

MongoDB 与 IoT



mongoDB

目录

IOT的发展趋势和挑战

MongoDB如何应对挑战

MongoDB成功客户案例

下一步

目录

IOT的发展趋势和挑战

MongoDB如何应对挑战

MongoDB成功客户案例

下一步

IoT 是当前CEO关注的重点领域之一

根据IDC对全球4500个组织和PwC对亚太区各行业CEO的调研

1. **56%**的被访者将IoT作为业务战略，亚太区**67%**CEO希望将IoT用于至少1个到3个业务上
2. 提高生产率，缩短上市时间和实现流程自动化是三大主要原因
3. **31.4%**的组织已经启动了IoT项目，另外**43%**将在未来12月内启动
4. 工业4.0加快工业企业部署IoT，亚太区**64%**的运输企业，**54%**制造业已经部署IoT
5. 针对IoT收益，工业最关注生产率和供应链，金融和消费业最关注用户体验和运营效率



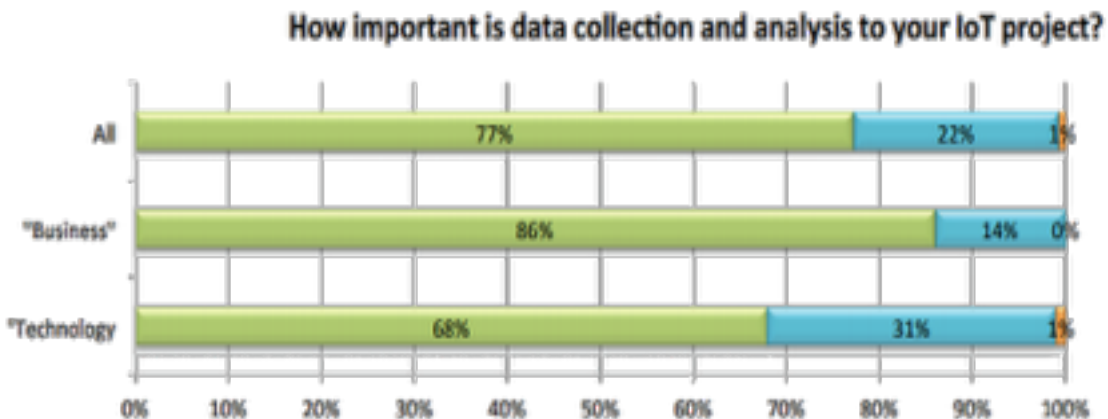
Source: IoT Talks: IDC's 2016 Global Decision Maker Survey

Source: PwC Internet of Things New wave of adoptions to push connected devices to center of business in Asia Pacific

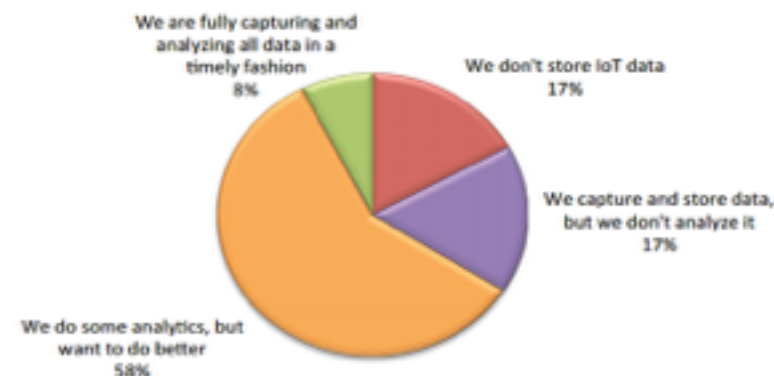
数据管理挑战是当前IoT项目面临的核心挑战

根据一项2015年针对当前IoT项目的技术团队和业务团队的调查

- 96%承认IoT项目面临挑战，其中41%认为及时采集和处理数据是最大挑战
- 86%的业务团队认为数据对项目非常重要，58%相信如果数据分析能更快将显著增加ROI
- 76%认为采集和存储成本制约了项目能力，如果成本能降低，会搜集更多数据
- 44%认为数据量太大导致无法有效分析
- 26%认为分析过程太慢无法有效响应
- 只有8%的项目完全实现了数据的及时采集和分析处理

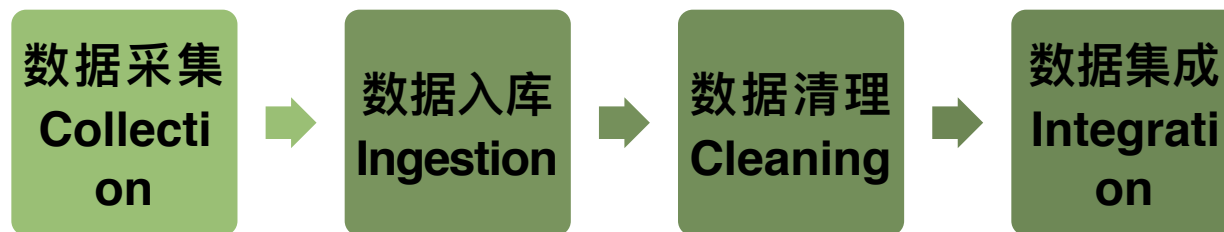


What level of data capture and analytics does your IoT project deliver TODAY?



