

## 第 2 章 企业级应用为何如此繁杂零乱

**摘要：**如今，即便是小型企业，也会在不同的地域开展业务，并为不同的行业提供服务。这会带来一系列的挑战，比如语言、货币、不同的法规要求和截然不同的行业预期。那些业务范围更广泛的大型企业则面临着更严峻的挑战。随着企业规模的不断发展，他们还需要跟踪不同业务领域中的海量信息。现代化的企业级应用需要及时满足这些需求。本章将简单介绍现代化的企业级应用，描述它们所执行的部分任务，以及所涵盖的业务领域。

企业级应用是指帮助企业运行业务的软件系统。它们可以提高业务流程实施的自动化程度，为规划、数据分析和数据管理等任务提供有力支持。企业级应用的主要功能在于，它能够集成和处理来自不同业务领域的的数据，提供涵盖整个企业运营状况的实时视图。

在理想状态下，企业级应用能够及时向用户提供指定环境下的所有相关信息，帮助他们做出有效的决策，并支持业务部门优化其运营。这使企业级应用明显区别于其他业务软件（如电子表格），因为后者无法自动从所有相关数据源中提取数据。企业级应用与其他业务软件的另一个不同点是，企业级应用只用于商业环境。

本章将介绍集成的企业级应用的范围和相关要求（请参阅第 2.1 节）。第 2.3 节将展示一些企业级应用的范例。第 2.4 节将介绍企业级应用的架构。第 2.5 节和第 2.6 节将介绍由此形成的数据访问模式。

### 2.1 当前的企业级应用

为了让大家了解一个集成企业系统所应涵盖的各种功能，我们在图 2.1 中展示了一家企业的各个部门。企业系统负责把分散在多个部门的业务流程集成起来（水平集成），它将计划和控制系统与事务系统连接起来（垂直集成）。

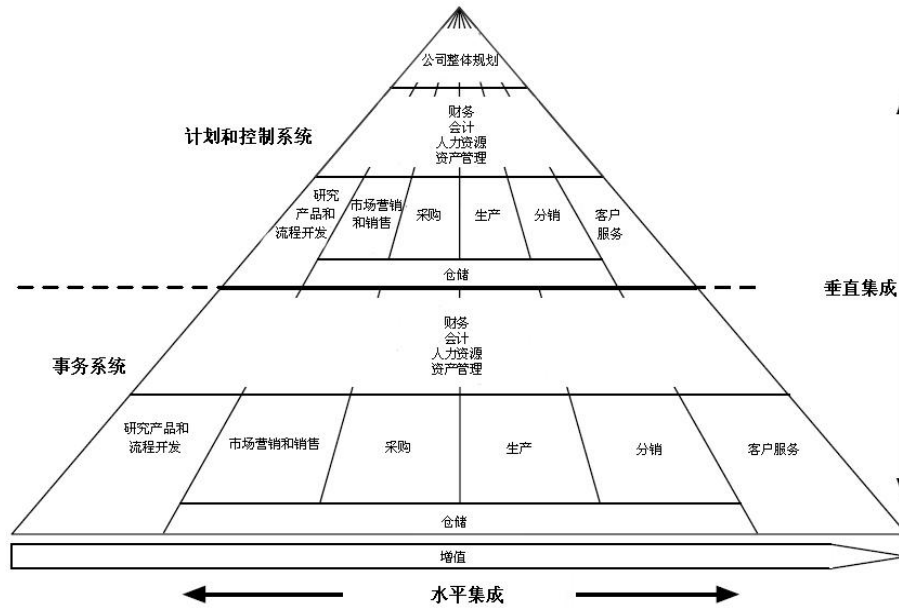


图 2.1 集成的企业信息处理（基于 [112] 和 [188]）

企业级应用不仅要集成同类公司的信息处理，还要迎合不同范围的客户群。例如，SAPBusinessSuite 是一款全球领先的集成企业资源计划（ERP）软件包，如今已广泛应用于各行各业，包括航空、国防、电信、银行业、工业机械和组件制造等各个领域。不同行业之间差异很大，有时不同客户的需求甚至是彼此矛盾的。例如，高科技行业的产品开发和上市生命周期不足 6 个月，而化工、铁路、或石油和天然气行业的相关生命周期可能长达数 10 年[50, 161]。各行业的公司都希望企业系统能根据自己的需求量身定制。要应对这种状况，成功的企业级应用必须实现参数化，并支持创建特定行业的解决方案，这具体包括，在当前系统没有包含客户需要的信息时，支持客户自己决定查询数据的方式，和扩展底层的数据表。企业软件的另一个重要的定制功能是，能够对现有业务流程进行调整，使其满足客户特定的需求。

跨国公司往往会横跨多个时区，这要求软件随时都是可用的，不能停机。企业级应用必须满足此要求。这包括提供从大规模故障（包括电力故障、数据丢失和用户输入故障）中恢复的机制。企业级应用如果在多个国家使用，则必须具有国际化特性，以使用户能够用母语与系统进行交互。此外，企业级应用还必须符合所有相关的国际国内法律法规。这意味着，软件要便于更新，以迎合定期修改的法律，如与所得税结构和税率有关的法律。另外，公司一般都需要保留数年的历史数据，企业级应用应当支持这种需要，同时确保不断增长的数据量不会对当前数据的处理产生影响。

