



# 选择一种数据库管理系统（DBMS）来迎接 第三代平台的挑战

一份IDC信息简报，由InterSystems公司资助 | 2017年3月



# 引言

在2016年10月，IDC执行了由InterSystems公司资助的一项研究，研究重点是全球企业对同时支持事务和分析工作负荷的数据管理解决方案的需求，以及这类解决方案将能带来的益处。

一共有来自澳大利亚、巴西、中国、德国、日本、英国和美国的502家企业接受访谈调研。

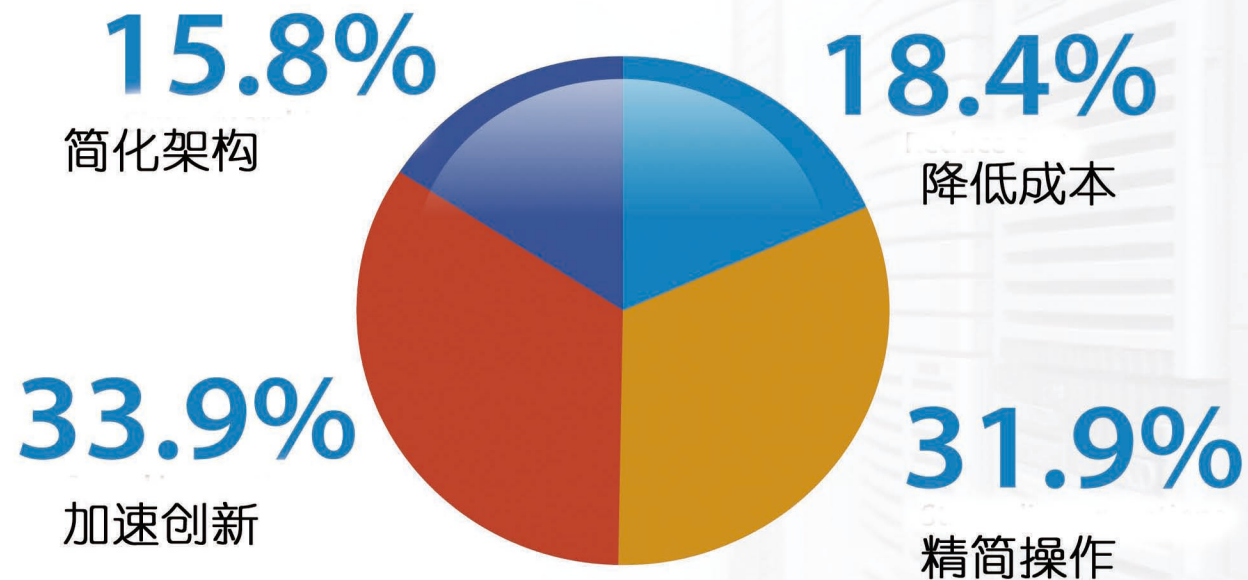
参与研究的公司来自多个行业且规模各异，因此具有良好的代表性。

根据这些公司对自己公司使用的事务数据库和分析数据库的熟悉程度，选择出合适的受访者。

这份报告呈现了这项研究的关键发现、分析和值得读者注意的要点。

# 公司希望加速创新和精简操作

今年你的首要业务目标是什么？