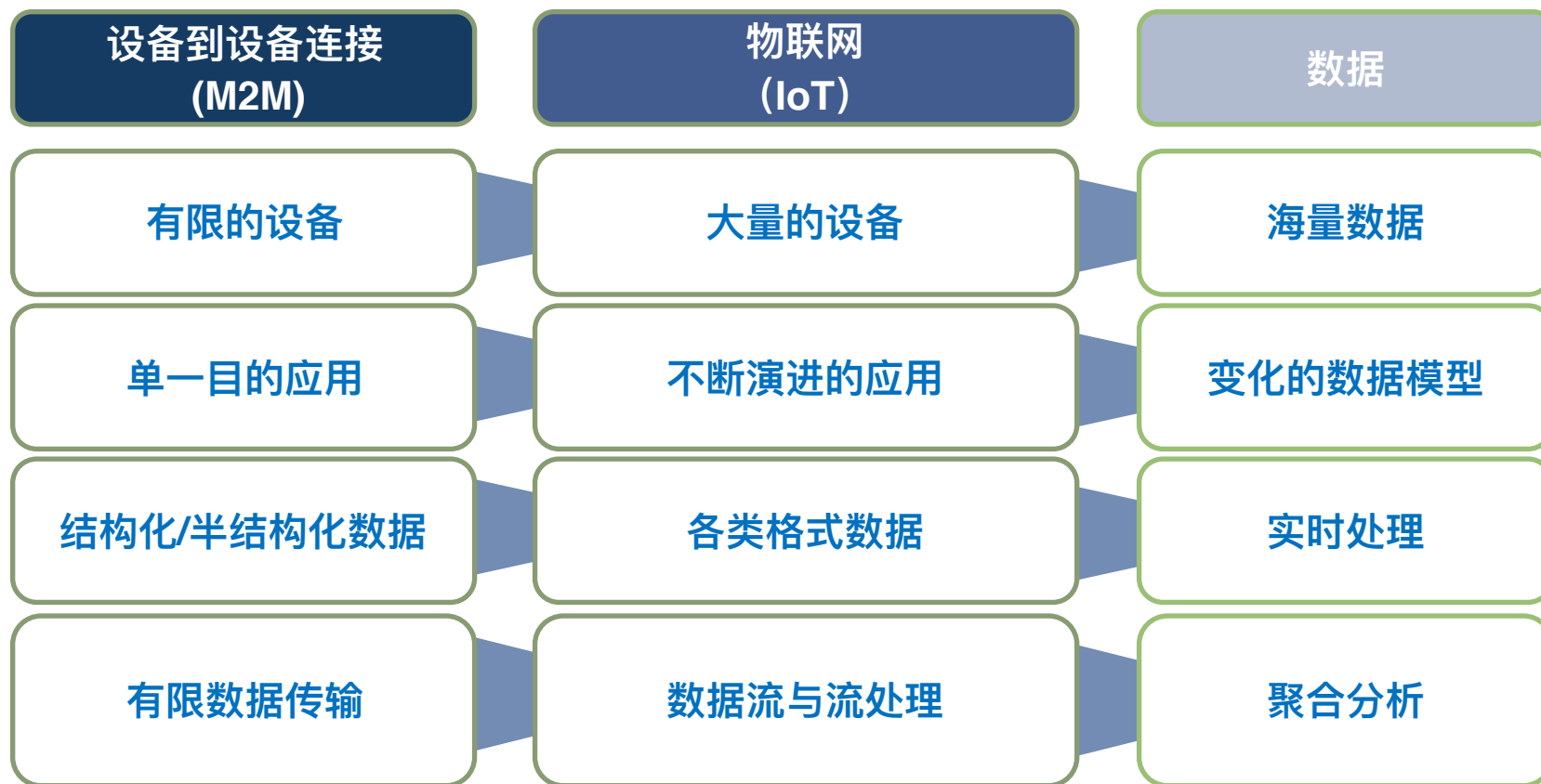
Source : Dimensional Research

在IoT场景下，数据处理流程的每一步都面临挑战



- **数据采集**— 需从各种各样的数据源，采集结构化、非结构化、半结构化数据
- **数据入库**— 需将大量各种格式的数据，快速导入统一的数据库
- **数据清理**— 需理解各种数据格式和内容；清理和标准化数据
- **数据集成**— 需关联、抽取、解析、融合、索引各类数据
- **数据分析**— 需具备查询、统计、预测和机器学习等各种分析能力
- **数据交付**— 支持查询、可视化等各种交付形式、并具备实时交付能力和高稳定性

