



借助物联网优化制造流程

借助物联网和大数据分析试点计划，
英特尔制造部门能够提升运营效率和增强盈利能力

通过使用大数据分析，
英特尔实现了预测维护，
节约了成本，从而改进了
自身的制造流程，并大幅
提高了产能，在物联网
方面取得了重大成功。

简介

当前，制造过程中所产生的数据的规模、类型和速度¹正在呈指数级增长，带来了巨大的发展机遇。通过认真分析这些数据，企业可获得竞争优势、快速响应不断变化的市场动态、显著提高制造利润、生产力和效率。工厂车间中的设备会生成成千上万种不同的数据类型，例如多个级别的单位生产数据、设备运行数据、流程数据、人工操作员数据等。这些数据可以在存储后用于分析目的。

尽管大型制造商多年来一直在使用统计流程控制和统计数据来分析优化生产，但当今数据的复杂构成为部署新方案、基础架构和工具带来了重大机遇。得益于更卓越的计算性能的涌现、开放标准的推出和广泛可用的行业专长，制造业已经准备好充分利用大数据。此外，在学术研究的推动下，技能娴熟的数据科学家数量庞大，他们将能够在使用大数据方面为制造业提供更出色的支持。借助全新的制造智能特性，制造商能够提高质量，增加产量，更准确地了解制造问题的根本原因，以及减少机器故障和宕机情况。借助新的业务价值和技术能力，制造商将能够改变业务模

式和实践，以优化设计，实现出色的可制造性，从而改进供应链管理，并采用定制的制造服务缩短专为各地精明消费者量身定制的产品上市时间。

本文概述了英特尔一家制造工厂的物联网（IoT）试点计划，以说明将数据分析应用于工厂设备和传感器可如何提升制造流程的运营效率并节约生产成本。在实施这一物联网大数据分析项目的过程中，英特尔与 Cloudera、戴尔、三菱电机和 Revolution Analytics 进行了紧密的行业协作，预计每年将能够节省数百万美元，并带来更高的投资回报。

挑战：如何从所有的制造数据中提炼价值？

大数据实质上是包含各种数据类型的庞大数据集，可以分为结构化、半结构化或非结构化，如表 1 所示。结构化数据适合放在格式整齐的表格中，因此相对容易管理和处理。结构化数据的优势是便于输入、存储、查询和分析。相关示例包括关系数据库中存储的制造数据，以及来自制造执行系统和企业系统的数据。另一方面，非

结构化数据（例如图片、文本、设备日志文件、人工操作员生成的值班报告和制造社交协作平台文本等）可能为原始格式，需要经过解码才能提取数据值。半结构化数据是结构化数据的一种，其格式不符合关系数据库或其他形式的数据表关联的数据模型标准结构，但包含用于分隔语义要素的标记或其他记号，并且数据中明确了记录和字段层次结构。在制造业，要想发挥大数据技术的强大作用，需要将这些数据集类型进行合并与关联，从中发现新的

洞见，从而创造出出色的业务价值。大数据技术的另一个价值定位就是，支持制造商以经济高效、可扩展的方式聚合并集中各类数据。

制造流程存在各种变量，如材料、流程方案和方法以及设备差异，这就要求制造商采用基于可扩展平台的大数据解决方案，这种平台能够随着业务增长和制造要求的提高而扩展。机器数据与产量和质量密切相关，因此能够为主动检测将要失去控制的流程提供重要信息。然而，一些类型的制造流程会产生大量数据文件（每个工具类型几天就可产生 GB 级数据，如表 2 所示），限制了使用传统方法从这些数据中分析、提取和存储有用信息的能力。如果不使用大数据技术，直观呈现来源广泛的大型数据集中的信息都非常困难。

制造业示例	实时、半结构化数据	非结构化数据	结构化数据
半导体	机器制造商标准，如 SECS/ GEM、EDA 或基于 COM、XML 的自定义标准	操作员值班报告	RDBMS 数据库 NoSQL* 企业数据仓库 存储在制造 PC 中的文件 电子数据表
电子		机器日志	
太阳能		错误日志	
机械		文本	
能源		视觉图像	
汽车		音频/视频广告	
航空		制造协作社交平台	
化工制药		RFID	
金属加工		直接来源于 PLC、电机和驱动器，直接来源于	
食品和饮料		运动控制器、机械臂	
制浆造纸		制造记录（时间序列数据	
服装纺织		结构）	
家具			

表 1. 制造数据示例

数据类型	数据大小（每周）	示例
机器参数和错误日志	每台机器约 5 GB	用于监控机器性能：点胶高度、位置（x、y 和 z）、传送带速度、流速、回炉温度、激光功率等。
机器事件	每台机器约 10 GB	用于测量加工时间：开始点胶、结束点胶、开始设置、结束设置
来源于视觉设备的缺陷图像	每单元约 50 MB 或每时隙 750 GB	用于确认故障模式、缺陷共性、缺陷映射的根本原因