问题：请列出接下来1年内你所在企业的最重要三个业务目标。

# 信息技术 (IT) 系统向第三代平台的转化带来了影响信息管理系统的新IT需求



来源: IDC

第三代平台是发展趋势和新技术的集中体现, 代表了IT行业的第三次主要发展浪潮。它能使企业对不断变化的业务环境做出敏捷、快速适应且及时的应对。

1

新工作负载需要信息管理平台支持更多数据类型、更大型数据集和具有更快的周转

2

问题的多样性需要多种数据管理工具才能给予解决。

3

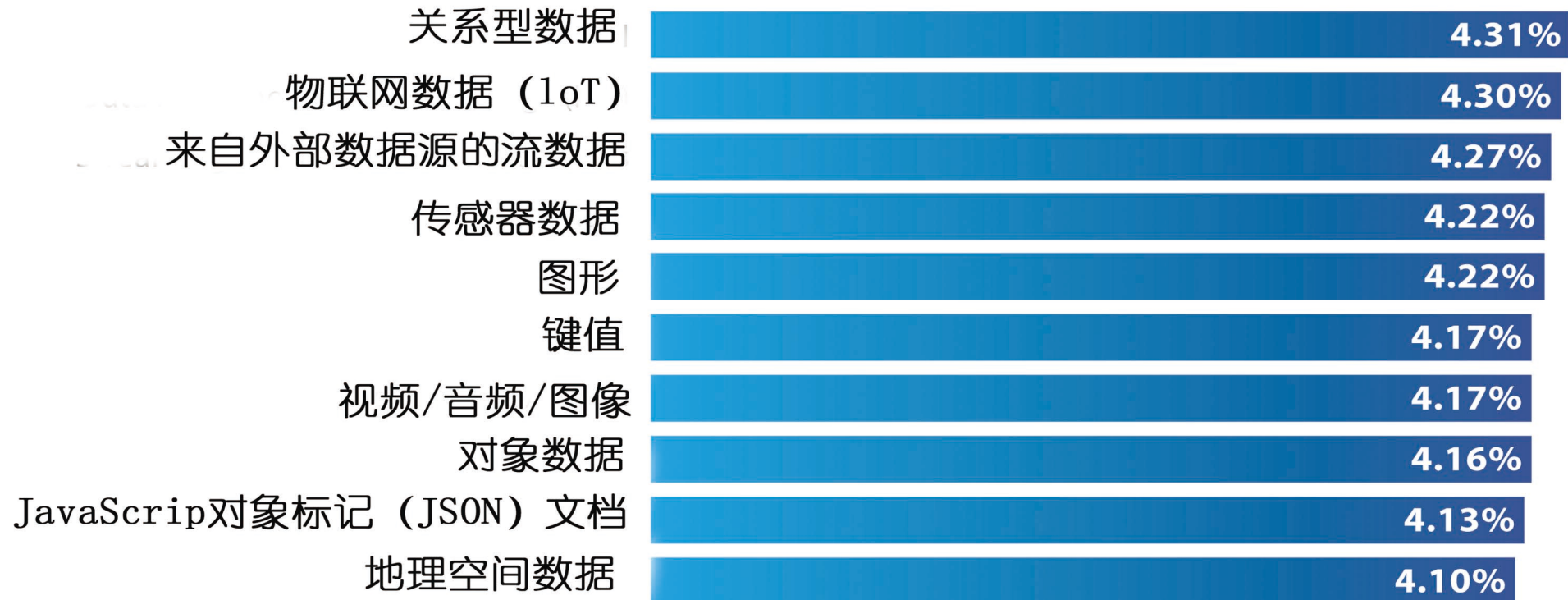
用户们逐渐意识到对更高性能、更好敏捷性和支持更多数据类型的平台的需求。

**问题: 请列出接下来1年内你所在企业的最重要三个业务目标。**

# 所有数据类型都是重要的

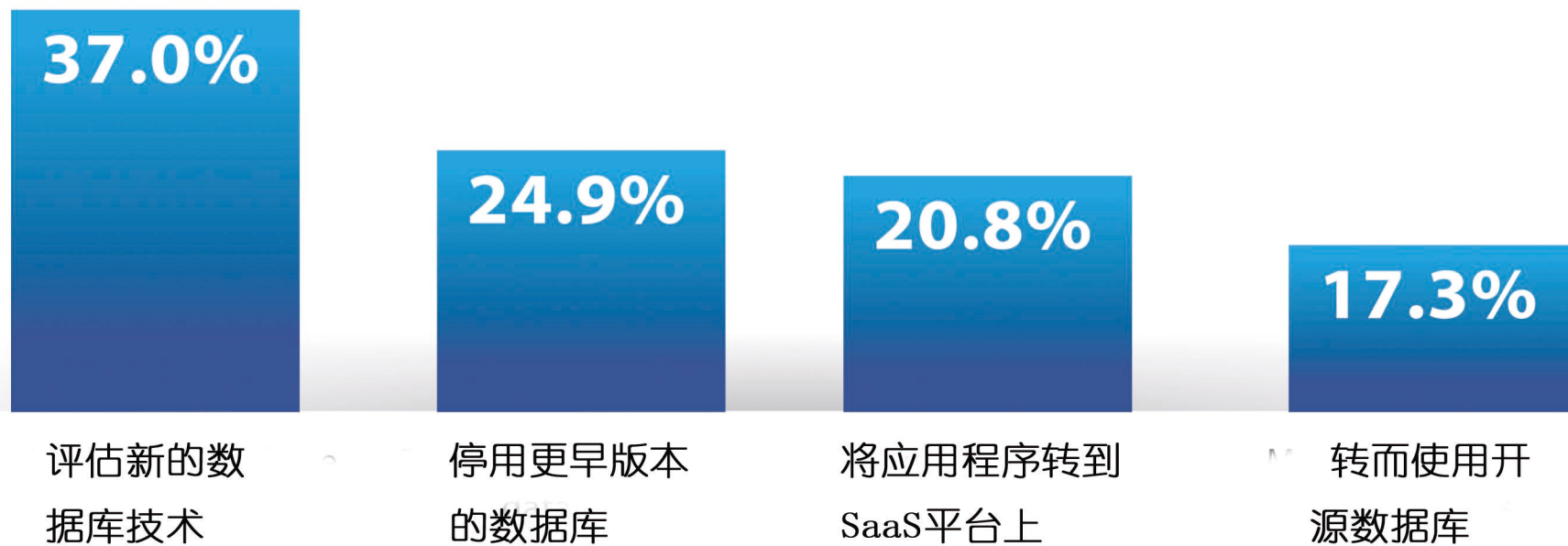
## 新的数据类型有多重要?