为支持IoT的数据管理要求，需要新的数据管理能力



所需具备的能力



Machina Research

目录

IOT的发展趋势和挑战

MongoDB如何应对挑战

MongoDB成功客户案例

下一步

IoT的海量数据要求低成本和易于扩展的数据库存储

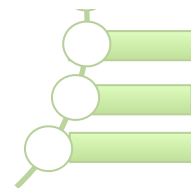


MongoDB
中文社区

IT大咖说
知识分享平台

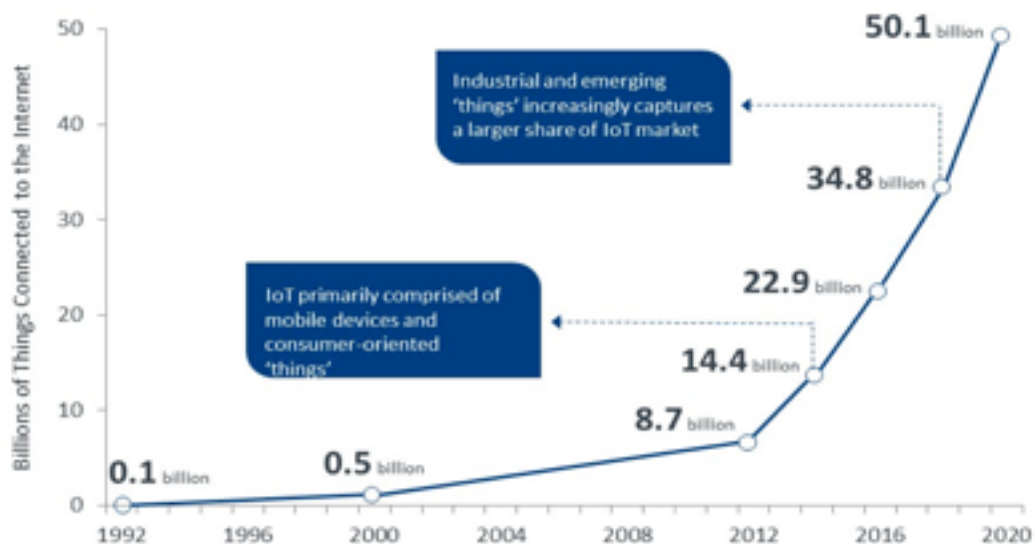
海量的IoT连接设备将产生真正海量的数据，根据IDC的预测，未来的10年，

1. 全世界每分钟新增1万5千个IoT连接设备，数量达到数百亿
2. 数百亿的IoT连接设备，将产生高达4ZB的数据需要处理



Projecting the 'Things' Behind the Internet of Things

From 2014-2020, IoT grows at an annual compound rate of 23.1% CAGR



CompTIA

Sources: Group SIR | Cisco | CompTIA

Connected
Devices
(by 2020)

Gartner
26B

IDC
32B

CISCO
50B

Data
Growth
(2013 vs 2020)

IDC

Total Data
4.4ZB → 44.4ZB

10x

IoT Data
.09ZB → 4.4ZB

49x

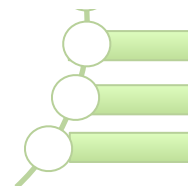
1 ZB = 1 PB x 1000000

传统数据的垂直扩展方式、成本无法承受，性能难



MongoDB
中文社区

IT大咖说
知识分享平台



面对IoT将产生的超海量数据，传统的数据库垂直扩展已经无法应对

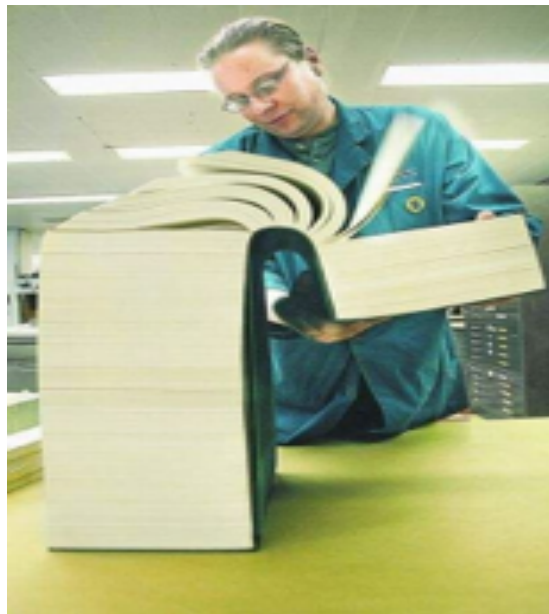
1. 继续扩展满足传统数据要求的高端服务器和存储，成本已经无法承受
2. 传统数据库集中存储方式，在如此大数据量下，性能也无法满足要求

■ 一个典型的航空IoT应用

- 一个航班，每分钟采集一次，每次采集 100 Byte:

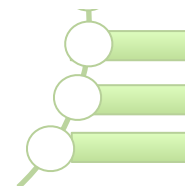
一个3小时航班产生： $3 * 60 * 100 = 18K$ 数据

- 全球每天航班数：100,000
- 每天产生数据量：1.8 Billion, 1.8TB/天
- 性能要求：21,000 QPS



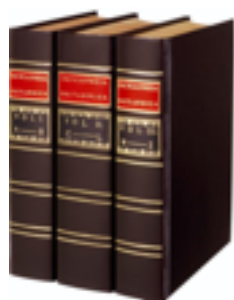
垂直扩展难以为继

MongoDB通过水平扩展应对大数据量挑战



MongoDB作为分布式数据库，采用水平扩展应对大数据量挑战

1. MongoDB部署在x86服务器中不依赖高端集中存储，可极大降低部署成本
2. MongoDB采用分布式架构和数据分片，分散了读写压力，可大幅度提升读写性能



第1版(1771) 三卷

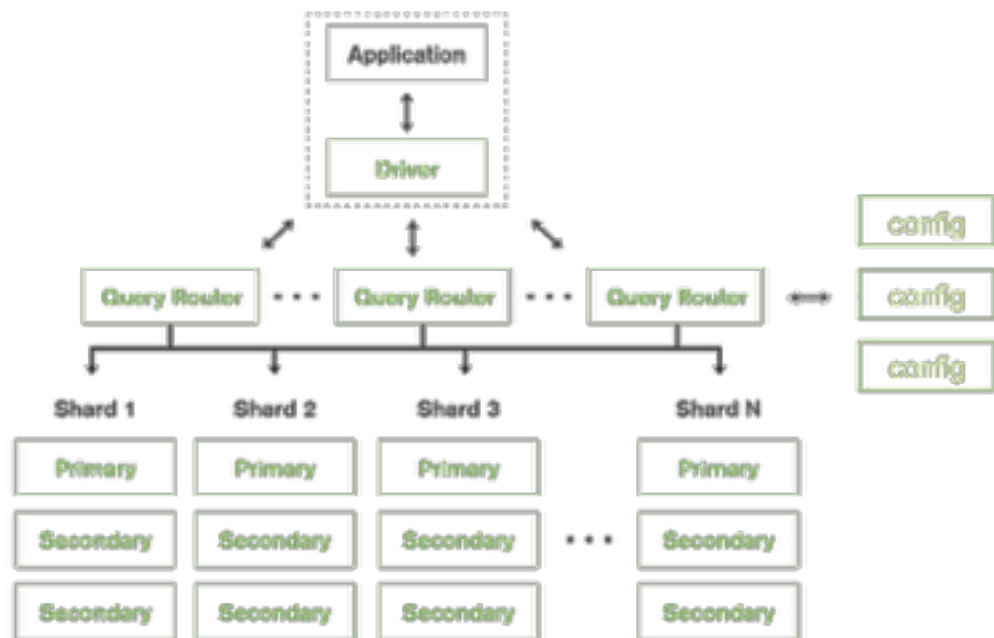
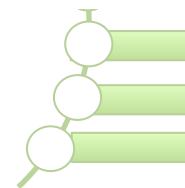


