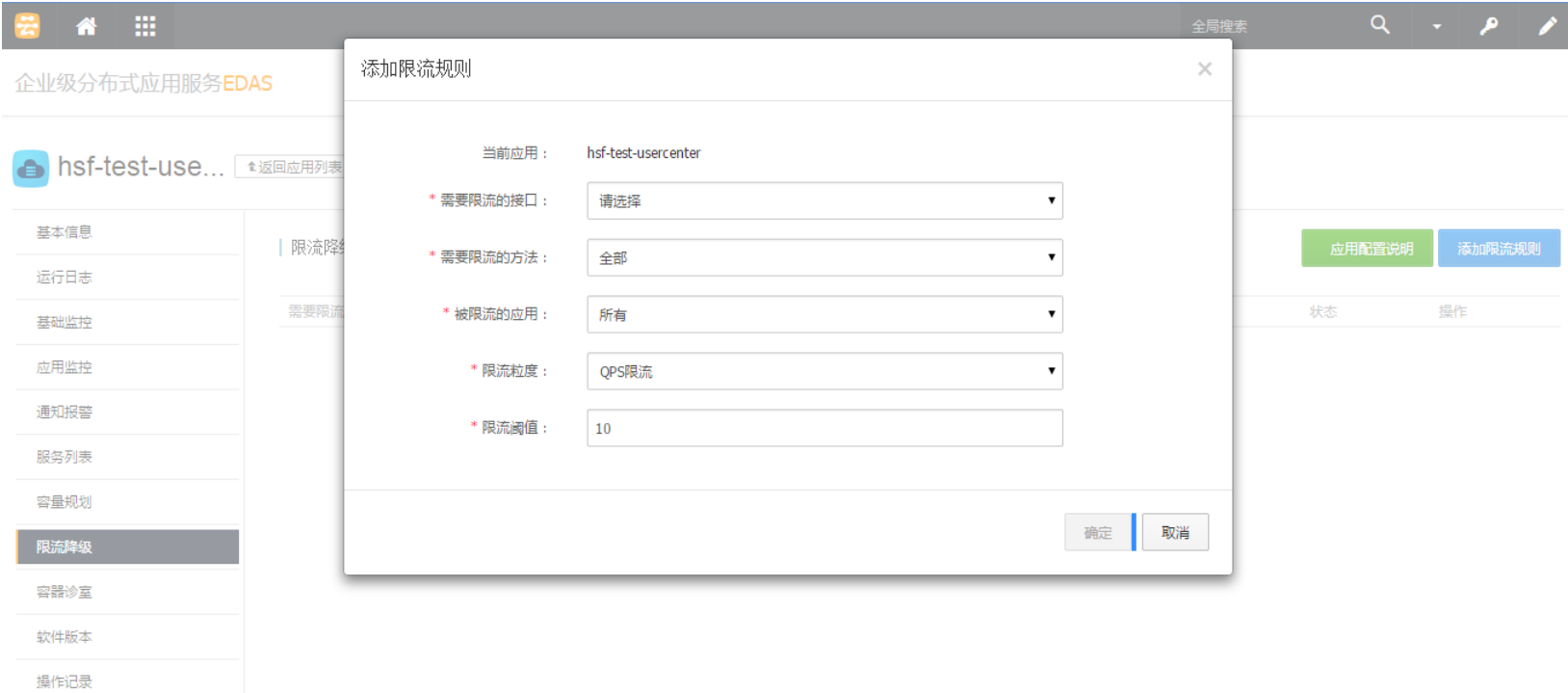
EDAS提供了微服务和应用运维管控所需的完整功能

-- Enterprise Distributed Application Service



服务综合治理——限流降级

- 基于阿里稳定性平台服务，基于单机QPS和并发数限流，在洪峰到来时保证应用系统的可用性

- The screenshot displays the Alibaba Cloud EDAS (Enterprise Distributed Application Service) console. A modal dialog titled '添加限流规则' (Add Rate Limiting Rule) is open over the 'hsf-test-usercenter' application configuration page. The dialog contains the following fields:
 - 当前应用 (Current Application): hsf-test-usercenter
 - * 需要限流的接口 (Interface to be limited): 请选择 (Please select)
 - * 需要限流的方法 (Method to be limited): 全部 (All)
 - * 被限流的应用 (Application to be limited): 所有 (All)
 - * 限流粒度 (Rate limiting granularity): QPS限流 (QPS Limiting)
 - * 限流阈值 (Rate limiting threshold): 10At the bottom right of the dialog are '确定' (Confirm) and '取消' (Cancel) buttons. The background console page shows a sidebar with navigation items like '基本信息', '运行日志', '基础监控', '应用监控', '通知报警', '服务列表', '容量规划', '限流降级' (highlighted), '容器沙盒', '软件版本', and '操作记录'. The main content area of the console shows the 'hsf-test-usercenter' application with a '返回应用列表' (Return to application list) link and buttons for '应用配置说明' (Application configuration instructions) and '添加限流规则' (Add rate limiting rule).

服务综合治理——容量规划

