

IT@英特尔

以数据智能化创新改变英特尔安全状况

英特尔全新网络智能化平台提供了一个背景丰富的环境，为整个信息安全部门带来价值，以数据优势推动信息安全保障机制变革。

“通过合理的数据布局和员工再培训，我们取得了事半功倍的效果。”

— Brent Conran
英特尔首席信息安全官

英特尔 IT 部门作者

Jac Noel

安全解决方案架构师

Todd Glasgow

产品负责人

Victor S. Colvard

信息安全工程师

Dennis Kwong

信息安全工程师

Ted Mahar

全球网络响应

Eric M. Monroe

数据架构师，数据科学家

Elaine Rainbolt

行业合作经理

Aubrey Sharwarko

数据科学家

执行概要

英特尔 IT 部门部署以 Splunk* 和 Apache Kafka* 等先进技术打造的全新网络智能化平台 (CIP)，推动英特尔信息安全 (InfoSec) 保障机制的变革。新平台从数百个来源和安全工具中提取数据，围绕数据提供背景丰富的可见性以及通用语言和工作界面。这极大地提高了英特尔整个信息安全部门的生产力，大大提升了工作效率和效能。平台提供的实时数据访问、流处理、机器学习工具、统一数据模型以及编排和自动化等诸多功能，能够有效降低识别复杂性日增的威胁及做出响应所用的时间，并最终加快获取洞察的速度，以尽早防御。

我们的团队仅在短短五周内就完成了这个大数据解决方案的部署，并立即开始实现商业价值。CIP 基础设施基于英特尔® 至强® 铂金处理器、英特尔® 3D NAND 固态硬盘和英特尔® 傲腾™ 固态硬盘，可为安全专家提供所需的计算能力，帮助他们更快、更智能地获取洞察，同时缩短安全工具之间切换所需的时间。

我们 CIP 的一些核心优势包括：

易于实施，促进人才快速发展 	横跨整个信息安全部门的通用工作界面 	数据分类、通用语言和即时搜索 	重要网络阵地信息安全部门已经为开发运维做好准备 
按需模式和自动数据归一化 	完整的威胁分类和杀伤链可见性 	轻松整合自定义的第三方安全工具 	连接至开源机器学习库 

目录

背景	2
英特尔 IT 部门的新 CIP	3
快速部署最佳实践	5
解决方案架构	8
未来举措	11
结论	11

英特尔首席信息安全官 Brent Conran 表示：“部署和扩展这种规模的网络数据分析平台并非易事。我们淘汰了有着 17 年历史的安全信息和事件管理系统以及用了 9 年的日志基础设施；引入全新数据湖，并实现了工具的现代化。通过合理的数据布局及员工再培训，我们取得了事半功倍的效果。当前，我们正在运用人工智能和机器学习大幅提高网络智能化的深度和速度。”

背景

英特尔 IT 信息安全 (InfoSec) 部门于多年前首次实施了旧版安全信息和事件管理 (SIEM) 系统及日志数据仓库。尽管系统在最初部署时非常先进，但是并没有随着现代行业标准的发展而演进或扩展，因此出现了一些问题：

- **可用性差：**开发新的检测逻辑、查询或高级统计分析需要大量专业知识。这意味着大多数部门必须依靠少数专家来创建更改任务、搜寻威胁或提取事件响应所需的数据。我们在新员工培训上投入了大量资金，但在他们精通这些工具后，有时会出现人才流失的情况。
- **数据瓶颈：**由于数据处理能力非常有限，即使是简单的数据转换、扩充或过滤也困难重重，甚至根本无法实现。容纳新的数据源也并非易事：随着诸如物联网 (IoT)、云和智能建筑等技术的飞速发展，每天都会涌现新的数据源。
- **数据不一致：**由于数据像孤岛一样存储在各处，因此不同的信息安全团队对数据往往会有不同的解读，从而导致为防御、检测和应对安全威胁所做的努力毫无章法。

- **巨大的技术债务：**数据湖因为数据没有归一化而沦为数据沼泽，由于自定义程度高导致难以维护。不同的信息安全团队使用多种不同的解决方案来分析相同的数据，往往导致结果不同和/或彼此冲突。

上述这些挑战使旧版系统无法跟上安全威胁瞬息万变的发展态势，无法针对安全威胁与漏洞实施有效防御、检测和响应，而且它仅能服务于信息安全部门中的一小部分用户。

新解决方案需要一个每天能从大量数据源中近乎实时地获取 TB 级数据，且富有弹性、高度可用、可扩展的平台。整个信息安全部门需要能够协同工作，快速轻松地使用和构建搜索查询、检测逻辑和仪表盘，由此更好地识别威胁、对其分类并减轻威胁，高效传达信息安全衡量标准。我们需要一个功能远超传统 SIEM 的新平台，以此提高信息安全部门在漏洞管理、安全合规和执行、威胁搜寻、事件响应、风险管理等各个方面的效力；我们需要一系列先进且开放的现代化功能，使数据对我们整个信息安全部门来说更有价值、更易用且更易访问。

英特尔纵深防御策略

自动化的防御、检测和响应功能可应对 99% 的威胁，使英特尔威胁搜寻专家能腾出时间专注于追捕余下 1% 试图渗透到英特尔环境中的高级威胁。

英特尔 IT 部门运用大量自动化工具和功能来保护英特尔免受安全威胁并对安全威胁做出响应。我们的策略是在多个层面实施保护，包括外围、网络、端点、应用和数据。我们专注于防御、检测和响应，同时也推动员工了解和参与有关信息安全的最佳实践。

我们的纵深防御策略可以成功检测大多数威胁并及时补救。但是，高级网络安全威胁的频率和复杂性正不断上升。要根除这 1% 的威胁（这些威胁太过复杂，因此能够逃过广泛的自动化安全环境），就需要为威胁搜寻专家提供从多个来源和工具收集的数据。我们新的网络智能化平台 (CIP) 提供了一种集成数据源和多种工具的方法。



- ➡ **了解更多信息：**“IT@Intel: Advanced Persistent Threats: Hunting the One Percent” (IT@英特尔：高级持续性威胁：狩猎百分之一)