表 2. 数据大小示例

应对挑战：英特尔制造部门使用大数据分析服务器和物联网网关的物联网试点计划

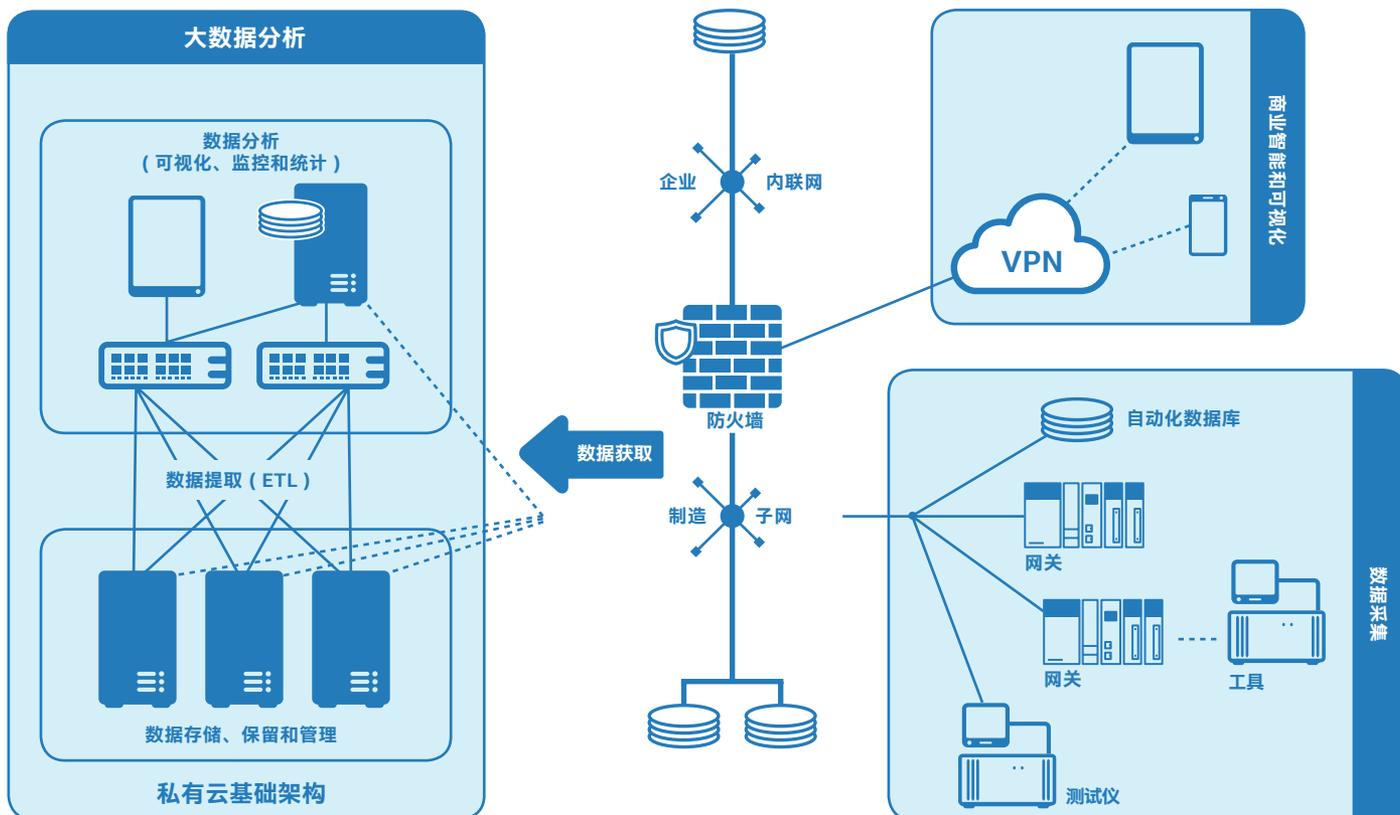


图 1. 支持为从工厂车间到数据中心的整个过程提供制造智能的端到端基础架构构建模块

图 1 所示为面向大小数据集的高级物联网制造架构，这些数据包含来源于制造车间和制造网络、可供收集和聚合的各类数据。该架构为通过数据可视化、监控和挖掘创建新型商业智能提供了全新可能。举例来说，该架构可以清理、提取和转换来自现有数据库的结构化数据、来自工具

传感器的非结构化数据以及来自传统设备的日志文件，并将它们整合在一个数据存储平台（即 Hadoop*）中。然后，运行在内部服务器（即数据存储服务器）上不同虚拟机的高级工厂应用可对数据进行可视化和分析。或者，用户可使用网络上的其他分析或监控应用访问这些数据。其他增

