模块化系统：可靠性的发展

第 76 号白皮书

版本 1

作者　Neil Rasmussen 和 Suzanne Niles

> 摘要

从自然的发展规律来看，模块化设计是复杂系统赖以生存和发展的关键要素。而这取决的一个重要因素是借助了*容错*这个关键可靠性优势，它可以确保模块化系统在故障模块维修期间改用其它性能完好的模块，以确保系统的正常运行。在数据中心，模块化设计已经植根于服务器和存储系统适用的全新容错架构中。随着数据中心不断演进以及受到大自然发展规律的启发，数据中心物理基础设施（DCPI）仍须不断完善，以为企业新的生存、恢复和发展策略提供支持。

目录

点击内容即可跳转至具体章节

APC 白皮书现收录于施耐德电气白皮书资料库
由施耐德电气**数据中心科研中心**发表，
DCSC@Schneider-Electric.com

Schneider Electric

简介

模块化是一项用于组织并简化复杂系统的成熟技术。从最简单的（手电筒电池）到复杂的（生物体细胞），模块化的应用取得了难以企及的成功。然而，对于处于从单块集成电路到模块化设计演进过渡边缘的人工系统，各方质疑声不断，而且进展缓慢，直到模块化的概念深入人心并展现出其久经考验的优势。

数据中心物理基础设施（DCPI）正处于这一过渡阶段。虽然采用结构单元架构的优势显而易见 — 它是可扩展的，灵活的，简易的、便携的 — 这点很容易理解，而且争议不大；然而，业内应用的模块化设计，却有一个要素成为了争议的焦点：可靠性。

对于这种全新的设计方式，使用传统而简单的可靠性分析方法（即“零配件越多，发生故障的风险则越大”的观念）说好听点是不具备全面性，说难听点根本是一种误导。本白皮书旨在通过举例的形式来说明模块化不仅体现了其最突出且易于理解的优势，而且还揭示了其最微妙、最鲜为人知、意义深远的可靠性优势：容错性。模块化设计特有的容错功能对于预防故障意义重大，从而其在复杂系统中也发挥了充分而出众的作用。

大自然进化案例：远古生物体

与数据中心或手电筒电池的诞生相比，模块化的历史更加悠久。极早期的非模块化系统 — 单细胞生物体 — 在三十亿年前便生活在地球上。这些生物体的化石记录表明，它们的外壳、触须、嘴、臂膀、钳子以及众多其他复杂的结构不断发展进化。有些生物体形庞大，宽度高达 15 厘米（6 英寸）。这些复杂的单体化单细胞设计数十亿年来主宰了地球的食物链。

大约五亿年前，多细胞生物体诞生。仅仅在几千万年间，它们进化迅速，赶超历经三十亿年进化历程的复杂单细胞生物体，取代它们成为了主导性的生命形态。

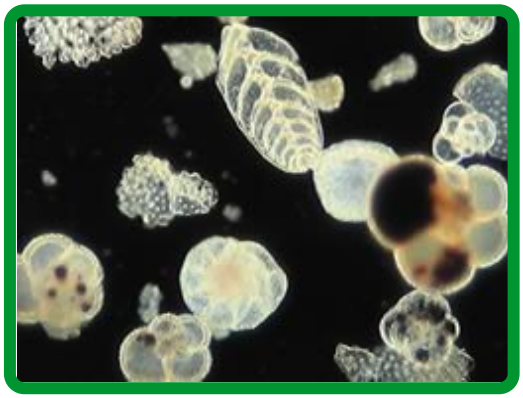


图 1

远古复杂的单细胞生物体

多细胞生物体的模块化优势

模块化多细胞生命形态何以战胜根深蒂固的单细胞生命形态？

- **繁衍和进化能力**：可通过标准接口添加可与现有模块交互作用的新模块（细胞），使系统得以从规模或者功能上扩展升级。
- **简化繁殖流程**：与繁殖单个复杂的细胞相比，繁殖大量较小的、较为简单的细胞相对较为简单、快捷，而且更加可靠。
- **模块功能专门化**：细胞司职的委派和专门化可实现团队合作时特有的效率和效能。在远古的多细胞生物体中，其中一种细胞负责运动，另一种细胞发挥防御作用，还有一种细胞负责觅食等等。
- **快速适应环境**：通过增加、减少或改造细胞，可加快检验分阶段设计变更的速度，要么通过，要么否决。