英特尔 IT 部门的新 CIP

替换旧版系统的第一步是明确关键要求、评估各种解决方案、教育引导决策者了解有关技术和商业价值，以获得他们的承诺和支持。

关键要求

早在 2017 年，我们就开始根据自身的键要求评估商用产品和开源技术：

- **强化安全运营中心的能力：**我们的事件响应和威胁搜寻人员需要具备按时间表开发和部署新检测逻辑所需的访问权限和知识，而无需等待专家或数据分析开发人员的参与。
- **进行扩展的能力：**我们拥有数量大、种类多的安全数据源。最初，我们需要平台每天能够提取约 12 TB 的数据，这相当于每天 220 亿个安全事件或每年 8 万亿个事件。但鉴于对数据增长的预期，我们需要平台在可预见的未来能够扩展到每天提取 50 TB 以上，同时大部分数据的留存期能达到 12 个月。
- **高效的数据共享和转换：**我们需要一条消息总线来大幅减少数据整合工作，推动安全系统之间的数据有效共享，实现数据过滤、数据充实和数据流转换。
- **高效的界面：**我们需要一个易于使用、类似于搜索引擎的用户界面，它具备搜索、报告、分析和可视化等现成功能，因而能促使整个信息安全部门迅速采用。
- **高性能：**无论事件发生在一小时前还是九个月前，信息安全方面的用户都必须能够快速找到所需的数据。
- **灵活性：**我们需要一个能够适应快速变化的平台，以便我们能够针对出现的变化（例如，出现新的数据源或因产品更新或配置更改导致日志记录发生变化）快速调整数据模型和解析规则。

- **符合行业标准：**我们需要一个被广泛采用的平台，以便能够利用由大量现成的安全工具和功能组成的大型生态系统。一个强大的生态系统包括用户社区、公共文档和共享软件、第三方培训、第三方集成和支持以及其他能够增加产品价值的元素。
- **可延伸：**英特尔致力于混合多云模型。解决方案必须能够轻松扩展到整个企业，并扩展到公有云和软件即服务 (SaaS)。
- **高度可用性：**英特尔信息安全保障工作从不停歇。解决方案必须能够全年不间断运行，并且不受站点和数据中心中断的影响。

解决方案概述

在评估了许多现有开源技术之后，我们为新 CIP 选择了以下主要组件：

- **Splunk Enterprise***，用于数据湖和通用工作界面。Splunk Enterprise 可从各种来源提取数据并建立索引，然后馈送至可搜索的存储库中，用户可从中生成衡量标准、报告、警报、仪表板和可视化信息。它包含一组围绕着防火墙、身份验证和数据防丢失等通用安全数据源设计的数据模型。Splunk Enterprise 的作用是我们 CIP 的“中枢神经系统”。
- **Splunk Enterprise Security*** 在 Splunk Enterprise 的基础上增加了 SIEM 功能。
- **Splunk IT Service Intelligence*** 为我们的重要网络阵地 (Key Cyber Terrain) 提供运营指导，由此检验我们的关键安全功能是否始终在按预期运行。
- **Splunk Phantom*** 专门用于安全编排和自动化响应，能够实现常见任务的自动化，从而使信息安全员工能够腾出时间从事更有价值和更复杂的活动¹。
- **Apache Kafka*** 用作企业级消息总线。Kafka 是一个发布订阅 (Pub/Sub) 消息传递系统，提供了一种高吞吐量、低延迟的解决方案，供提取和生成数据源以及执行流内数据转换（也称为流处理）之用。Kafka 充当那些跨信息安全系统的数据的“循环系统”。
- **高性能英特尔® 架构**为我们的 CIP 提供了必需的硬件基础，其中包括英特尔® 至强® 可扩展处理器和英特尔® 固态硬盘（英特尔® SSD）。

更多信息，详见“[解决方案架构](#)”。

¹ 英特尔的信息安全部门目前正在进行 Splunk Phantom* 的生产试点。

除了支持数据充实和过滤外，Kafka 还允许数据“一次获取，多次使用”，从而在我们所有安全功能上实现规模经济。Splunk 是众多的数据使用程序之一，主要用于数据的分析、可视化和报告。目前，我们已将上百个不同数据源输入到 Kafka 和 Splunk 中。这些数据源的例子包括超过 200,000 台客户端设备、800,000 台服务器、数百个防火墙、Web 代理服务器、网络监控工具，以及许多其他类型的上下文数据，例如 IP 地址映射、地理位置数据和人力资源数据。

如图 1 所示，我们的 CIP 为整个信息安全部门提供支持。我们发现，实现工具现代化和将数据移至集成了数据分析功能的平台可以提高信息安全员工的效力。我们的 CIP 是整个组织的增效器。



解决方案的优势

CIP 在网络安全威胁日益复杂的环境中为英特尔带来更高的安全性。我们教育引导决策者认识到平台具有以下优势：

- **敏捷：**CIP 使我们能够在几小时甚至几分钟内完成事件检测并做出响应。而之前以较旧技术构建、采用点对点集成并通过串行方式传输数据的 SIEM 系统往往需要数日或数周才能完成相同的任务。例如，现在我们从客户端设备收集数据时，不必等待端点

检测和响应 (EDR) 系统来处理数据。实际上，最近我们在已完成数据收集的 EDR 系统意识到威胁（黑客工具）之前，就已根据关联规则检测到了它。正常情况下，可能要花费 EDR 系统数小时才能检测到的威胁，我们的 CIP 几分钟内就已检测到。

- **高效：**得益于我们的 CIP，威胁情报工作也得到增强。我们可以在消息总线上发布入侵威胁指标（例如 IP 地址或文件哈希值），这会使所有安全系统（例如防火墙和代理服务器）都有机会自动使用总线中的威胁信息更新其内部威胁表单。
- **易于实施，促进人才快速发展：**Splunk 是日志管理的行业领头羊，部分原因是其能提供出色的用户体验、有益的支持文档和供应商提供的培训课程。即使是之前从未使用过 Splunk 的人员，生产力也能得到提高。此外，由于 Splunk 在行业中普及率较高，我们还可以聘请外部人才来帮助我们加速平台的实施和采用。
- **横跨整个信息安全部门的通用工作界面：**该平台大幅减少了事件响应者在执行数据分析时所需的控制台数量，由此大大缩短了事件响应时间。威胁形势正不断变化；由于 CIP 优化了数据访问，因此我们所有事件响应者都可以快速创建或更新用于识别感兴趣事件的关联搜索和检测逻辑。他们可以访问大量历史数据以快速开发和测试检测逻辑。在检测到事件后，他们可以