(评价量表, 1=不是非常重要, 5=非常重要)



# 接近2/5受访者在考虑新的数据库平台, 同时有少数受访者逐渐转而使用开源数据库

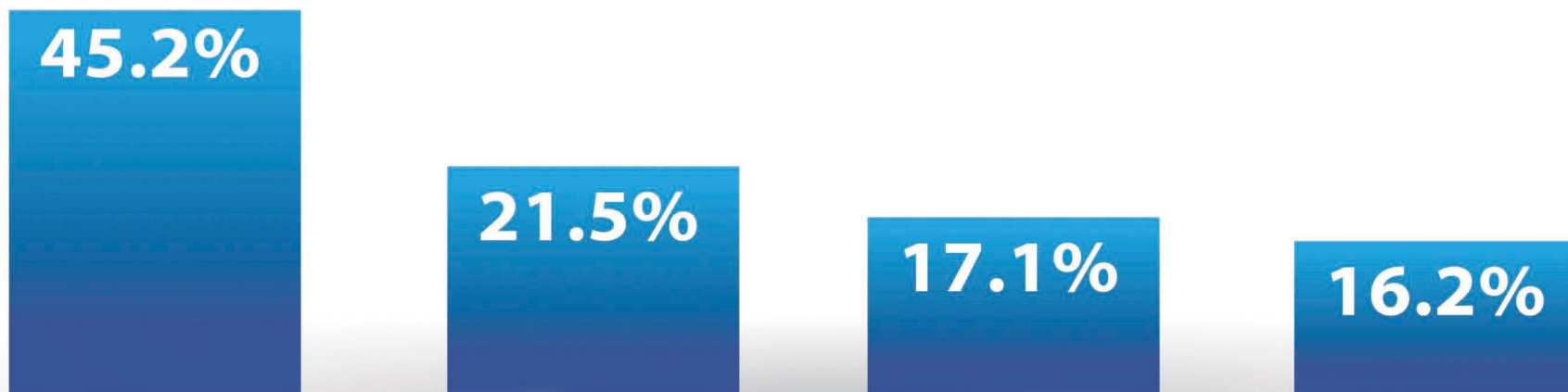
你今年的首要技术目标是什么?



来源: 第三代平台信息管理需求调查, IDC, 2016年10月, n=502

# 开源有多重要？

通常人们认为, 新的第三代平台工作负载需要以开源DBMS作为基础。我们的调研发现事实并非如此。



所有新的DBMS技术都必须受到产权保护 (也就是不是开源)

我们倾向于选择开源的新DBMS技术, 但也不排斥其他技术

所有新DBMS技术都必须是开源的

我们只想选择能最佳地满足技术特性和总体成本要求以及能最好地解决现有问题的DBMS

问题: 下列哪一句陈述最符合你对开源的看法?

# 大分水岭:

## 事务 vs. 分析



### 事务

- ✓ 面向记录处理
- ✓ 驱动业务操作
- ✓ 为快速写入而设计, 而不是检索
- ✓ 最小化操作schema

### 分析

- ✓ 包含来自多个事务数据库的数据
- ✓ 针对检索速度而做的优化
- ✓ Schema被设计为帮助回答关键问题



# 实时决策

## 战略

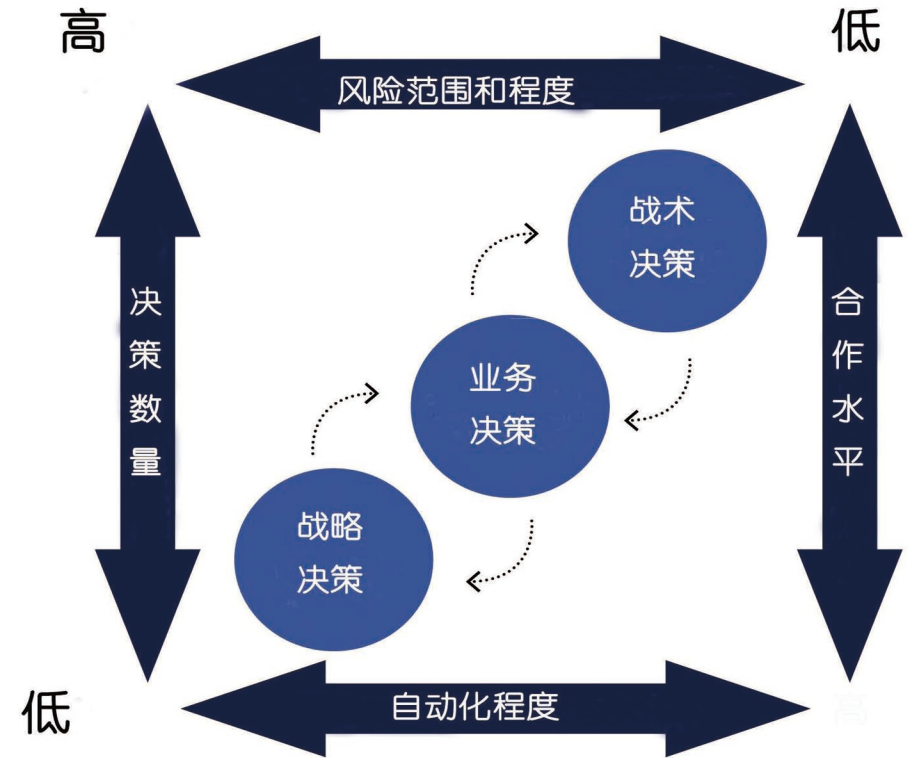
- ✓ 深度分析模型和数据仓库报告能为执行人员决策提供支持
- ✓ 数据分析师梳理海量数据、执行管理人员进行决策

