

EMQ

「面向物联网的数据基础设施」

白皮书

www.emqx.com

目录

前言	3
物联网时代数字化转型的技术趋势	4
云计算从云转型到云原生	4
从云计算到边缘计算	5
物联网时代数据处理和分析的范式转变	6
生产主体：从人类行为到物理状态感知与记录	6
产生方式：从有波峰波谷到稳定持续产生	7
数据类型：从交易事务型表数据到实时分析型流数据	7
数据体量：从 TB 时代到 PB/EB 时代	7
处理方式：从批处理到实时流式计算	7
架构模型：从 Schema on Write 到 Schema on Read	8
从云原生到面向物联网的架构设计	9
面向物联网 (IoT-Oriented) 架构设计原则	9
面向物联网 (IoT-Oriented) 架构的功能特点	10
面向物联网架构的数据基础设施	13
架构范式	13
连接 — 弹性可靠多协议，连接海量物联网设备	13
移动 — 实时消息引擎，双向数据移动与分发	14
存储 — 低延时、动态扩展的云原生流数据存储	15
处理 — 云边协同的多级数据一站式实时处理	15
分析 — 实时数据分析与业务洞察，即刻决策	16
Data Infrastructure for IoT 架构范式下的产品组合	17
EMQ X — 云原生分布式 MQTT 消息服务器	17
HStreamDB — 云原生分布式实时流处理数据库	17
NanoMQ — 超轻量边缘 MQTT 消息服务器	18
Neuron — 物联网边缘工业协议网关软件	18
eKuiper — 轻量边缘流式处理引擎	18
结语	19

前言

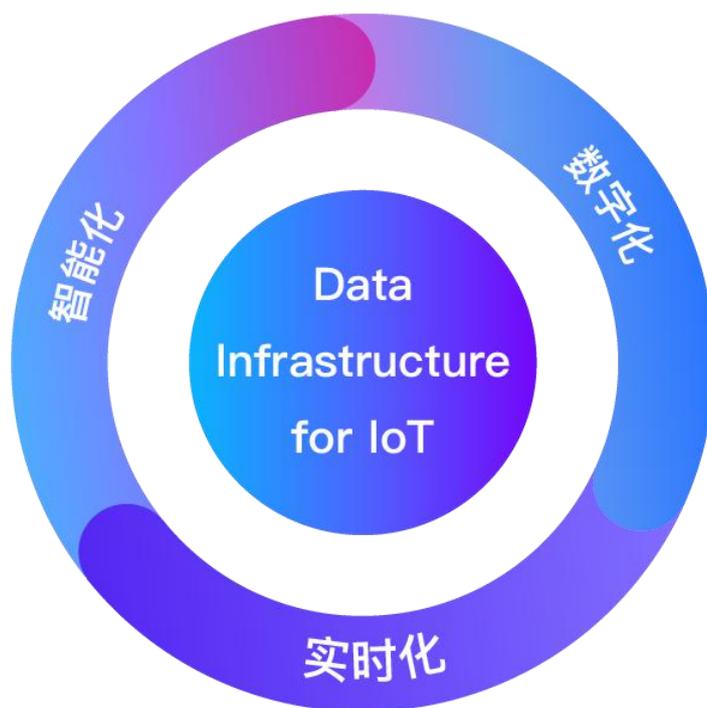
现在，是时候思考物联网时代的企业关键业务挑战与机遇。

在过去的近十年中，EMQ 一直围绕物联网的「连接」与「数据」开发开源基础设施软件。如今，EMQ X 中间件已解决了「海量连接」的挑战，HStreamDB 流数据库则正试图解决海量物联网数据的存储、处理与实时分析。

在多年为企业交付物联网基础软件过程中，我们发现企业大多以场景优先的方式构建单一物联网应用。

EMQ 始终认为，物联网场景与应用的核心逻辑是以「数据」为中心，从数据获取到业务洞察，从而创造价值。特别是中大型企业，更应以「数据」为中心，从「数据」出发思考物联网业务，在融合物联网数据与传统的企业经营数据基础上，灵活构建多样化的创新型业务，以实现企业的数字化、实时化、智能化转型。