第15版(2010) 三十二卷

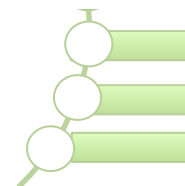
面对数据量持续增长，300年来，水平持续扩展成功案例：大英百科全书

MongoDB通过水平扩展应对大数据量挑战

1. MongoDB通过分片将数据保存在不同x86服务器上，分散读写负载
2. MongoDB通过片键确定数据位置，并自动实现数据存储分布的均衡
3. 对不同服务器上数据的查询，MongoDB自动组装，对应用完全透明



IoT需要灵活的数据模型



复杂的业务信息、多样的传输协议、不断升级的传感器，都需灵活数据模型来应对

1. 在一个应用中处理，多样且复杂的业务信息，需要灵活的数据格式
2. 在一个应用中连接和集成，多种的不同传输协议，需要灵活的数据格式
3. 在一个应用中应对，同时存在且不断升级的终端设备，需要灵活数据格式



一个互联网汽车有几十个功能不同的传感器，传输从结构化到非架构化的各种数据



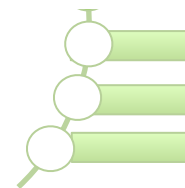
完成端到端的数据处理
IoT经常要面临十几种传输协议
MQTT, CoAP, AMQP等等



连接设备和传感器升级周期越来越短，
IoT需同时应对不同版本的设备
同一个应用同时处理3代设备也不罕见

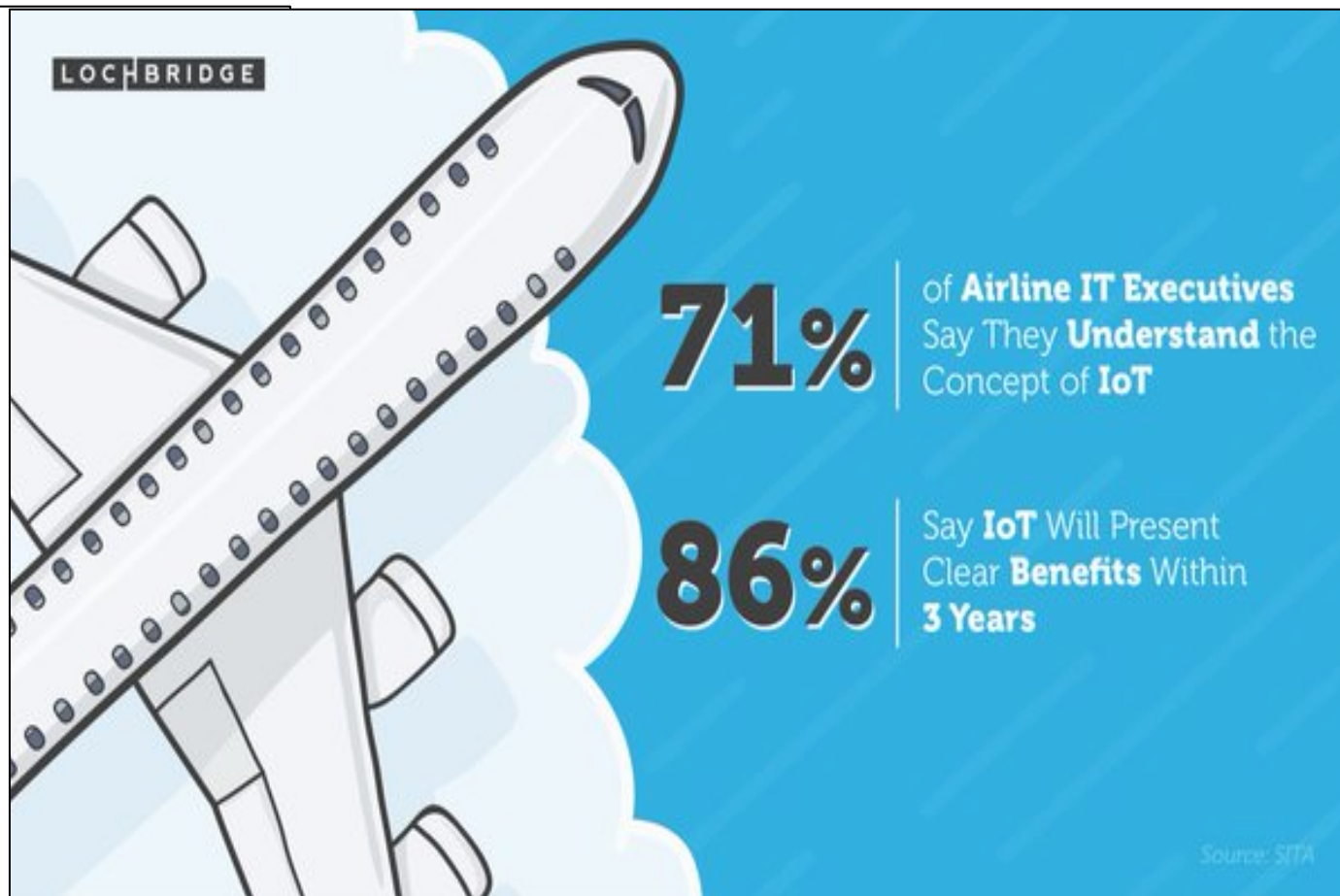
传统数据库难以高效处理复杂的业务信息

现代IoT应用中需要处理，逻辑层次复杂的业务信息，以喷气发动机监控为例

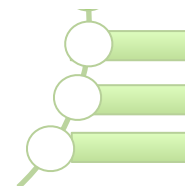


1. 位置信息: [38.2031, -120.4904]
2. 速度信息: 750
3. 高度信息: 29384
4. 引擎信息:
 - ① 燃油剩余: 78%
 - ② 温度信息: 89
 - ③ 排/进气压力比 (EPR) : xx
 - ④ 关键设备参数 N-value:
 - a) N1: xxx,
 - b) N2: xxx,
 - c) N3: xx

...