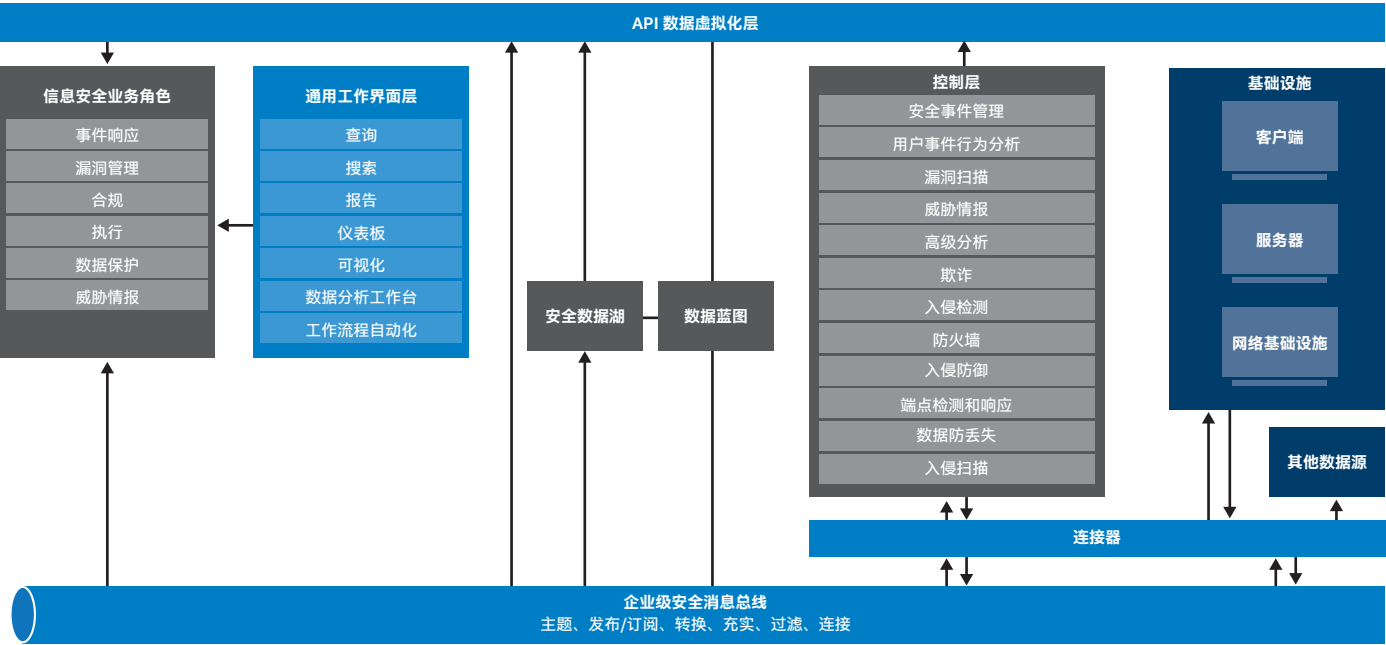


图 1. 我们的 CIP 通过将所有数据源集成在单一的通用工作界面，为整个信息安全部门和所有安全管控职能部门提供支持。

轻松搜索 TB 级的数据，从中找出事件发生的根本原因。我们还可以快速生成衡量标准，帮助我们了解关联搜索的有效性，而这正是上一代 SIEM 工具所缺乏的功能。

- **数据分类法、通用搜索语言和按需模式：**整个信息安全部门可以使用相同的语言，轻松地从我们的 CIP 中检索数据。信息安全部门的每个人都有能力，同时也具备相关技能，可以直接利用数据开展工作，从而提高其工作效力和整个部门的效率。
- **开发运维准备就绪：**Splunk 支持数据源的逐步融合以及集成自动数据归一化的按需模式功能。这些特性正是我们实现“提速”的关键因素。更多信息，详见“[敏捷的开发运维流程](#)”。
- **减少技术债务：**通过将旧版应用迁移到 CIP，我们可以大幅减少过时、冗余或自定义的应用。这可以让我们把技术资源集中在价值更高的解决方案上。
- **精细控制：**使用 Splunk 索引，我们就可以基于“必知 (need to know)”的业务案例来授予选择部分数据存储库的权利。我们还可以管理信息安全以外团队的查看权限。例如，业务部门风险经理可以访问与其部门相关的信息，而不必访问任何其他安全数据。

这些优势结合起来（见图 2）使我们能够比以往更快地检测到复杂的威胁并做出响应。我们使用 CIP 提供近乎实时的响应，这是我们确保英特尔合规且安全的关键组成部分。

快速部署最佳实践

我们运用了大量最佳实践，因此仅用五周就将 CIP 部署到了生产环境中。以下各节概述了我们完成此任务的过程。

广泛的范围：人员、数据和技术

虽然我们的 CIP 本质上是一项技术，但依托技术带来价值的是人员和数据。如图 2 所示，我们通过整合技术、人员和数据实现信息安全保障机制的变革。

我们使用功能丰富、面向未来、扩展能力远超我们当前所需的先进技术，帮助我们发现和应对新型威胁。大家都知道如何利用这些行业标准技术，因此它们可以事半功倍地推进协作、效率和采用。例如，Splunk 有着庞大的用户社区，因而雇用对 CIP 的关键组件有使用经验的人士较为容易。在英特尔，知道如何使用 Splunk 的人远远超过之前知道如何使用旧版 SIEM 系统和日志数据仓库的人。配备应用商店和连接器的开放生态系统是加速平台协作和落地的催化剂，有助于和其他标准及框架（例如开源机器学习库）保持一致。

在整个信息安全部门中使用通用查询语言和通用数据湖为新 CIP 在各信息安全团队中的广泛采用提供了便利条件。此外，Splunk 产品具有高度集成的功能，例如能够与我们现有的开发运维事件管理和 IT 警报平台集成。解决方案的实时流式传输功能使我们能够更快地检测和响应，并获得更多出色洞察。