在本白皮书中，EMQ 正式提出「面向物联网的数据基础设施」架构范式，与产业一起迎接物联网时代企业的关键业务挑战，实现企业业务创新与价值创造。



物联网时代数字化转型的技术趋势

随着物联网技术的逐步成熟，海量设备产生的数据为企业数字化提供了更加丰富的数据基础。如何将海量数据与企业生产管理系统、业务应用系统以及运维监控系统有机结合，是企业在未来实现差异化竞争与业务创新的关键。在以物联网数据为基础的数字化转型过程中，云原生、分布式云等新技术成为催化剂，并呈现以下新的趋势：

云计算从云转型到云原生

云的本质

定义：云计算是一种模型

模型主体如下：大量计算机构成的资源池；动态伸缩的廉价计算服务

理念：使用比拥有更重要

云转型

在数字化转型过程中，云计算可以为企业降低成本、提升效益，为企业业务创新以及 AI 等新兴技术的应用创新提供有力支撑，还能通过虚拟化技术降低其对硬件设备的要求，提升基础设施的自主可控水平。

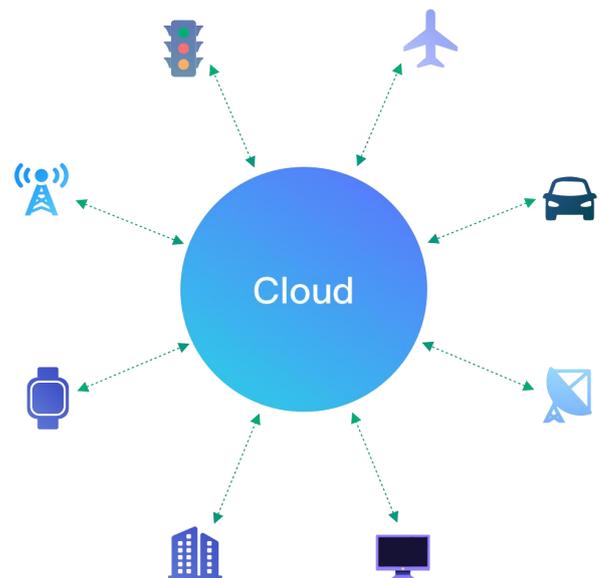
云转型强调资源优化，即数据与应用上云，充分享受云的虚拟化带来的资源集约优势。

云原生

如果说企业上云是将云计算当作工具，开始利用云获取价值，那么云原生时代的来临，则意味着云的理念开始深入企业业务的核心逻辑，真正成为企业赖以提升效率的必需途径。

云原生强调效能提升，即基于云模型重新设计基础设施软件架构以及应用开发模式，从而革命性地提高数字化效能，真正享受云计算红利。其具有极致弹性、服务自治、故障自愈、规模复制等特点。

由于过去所依赖的基础架构软件的系统设计的边界被打破，可以在全新的 Tradeoff 下完成设计，因而诞生了从软件产品到开发方式的一系列变革。如 Docker 与 Kubernetes、DevOps 理念等。



从云计算到边缘计算

边缘计算的出现

传统的云计算模型将所有数据通过网络上传至云计算数据中心，利用云计算中心的超强算力来集中解决应用的计算需求。然而，在万物互联时代，云计算的集中处理模式有三方面的不足：

- **实时性需求与网络覆盖：**随着边缘设备数量增加，设备产生的数据量持续激增，导致网络带宽和延迟逐渐成为云计算的瓶颈。保证全场景网络覆盖的高昂成本也迫使企业下放一部分计算到边缘。
- **数据安全与隐私：**随着数据采集设备的普及，直接将采集的数据上传至云计算中心会增加泄露用户隐私数据或企业核心数据资产的风险。大量搜集客户群体画像信息上云也带来个人隐私泄露问题，与隐私相关的数据将更多在边缘分析。
- **能耗较大：**随着云服务器运行的应用程序越来越多，未来大规模数据中心对能耗的需求将难以满足。提高能耗使用效率本身并不解决根本问题，万物互联时代问题更加突出。