强功能包括在 Hadoop 或其他类型文件系统中进行分析或在内存中进行分析以提升性能。分析结果可通过网络商业智能层中的直观可视化功能展示给用户。

英特尔制造部门试点计划中的大数据分析物联网服务器设置

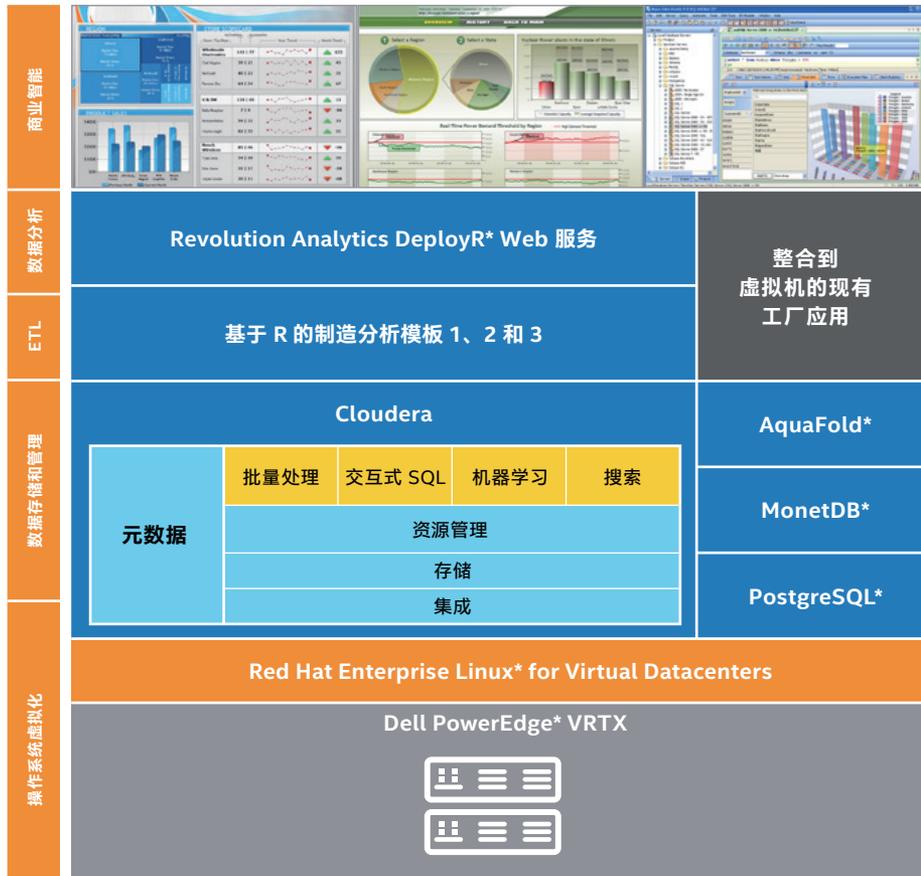


图 2. 大数据分析服务器

英特尔制造部门在其试点部署中使用的大数据分析服务器如图 2 所示。戴尔的紧凑型系统 PowerEdge* VRTX² 被选作内部服务器，以在私有云环境中托管大数据和分析软件。该系统包含一个 Dell PowerEdge VRTX 机箱，配备 25 块 900GB 硬盘和 2 台 Dell PowerEdge M820 刀片式服务器，每一台服务器配备 E5-4600 产品家族中的 4 颗英特尔® 至强™ 处理器。英特尔至强处理器 E5-4600 产品家族提供了一款成本适宜的紧凑型四路处理器解决方案，最高支持 8 个内核、20MB 末级高速 (L3) 缓存和 1.5TB 内存容量，同时具备可快速转移数据的信道。

两台 M820 服务器可托管分析和应用软件，以及在多台虚拟机中运行的 Hadoop 节点。Red Hat Enterprise Linux* for Virtual Datacenters 操作系统为针对可扩展、完全虚拟化数据中心而设计的服务器提供了一种全面的虚拟化软件解决方案。

分析和应用节点

图 3 显示了软件被如何分配至不同虚拟机 (VM)。该节点托管着 5 台虚拟机，分别运行着不同的分析和应用工作负载。具体包括：

来自 Revolution Analytics 的 Revolution R Enterprise*

通过链接到高性能英特尔® MKL 多线程数学库，Revolution R Enterprise 可充分利用多处理器的强大功能，加快常见矩阵运算速度。此外，Revolution R Enterprise 的并行分布式算法可支持对大数据进行统计分析，采用多核服务器、高性能计算群集、大数据设备和 Hadoop* 实现线性扩展。

- 来自 Revolution Analytics 的 Revolution R Enterprise* 是一款基于强大的开源 R 统计语言的分析软件。该软件可在分析解决方案和企业软件之间提供无缝、安全的数据桥，可解决采用基于 R 的分析功能及现有 IT 基础架构的企业所面临的关键集成问题。³
- MonetDB*: 一种面向列的开源数据库管理系统，旨在帮助高效完成对大型数据库进行的复杂查询任务，如将表格与数百个列和数百万行进行组合。MonetDB 已使用在各种需要高性能的应用中，如数据挖掘、在线分析处理 (OLAP)、地理信息系统和流数据处理等。⁴
- PostgreSQL*: 一款强大的开源对象关系型数据库系统，用于在线交易处理 (OLTP)。⁵