- **容错性**：有了细胞冗余，即使在单个细胞可能发生故障后，也不会减弱系统的机能，与此同时，可让细胞进行修复，而且不会导致系统发生故障（这里所指的是伤残甚至灭亡）。

图 2

远古多细胞生物体

上面提到的最后一个属性，即容错性，是模块化系统优于单体化系统的一个重要的可靠性优势。模块化将系统拆分成一个有许许多多小型组件构成的装置，这将有助于优化组件的冗余，因此，当系统中的一个，甚至多个组件发生故障时，也不会对系统的正常运行产生不利影响。轻微划伤时，人体皮肤细胞将会损失几百个细胞，但是却不会影响到我们的健康。因为在受损细胞进行修复时，其他细胞会正常的“各司其职”。模块化并不是人类发明的产物 — 我们本身就是一个模块化的机体。由于人均有数以万亿个模块（细胞）构成，因此我们无时无刻不盛享容错性带来的好处。

IT 演进案例：磁盘驱动器

在大型数据中心盛行的年代，存储设备使用的是许多个 14 英寸金属盘堆叠而成的大型专用硬盘，其读/写机制复杂精密，机柜的大小与洗衣机不相上下。1978 年，IBM公司提出了划时代的构想，即使用小型磁盘阵列，但是他们并没有将此构想付诸实践，因为他们认为它的可靠性无法与传统的单体化设计相提并论。再加上容错性研究与实践刚刚起步，主要的局限还来自于电子系统元件故障可能需要以付出生命为代价的航空航天行业。[1]

1987 年，伯克利分校的研究人员注意到计算速度和存储访问速度之间的差距不断拉大，外部磁盘驱动器开始被应用于个人电脑中，他们将此视为将它们用作系统的构件以提高数据传输速度的契机。一年后，他们发表了具有里程碑意义的学术论文《廉价磁盘冗余阵列方案》，提出了多个有关将此类阵列用于存储、检索和恢复数据的数据写入计划（“RAID 级别”）。1990 年，通过将 5.25 英寸磁盘应用于个人电脑中，理论和硬件紧密地结合在了一起，磁盘发展到了一定高度，它们具备了可在首个 RAID 阵列中使用的容量、性能和可靠性。这些新型的模块化存储设备在冗余和读/写速度之间取得了平衡，它们取代了大型存储设备，成为了地板空间的组成部分。

图 3

RAID 阵列

[1]如今，当 IT 系统发展成为几乎所有行业（包括医疗卫生和军队）的核心，数据中心变成了关键任务的应用环境，故障可能会导致整个数据中心损坏。因此，容错性变成了数据中心设计的首要考虑因素，甚至超过了原本首要首当其冲的经济利益。

RAID 阵列的模块化优势

模块化 RAID 阵列何以战胜传统的单体化存储设备？

- **扩展和扩容能力**：可通过为每个阵列增加模块数量或添加阵列来扩大存储容量。
- **简化制造流程**：与制造传统的复杂大型驱动器相比，制造多个用作 RAID 模块的小型驱动器的工序较为简单。
- **模块功能专门化**：阵列中的驱动器可用于增加存储容量，提高访问速度，或增加冗余，所有这些均视该阵列的 RAID 级别而定。此外，RAID 阵列本身可用作更高层面的模块，在此情况下，需要将不同的应用程序分配给每一个 RAID 阵列。
- **快速适应环境**：可添加或拆卸驱动器；可根据容量、速度和冗余之间的预期平衡轻松更改 RAID 级别。
- **容错性**：RAID 数据写入计划包括可在其中一个驱动器发生故障的情况下恢复数据的冗余。

令 RAID 设计者讶异的是，RAID 在市场上广受欢迎的原因并不在于其增速的优势 — 这原本是设计 RAID 的初衷 — 而是其兼具的容错能力的高可靠性。直到 1988 年发表的学术论文《廉价磁盘冗余阵列方案》的作者提到 RAID 设计可能具有容错性 — 在现场演示时，他们拆除了一个驱动器，然而阵列仍然如常运行 — 当时普遍对可靠性的预容错能力的具有代表性的理解都是一种误解：即由于多驱动器系统的零配件较多，因此，它的可靠性较低。