## 业务

- ✓ 收集和分析围绕项目或产品的数据
- ✓ 终端用户检索数据、调整他们的项目或产品计划

## 战术

- ✓ 呈现即刻可得的现行数据和流数据
- ✓ 领域员工或自动化电脑算法做出决策



依赖于实时运算的战术决策每天驱动着业务

# 精简业务需要实时分析

## 慢速数据整合

转换和合并来自多个数据源的数据需要较长时间, 这会阻碍其出现在决策窗口期。

对任何数据源进行的任何更改都需要改变数据转换, 很可能也需要改变分析数据库schema, 这会导致不灵活性。



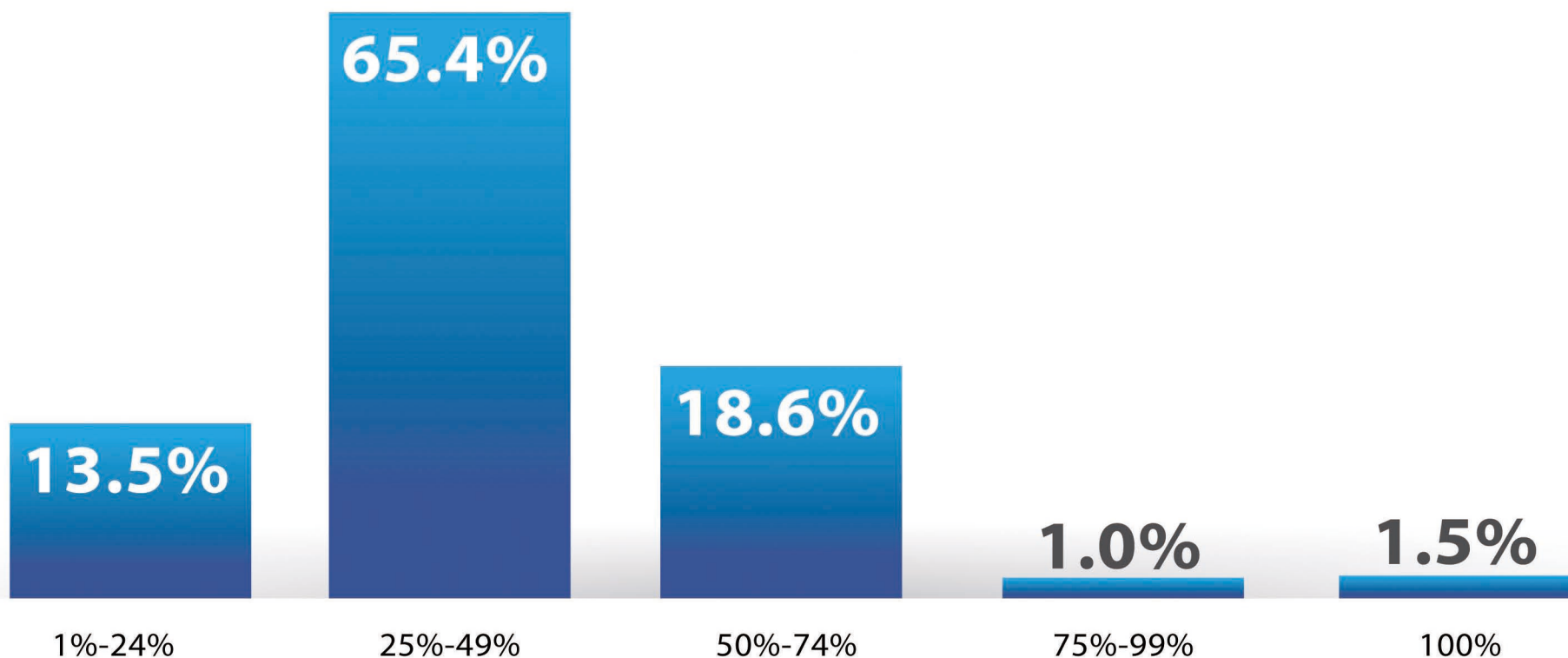
## 两种独立的数据库类型

针对事务进行过优化的数据库通常不能及时执行复杂的分析查询。

而针对分析进行过优化的数据库的事务处理速度通常太慢, 不能满足应用程序的吞吐量需求。

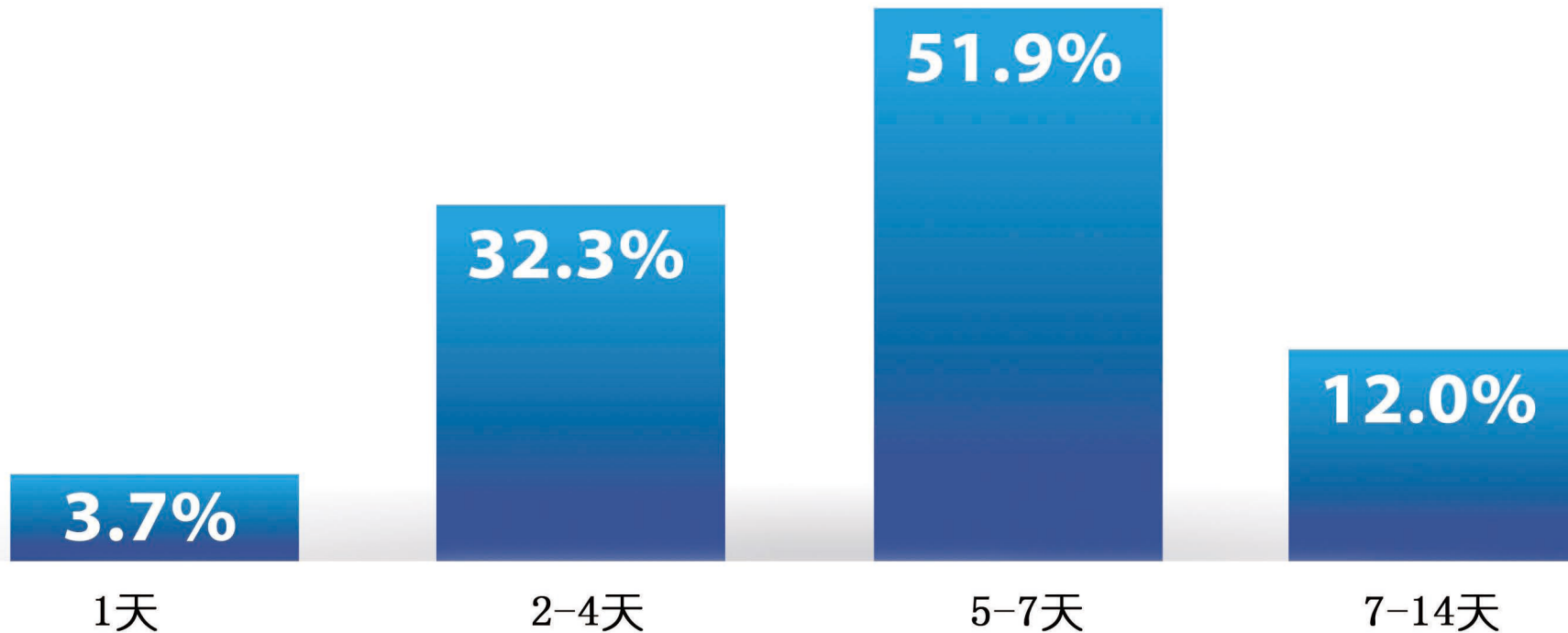
不协调统一的业务操作

# 大量事务数据是通过提取/转换/加载 (ETL) 技术 以较慢速率传送的



问题: 通过提取/转换/加载 (ETL) 技术在事务系统和分析系统之间传送的信息占比是多少?