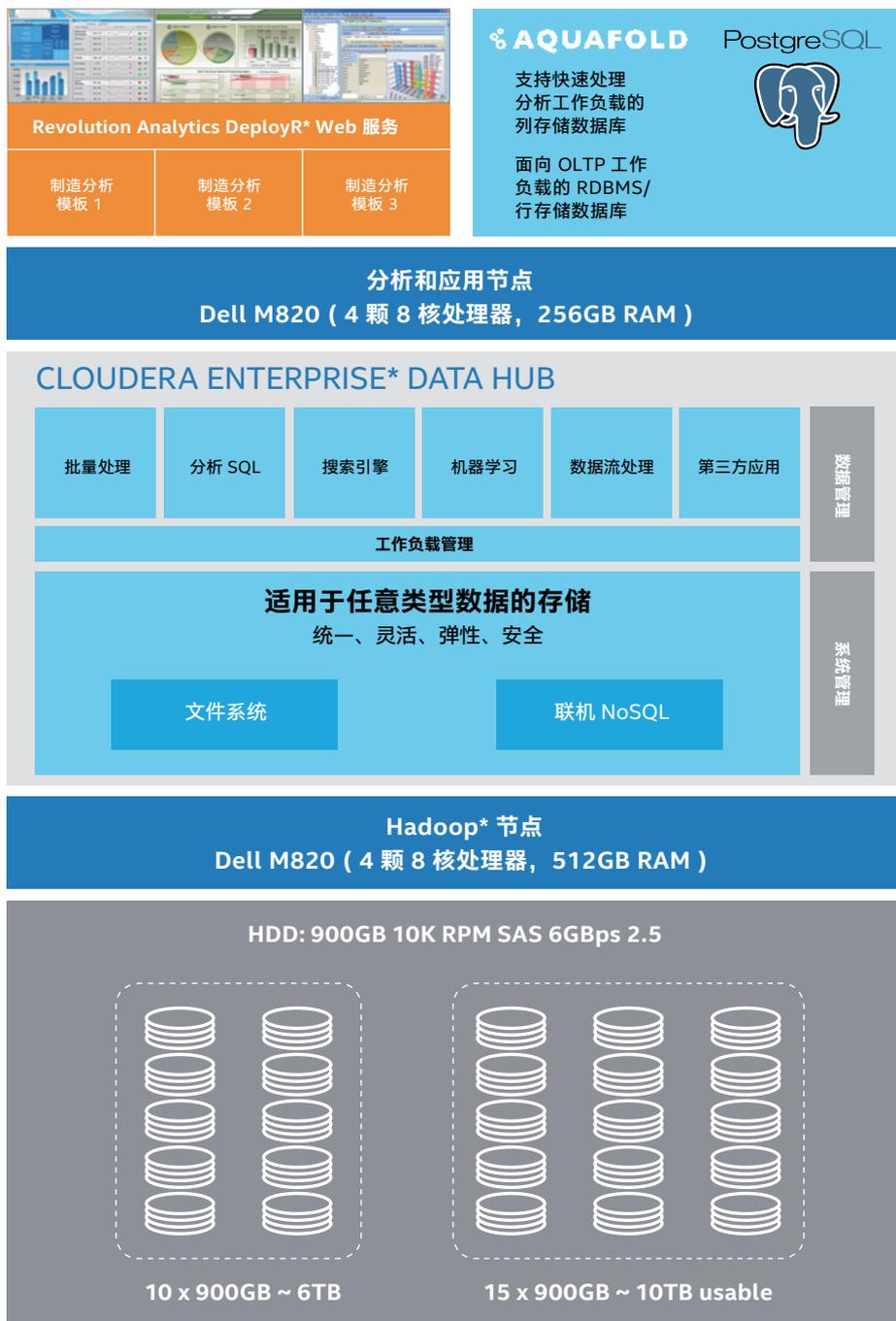


图 3. 大数据分析服务器上软件分配至虚拟机的情况

- **AquaFold***: 一种应用服务器，用于快速构建和部署生产质量数据库和报告应用。它包括丰富功能，如多源数据同步，跨数据库数据迁移，数据仓库提取、转换和加载，调度数据导出/导入和自定义商业智能仪表盘等。支持与网络中的各种生产数据库通信，如 PostgreSQL、Microsoft SQL*、Oracle* 以及具有 ODBC* 和 JDBC* 连接的 IBM DB2*。⁶