- 在真实线上环境基础上，通过调整服务器权重，真实模拟压测情况，评估单机最大服务能力，提供吞吐能力数据以供性能优化参考
- 容量评估：自动计算前端关键请求与后端机器数量的对应关系，对加减机器进行预测

企业级分布式应用服务EDAS

资源管理

应用管理

服务治理

链路分析

账号管理

hsf-test-use... 返回应用列表

基本信息

运行日志

基础监控

应用监控

通知报警

服务列表

容量规划

压测配置

压测结果

容量数据

限流降级

容灾诊室

压测配置

启动压测

修改

删除

压测目标ECS实例(实例ID/名称/IP):

i-287lvbz7b/iz287lvbz7bZ/10.144.3.214

手动控制:

权重倍数 ?

第1分钟

2

第3分钟

3

第5分钟

4

第7分钟

5

第9分钟

6

压测限制 ?

CPU:

80

%

Load:

10

QPS:

RT:

50

压测接口:

接口名称

版本

服务综合治理——弹性伸缩

- 根据cpu, load, rt三个指标做应用的自动扩容或缩容

扩容规则: ☐

触发指标:

CPU \geq 1 %

RT \geq 1 ms

Load \geq 1

触发条件:

任一指标

持续时间超过:

1 分钟

每次扩容的实例数:

1 台

最大实例数:

3 台

缩容规则: ☐

触发条件:

CPU $<$ 10 %

RT $<$ 10 ms

Load $<$ 10

触发条件:

任一指标

持续时间超过:

1 分钟

每次缩容的实例数:

1 台

最小实例数:

1 台

数据化运营——鹰眼监控

- 基于阿里鹰眼监控平台，提供应用响应时间和吞吐量信息，并提供全链路分析功能，找出系统热点和瓶颈

2013-07-20 05:25:28.179 ERROR taobao.hsf -
基于 RPC 协议调用服务
[com.taobao.wireless.trade.api.tmall.hsf.TmallBagInterface:1.0.0]的
[buildConfirmOrder]方法时出现错误：

所调用的服务目标地址为：[...]
参数信息为：[...] Traceld=ac18287913742691251746923
错误原因为超时，请查看服务器端的执行日志是否也超时，执行时间为：3000
毫秒。

HSFTimeoutException
at
com.taobao.hsf.....HSFResponseFuture.getResponse(HSFResponseFuture.java:52)
at
com.taobao.hsf.....SyncInvokeComponent.invoke(SyncInvokeComponent.java:51)
at ...

调用链ID: ac18287913742691251746923 时间: 2013-07-20 05:25:25.174, 调用链总时长: 16s262ms 日志原文