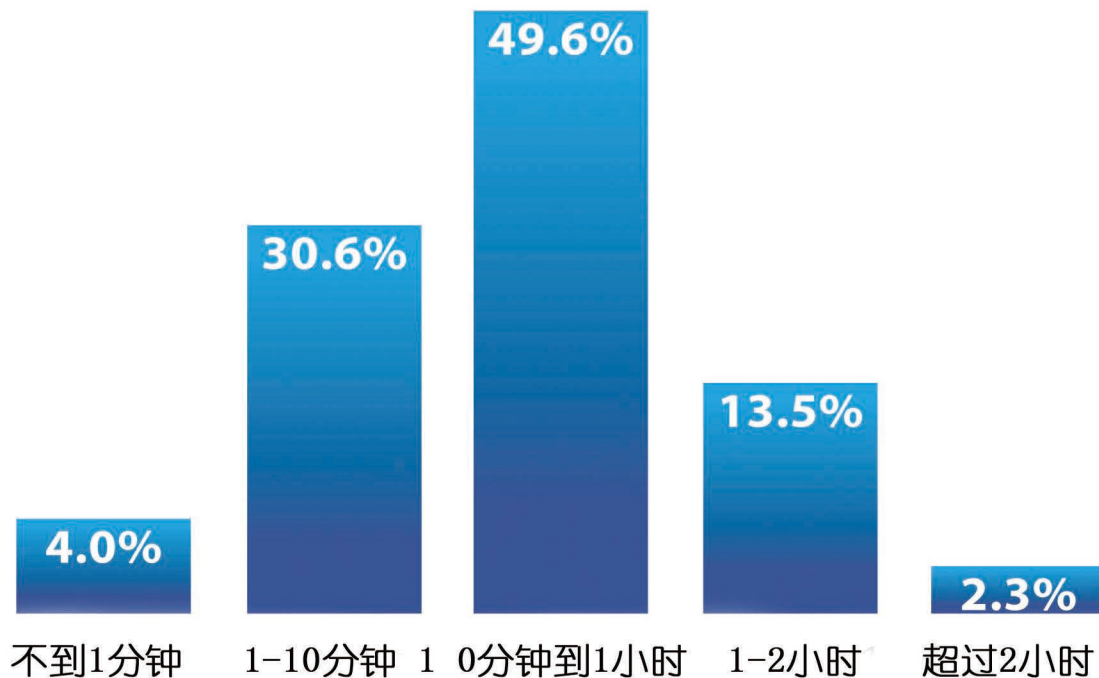
## 通过ETL传送的数据速度不够快



问题: 平均而言, 对每一项ETL任务, 当数据到达分析数据库的时候距离其产生已经有多长时间了?

# 你可以“即刻”使用实时数据做出决策吗？

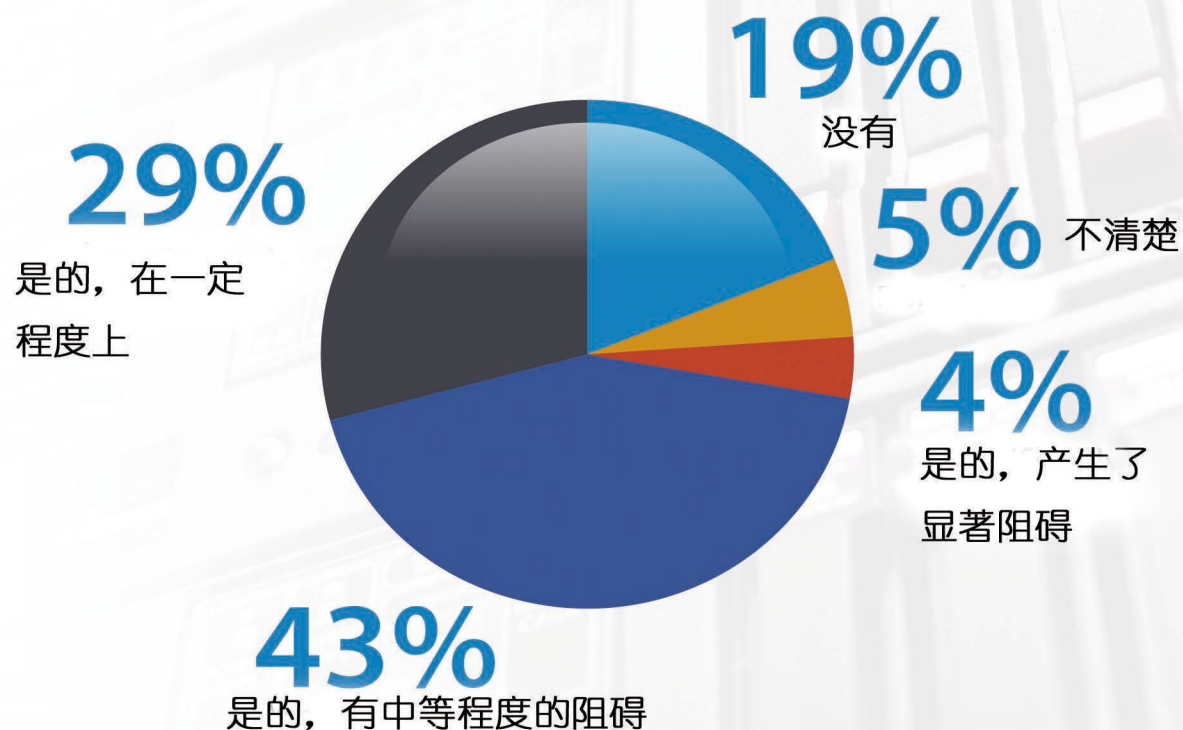
- ✓ 如果你希望追踪业务进展, 一定的延迟也许是可以接受的
- ✓ 基于实时数据做出决策需要数据库具备将分析性查询融入事务处理的能力。
- ✓ 这一种能力需要以零延迟为基础
- ✓ 即使变更数据捕获 (CDC) 都不能实现零延迟



问题: 平均而言, 每一次CDC连接, 数据被传送到分析数据库中的延迟是多长时间?