边缘计算模型中，边缘设备具有执行计算和数据分析的处理能力，将原有云计算模型执行的部分或全部计算任务迁移到网络边缘设备上，降低云服务器的计算负载，减缓网络带宽的压力，提高万物互联时代的数据处理效率。

边缘计算的挑战

边缘计算解决了诸多云计算集中化处理带来的问题，但也同时衍生了新的挑战：

- **云边协同难：**边缘计算在延迟和可靠性等方面的优势明显，但当企业试图全局管控所有边缘节点时，云所擅长的地方边缘计算大多无法满足，且两者协同工作的运维复杂度指数级增加。
- **边缘设备互操作难：**过去的云上业务可通过微服务化或者服务网格实现业务间或者 API 间的互操作，但 IoT 产业数据基础设施大多数是纵向数据而非横向数据连接，边缘设备之间的互操作困难重重，极大的制约了边缘计算的应用开发潜力。



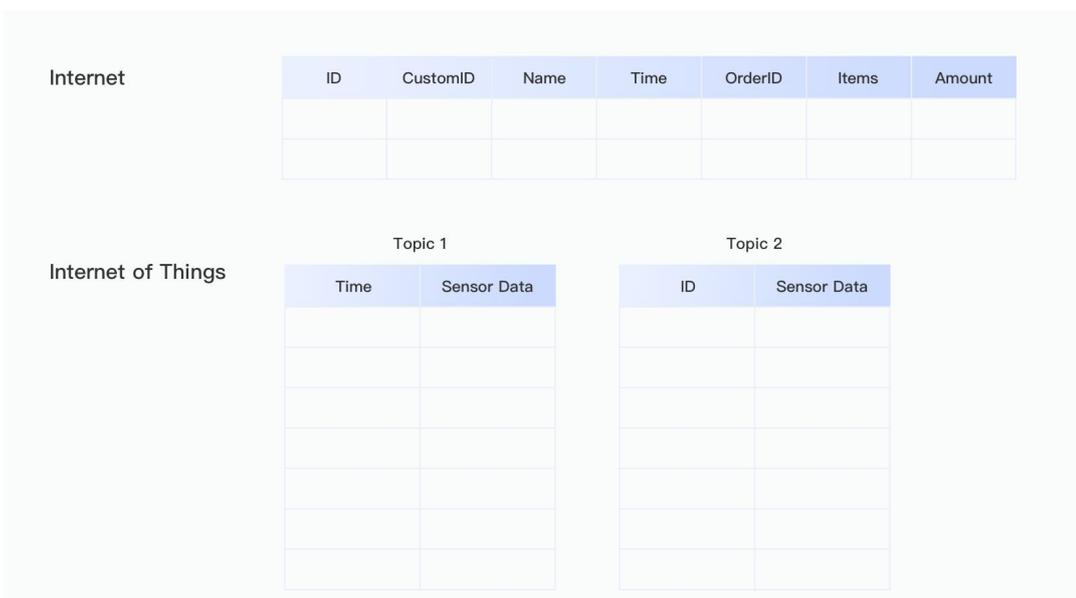
物联网时代数据处理和分析的范式转变

生产主体：从人类行为到物理状态感知与记录

互联网时代，数据产生的源头是人类基于 PC 机操作的行为和触发的事件等相关的记录，该类数据范式与交易型数据非常类似。

移动互联网时代，由于人类开始使用手机，使用时长相比 PC 机时代大幅增加，数据源自基于手机的操作行为和触发的事件等相关的记录；不仅如此，在手机上的多个传感器开始源源不断的产生时序数据，催生消费物联网的第一批应用。

随着企业数字化转型与数据驱动转型的深化，越来越多的企业开始基于可连接传感器设备实现企业全生命周期的数据化，产业物联网正式崛起。



当数据生产主体从人类转移到机器时，数据基础设施的技术架构思路也会发生相应变化。主要变化如下：

- 从行存储到列存储到行列混排与向量化计算；
- 从应对中低频间断产生型数据到应对高频持续产生型数据；
- 从 Schema on Write 到规则引擎到复杂函数式流计算。