## 2.2 企业级应用范例

以下列出的是一家公司可能会用到的部分企业级应用。我们会在本节后面作详细阐述，帮助非业务用户的读者大致了解企业级应用的功能。为了举例说明，我们参照的公司具有以下的使用情况。

- 需要为以后的计划期间预测产品需求。预测结果将被用来制定更详细的生产计划（需求计划）。
- 客户联系公司并订购产品（销售订单处理）。

- 每次接到新订单时，检查所需的产品，并通知客户能否装运以及何时装运产品（可承诺量）。
- 有些客户收到产品后没能及时付款。因此，公司必须跟踪付款情况，并针对未清发票生成催款提醒（催款）。
- 公司每周进行一次销售分析（销售分析）。

第 6.2 节将介绍目前尚未实现的创新型企业级应用。

## 需求计划

需求计划通过整合多个数据源来对未来销售做出预测。这些数据源包括当前和历史销售情况、新产品的介绍、产品停产、市场预测以及其他可能对购买行为产生影响的事件。由此而产生的结果将影响生产计划，生产计划用来确定产品制造的数量和日程表。

需求计划应用程序具有以下特性：公司运行一个单一的需求计划就会涉及大量的数据（较大公司的数据量可能多达 100 GB）。这些数据是基于成百上千种不同的产品，和精细到天或小时级别的计划层次而创建出来的。每类产品还包含各自的子类和不同的配置。

每次运行计划时，所有当前的数据都会聚集在一起，因此，对这些数据的主要操作是聚集和分解。然后，执行计划层次中更高级别的更改，并将这种更改逐层分解至最细的粒度。由于要读取所有数据，而只修改部分数据，所以这个步骤的读操作远多于写操作。由于计划人员需要以交互的方式在系统上持续工作，底层数据库系统执行上述操作时必须达到亚秒级的响应时间。此外，由于存在多用户并发访问的情况，必须为每一位用户创建独立的环境以确保数据视图的一致性。这会成倍提高对存储和计算能力的要求。

我们可以看出，上述应用场景是和分析相关，但它也需要实时访问事务数据。因此，事务型和分析型的解决方案必须合二为一。

## 销售订单处理

销售订单处理的主要目的是捕获客户订单。销售订单由客户名称、开票地址和支付方式等基本部分组成。订购产品的详细信息以行进行存储。销售订单处理包括对事务数据的读取和写入操作。读取操作包括核对库存、产品位置和价格，这些只是简单地读取操作，而写操作则需要插入新销售订单以及对应的行项目。读取操作比写操作更频繁，所有的交易都会涉及由高度可预测的查询来处理的少量数据。由于员工需要以交互的方式使用应用，销售订单处理必须实现快速响应。

## 可承诺量

要决定是否接受产品订单，必须先成功地完成可承诺量（ATP）的检查。ATP 应用程序可以确定在指定的交付日期[113]，现有库存和计划库存中的产品数量是否充足。如果要实现灵活一些的操作，比如用优先级高的订单替代已经接受但优先级低的订单，ATP 就必须汇总所有处理中的订单，而不只是考虑库存水平。ATP 要对大量数据执行读写操作。其中主要是读取操作，用来汇总各个时间序列。写操作在细粒度事务层次中执行，例如按照对用户的承诺申报产品，其数据类型和记录的变化很小，只有在添加新产品、客户或地址时，数据量才会增加。

## 催款

催款是指扫描未清发票、找出逾期未付款的发票、为相应的客户生成提醒通知，以及跟踪

通知发送情况的整个过程。催款涉及对大量交易数据的读取操作。读取操作可以用批量方式进行（例如逐个客户处理未清项目列表），也可以用分析查询方式进行。分析查询方式会导致数据聚集，例如，确定所选客户组在某会计年度逾期未付款的发票。读取操作只访问未清项目的少量属性。若要跟踪已发出的付款提醒，则需执行插入操作。在运行催款之后需要通过执行更新操作，修改客户的主数据，如当前的催款级别。

## 销售分析

销售分析提供历史销售数据的总体概况，其分析结果通常用于企业决策。在分析过程中会对大量交易数据和预整合数据进行只读操作。例如，销售经理可以按产品组和国家/地区分析销售量，并与上一年的数据做比较。销售分析包括对近期数据和历史数据的分析，其查询具有多维性，往往需要扫描和汇总大量数据。

## 2.3 企业级应用架构

当前的企业级应用由客户端设备、应用程序服务器和持久层组成。不同的企业级应用供应商在不同层中实施业务逻辑。例如，PeopleSoft 倾向于在客户端设备中实施业务逻辑，而 Oracle 则依赖于靠近数据库的存储过程，在持久层中实施业务功能。SAP 则采取了第三种方式，即：建立瘦客户端，将所有业务逻辑存放在可扩展的应用程序服务器中。SAP 方法的原理如图 2.2 所示。