	类型	状态	大小	服务/方法	
mtop	TRACE	OK	-	http://api.m.taobao.com/rest/api3.do	3s8ms
wdc	HSF	OK	8.5KB		3ms
⊞ sirius	HSF	TIMEOUT	6.9KB	wireless.TmallBagInterface@buildConfirmOrder~P	3s1ms
(tair@	TAIR	NOTEXST	65B		1ms
(tair@	TAIR	OK	57B		0ms
⊞ buyaj	HSF	OK	11.6KB		3s143ms
cart	HSF	OK	2.4KB		3ms
deli	HSF	OK	844B		1ms
trac	HSF	OK	1.3KB		2ms
⊞ invc	HSF	OK	6.1KB		3ms
⊞ invc	HSF	OK	8.9KB		3ms
⊞ invc	HSF	OK	5.8KB		3ms
⊞ deli	HSF	OK	6.5KB		3ms
⊞ deli	HSF	OK	6.2KB		13ms
⊞ deli	HSF	TIMEOUT	6.2KB	delivery.DeliveryTradeService@getItemSupportPost~LL	3s9ms
trac	TAIR	CONNERR	-	GET:group_1:214	3s9ms
trac	HSF	OK	727B		1ms
⊞ logi	HSF	OK	805B		3ms
umj	HSF	OK	13.6KB		16ms
⊞ deli	HSF	OK	11.4KB		15ms
trac	HSF	OK	9.4KB		21ms
trac	HSF	OK	705B		2ms
trac	HSF	OK	815B		2ms

- ☆ 完整记录所有故障
- ☆ 准确定位故障源

EDAS链路分析

海量调用链进行统计，得到链路各个依赖的稳定性指标

层次	名称	QPS	峰值 QPS	调用 比例	被调用 均值	平均 耗时	耗时 比例	出错 率	同机房	标记
0	http://	82.66	8700	1.0000	1.0	312ms	16.52%	0.00%	0.0%	瓶颈
1		1.8	34450	0.0221	11.07	0ms	0.01%	0.00%	100.0%	
1		439.22	22200	5.3138	5.45	0ms	0.86%	0.02%	99.98%	
1		0.21	3660	0.0025	1.02	2ms	0.00%	0.00%	100.0%	
2		0.21	3660	0.0025	1.02	0ms	0.00%	0.00%	100.0%	
2		0.13	20	0.0016	1.0	1ms	0.00%	0.00%	100.0%	
1		80.61	8480	0.9752	1.0	5ms	1.45%	1.35%	100.0%	
2		0.0	130	0.0000	1.0	10ms	0.00%	0.00%	98.0%	
2		0.01	190	0.0001	2.13	10ms	0.00%	0.00%	81.25%	
2		0.01	130	0.0001	2.27	0ms	0.00%	0.00%	100.0%	
2		0.01	190	0.0001	2.13	0ms	0.00%	0.00%	100.0%	
1		79.45	8440	0.9612	1.0	0ms	0.08%	0.00%	100.0%	
1		0.85	60	0.0103	1.09	5ms	0.01%	0.01%	100.0%	强依赖 错误阻塞
1		0.15	520	0.0018	1.0	107ms	54.29%	0.00%	100.0%	瓶颈
1		0.08	30	0.0010	1.0	2ms	0.00%	0.00%	100.0%	
1		0.15	520	0.0018	1.0	0ms	0.00%	0.00%	100.0%	

对依赖的压力

易故障点

瓶颈点

DRDS让关系数据库具备完全线性扩缩能力

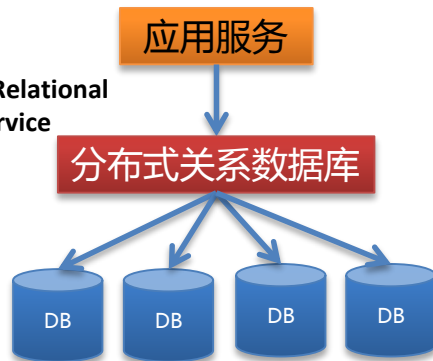
数据库不再成为性能瓶颈

- ✓ 由廉价硬件构成的关系数据库服务集群
- ✓ 可支持的TPS/QPS取决于机器性能和机器数
- ✓ 持久化存储的数据量取决于磁盘容量

对应用程序无要求

- ✓ 与MySQL协议 100% 兼容
- ✓ 支持JAVA, Lua, PYTHON, RUBY等任意语言
- ✓ 支持JDBC, myBatis, Hibernate等持久化框架