IT 演进案例：刀片服务器

正如本白皮书所述，刀片服务器是从单体化设计过渡到模块化设计的进程的核心。多年来，传统的独立服务器不断扩容和提速，并随着网络计算的扩容而担负了越多越多的任务。随着需求的增加，新型服务器被添加到数据中心中，这通常只是权宜之计，而且缺乏协调和规划；对于数据中心操作人员而言，在不知情的情况下，通常难以发现服务器被添加到数据中心中。这样一来，将会增加配线盒和布线的复杂性，从而不可避免地造成混乱，导致出错而且缺乏灵活性。

第一台刀片服务器于 2001 年首次面市，这是一个纯粹简单的模块化架构的典型例子 — 放置在机柜中的刀片服务器实际上都是相同的，它们配备了相同的处理器，随时可供配置而且可由用户确定使用意图。它们的推出将模块化的诸多优势引入到了服务器环境中 — 即可扩展的、制造简单的、功能专属的和可适应性。

不过，虽然这些典型的模块化优势使刀片服务器广泛应用于数据中心，但是，模块化设计的另一大关键功能却有待充分发挥其潜力：容错性。具有容错功能的刀片服务器 — 在某些刀片服务器发生故障的情况下可通过内置的“故障转移”逻辑改用其他刀片服务器 — 这是最近才开始推出的而且具有很高的性价比。具有容错功能的服务器的可靠性远胜于最新的冗余软件和服务器群组提供的可靠性，使刀片服务器成为了数据中心中占有主导地位的服务器架构。随着自动化容错功能的兴起，行业观察家预测，在未来五年内，刀片服务器将会普及应用于数据中心中。

图 4

传统的服务器

刀片服务器的模块化优势

模块化刀片服务器何以战胜大型独立的服务器？

- **扩展与扩容能力**：可通过添加模块（刀片服务器）轻松地扩展计算容量。
- **简化制造流程**：与制造整个服务器相比，制造多个小型刀片服务器的工序较为简单。电源、制冷风机、网络接口和其他支持组件均集中在机柜内并为多台刀片服务器所用，从而简化刀片服务器的结构。
- **模块功能专门化**：用户可根据需要通过软件应用程序配置刀片服务器。
- **快速适应环境**：可以根据业务需要或财政预算添加或拆除刀片服务器，也可以重新配置刀片服务器，以运行不同的应用程序。
- **容错性**：可通过内置的“故障转移”逻辑自动处理刀片服务器的故障，并无缝地改用其他刀片服务器。

图 5

刀片服务器（装有 10 台刀片服务器的机箱）

IT 系统故障定义的变化

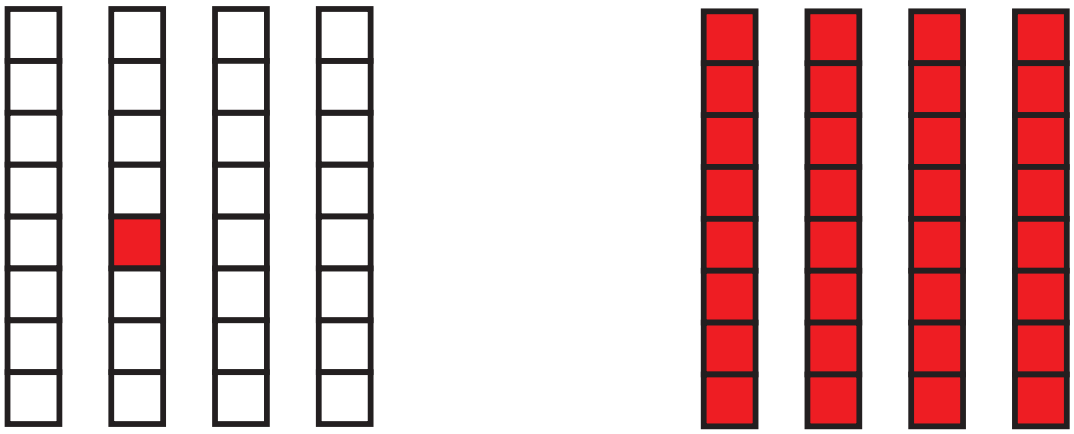
上面举的三个例子阐述了模块化设计远胜于复杂单体化设计的原因，这些因素其实都是模块化最根本的且固有的特性。其中一个便是容错性，这对于数据中心的未来意义深远。一旦数据中心中的服务器和存储设备变成了具有容错能力的设备，那么，它将会改变 IT 故障的定义方式。