传统数据库难以高效处理复杂的业务信息



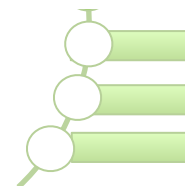
传统关系型数据库设计方案1：单表多列设计方案

EVENT_ID	PLANE_ID	TIMESTAMP	LAT	LONG	ENG-TEMP	FUEL	SPEED
100001	3902	1437297148810	38.2031	-124.4904				
100002	3902	1437297149213						750

存在问题

1. 不同时段和信息源采集数据，会存在大量空值，浪费存储空间
2. 需要增加新的采集数据，必须修改表，改动和升级成本非常高

传统数据库难以高效处理复杂的业务信息



传统关系型数据库设计方案2：主从表设计方案

EVENT_ID	PLANE_ID	TIMESTAMP
100001	3902	1437297148810

EVENT_ID	METRIC_NAME	METRIC_VALUE
100001	LAT	138.2031
100001	LONG	-124.4904
100002	SPEED	750

存在问题

1. METRIC_VALUE参数值为保持统一只能定义为Text型，分析处理复杂
2. 为提高查询效率，给特定参数构建索引会非常复杂，导致查询困难
3. 针对飞机的任何参数查询都需要JOIN操作，I/O操作成本高，查询速度慢

MongoDB非常适用高效存储IoT的复杂的业务信息



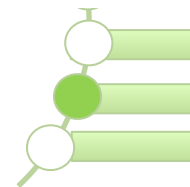
MongoDB的在设计IoT数据模型上，有明显的优势：

1. 与业务层次逻辑保持一致，易于理解与处理
2. 直接提供支持地理信息的存储、索引与查询
3. 数据存储节约空间、且后续扩展简单又便利

1. 位置信息: [38.2031, -120.4904]
2. 速度信息: 750
3. 高度信息: 29384
4. 引擎信息:
 - ① 燃油剩余: 78%
 - ② 温度信息: 89
 - ③ 排/进气压力比 (EPR) : xx
 - ④ 组件设备参数 N-value:
 - a) N1: xxx,
 - b) N2: xxx,
 - c) N3: xx

```
{  location: [ 38.2031, -120.4904 ] ,  
  speed: 750,  
  altitude: 29384,  
  engine: {  
    fuel_level: 78% ,  
    temperature: 89,  
    EPR: xx,  
    N-value: { N1: xxx, N2: xxx, N3: xx}  
  }  
}
```





IoT数据分析的目标与所需具备的能力

分析IoT的数据来产生商业价值，是IoT创造价值的途径

- 分析的业务目标：描述分析、预测分析、诊断分析、指导分析、自动化分析
- 所需的技术能力：实时分析、灵活查询、趋势分析、批处理集成、机器学习集成

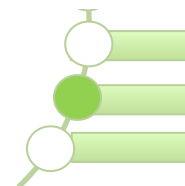


- 技术能力**
- 实时分析：支持实时（小于1秒延迟）仪表盘（dashboards）展示能力
 - 灵活查询：支持即兴和特定（Ad-hoc）的查询能力
 - 趋势分析：支持历史和趋势分析报告能力
 - 批量处理：支持集成数据批处理能力
 - 机器学习：支持集成数据机器学习能力