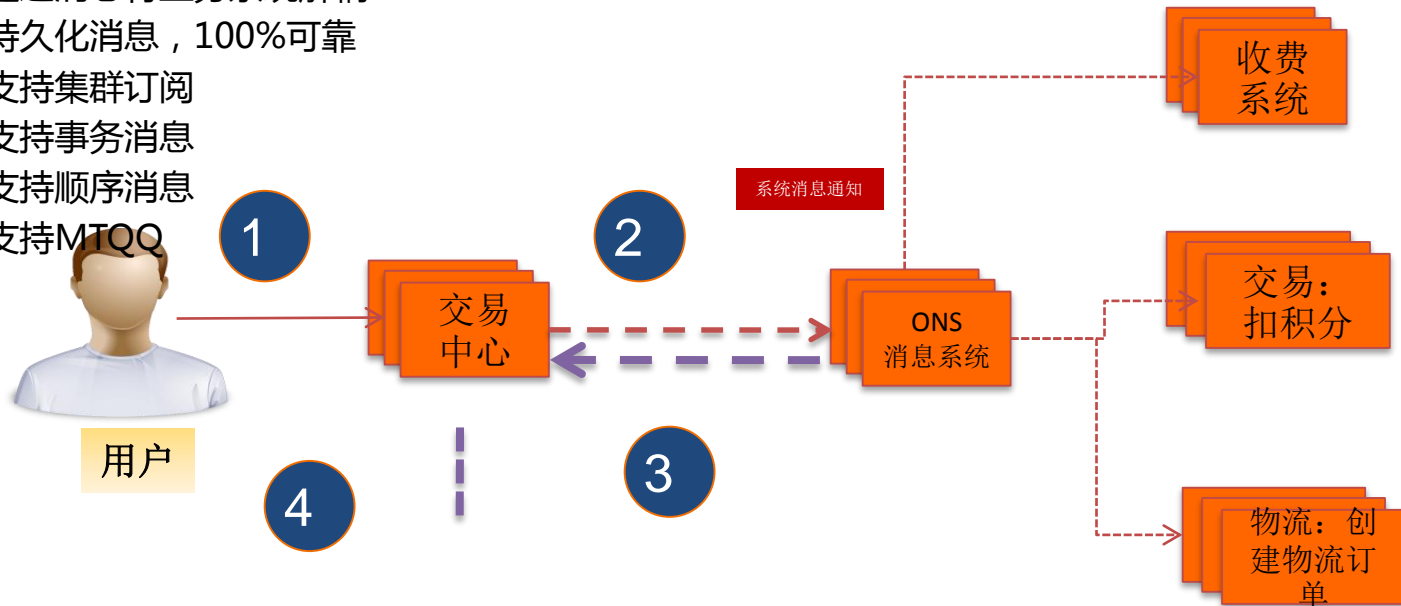
DRDS -
Distributed Relational
Database Service



消息系统(MQ)

- 分布式消息通知系统 (Notify/MQS)

- 通过消息将业务系统解耦
- 持久化消息，100%可靠
- 支持集群订阅
- 支持事务消息
- 支持顺序消息
- 支持MTQQ



交易系统使用ONS的案例

消息系统(MQ)

MQ特点

4种消息类型

普通消息、定时消息、事务消息、顺序消息

3种消费方式

HTTP、MQTT、SDK(JAVA/C++/.NET PHP)

消息
丰富

管理
多维

性能
优越

服务
健壮

百亿级堆积能力

单机TPS可达10W

毫秒级投递延迟

支持万级节点高并发

高性能集群真正水平扩展

消息按ID\KEY\TAG查询

回溯重新消费3天内消息

消息轨迹跟踪

监控报警机制

99.99%数据可靠性

99.9%服务可用性

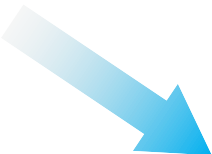
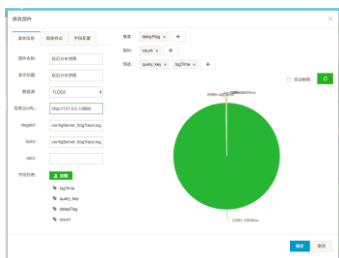
消费失败定时重试

阿里多次双十一的真实场景考验

消息防篡改，加密

实时应用监控大盘(ARMS)

基于数据魔方的报表大盘和报表配置:



报表:



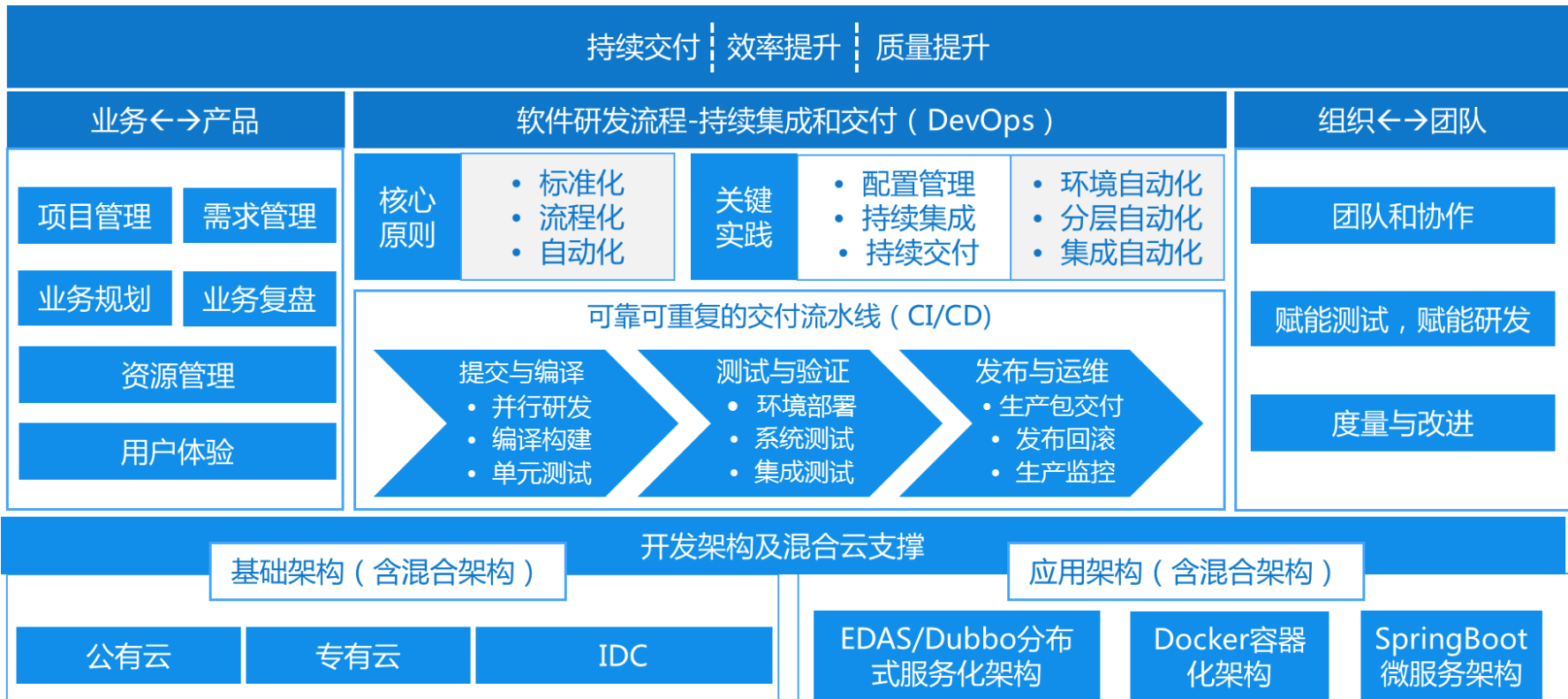
大盘:

告警:



城市	类目	销量 - 女性	销量 - 男性	收入	成本	利润	利润 - 环比
上海	学习	1,851	1,567	CNY19,887.24	CNY12,877.64	CNY7,009.60	2.71%
	电器	1,494	1,700	CNY18,399.60	CNY11,385.87	CNY6,413.73	-2.19%
运动	食品	1,616	1,809	CNY19,736.87	CNY12,968.46	CNY6,768.41	-20.27%
	学习	1,634	1,770	CNY18,943.55	CNY12,293.20	CNY6,650.35	-6.32%
电器	运动	1,403	1,742	CNY19,805.17	CNY12,743.55	CNY7,061.61	-4.61%
	运动	1,401	1,579	CNY19,038.63	CNY12,354.93	CNY6,683.70	-3.14%

云效价值交付



Thank you!

花名： 平安

TEL : 18858189249