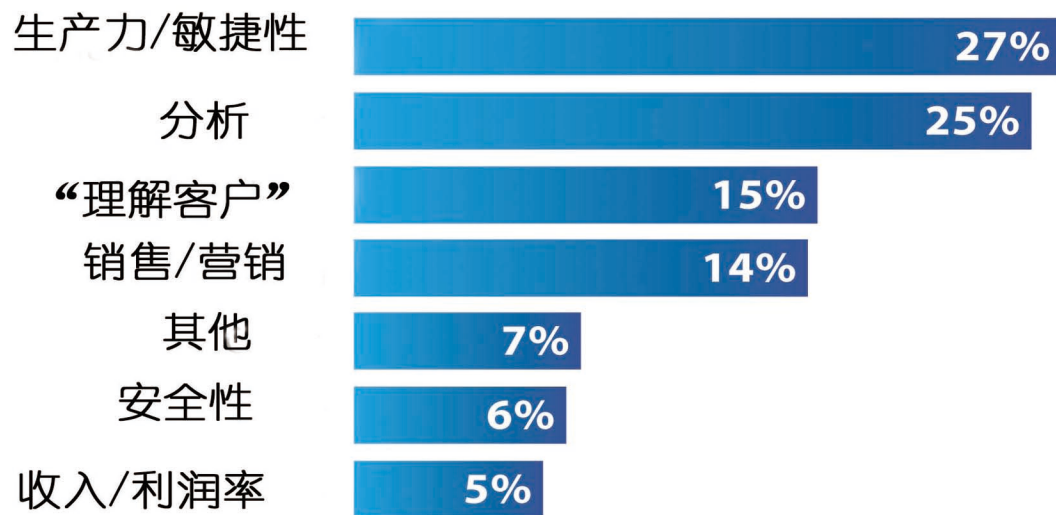
CDC数据的延迟时间 (%)

# 超过3/4受访者称不及时的数据已经阻碍了商业机会



问题: 不能分析当前实时数据是否已经阻碍了你所在企业获得商业机会的能力?

# 不及时的数据会在很多方面造成阻碍



问题: 请简短描述不能分析当前实时数据在哪些方面已经阻碍了你所在公司获得商业机会的能力?

## 及时数据的益处——来自客户的声音

“能够即刻察觉欺诈, 从而采取适当措施减少损失”

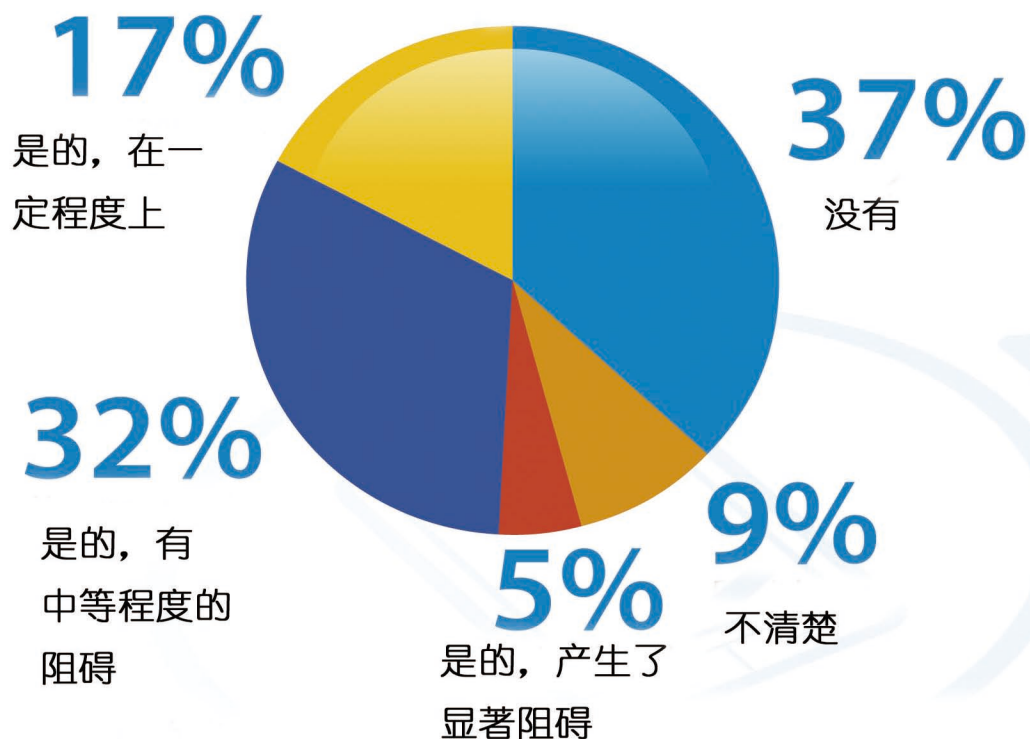
“能进行我们所需的强有力研究和开发”

“实时数据分析能从多个方面加快决策过程”

“能立刻发现公司内部出现的差错”

“我们可以更好的了解我们的客户”

## 不及时的数据也会限制操作效率



问题: 不能分析当前实时数据是否已经限制了你在公司改善经营效率的能力?