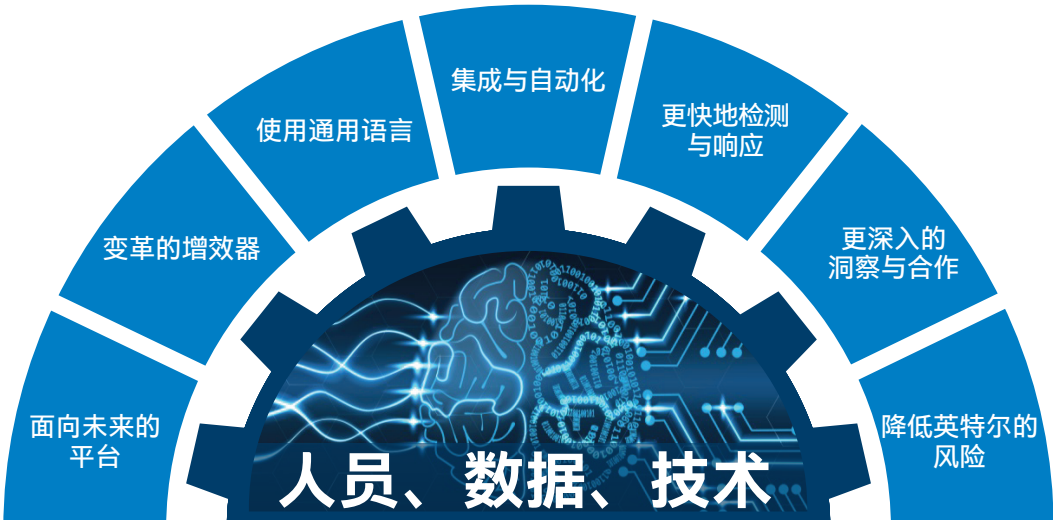


图 2. 我们的网络智能化平台可为英特尔带来各类优势。

消息总线使我们能够连接不同的数据源，推动整个信息安全部门认知过程的发展。从本质上看，我们的 CIP 提供了一个反馈循环，它集思广益，带来能够实现更有效防御的点子，有助于保护英特尔的品牌和声誉。

人员和数据准备

尽管保护信息安全是信息安全部门的主要职责，但与英特尔各业务部门合作始终是我们成功的关键。每天流入 CIP 的数据大都来自这些业务部门。

实现快速发展的关键因素是准备利益相关者数据。我们并没有尝试一次性准备所有数据源。相反，对于每个数据源，我们都遵循一个六步流程（见图 3）：

- 1. **协调资源**：我们与数据业务管理专员紧密合作——他们必须致力于此。
- 2. **培训**：我们培训这些数据管理专员，让他们了解 Splunk 语言。
- 3. **业务工作流程**：我们在英特尔确定了一个关键合作伙伴，由其解释数据与业务成果之间的关系，即，数据如何与其他数据结合生成一个决策，以及我们如何实现该决策的自动化。
- 4. **数据表征**：我们通过不懈的努力，了解数据的详细信息。数据是否需要连接、过滤、解析或充实？新衍生变量是否需要计算？
- 5. **数据建模**：确定数据在逻辑上进入业务决策并映射到通用信息模型的方式。此时，对 Splunk 筹备团队来说数据源已准备就绪。
- 6. **并入我们的 CIP**：一旦完成了前五个步骤，对 Splunk 和 Kafka 筹备团队来说数据就已准备就绪。团队首先开发了 Kafka 连接器，或者从 Splunkbase 中选出现成的连接器。然后将数据流元素映射到 Splunk 数据模型中。

敏捷的开发运维流程

我们的 CIP 工程和部署团队规模不大，因此必须巧妙地利用时间。如上一节所述，我们运用了敏捷流程在人员和数据上做好准备，而 Splunk 能够很好地支持这种方法。

对于传统的关系型数据库管理系统 (RDBMS)，分析人员需要对完整的结束状态进行规划并定义键值；确定需要多少个表和字段；加载数据；然后执行优化。这意味着需要很长时间才能获得商业价值。Splunk Enterprise 则完全不同。它支持通过按需模式（也称为“读取模式”、“需求模式”和“使用模式”）获取数据，并在过程中了解各项需求。即使情况发生了变化，我们也可以灵活地转换和移动。在确定优势之前，我们并不需要拥有包含所有数据源的完整库。相反，我们优先考虑的是通过哪些数据源来创建最小可行性产品。我们在几天内完成了五六个数据源的准备工作，并展示了 CIP 的功能。这种渐进式的胜利激发了大家对于新平台的热情，并带来了直接的商业价值。

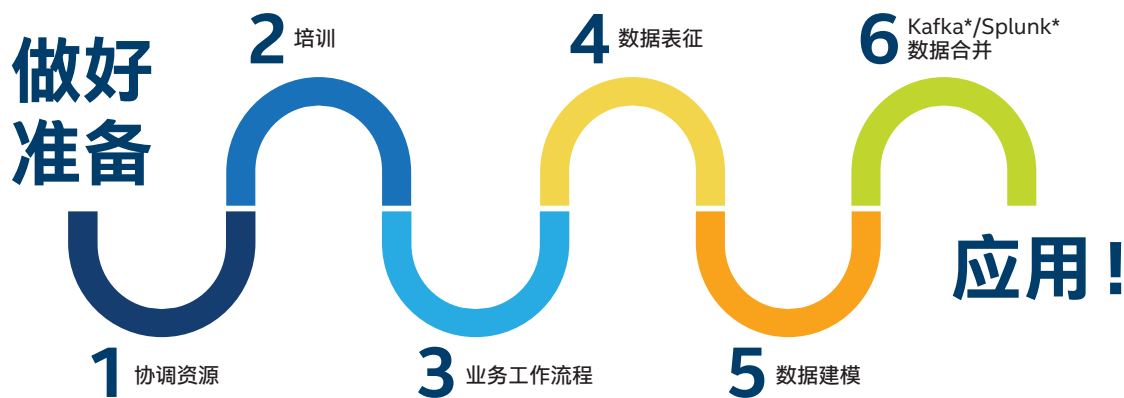


图 3. 为了能够快速部署我们的 CIP，我们使用敏捷流程来与英特尔各业务部门合作，做好人员和数据方面的准备。

为了进一步加快平台落地并迅速见成效，我们在事件响应团队中设下一项 30 天“挑战”，目标是完成正式培训并达到基本能力水平。随后，其他信息安全团队也纷纷设立了各自的挑战目标。这些激励举措加强了我们发现平台优势的能力。

BDAT 框架

开放群组架构框架* (TOGAF*) 定义了四个层：业务、数据、应用和技术 (BDAT)。我们将 CIP 映射到此框架中，如下所示：

- **业务：**Splunk 的通用工作界面意味着信息安全部门的员工不必登录多种工具；单一工具即可满足所有需求。
- **数据：**我们逐步了解了数据与业务能力之间的映射关系，并能根据需要搜索数据。
- **应用：**Kafka 提供了一种高效的数据分发方式（一次获取，然后由许多不同的应用使用）；即使情况发生变化，我们也只需在一处进行更改即可。它提供的抽象层可以面向许多不同的技术。Splunk 则提供了一组先进的数据操作功能。Kafka 和 Splunk 结合，可支持自动化运维剧本编排：在发生特定事件或出现告警时，会触发相关决策，进而产生特定、已定义的反应。
- **技术：**我们使用的是极其出色的软件技术，例如 Splunk、Kafka 和开源机器学习功能。我们鼓励供应商为 Kafka 和 Splunk 提供原生支持，以促进新数据源的就绪和系统集成。必要时，可在 Kafka 消息总线上通过流处理来扩充数据源，然后再由 Splunk 和其他应用使用。