Hadoop* 节点

基本的 4 节点 Cloudera Hadoop 集群通过四个虚拟机来运行，包含一个主节点和三个工作节点。

- **Apache Hadoop** 是一种面向可扩展分布式计算的开源分布式软件平台。它在 Java 中进行编写，运行在配置有直接连接存储的行业标准服务器集群上，通过向集群中添加成本较低的节点便可经济高效地提升性能。
- **Cloudera Enterprise Data Hub* (CDH)** 为大数据提供了一个统一平台，支持在单一位置存储、处理和分析所有数据，能够在提升现有投资价值的同时，支持采用重要的新方法从数据中获取价值。CDH 是一个完全由 Apache 提供许可的开源软件，在提供统一的批处理、交互式 SQL、交互式搜索以及基于角色的访问控制方面具有独特优势。⁷

Cloudera 堆栈详情

Cloudera Enterprise Data

Hub* (CDH) 提供了 Hadoop 的核心元素, 可扩展的存储和分布式计算能力, 同时还提供了用户界面等附加组件, 以及安全性等必要的企业功能, 并集成了广泛的硬件和软件解决方案。

CDH 能够帮助企业转变并加速使用大数据的方法。当与基于英特尔® 至强™ 处理器的数据中心架构组合使用时, Cloudera 和英特尔为工业和制造企业提供了一种综合方法, 可帮助他们通过分析数据提高工厂运营效率。

物联网 (IoT) 网关

三菱电机 MELSEC-Q 系列* C 语言控制器是一款基于英特尔® 凌动™ 处理器的网关*,⁸ 用于聚合并安全摄入数据至大数据分析服务器。数据摄入指验证、过滤和重新格式化数据, 以便大数据分析软件更轻松地对其进行分析。

三菱电机 MELSEC-Q 系列 C 语言控制器是具备众多智能系统功能特点的嵌入

式解决方案, 包括处理从传感器或通过网络收集的大量数据所需的稳定网络连接和卓越计算性能, 同时支持复杂的系统控制和操作。这一控制器的核心是采用英特尔® 架构和 Wind River VxWorks* 实时操作系统的硬件平台。

三菱电机开发了 MELSEC-Q 系列 C 语言控制器, 旨在满足多种工厂自动化要求, 具备卓越的可靠性, 可适应苛刻的环境, 且长期可用。这些特性助它成为一款坚实可靠的产品, 几乎无需维护物联网制造应用。

MELSEC-Q 系列 C 语言控制器取代了传统可编程控制器中使用的梯形逻辑, 使用国际标准 C 语言 (C 和 C++) 提供了更灵活的编程能力。它能够支持用户充分利用现有的 C 语言软件和开发资源。

CIMSNIPER* 是一款适用于三菱电机 MELSEC-Q 系列 C 语言控制器的数据获取和处理软件包。它能够收集流程数据 (包括 SECS 信息) 和制造设备错误, 而无需修改现有系统。

英特尔制造部门的大数据分析案例

在过去的两年中, 英特尔开发了十几个大数据项目, 巩固了运营效率和收入。下方列出了几个示例:

缩短产品测试时间

英特尔® 制造的每一枚芯片都要经过彻底的质量检查, 通过多种复杂的测试。

英特尔发现, 利用制造过程中收集的历史信息可以减少测试所需的产品数量, 从而缩短测试时间。作为概念验证实施的这款解决方案在 2012 年为英特尔® 酷睿™ 系列处理器节约了 300 万美元的测试成本。将这款解决方案扩展到更多产品后, 英特尔预计会节约 3,000 万美元的成本。⁹

改善制造监控

数据密集型流程也用于帮助英特尔在生产线上检测故障, 这是一个高度自动化的环境。英特尔从整个工厂网络的制造工具和测试工具中提取日志文件, 每小时多达 5 TB。通过捕获和分析这些信息, 英特尔能够确定在制造流程中的哪个步骤开始偏离常规容差。¹⁰

针对此处讨论的当前端到端平台试点部署, 英特尔与三菱电机、Cloudera、Revolution Analytics 和戴尔进行了紧密合作, 成功开发了许多卓越功能, 在使用数据挖掘科学解决实际制造问题方面取得了巨大进展, 并通过成本节省和决策改进为英特尔节约了数百万美元。该项目的主要目标是发掘数据和数据分析的价值, 通过预测制造获得更深刻的洞察力, 并降低制造成本, 同时不影响生产量或质量。

下面详述了英特尔通过集成面向制造物联网的大数据分析和技术获得了一些突破性成就和发现。

试点计划成效:

使用案例 1: 通过监控和分析设备参数值并在部件发生故障之前及时更换, 减少不必要的产能损失。

背景

自动测试装置 (ATE) 是专门用于在不同设备上执行测试的机器, 这些设备被称作被测设备 (DUT)。ATE 使用控制系统和自动化信息技术快速执行测试, 测量并评估被测设备。¹¹ ATE 系统连接到称为处理工具的自动更换工具上。该工具能够物理更换测试接口单元 (TIU) 上的被测设备, 以便接受该装置的测试。