以数据中心的两种故障情形为例（**图 6**）。右图是全部机柜均发生故障，原因是单台保护整个数据中心的大型 UPS 发生了故障并切断了负载供电。左图是单个机柜发生故障。在传统的数据中心中，这两种故障情形都被 IT 经理归类为发生故障，这是因为单个机柜发生故障时，服务器、磁盘阵列、交换机和路由器之间的相互依赖性很可能会引起连锁效应，从而导致整个数据中心宕机。

由于在计算和存储设备使用的新型阵列式设计的概念深入人心，因此，左图的故障形式 — 即单机柜发生故障 — 开始被 IT 经理视为一种“理想的”故障，因为即使单个装置发生故障，资源的冗余目前也可确保数据中心正常运行。由于容错架构广为应用，因此，数据中心将具备多设备容错功能，而且不会导致整个系统宕机。当刀片服务器履行其早期的无缝容错承诺时，一个、两个、三个，甚至更多个的机柜故障将不会对整个数据中心的正常运行产生任何影响。

图 6

数据中心的两种故障情形（俯视图、4 行机柜，每行 8 个）



单个机柜发生故障

全部机柜发生故障

DCPI 的影响

这种新的故障管理模式 — 认定某些模块必然发生故障，而为确保其正常运行已经做好万全准备 — 预示着新的 IT 架构应当受到数据中心物理基础设施的保护。比如，随着数据中心 IT 层面的容错功能性增强，因此，由单台大型 UPS 保护电源变得不甚理想，原因在于，如果 UPS 发生故障，那么，整个系统将会停止运行 — 而对于具有容错功能的数据中心，如果单个机柜发生故障，并不会遭遇此问题。如果数据中心每个机柜使用一台 UPS，那么，单台 UPS 发生故障将只会导致单个机柜发生故障，而不是整个系统。而且，增加 UPS 的数量，仅仅只是提高了单台 UPS 发生故障的可能性，此类故障对于整个系统来说是可接受的。比如，如果三个故障机柜即会导致系统停止运行，那么，当三台 UPS 同时发生故障时，便可能会导致系统宕机，而这种情况发生的可能性很低 — 其概率远低于单台大型 UPS 的故障发生率。基于此，鉴于 IT 系统的容错能力不断增强，因此，主张可靠性理论更趋向于采用模块化分布式供电和制冷架构。

单体化和模块化 DCPI

自三十年前数据中心推出以来，数据中心物理基础设施（DCPI）的架构基本保持不变。从最小型的计算机机房到最大规模的企业设施，物理基础设施的主流发展一直固守于用于保护电源和制冷系统的集中式“装置”。这种设计方式要求对设备和接口进行特殊的单体化配置。通过以模块化设计取代这种架构，DCPI 不仅可以为模块化容错 IT 设备提供适当的支持，其设备本身还可以尽享容错这一显著的模块化优势。

图 7

集中式单体化 UPS

DCPI 的模块化优势

模块化 DCPI 何以取代传统的单体化 DCPI？

- **扩展和扩容能力**：可根据数据中心当前的 IT 需求调整模块化 DCPI 的大小，而且 DCPI 可随着需求的增长进行扩容。此优势对于 DCPI 至关重要，传统的方法则是一次性将用以支持预计最大 IT 需求的供电和制冷系统部署完毕，这会造成投入成本和运营支出庞大。
- **简化制造流程**：模块化设计意味着须制造大量的小装置，而不是少量的大装置。产量越大，缺陷则越少，设计越小、越简单，则制造过程中的自动化程度越高，涉及的手工活越少，因此，缺陷就越少。
- **模块功能专门化**：可针对数据中心不同组件的可用性和制冷要求根据各种配置制造电源保护和制冷装置。
- **快速适应环境**：由于新添置的设备和 IT 设备每 2 至 3 年须更换一次，因此，必须不断地更换数据中心内的布局。新设备的尺寸或形状、供电或制冷要求以及插头等等可能都会不同。对模块化 DCPI 进行扩容或重新配置较为简单，可以满足不断变化的 IT 需求。
- **容错性**：由于具有容错功能的 IT 设备可允许数据中心在 IT 组件发生故障的情况下继续正常运行，因此，具有容错功能的 DCPI 设备可允许供电或制冷系统在 DCPI 组件发生故障的情况下继续正常运行。可通过 DCPI 装置的冗余或 DCPI 装置内的组件内部冗余发挥容错功能 — 例如，通过将其他功率模块添加到 UPS 中来实现容错功能。