数据分析

我们专注于引入干净、丰富的数据，然后尽可能多地使用开箱即用的工具来生成数据驱动型洞察。这样我们就不需要维护大量的自定义代码，因此有助于规避技术债务。我们还使用威胁情报源、开源机器学习库以及持续集成/持续部署自动化来加快数据分析速度。

根据数据和目标，我们的数据分析技术包括分类、聚类、特征选择或所有这些手段的组合。最终，我们的数据分析工作能在运营、系统运行状况、工作流程告警和异常检测方面带来业务价值。

Splunk 对于按需模式的支持（先获取数据，然后再进行结构化处理）是我们实现数据分析的重要助手。

Splunk* 与英特尔的合作为客户带来的益处

在高科技行业中，几乎每家公司都希望提供更出色的人工智能 (AI) 和机器学习解决方案，以应对大数据带来的挑战。但是，“更出色”的解决方案通常需要各类专家的密切合作。为此，Splunk 和英特尔于今年年初开始合作，通过更高性能、更低成本的客户体验，为双方的共同客户带来更大的商业价值。几位英特尔和 Splunk 解决方案架构师和产品工程师组成了一支拥有广泛专业知识的团队，并确定了多个跨产品和技术的机会。

其中之一就是从英特尔 IT 内部信息安全 (InfoSec) 部门收集洞察。该部门将 Splunk 产品用于其 CIP。这一协作团队正在整合来自安全运营分析师、数据科学家、数据架构师、安全架构师和安全工程师等多个英特尔 IT 信息安全岗位的人员的反馈。结果就是，协作团队了解了并将继续了解客户面临的挑战、他们的目标以及衡量成功的标准。

协作团队已准备好发布第一个联合参考架构，其中 Splunk Enterprise* 将运行在基于第二代英特尔® 至强® 可扩展处理器、英特尔® NVMe* 固态硬盘和英特尔® 傲腾™ 固态硬盘的服务器上。此参考架构可为客户带来两个重要益处：

- 它方便客户针对自己的 Splunk 实施方案选择合适的解决方案组成部分（计算、存储和网络）。
- 它为确定 Splunk 集群规模提供了明确的指导。

通过与英特尔和 Splunk 的共同客户共享优化成果，该协作团队可以帮助客户改进业务流程，同时帮助客户节约时间、资源和金钱。

借助性能更强的平台，客户就能更快地获得关键洞察。与英特尔合作的 OEM 也能从中受益：他们可以访问技术文档，了解如何设计和部署高效的解决方案来满足客户的特定需求。



解决方案架构

在我们着手对替换 SIEM 和日志系统进行评估时，就无意做“新瓶装旧酒”的事。相反，我们采取了一切从头开始的“绿地方法”。我们需要通过新 CIP 改变信息安全部门使用数据和从中获取洞察的方式。我们不想简单地迁移旧版规则、逻辑和过时的安全实践。我们选择的是转型而非迁移。其结果就是一个专为变化而构建的平台。随着数据源和数据量的增长、威胁和漏洞的演进以及业务需求的变化，我们的 CIP 将在未来不断适应并持续带来业务价值。

图 4 说明了我们的 CIP 架构。基础部分是高性能的英特尔® 硬件技术，例如高内核数英特尔® 至强® 可扩展处理器、英特尔® 固态硬盘和 10 GbE 网络。硬件之上是集成到 Kafka 消息总线中的数据源。消息总线采用发布订阅模型将我们所有的安全功能与数据绑定在一起。Splunk Enterprise 则提供了大量面向用户的功能。其他 Splunk 产品（例如 Splunk ES、Splunk IT Service Intelligence、Splunk Machine Learning Toolkit、Splunk Phantom）和第三方应用在 Splunk Enterprise 之上提供了各类附加功能和价值。

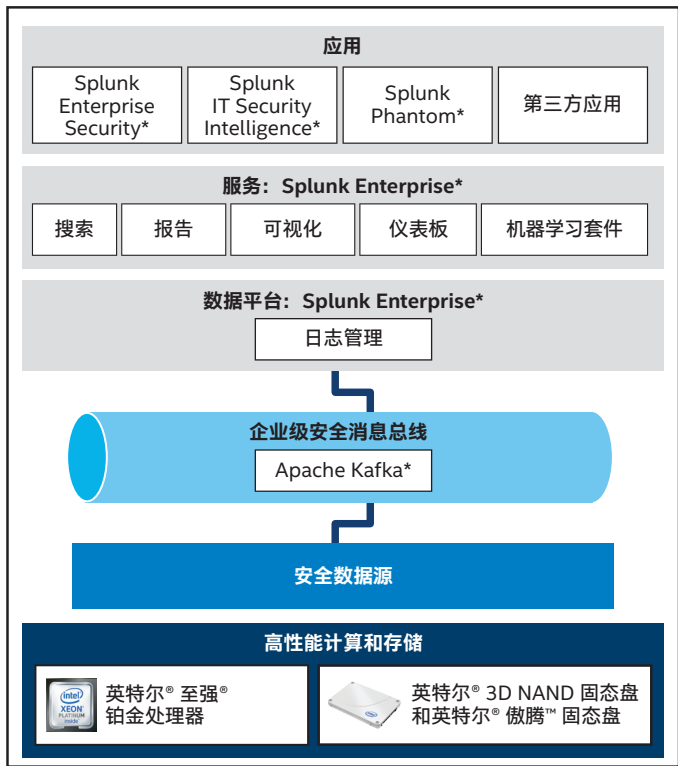


图 4. 以先进技术打造的高性能 CIP 使我们能够识别和应对复杂的威胁。

以下各节更详细地描述了我们如何使用 Splunk 和 Kafka，并讨论了确定计算和存储资源大小的因素。

使用 Splunk 提取和处理数据

Splunk Enterprise 是一个以时间为索引的事件数据库，它收集各种数据，供搜索、报告、可视化、机器学习等使用。Splunk 帮助我们创建易于理解、数据驱动、具有上下文和相关性的叙述。它能与广受欢迎的机器学习库集成，使我们得以快速、轻松地进行预测和测试。