问题陈述

有缺陷的测试接口单元将会错误地将良品分类为次品, 会对英特尔制造运营成本造成负面影响。有缺陷的测试接口单元会导致对被测设备进行错误分类, 包括拒绝良品。英特尔制造部门的目的是检测潜在的 TIU 缺陷, 以便及时修理或更换单元, 防止它们被错误分类。如果有缺陷的测试接口单元将良品错误分类为次品, 该单元将作废。在定期预防性维护期间, 一些即便仍然运转正常的组件也会使用备件更换, 以避免发生此类问题。

结果和优势

分析功能可在现有工厂联机流程控制系统触发之前, 预测出高达 90% 的潜在测试接口单元故障。在此处的情况下, 这可帮助及时更存在缺陷的 TIU 以免造成过度拒绝良品, 将产能损失降低了 25%。¹²

此外, 英特尔还减少了在预防性维护过程中提前更换尚未故障的备件的需求, 从而预计可以降低 20% 的备件成本。

使用案例 2: 通过在焊球焊接设备中消除和减少错误焊球装配情况降低产能损失

背景

焊球焊接模块是为基片表面涂抹焊膏的部件。焊球被放置到焊球焊接表面, 然后焊膏将其固定在相应位置上。整个封装通过一个回流焊炉, 将在其中融化基片表面上的焊膏和焊球。

焊球被真空吸附到贴装头的小孔上。系统将检查该贴装头上的焊球是否过多或缺失。当贴装头与基片对齐时, 焊球被放置在基片的焊膏上。释放焊球后, 将检查贴装头上是否残留有任何焊球。最后, 摄像头成像系统将检查基片上是否缺失焊球或焊球位置是否存在偏移。

问题陈述

缺失焊球的单元为有缺陷的材料, 会造成产能损失。多种场景会导致单元缺失焊球, 包括真空压力不足等。

成效和优势

通过可视化处理传感器读数并将其与各种机器数据和执行系统数据进行关联, 英特尔成功降低了产能损失, 优化了维护成本, 并避免了设备突然宕机。¹² 这可帮助技术人员主动解决问题, 朝着创建预测维护功能而努力。

使用案例 3: 使用图像分析确定良品或次品

背景

机器视觉设备是一种模块, 可筛选单元并将其分类为良品和边际单元。良品将传送到加工流程, 而边际单元将接收制造专员的检查并确定其优劣。该手动流程较为耗时。

问题陈述

人工检查并分类边际单元的流程非常繁琐, 有时需要约 8 小时才能成功将边际单元与真正的不良品隔离。

这是因为单元送达操作员、然后传送到隔离模块、最后进行隔离比较费时。图像分析可帮助迅速识别检测模块检测到的不良品。

成效和优势

机器视觉设备模块中记录的边际图形经过预处理。每一张图像都是非结构化数据, 需要修改尺寸、剪裁并转化为灰度模式, 然后将每个像素转化为二进制格式。此流程的下一阶段涉及特征选择, 非结构化图像将由一系列不同的值定义。然后, 将这些值赋予各种机器学习算法, 用于区分真正不良品和边际不良品。