# 同时运行多个数据库的成本

大多数受访者同时使用着超过5个  
生产事务数据库

25%受访者同时使用超过10个

**超过60%**受访者同时使

用着超过5个分析数据库

同时, 超过30%受访者同时使用  
超过10个分析数据库

而且无一例外, 他们为每一个数据库配备了  
1-2名数据库管理员 (DBAs)。

## 不断增高的复杂性

- 有限的资源 (服务器和存储设备)
- 每一次改变数据分配都是一个工程
- 应用程序相互孤立导致系统过度饱和

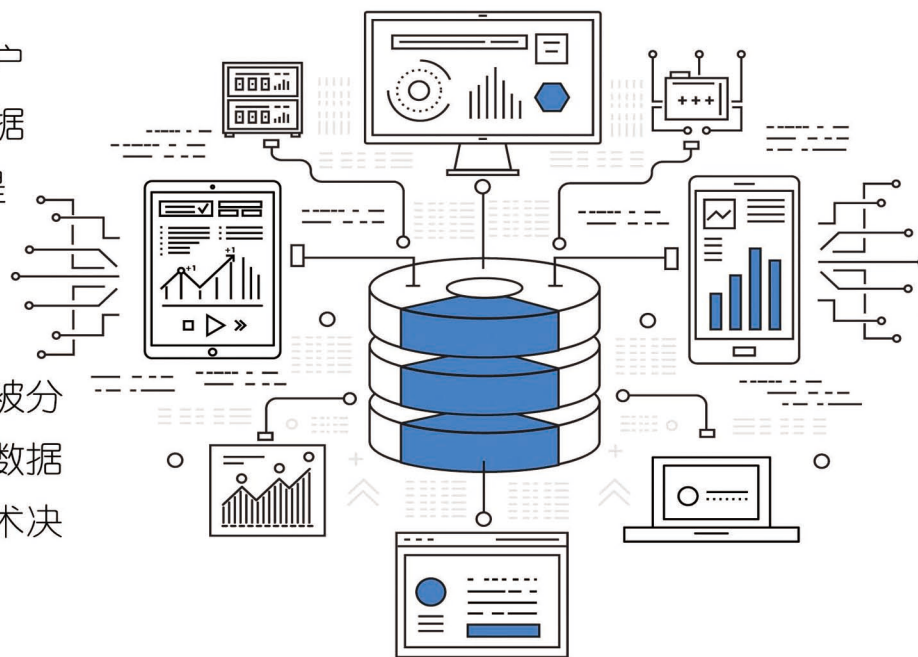
## 数据库管理的安装成本

- 传统数据库部署=固定成本
- 部署是以最高工作负载量为标准而进行
- 结果: 持续存在的硬件利用不足
- 员工成本随着服务器数量增加而逐步增高, 系统也是根据最高工作负载量进行分配

## 合并数据库操作的需求

我们已经观察到, 大多数用户需要支持多种数据类型的数据库, 远超过本地RDBMs所能提供的支持。

我们观察到, 只要数据仍然被分别隔离在事务数据库和分析数据库中, 就不能良好地支持战术决策。

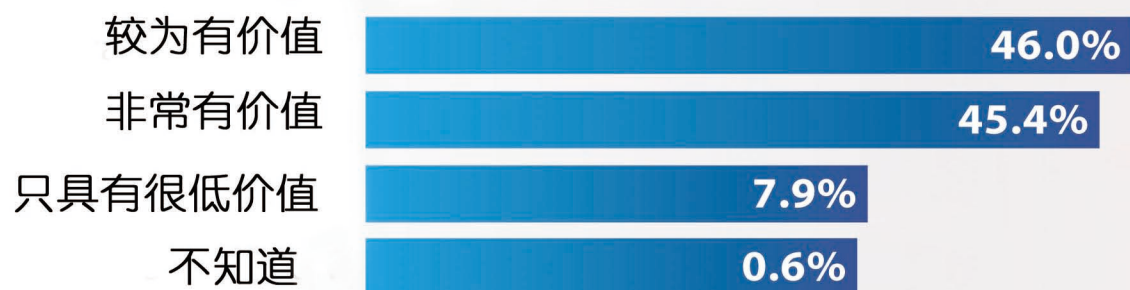


我们也观察到, 同时维护多个数据库会导致过高成本和数据中心工作复杂性。

对关键问题的答案是, 将数据紧密连接并进行联合分析处理和事务处理, 最理想的情况是能包含各种数据类型以进行数字化转型。

受访者们认为在一个DBMS中混合多种数据管理能力的需求有多重要? 请看下一页。

# 超过90%受访者认为混合特性非常有价值



- ✓ 灵活的可扩展性
- ✓ 混合关系数据和非关系数据
- ✓ 在不影响事务或操作性能的同时, 执行对实时操作数据的复杂解析查询

问题: 你相信一种具有下列特性的数据库系统对你所在公司而言有多大价值?



# 结论

第三代数据平台技术开启的数字化转型时代带来了客户对分析大量数据类型以及对近乎实时地分析事务数据的需求。

很多公司目前仍然在使用ETL数据传送方法, 这不能满足实时数据分析所需。

不及时的数据显著限制了公司的竞争能力和保持敏捷的能力  
对上述问题的解决方案是在一种具有混合特性、能合并数据库操作、同时具有执行事务和分析能力的单一DBMS。