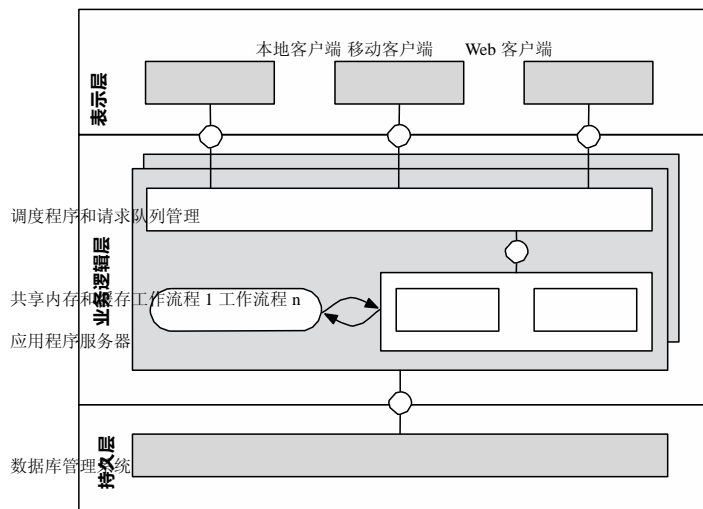


图 2.2 SAP R/3 的三层架构

在实际应用中，此图会更加复杂，因为公司往往会在数据库管理系统和应用程序服务器中添加系统冗余，以提高性能。20年前，当SAP R/3引入此模型时，其理论基础是：数据库服务器是一种限制因素，所以应该尽量将负载从数据库服务器转移至多个应用程序服务器中。通常企业级应用程序支持多个数据库管理系统，这使得只有部分被支持的系统可被用于应用程序开发。这可导致应用程序服务器变得既繁琐又复杂。一个主要的原因是，为了最大程度地降低数据库的利用率，通过引入应用程序缓存，存储数据库的临时结果，并把所有计算都放在这一层中执行。

## 2.4 企业级应用中的数据处理

企业级应用处理的数据存在多种不同的形式。我们的客户分析显示，中型企业的系统目前包含100 GB的事务数据和1 TB的只读分析数据。大型企业的系统包含35 TB以上的事务数据和40 TB的只读分析数据。

无论是出于计划和控制目的，还是为了创建管理层所需的报表，处理如此庞大的数据无疑会对系统的整体性能产生重大影响 [124]。为了减轻这种影响，目前大部分企业级应用都采取了双管齐下的方法：一方面，将其中部分任务分到BI应用程序中；另一方面，将另一部分任务转到日常事务处理应用程序中。

BI应用程序的特点是：对整个企业系统中的大部分数据进行读取密集型操作，包括定期运行预定义报表或执行复杂的、往往需要汇总大量数据的即席查询。这类查询可回应管理层所提出的关于公司当前状况或关注领域的问题，例如公司目前在某个国家/地区或区域的发展状况。在传统事务数据库模式下，这类查询往往效果欠佳，因为传统模式是专门为写操作而优化的。这意味着对各种表中存储的数据进行分析时，需要将这些表连接在一起，而这会严重影响性能。

BI应用程序先对数据进行预整合，再将整合结果存储在读取优化的特殊模式下，从而轻松解决上述问题。数据分析只能沿着预先选定的维度进行 [170]。在这些读取优化的模式下，特定属性会在多个表中重复出现，以减少所需的连接数。预整合的数据会定期更新，具体过程是：先从事务系统中提取数据，再根据聚集规则和读取优化模式对数据进行转换，最后将数据加载至BI应用程序中 [93]。

这样将企业系统分为执行OLTP和执行OLAP的系统，进而将企业级应用精确划分为事务处理应用程序和分析应用程序，这是与实际情况相悖的做法，因为我们刚才提到的应用程序和其他应用程序均同时执行事务处理和分析任务，并没有将二者分开。

## 2.5 企业级应用中的数据访问模式

提到对底层内存数据管理系统的影响，需要学会区分对企业实体的单一实例的处理和对一组实例的属性的集合处理。单一实例可以指一个销售订单或一个客户。如果需要读取所有逾期未付款的发票，或者按销量显示前十名客户，则需要采取集合处理。尽管事务系统仅对对象的单独实例进行操作，但在集成企业级应用中，大部分数据都是通过集合处理被使用的 [99]。

企业级应用涵盖运营系统以及计划和控制系统的（见图 2.1）。因此，我们可以这样总结：企业级应用具有混合的工作负载 [100]。

## 2.6 结论

从本章中，可以明显看出，企业级应用需要执行各种各样的任务并处理海量的数据。这样的数据量会限制分析功能获得数据的时效性，从而导致我们需要单独的分析系统。但是，由于技术原因而将企业级应用分为两个独立系统的做法存在很多缺点，因此本书的核心主题是

探讨如何运用各种技术和技巧，避免建立单独的分析系统。这样，企业级应用的用户才能实时访问企业系统中的所有数据，并获得单一可靠的信息源。