图像分析缩短了从大量边际单元中分离真正不良品的时间。图像分析确认不良品的速度大约比人工方法快 10 倍。¹²

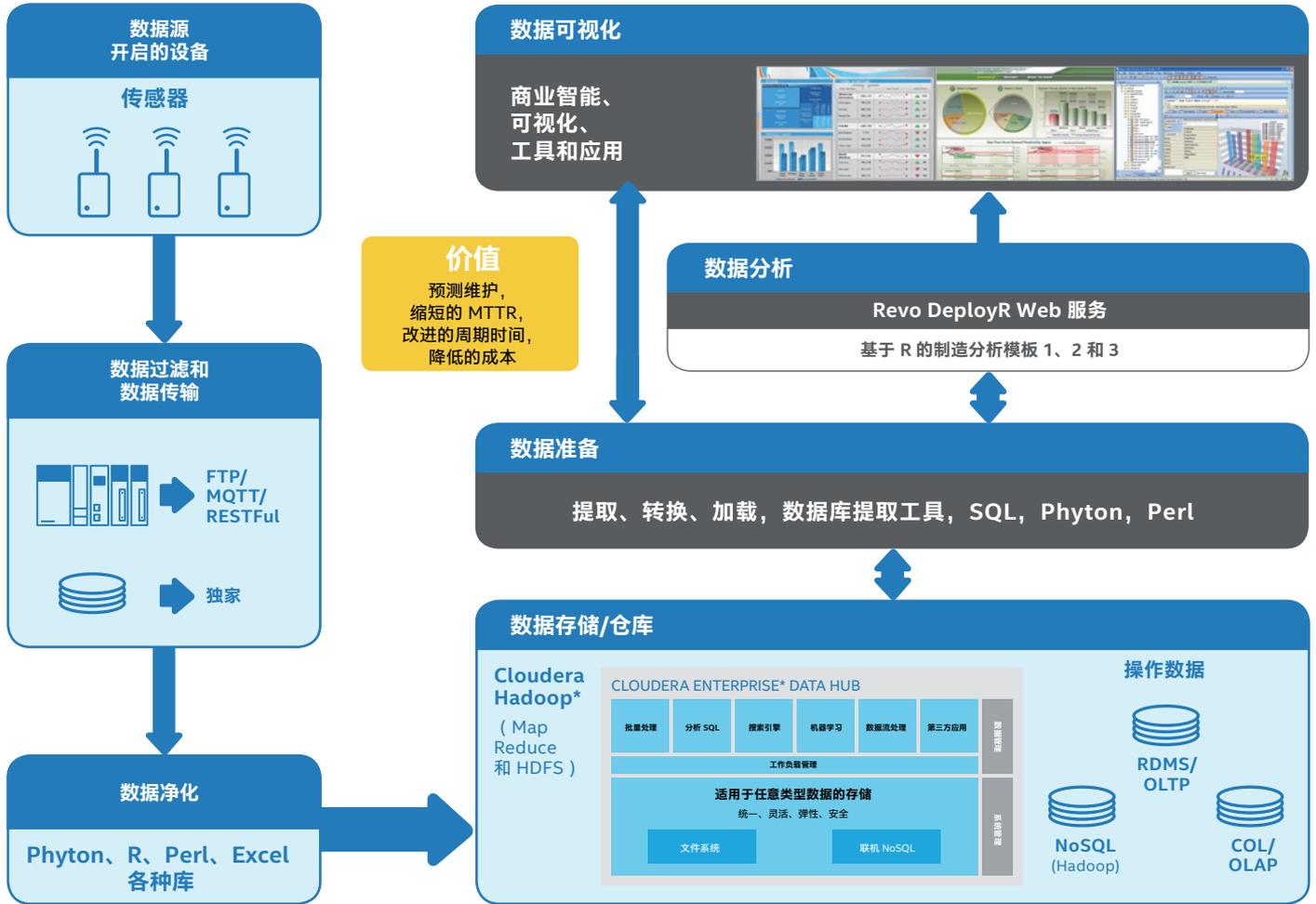


图 4. 高级数据流

数据流

图 4 显示了上述使用案例中的数据如何传输。

- 基于英特尔凌动处理器的网关将实时获取的机器数据和传感器数据发送至大数据分析服务器（BDAS）。例如，焊球焊接模块的机器数据通过机器界面端口收集，来自模拟传感器的机器数据以每秒数 MB 的速度进行流传输。

- 收集到的所有工厂数据都存储在 Hadoop* 中。以下不完全列表显示了使用 Hadoop 的现有功能可实现的可能性：
 - HTTP：大数据分析服务器展示一个支持对 HDFS 进行操作的经验证的 REST HTTP 端点。
 - Apache Sqoop* 提供了连接工具，用于将非 Hadoop 数据存储（如关系数据库和数据仓库）中的数据迁移至 Hadoop。
 - Flume* 能够接收持续的日志数据。
- 本部分上述使用案例中的特定数据格式为逗号分隔值（CSV）文件或原始图像。尽管 Hadoop 生态系统包含多种数据摄入途径（如上所述），但一些工厂机器的网络传输能力有限，要求通过定制工程向 HDFS 提交数据：

- FTP: 物联网网关拥有一个 FTP 客户端, 定期连接到大数据分析服务器, 并将最新获取的数据直接传输到 HDFS。也可以根据实时数据流和分析要求使用 MQTT 和 REST 等其他数据流协议。
- CIFS 共享 (Windows 共享) : 大数据分析服务器提供了一种 Windows/CIFS 共享目录, 网关可将文件复制到该目录中。
- CSV 文件被使用 Pig 直接导入 HBase*, 而原始图像会先使用计算机视觉技术和 map-reduce 任务进行预处理, 生成图像的文本数据表示。
- 根据操作要求, 数据将在三个数据库中的一个进行存储: NoSQL (Hadoop)、RDMS/SQL 或 Coli/OLAP。
- 数据库可使用各种工具访问和处理, 包括 AquaFold、专门报告、工作流调度、ETL 和数据库集成。同时, Cloudera Distribution for Hadoop 可对数据实施各种操作。
- 处理过的数据使用专为工厂应用设计的 Revolution Analytics 工具进行进一步分析。
- 这些数据在易于理解的仪表盘中向用户展示。