为了保持一致性，我们建立了跨业务部门使用的运营衡量标准和指标。这有助于每个人都能在相同的基础知识和衡量标准上开展业务。这一任务由我们的知识对象管理器进行协调和执行，并在规划环节重新评估，做好用例准备。此外，Splunk 还使每位信息安全部门员工都能够随时创建临时衡量标准。这种灵活性如此之大，它让信息安全部门员工能够利用数据构建自定义分析。

在 CIP 中使用 Splunk 的关键优势在于它保证了真实情况的一致性。整个信息安全部门都可以就数据和查询/搜索开展交流、进行共享。这样就能在提高效率的同时，生成更一致、质量更高的数据产品。

以时间为索引的事件数据库

使用 Splunk Enterprise* 的关键优势在于，它保证了真实情况的一致性。整个信息安全部门都可以就数据和查询/搜索进行共享。

使用 Apache Kafka* 进行数据转换

长久以来，数据是以点对点集成的方式连接到信息安全应用的（见图 5 左侧）。这种方法有几个缺点：连接非常复杂、难以维护且因解决方案紧密耦合、集成较为脆弱，假以时日必然会发生故障——所有这些都会导致技术债务增加；添加新数据源或将现有数据源连接到新功能非常耗时；目前尚无编排。因此，如果系统中发生某些更改，则必须得多处添加自定义代码；此外，监控和管理也非常困难。

相比之下，有了 Kafka 消息总线，数据仅需获取一次，然后任何数量的应用都可以从总线中使用数据。借助 Kafka 这样的消息总线可为上游数据生产者和下游数据使用者创建数据抽象层，减少维护工作量。这样就能在不中断整个系统的情况下，添加或删除数据源和应用。

其他核心优势还包括能够接近实时地对流内数据进行切片、过滤、充实、聚合及归一化。这些数据转换的示例包括：

- 将 IP 地址与主机名相关联（充实）
- 将包含用户帐户的事件与包含工作人员特定信息（例如来自某业务部门的信息）的其他数据结合起来（连接）
- 根据日期、站点或设备类型等详细信息选择数据（切片）
- 删除无关或多余的数据（过滤）
- 评估文本字符串以测试是否符合结构型模式，例如验证电子邮件地址是否具有正确的 abc@xyz.com 格式（解析）

Kafka 使我们能够在单一平台上协调多个活动，由此监控和管理工作要比使用点对点方法时容易得多。

调整 CIP 基础设施规模

我们的 CIP 需要高性能计算和存储，以随着信息安全环境中不断增长的数据量而进行扩展。我们设计的 CIP 支持无限扩展，因此能让我们从容面对未来所需。

请注意，调整 Splunk 和 Kafka 实施方案的规模涉及许多工作，而不仅仅是设置要提取或发布的数据量。例如，搜索的数量和类型，以及转换每个 Kafka 数据流的次数都是关键因素。以下各节阐述了我们对如何调整 CIP 算力和存储大小的认识。

算力要求

Splunk 和 Kafka 对 CPU 和存储性能的要求都很高。当前，我们的 CIP 由 275 台基于英特尔® 至强® 铂金处理器（大约 10,000 个高性能内核）的服务器组成。我们最近部署了更多服务器，以便随着新数据源、新应用和新用户的出现，不断满足增加算力的需求。我们的生产和预生产 CIP 环境存储超过 10 PB 的数据，所有数据都保存在英特尔® 固态硬盘上。

- **确定 Splunk 算力时考虑的因素：**对于每个查询，Splunk 都需要一颗高性能内核。这是因为 Splunk 搜索层会将查询发送到 Splunk 索引器（事件数据在此建立索引并存储），并生成一个线程来运行搜索。线程的示例包括为数百个用户和众多应用提供的数据库模型加速、机器学习、相关搜索、即时搜索、计划搜索、报告和仪表板生成。

Splunk 搜索头和管理服务器是搭载 24 核英特尔® 至强® 铂金 8168 处理器的双路服务器。Splunk 索引器配备的是双路 24 核英特尔® 至强® 铂金 8160 处理器。这些处理器可提供每天提取超过 12 TB 数据并支持至少 48 个并发搜索所需的算力。在我们看来，以每服务器机箱最大程度容纳高性能内核，实现最高存储能力非常重要。双路配置让我们能够达到这一出色效果。

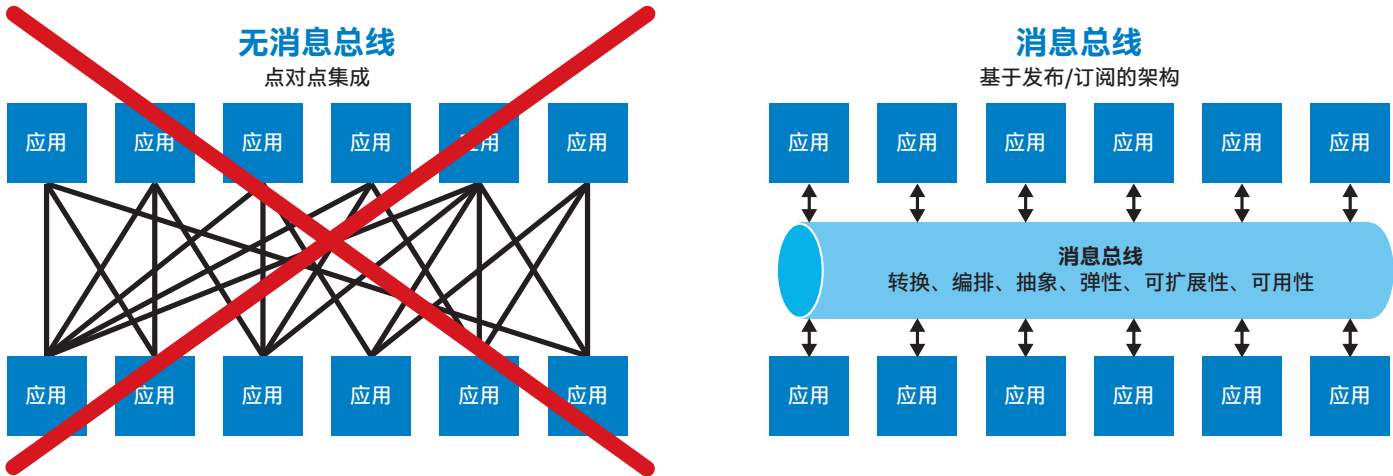


图 5. 我们的 CIP 摒弃了难以管理的点对点集成，而是使用消息总线（例如 Kafka*）来提高我们转换数据、编排活动以及优化弹性和可扩展性的能力。

通过横向扩展 Splunk 的索引层，每个服务器检索的数据量会减少，搜索速度会加快。举例来说，如果使用 10 台服务器搜索 1 TB 的数据，则每个服务器需要检索 100 GB 数据；如果使用 100 台服务器搜索相同数量的数据，则每台服务器仅需检索 10 GB 数据。

