

戴尔科技助力企业 数字化转型和降本增效实践

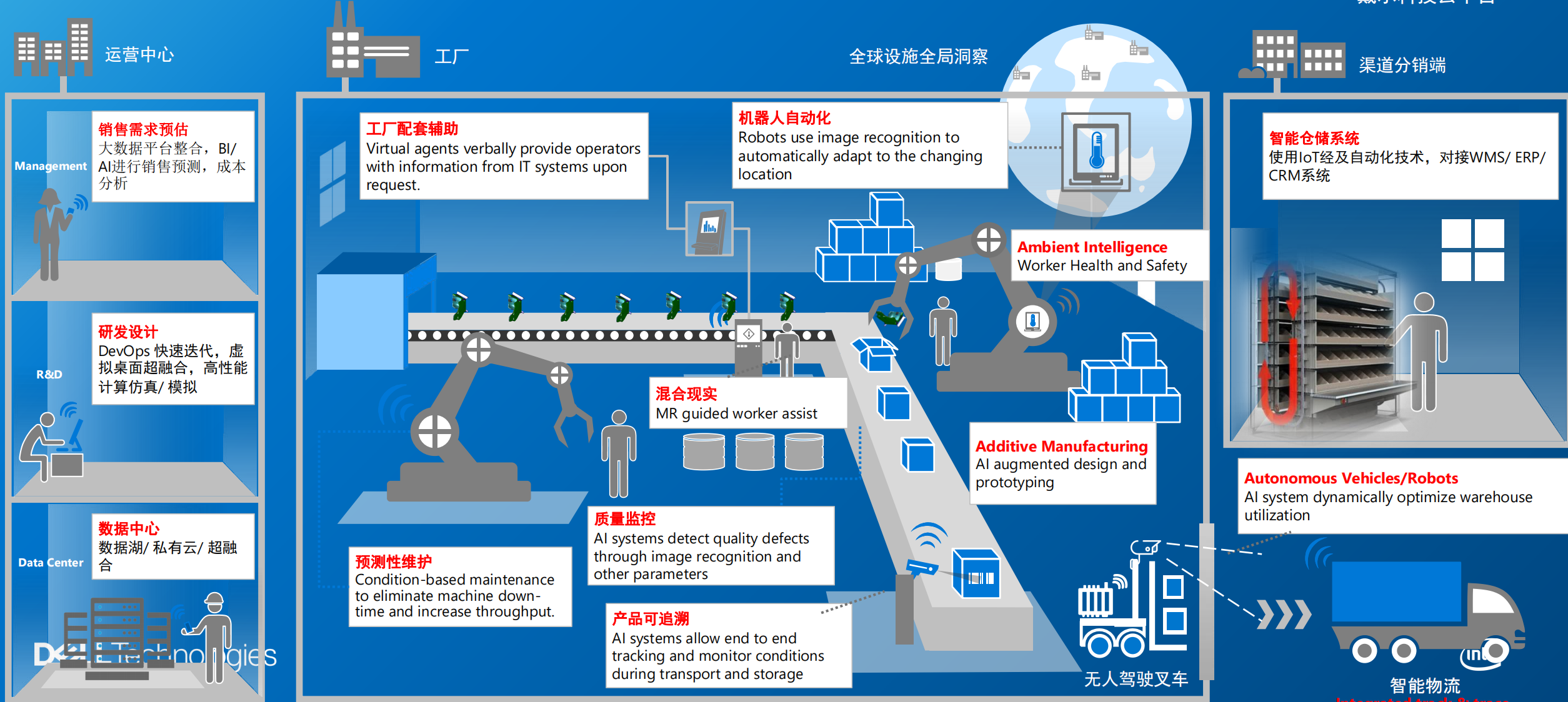
林帆

戴尔科技大中华区售前系统工程部
制造业首席技术官

Dell科技智造科技环节举例 – 汽车配件超级工厂

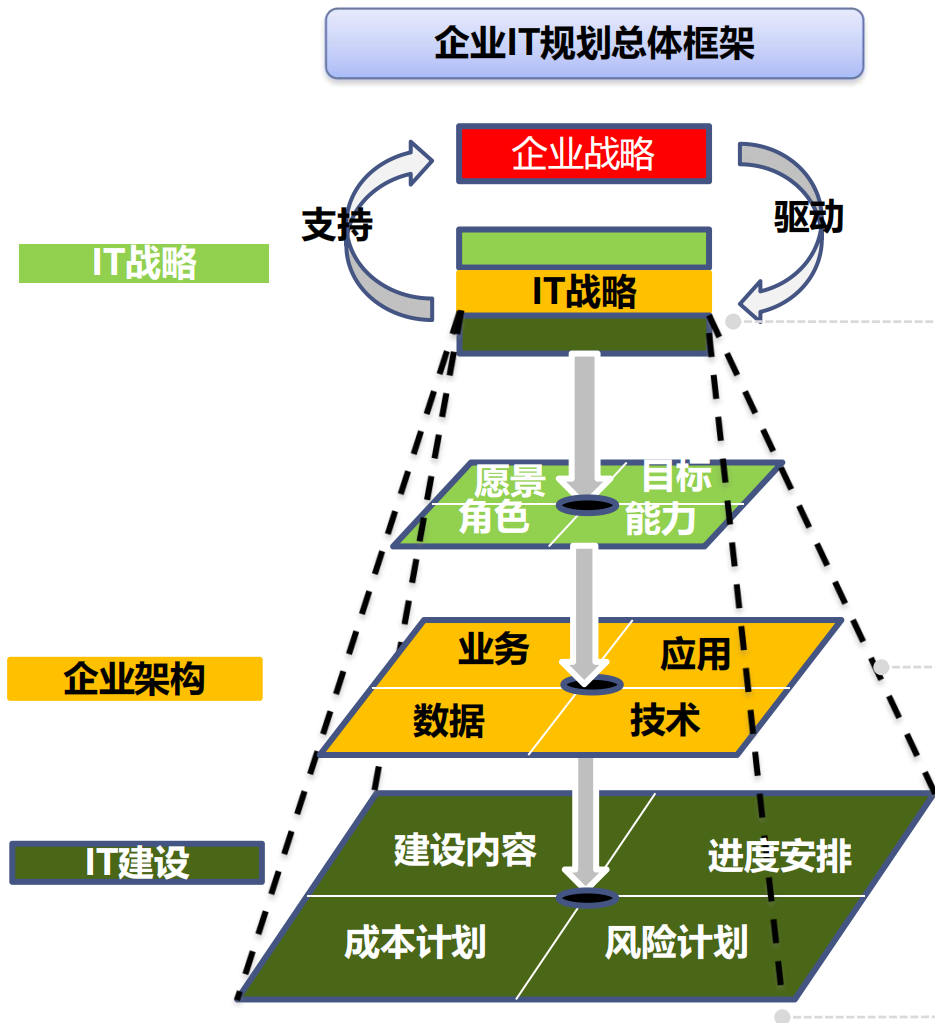


戴尔科技云平台



戴尔咨询助力企业数智化落地/预算, 规划, 效果

看五年, 想三年, 做一年/ 做五年, 看三年, 想一年的迷思



《顶层设计报告》

在深刻理解企业业务战略的基础上, 提出企业IT战略规划, 描绘信息化建设蓝图、建设方向与目标, **指导企业十四五信息化建设, 承接省、市政府信息化规划与监管, 是政府预算投入的依据。**

《架构设计报告》

设计出企业信息化业务架构与应用架构资产体系, **实现由信息化由规划转向实现的桥梁, 是企业信息化建设的标准, 指导与规范各分公司、业务部门信息化项目立项、设计与验收评审等。**

《项目规划报告》

规划出企业重点需要推进的项目建设内容、进度、预算及组织推进方案, **是推动企业信息化由战略转向实现与落地的重要抓手, 指导企业各分公司信息化招投标等建设管理。**

戴尔科技集团数字化工厂与AI 大数据平台打造能力

建立立起研发、采购、制造、储运、销售及服务一体化的数字化工厂，实现工厂的纵向集成和企业价值链的横向集成

智造 IT技术



智造 IT服务

企业战略与变革

- 企业战略
- 业务流程再造
- 组织结构设计
- 关键绩效指标设计
- 变革管理
- 运营体系设计
- 精益生产管理

IT战略及系统整合

- 信息战略与规划
- 企业信息治理
- 信息服务管理
- 企业数据规划、建模