MongoDB支持IoT应用所需的各种分析能力



IT大咖说
知识分享平台

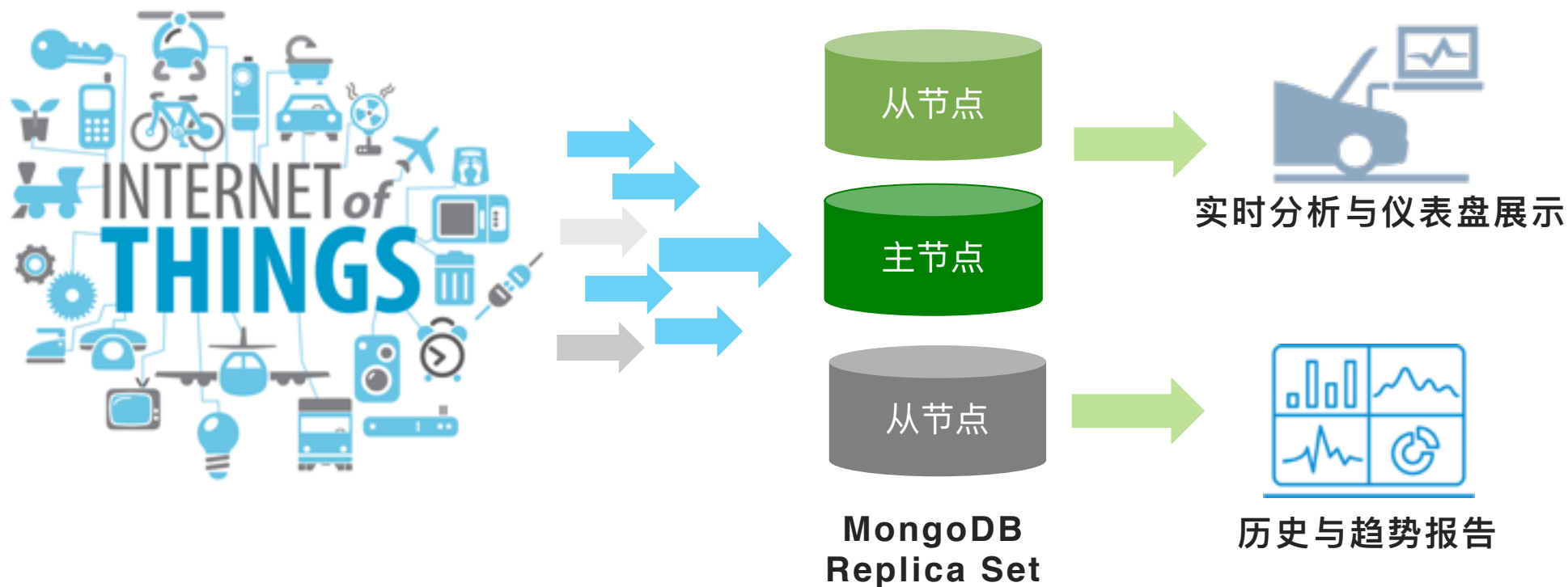


MongoDB通过自身框架和集成能力，可支持IoT应用所需的各类分析能力

- 支持实时（小于1秒延迟）仪表盘（dashboards）展示能力：Aggregation framework
- 支持即兴和特定（Ad-hoc）的查询能力：Aggregation framework
- 支持历史和趋势分析报告能力：Aggregation framework 或者 BI Connector
- 支持集成数据批处理能力：Hadoop connector
- 支持集成数据机器学习能力：Spark connector

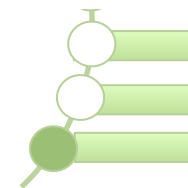
MongoDB支持同时执行多种分析，而互不影响

MongoDB分布式架构，同时支持应用执行多种分析，并将彼此影响降到最低



IoT的数据平台需要稳定、高可用和全球部署能力

- IoT业务7x24服务属性，要求数据平台提供7x24x365的高稳定、可靠的服务
- IoT数据量的快速增长，要求数据平台能够对应用透明的情况下快速、平滑扩展
- IoT业务物理分布特点，要求数据平台能够支持多中心、全球和云中部署



保持系统稳定、健康，运维挑战面临三大调整

- 如何定位性能问题：面对海量的数据操作，如何快速定位到底有哪些慢操作
- 如何发现潜在故障：分布式系统数据库服务节点众多，如何早于用户发现系统问题
- 如何快速恢复问题：在数据或系统出现问题的时候，如何快速的恢复数据服务