- **确定 Kafka 算力时考虑的因素：**Kafka 代理服务器和 Kafka Connect* 工作服务器使用双路 14 核英特尔® 至强® 金牌 6132 处理器（每个服务器 28 个物理 CPU 内核）。现有的 Kafka 流处理器服务器使用的也是双路 14 核英特尔® 至强® 金牌 6132 处理器。不过，随着我们不断扩展流处理能力，我们正在为新的 Kafka 流处理器服务器配备双路 24 核英特尔® 至强® 铂金 8160 处理器（每个现有服务器物理 CPU 内核数量为 28 个，每个新服务器的物理 CPU 内核数量为 48 个）。

Kafka 的设计旨在利用并行处理技术；因此，随着数据量的增加，以多服务器和多 CPU 内核实现的横向扩展可以加速数据流实时处理。例如，我们发现在为 10,000 个 Kafka 主题分区提供服务进行工作负载分配时，与使用 3 台代理服务器相比，使用 10 台代理服务器能够显著提高分配速度。

存储要求——分秒必争

在我们目前的 Splunk 基础设施中，每个索引器使用 24 个基于 SATA 的英特尔® 固态硬盘 S4500 系列（每个 3.8 TB），提供总计 72 TB 的数据存储量。因为要实现符合英特尔一年数据留存要求所需的每 Splunk 索引器服务器存储总量和数据量必须使用 RAID 控制器，因此我们选择了 SATA 固态硬盘。我们还想确保每个索引器上存储的 72 TB 数据具有 RAID 提供的硬件冗余。借助 RAID 冗余，单个 3.8 TB 固态硬盘出现故障并不会影响 72 TB 的卷，也不需要数据进行恢复。Splunk Enterprise 应用在软件层具有出色的弹性，但是我们希望硬件层也具有这种弹性，以抵消丢失整个索引器带来的风险（这可能会在整个平台上对性能产生一连串负面影响）。

无论是等待搜索结果的最用户，还是负责部门网络分析平台的工程师，搜寻安全威胁（也称为狩猎）必须分秒必争。工程师通常希望：

- 减少搜索所需的平均总耗时，包括启动搜索的时间和搜索运行的时间。
- 即使用户数量和数据量不断增加，也能增加整个平台的搜索并发量。

图 6 说明的是 Splunk 在 2018 年年中平均每天大约进行 56,000 次搜索。随着我们不断添加用户，搜索量增加了一倍多，达到每天约 115,000 次搜索²。在此期间，我们的数据提取速率从每天平均约 5.5 TB 提升到约 12 TB³。

面对如此大的数据量和搜索量增幅，我们的 CIP 工程团队必须不断寻找在软件堆栈各个层面（包括操作系统、Splunk Enterprise 内核及 Splunk 应用）上进行优化的方法。但同时我们也想提高平台的速度和效率。特别是，我们认为必须减少平台的存储 I/O 等待时间，只有这样才能充分利用运行高性能英特尔® 至强® 铂金 8168 处理器的平台服务器。

我们从 Splunk Enterprise Security 搜索头 (ESSH) 开始存储优化工作。我们在评估了多种新型存储技术后决定在我们的生产环境中添加英特尔® 傲腾™ 固态硬盘 P4800X 系列 (750 GB)，用于主动应用存储。

我们下一步计划在通用搜索头 (GUSH) 中部署英特尔® 傲腾™ 固态硬盘。与此同时，我们还在评估英特尔® 傲腾™ 持久内存，以期进一步提高系统性能。Splunk 索引器上的主动存储可以利用这一技术提供的大容量、持久和高性价比的内存。

² 截至 2019 年 7 月 24 日的 Splunk 搜索次数。

³ CIP 生产环境于 2018 年 7 月 5 日开始为 Splunk 数据湖提取数据。截至 2019 年 7 月 24 日，Splunk 数据湖每天提取约 12 TB 数据。