- 企业安全及合规
- 企业信息标准
- 主数据管理

企业应用与解决方案

- 智能制造数字化工厂方案及实施
- 生产管理
- 供应链管理
- 营销销售管理
- Sap、Siemens定制化实施
- 架构设计、需求分析、设计
- 开发实施

IT技术服务与基础设施

- 技术架构设计
- 基础设施架构设计
- 云平台 (IaaS/PaaS/SaaS)及实施服务
- 大数据平台及实施服务
- AI平台及实施服务
- 物联网集成及服务
- IT运维

戴尔供应链工厂标准



全面的数字化工厂转型方案

采用独特的方法论，我们对客户具体情况快速制定针对性的解决方案，且任何两个方案都不会相同。



智能制造数字化工厂项目交付

采用敏捷项目管理，支持传统应用开发交付和PaaS云原生微服务开发交付，助力企业数字化转型。

- 作为全球领先的咨询服务与IT产品提供商，戴尔服务能帮助客户应对瞬息万变的外部环境，设计供应链协同模式，有效解决棘手企业智能制造相关业务和技术问题。我们与客户携起手来，拓展密切协同的伙伴合作关系，籍由精深专业经验、久经实践验证的最佳实践方法和工具，能有效协助客户实现战略目标。

XX轮胎智能工厂项目（6期）

主数据管理MDM项目介绍

在大数据时代，数据得到了各行各业的空前重视，数据增值价值成为共识，数据驱动业务成为企业IT发展的方向。国家层面推动两化融合、中国制造2025、智能制造过程中，数据提升到战略性基础资源的地位。对企业而言，主数据管理是做好数据治理、管理企业数据资产、挖掘数据价值的基础和先行工作，主数据管理平台致力于企业主数据的透明可视、触手可及，同时解决数据一致性、数据质量、数据共享等问题，汇聚并整合主数据，使主数据可见、可信、可用，形成企业资产并发挥数据价值。