一个真实的IoT项目：

慢查询到底在哪里？



管理1000万连接设备，
需处理**5亿条消息/每天**

到底哪里会出问题？



50个的数据节点部署
在不同城市

怎么恢复这么大的数据？

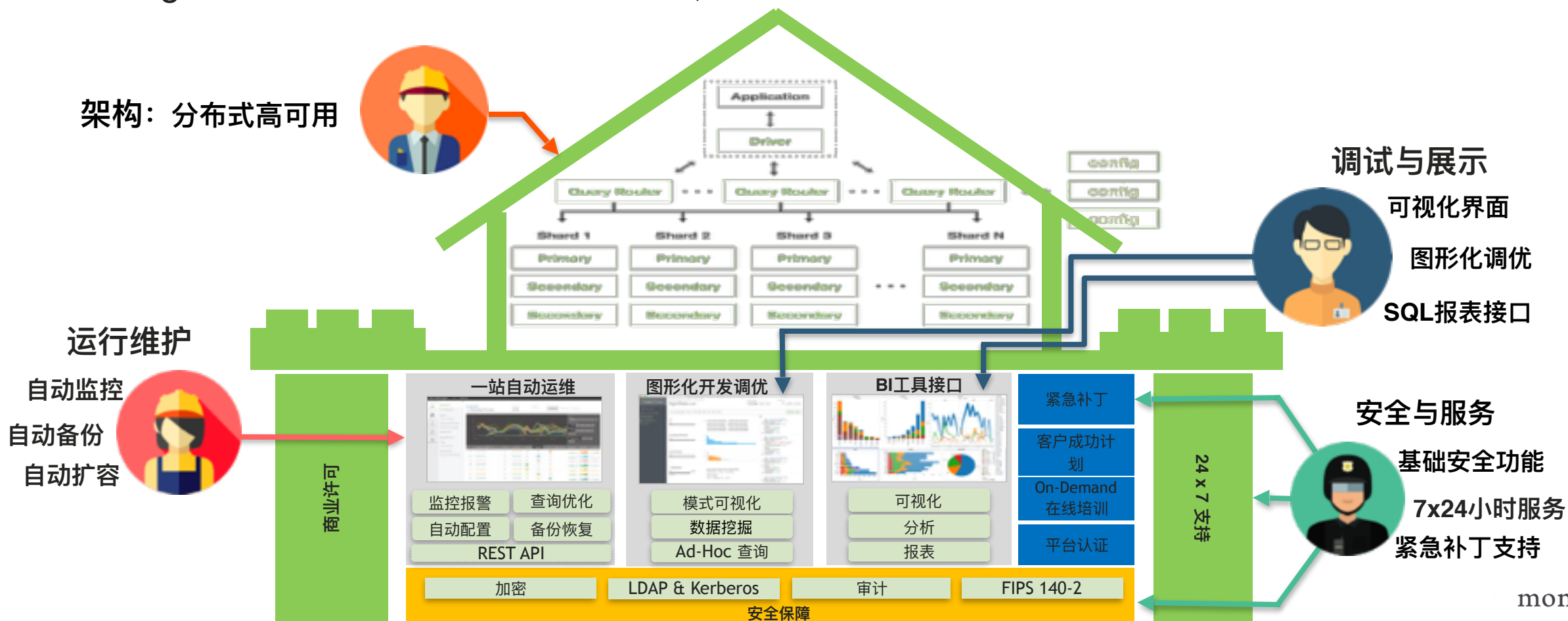


5TB数据/45天

MongoDB从架构和工具两个层面保证系统稳定、健

IoT对系统稳定要求同时满足：系统高可靠、高性能、高安全、易维护四个条件

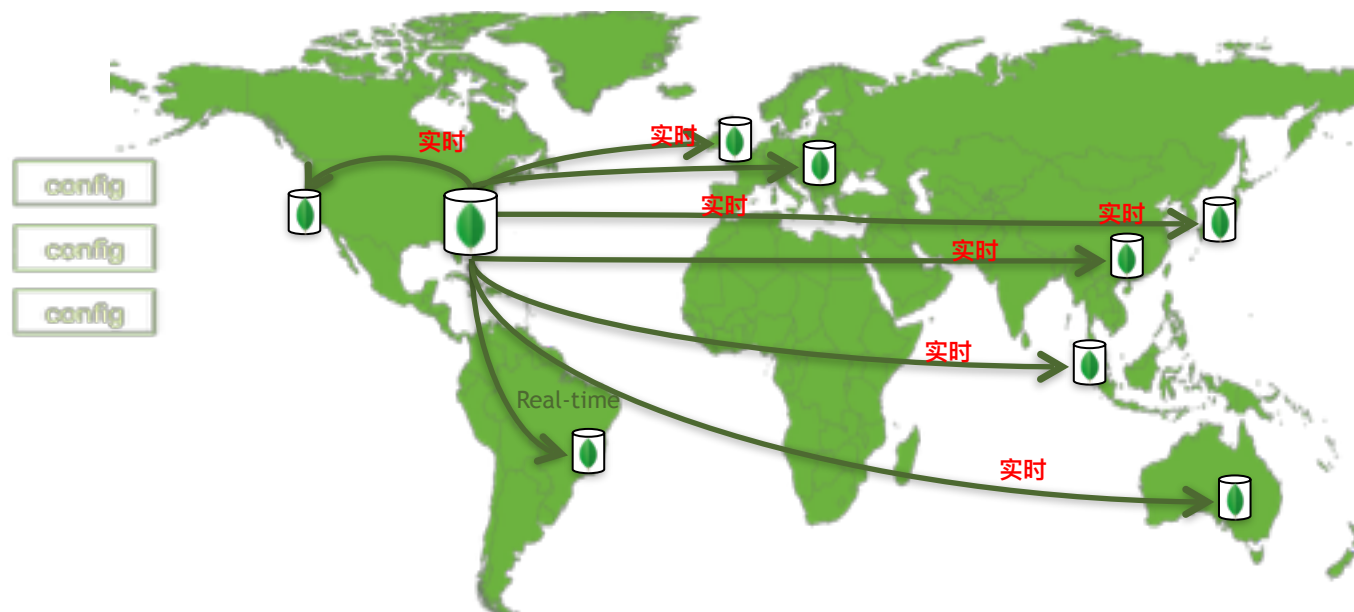
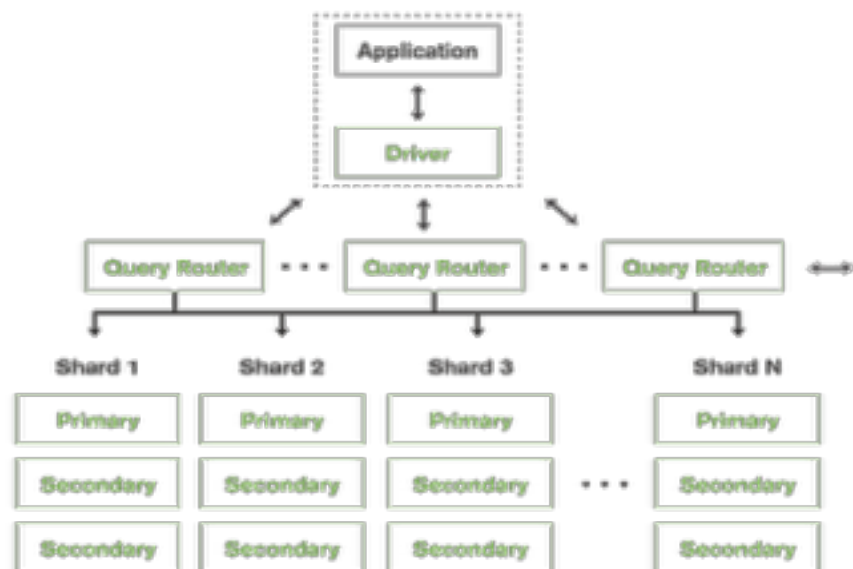
- MongoDB复制集和分片，从上部署架构层面，保证系统任何无单点故障风险
- MongoDB企业版从工具和服务层面，保证系统安全、健康和易于快速扩展



MongoDB的分布式架构支持多中心和云中部署

MongoDB的分布式架构不仅消除单点故障，还使得多中心部署更简单

- MongoDB不仅支持多中心部署，还有众多大型机构的全球部署案例
- MongoDB完全支持各类云部署，当前有60%的MongoDB运行在各类云中



MongoDB企业版助力系统稳定、健康运行

在大访问量下如何快速掌握、分析、定位问题和恢复服务，是保证系统稳定的关键

- 定位慢操作：Ops Manager可以快速定位慢查询，并帮助分析慢查询的可能原因
- 发现有问题的：Ops Manager可以监控100多个关键指标，提前发现和避免重大问题
- 恢复原服务：Ops Manager可实现多分片环境下的增量备份，可恢复到指定时间点



IoT平台运维的实际案例

在某大型IoT平台中的实际案例中，Ops Manager可帮助客户明显提升系统稳定性



定位性能问题

手工方式

5个小时

VS

Ops Manager

15分钟

- 1.从多节点采集数百M日志
- 2.根据关键字寻找慢查询
- 3.手工制作图表寻找规律
- 4.定位可能慢查询点和事件
- 5.查询相关时间段其他操作
- 6.关联分析确定可能原因