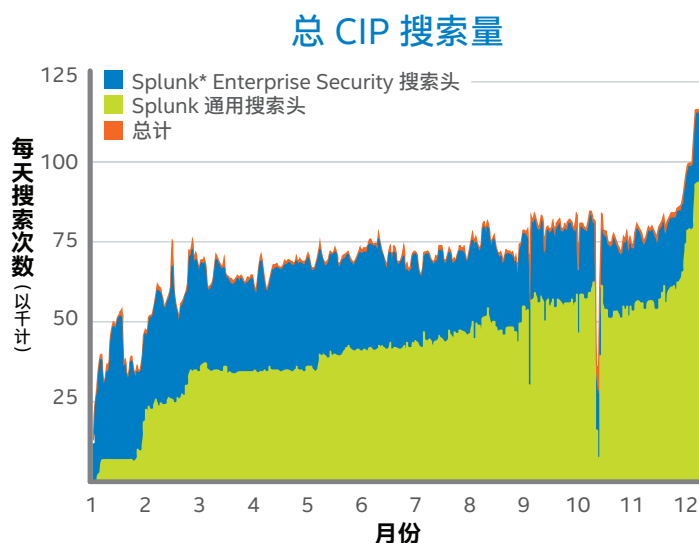


图 6. 在过去一年中，在我们的 CIP 上执行的搜索数量增加了一倍以上。

未来举措

我们的 CIP 具有出色的业务价值。它以一系列出色的专业功能迅速获得整个信息安全部门的青睐。我们计划在未来几个月内继续提高采用率并添加新功能。示例如下：

- **提高采用率：**我们最初部署 CIP 的重点是替换我们上一代 SIEM 和日志管理系统。上一代系统主要由包括安全运营中心在内的事件响应团队使用。现在包括漏洞管理、补丁/合规、风险管理和治理在内的其他团队也在使用我们的 CIP。我们不断寻找机会为 CIP 增加功能和增添价值，包括将旧版应用迁移到 CIP 中，以此减少我们的技术债务。
- **处理数据：**我们正通过在 Kafka 消息总线上进行数据充实、转换、解析、时间切片和过滤来增加数据流处理用例。这不仅能提高网络智能化，还可以帮助降低下游系统处理数据的成本。
- **整合机器学习：**实现 CIP 与其他机器学习工具的集成可以增加威胁情报的获取概率，并帮助我们创建可在运营、系统运行状况、工作流程和警报编排等许多其他领域实现业务价值的工作流程，同时也为事件响应团队带来业务价值。目前，我们已在 Splunk 数据湖及应用中使用了一些机器学习模型。未来，我们计划探索直接在 Kafka 中部署经过训练的机器学习模型的可能性。
- **增强性能：**我们计划在其他需要快速 I/O 的用例中添加英特尔® 傲腾™ 固态硬盘。例如，Splunk 索引器上的热存储或温存储，以及 Splunk Enterprise 和 Splunk IT Service Intelligence 搜索头中的临时工作数据集。将基于 SATA 的固态硬盘替换为使用 PCIe* 接口的非易失性存储器* (NVMe*) 的固态硬盘，应该有助于减少因 RAID 控制器造成的性能瓶颈。而且，根据固态硬盘的大小，我们或许可以减少每台服务器使用的固态硬盘数量。
- **提高可靠性：**英特尔® 傲腾™ 固态硬盘的另一项优势是提高耐用性。CIP 中的大多数数据只写入一次但需要读取多次，因此总体来说，我们对于固态硬盘耐用性的关注程度要小于对 Splunk 索引器性能的关注。但是，Splunk 搜索头的情况有所不同。搜索是从多个搜索头连续发起的。搜索头会查询数据，结果会写入临时区域，并创建摘要数据。因此，虽然搜索头所用的固态硬盘不必和索引器所用的固态硬盘容量一样大，但它们确实需要高耐用性。英特尔® 傲腾™ 固态硬盘专为写入频繁的环境设计，具有超高的耐用性。

“英特尔信息安全保障机制比以往任何时候都更加敏捷。但是我们需要继续磨练技能。”

—— 英特尔首席信息安全官 Brent Conran

结论

我们的 CIP 帮助公司提高了生产力和工作效率。我们改变了整个信息安全部门运用技术和数据来保障信息安全的机制，使员工有了通用语言和工作界面。我们使数据对于整个信息安全部门都是可访问、可用和有价值的。正如我们首席信息安全官 Brent Conran 所说：“英特尔信息安全保障机制比以往任何时候都更加敏捷。但是我们需要继续磨练技能。由于人工智能和机器学习的加入，数据量激增，这意味风险的加剧。但同时这也能带来更多回报……不仅能避免增加成本……还能切实节省资金。”

在构建 CIP 的过程中，我们将英特尔® 架构、Splunk 和 Kafka 相结合，这使我们能够更快地应对威胁、提供防御威胁所需的洞察，并帮助降低风险。

CIP 的一个关键方面是能够将机器学习与流处理和基于规则的逻辑相结合，推动一般事件的编排和自动化，滤除误报（每天可能有多达数千甚至数百万个误报）。我们正在向信息安全分析师或下游系统提供上下文相关的丰富数据和真实的威胁检测结果。最终，我们的 CIP 不仅会使我们行动速度更快，还能帮助我们不断提高整个信息安全部门的效力。在我们继续开发新产品、进入新市场、支持新客户的过程中，这种敏捷性对于确保英特尔的合规和安全至关重要。

有关英特尔 IT 部门最佳实践的更多信息，请访问 intel.cn/IT。

IT@英特尔

我们促进了英特尔内部 IT 专业人员之间的交流。英特尔 IT 部门解决了一些当前极其复杂棘手的技术难题，我们希望在公开的同行交流论坛上，直接与其他 IT 专业人员分享这些经验教训。

我们的目标非常简单：提高整个部门的效率，提升 IT 投资的商业价值。

请关注我们并加入对话：

- [Twitter](#)
- [LinkedIn](#)
- [#IntelIT](#)
- [IT Peer Network](#)

相关内容

如果您喜欢本白皮书，那么您也可能会对以下相关内容感兴趣：

- Advanced Persistent Threats: Hunting the One Percent (高级持续性威胁：狩猎百分之一)
- Security Architecture Enables Intel's Digital Transformation (安全架构为英特尔实现数字化转型提供支持)
- Enterprise Architecture: Enables Intel's Digital Transformation (企业架构：为英特尔实现数字化转型提供支持)
- Securing the Cloud for Enterprise Workloads: The Journey Continues (保护云安全，支持企业工作负载：旅程还在继续)
- Enterprise Technical Debt Strategy and Framework (企业技术债务战略与框架)

缩写

BDAT	业务、数据、应用和技术
CIP	网络智能化平台
EDR	端点检测和响应
ESSH	Enterprise Security 搜索头
GUSH	通用搜索头
InfoSec	信息安全
IoT	物联网
RDBMS	关系型数据库管理系统
SIEM	安全信息和事件管理
SSD	固态硬盘

编著者

- Bill Brasse, 英特尔 IT 部门项目经理
- Frank Ober, NVM 解决方案部门企业架构师
- Jeff Sedayao, 英特尔 IT 部门行业合作经理
- Merritte Stidston, 销售和营销部门技术专家
- Jerome Swanson, 英特尔 IT 部门信息安全工程师
- Sandeep Togrikar, 数据中心部门企业架构师

英特尔技术特性和优势取决于系统配置，并可能需要支持的硬件、软件或服务得以激活。产品性能会基于系统配置有所变化。更多信息请从原始设备制造商或零售商处获得，或请见 [intel.cn](#)。

英特尔处理器编号不是性能指标。处理器编号用于在每个处理器家族中区分不同功能，不能跨越不同的处理器家族进行比较： [关于英特尔处理器编号](#)。

本文中的信息旨在提供一般性说明，而非具体的指导。其中给出的建议（包括可能的成本节省）基于英特尔自身的经验，仅为预测。英特尔不保证其他公司会获得相似的结果。

本文档中提供的信息与英特尔产品和服务有关。本文并未（明示或默示、或通过禁止反言或以其他方式）授予任何知识产权许可。除英特尔在其产品的销售条款和条件中声明的责任之外，英特尔概不承担任何其他责任。并且对于英特尔产品和服务的销售和/或使用，英特尔未做出任何明示和默示的保证，包括但不限于，关于适销性、适合特定目的及不侵权的默示保证，以及履约过程、交易过程或贸易惯例中引起的任何保证。

本文并未（明示或默示、或通过禁止反言或以其他方式）授予任何知识产权许可。

英特尔、英特尔标识以及其他英特尔商标是英特尔公司或其子公司在美国和/或其他国家的商标。

* 其他的名称和品牌可能是其他所有者的资产。

© 英特尔公司版权所有。保留所有权利。 0919/WWES/KC/PDF