和前面几个举得模块化例子一样，前四个特点是模块化设计取得成功的前提，但是第五点 — 容错性 — 是至关重要的。此外，由于数据中心的运行完全依赖于供电和制冷，因此，容错可靠性对于 DCPI 及其保护的 IT 设备一样重要。无容错 DCPI 的容错数据中心就好比是地基坚固而缆索脆弱的吊桥。

图 8

机柜式模块化 UPS

结论

从单体化设计到模块化设计的过渡是复杂系统必然的演进，因为模块化在效率、灵活性和可靠性方面极具优势。通过对各个成功案例的研究，我们可以更清楚地认识到模块化具有的潜力，以对自构建以来一直采用单体化设计而从未用其他方式方式进行思考的系统带来重大的、甚至创新性的改进。模块化的容错性和其他关键的特性 — 扩展性、适应性、专门化、制造简单 — 这是模块化的固有特性，在模块化的人工系统中得以淋漓发挥，是系统发展的必然结果。

在 IT 的世界里，模块化设计在存储和计算方面（RAID 阵列和刀片服务器）彰显的优势不容忽视。更为重要的是，数据中心正准备追随诸如航空航天等行业的发展步伐，在系统的部署过程中充分发挥模块化的优势 — 自上世纪七十年代以来，模块化的容错性特点被广泛应用于关键任务的应用系统中。透过容错性，我们认识到严格控制组件的质量仅仅是确保系统可靠性的第一步，而在组件发生故障时维持系统的正常运行才是可靠性的最终策略。

由于模块化和容错性成为了数据中心设计的新方向，因此，数据中心物理基础设施的发展也必须朝此方向，如此才能有效地保护这些数据中心并发挥模块化在效率、灵活性和可靠性方面的优势。

关于作者

Neil Rasmussen 是施耐德电气旗下 IT 事业部—APC 的高级创新副总裁。他负责为全球最大的用于关键网络设备（电源、制冷和机柜等基础设施）科技方面的研发预算提供决策指导。

Neil 拥有与高密度数据中心电源和制冷基础设施相关的 19 项专利，并且出版了电源和制冷系统方面的 50 多份白皮书，其中大多白皮书均以 10 几种语言印刷出版。近期出版的白皮书所关注的重点是如何提高能效。他是全球高效数据中心领域闻名遐迩的专家。Neil 目前正投身于推动高效、高密度、可扩展数据中心解决方案专项领域的发展，同时还担任 APC 英飞系统的首席设计师。

1981 年创建 APC 前，Neil 在麻省理工学院获得学士和硕士学位，并完成关于 200MW 电源托克马克聚变反应堆的论文。1979 年至 1981 年，他就职于麻省理工学院林肯实验室，从事飞轮能量储备系统和太阳能电力系统方面的研究。

Suzanne Niles 是施耐德电气数据中心科研中心的高级战略研究员，加入数据中心科研中心之前，Suzzane 在卫斯理女子学院（Wellesley College）从事数学方面的研究，而后 在麻省理工学院（MIT）获得计算机科学学士学位，并发表关于手写输入识别的毕业论文。Suzzane 拥有超过 30 年针对不同阶层听众，包括上至软件说明书，摄影图片，下至儿歌的多元化的教学经验.

资源

点击图标打开相应参考资源链接

数据中心物理基础设施：优化业务价值
第 117 号白皮书

数据中心物理基础设施的标准化和模块化
第 116 号白皮书

浏览所有 白皮书
whitepapers.apc.com

浏览所有 TradeOff Tools™ 权衡工具
tools.apc.com

联系我们

关于本白皮书内容的反馈和建议请联系：

数据中心科研中心
DCSC@Schneider-Electric.com

如果您是我们的客户并对数据中心项目有任何疑问：

请与您的 **施耐德电气**销售代表联系