一目了然



部署新的环境

手工方式

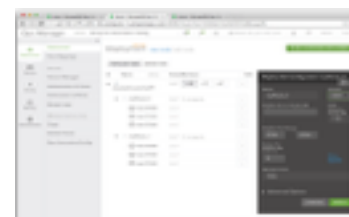
1天

VS

Ops Manager

30分钟

- 安装配置：
150多个步骤
- 版本升级：
100多个步骤



一键完成操作



备份&恢复

手工方式

RPO: 1天

VS

Ops Manager

RPO: 15分钟

- 每天系统快照
- 全量备份
- 手工恢复
- 实时增量备份
- 按时间点恢复
- 自动恢复

总结：MongoDB是IoT数据平台的领导者

- 因为MongoDB所具备的大数据处理能力、灵活的文档结构、高读写性能、极强数据分析能力和企业级的稳定性可靠性，MongoDB非常适用于IoT平台数据库
- 为此，2016年Forrester将MongoDB列为大数据NoSQL的最领先领导者

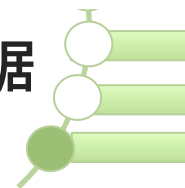
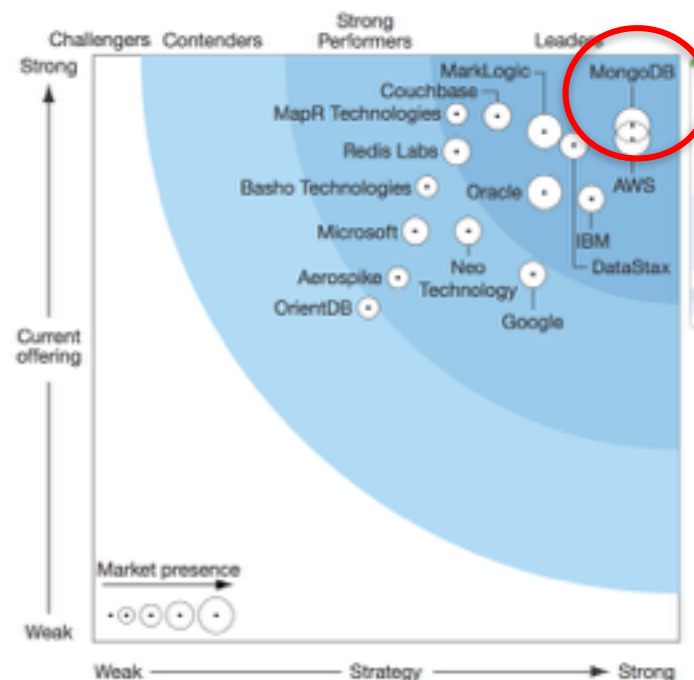


FIGURE 3 Forrester Wave™: Big Data NoSQL, Q3 2016



目录

IOT的发展趋势和挑战

MongoDB如何应对挑战

MongoDB成功客户案例

下一步



BOSCH

Invented for life



MongoDB
中文社区

IT大咖说
知识分享平台

制造业IoT案例

全球最大的零部件制造商采用MongoDB构建通用IoT平台

挑战

Bosch期望开发一个通用的IoT平台，支持从制造到零售，从改善客户体验到提升生产安全的各种场景的IoT应用

为支持近乎无限的业务场景，Bosch发现传统的数据库的严格数据模型无法满足IoT对灵活数据格式的要求，性能无法满足业务要求，需要寻找新技术解决这个问题

方案

因为MongoDB灵活的数据模型、易于扩展、低成本的特点，Bosch选择MongoDB作为其IoT平台的数据库

同时Bosch邀请MongoDB的专业服务团队，帮助检查系统部署和分析开发中遇到的问题，提升系统性能。

收益

Bosch将所有的IoT应用都统一采用MongoDB平台

平台成功应对了从3百万车辆到3亿车辆的增长

基于2.5T数据的聚合分析查询性能从10个小时，提高到3秒

能源行业IOT案例

全球知名电器制造商采用MongoDB构建能源IoT平台

挑战

西门子需要开发的能源管理系统，需要同时支持多种不同的协议，支持云中部署，并具备数据聚合分析能力，并具备实时分析性能

到底是采用传统数据和还是采用No-SQL，或者一起使用是开发团队面临的重大选择

方案

因为MongoDB同时满足，西门子对数据库的IoT的三个关键要求：

- 对JavaScript架构友好，JSON格式
- 高性能写入，支持灵活的聚合分析
- 能够在云中部署，并易于扩展

最终，西门子决定只使用MongoDB作为单一数据库

收益

成功支持了米兰2015年世博会，实现对园区电网、智能建筑、公共照明和电动车的能源IoT管理，6个月会期

- 实现采集1.2亿条数据采集
- 实现采集数据45G
- 生成历史与趋势数据17G
- 生产聚合分析数据5G



智能家居IoT案例



全美最大的智能家居服务提供商采用MongoDB，12个月重构智能家居IoT平台

挑战

Vivint作为全美最大的智能家居服务商业业务发展迅速，为应对业务的增长，并保持其服务的低价，以获取市场竞争力，

2012年决定将原来使用关系数据库的平台重构，但整个项目必须在一年内完成。

方案

“我们选择MongoDB有很多原因，但第一条是，我们时间有限”“MongoDB的先进架构和灵活数据模型，让我们的数据采集层、应用消费层的数据都是对象模型，这极大的减少了开发时间”

第二是可水平扩展性，“你当然可以垂直扩展Oracle，但成本太高也无法应对我们需要处理的数据量”“简单的说垂直扩展对我们不可行”“MongoDB扩展非常快”

---Nicholas Brown, Vivint首席软件架构

收益

- 12个月内完成了平台重构
- 只需5%传统数据运维人力
- 管理1亿连接设备
- 每天处理5亿条消息
- 每45天生成5TB的数据
- 数据节点已经增长到50个



MongoDB
中文社区

IT大咖说
知识分享平台

您的下一步？