经验总结和结论

英特尔使用物联网网关, 借助从英特尔自身制造网络和设备及传感器提取的数据, 集成并验证了大数据分析内部服务器解决方案, 从而证实了物联网对于制造业的出色业务价值。采用了三菱电机 MELSEC-Q 系列 C 语言控制器。试点计划的实施有赖于工厂工程师、IT 部门和来自 Cloudera、戴尔、三菱电机和 Revolution Analytics 的行业专家的通力合作。团队首先采用现有机器性能和监控数据, 然后使用大数据分析和建模功能获取用于预测潜在偏移和故障的数据。机器组件故障预测功能可支持工程师修复并防止偏移, 从而通过避免浪费生产部件、缩短维修时间和减少机器备件使用量节约大量资金。

使用了大数据分析服务器和物联网网关上各种软件构建模块的集成套件。尚未开始利用制造数据中包含的智能优势的制造商, 可以应用并实施此框架。已经使用数据提升效率的制造商可以进一步增强现有能力, 以提高数据挖掘和分析能力。

对于英特尔而言, 此试点计划预计每年可节约数百万美元, 并能够带来更高的投资回报。优势包括延长设备组件正常运行时间、最小化良品和次品错误分类 (从而提高产能和生产力), 支持预测

在制造业, 大数据分析和物联网可用作端到端平台, 是实现智能制造的中坚力量。该平台可扩展, 使用当前行业标准构建模块可提供各种配置。

维护, 同时减少组件故障。英特尔的制造环境和设备中还有多种其他类型的参数、度量法、产品和设备数据 (结构化和非结构化), 可以通过挖掘和分析这些数据获取新的业务价值。通过利用这一机会, 英特尔将能够进一步提高工厂的效率和生产力, 建立出色竞争优势。

1. Gartner 的分析师 Doug Laney 于 2001 年在 MetaGroup 学术刊物上介绍了 3V 概念, 3D 数据管理: 控制数据数量、多样性和速度。http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.
2. www.dell.com/us/business/p/poweredge-vrtx/pd
3. www.revolutionanalytics.com/revolution-r-enterprise
4. en.wikipedia.org/wiki/MonetDB
5. www.postgresql.org/about
6. www.aquafold.com/
7. www.cloudera.com/content/cloudera/en/products-and-services/cloudera-enterprise.html
8. www.mitsubishielectric.com/fa/products/cnt/plcq/items/index.html
9. 资料来源: “2012-13 年英特尔 IT 年度报告”, www.intel.com/content/www/us/en/it-management/intel-it-best-practices/intel-it-annual-performance-report-2012-13.html.
10. 资料来源: “英特尔借助大数据削减制造成本”, 《信息周刊》, 2013 年 3 月 18 日, www.informationweek.com/software/information-management/intel-cuts-manufacturing-costs-with-big-data/d/d-id/1109111。
本文件中所描述的总体拥有成本或其他成本降低情形旨在让您更好地理解购买特定英特尔产品与许多特定情况变量结合后, 将会如何影响您今后的成本节省。本文件中的任何内容不应解释为特定成本的承诺或合约。
11. 资料来源: http://www.techopedia.com/definition/2148/automatic-test-equipment-ate
12. 根据制造过程中的封装大小、工艺和使用的设备, 结果可能不同。

本文件中包含关于英特尔产品的信息。本文件不构成对任何知识产权的授权, 包括明示的、暗示的, 也无论是基于禁止反言的原则或其他。除英特尔产品销售的条款和条件规定的责任外, 英特尔不承担任何其他责任。英特尔在此作出免责声明: 本文件不构成英特尔关于其产品的使用和/或销售的任何明示或暗示的保证, 包括不就其产品的 (i) 对某一特定用途的适用性、(ii) 适用性以及 (iii) 对任何专利、版权或其他知识产权的侵害的承担任何责任或作出任何担保。除非经过英特尔的书面同意认可, 英特尔的产品无意被设计用于或被用于以下应用: 即在这样的应用中可能因英特尔产品的故障而导致人身伤亡。

英特尔有权随时更改产品的规格和描述而毋需发出通知。设计者不应信赖任何英特尔产品所不具有的特性, 设计者亦不应信赖任何标有“保留权利”或“未定义”说明或特性描述。对此, 英特尔保留将来对其进行定义的权利, 同时, 英特尔不应为因其日后更改该等说明或特性描述而产生的冲突和不相容承担任何责任。此处提供的信息可随时改变而毋需通知。请勿根据本文件提供的信息完成一项产品设计。

本文件所描述的产品可能包含使其与宣称的规格不符的设计缺陷或失误。英特尔提供最新的勘误表备索。在发出订单之前, 请联系当地的英特尔营业部或分销商以获取最新的产品规格。索取本文件中或英特尔的其他材料中提的、包含订单号的文件的复印件, 可拨打 1-800-548-4725, 或登陆英特尔网站 www.intel.cn。

英特尔 © 2014 年版权所有。所有权保留。英特尔、Intel 标识、Intel Atom、英特尔凌动、Xeon 和至强是英特尔在美国和/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。