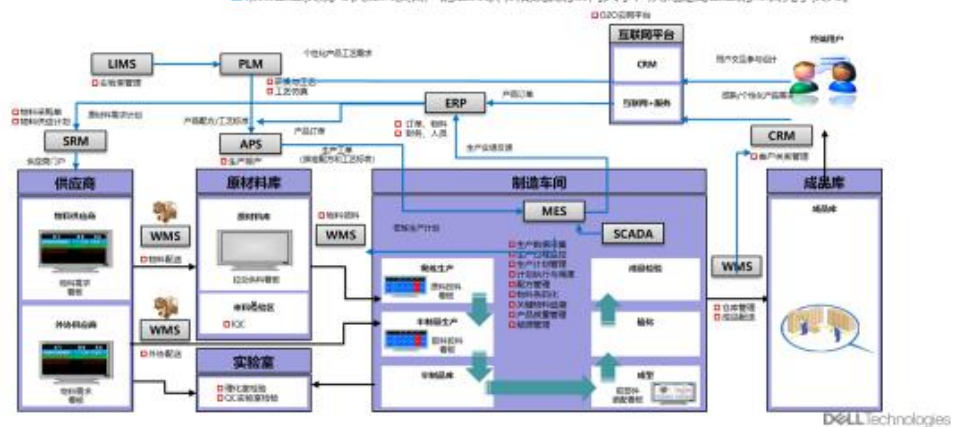
Internal Use - Confidential

11 of 17 © Copyright 2019 Dell Inc.

Dell Technologies

XXESB项目

- 提高XX业务流程执行效率30%以上，实现业务之间数据的实时共享。
- 帮助XX建设了企业级统一的服务总线，实现了产供销人财物各业务之间的横向集成。
- 是企业数字化转型中网络化的重点应用。
- 制定了企业的应用服务标准，为信息应用系统建设提供了规范和依据，避免了应用的重复建设，降低了系统建设成本。
- 帮助企业实现与供应商及客户的业务联动和数据协同共享，从而提高企业的综合竞争实力。



Internal Use - Confidential

13 of 17 © Copyright 2019 Dell Inc.

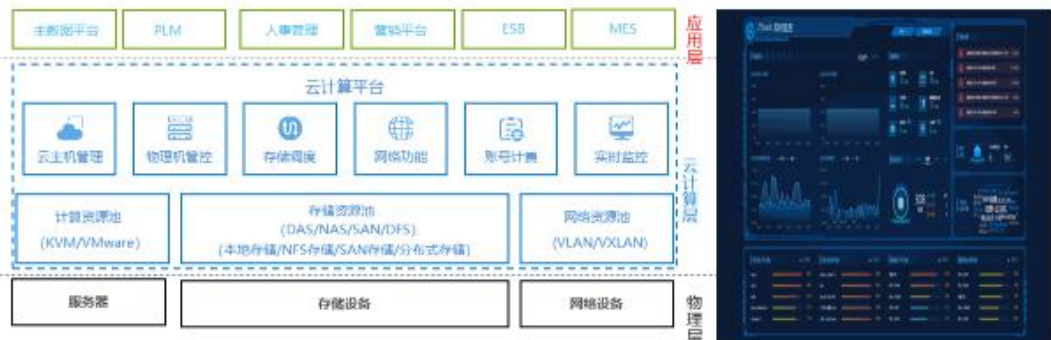
Dell Technologies

XX混合云项目

成本降50%以上

物理机29台服务器

虚拟机8台新购+8台利用



Internal Use - Confidential

15 of 17 © Copyright 2019 Dell Inc.

Dell Technologies

XXX全局能源管理项目

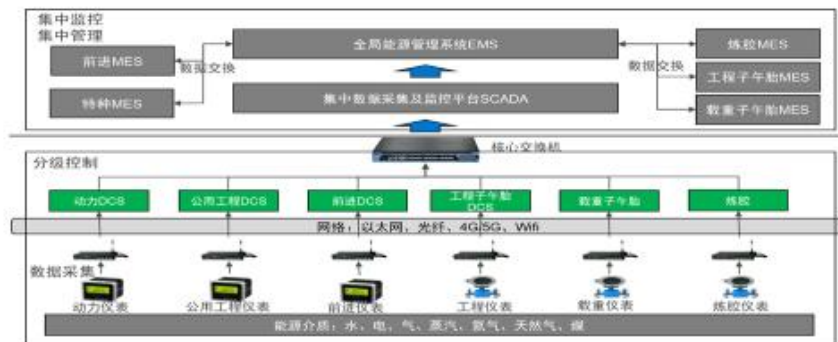
背景：XXX每次停电1小时损失¥70W收入，非计划性停机正在生产的轮胎会转为次品，造成5~10W损失。

目标：低碳生产，节能降耗，降本增效，绿色轮胎

科技手段：产线仪表IoT（苦力活），对接全局能源管理系统EMS，对接多个核心MES，实时监控，集中管理。



电力成本占整体运营成本近50%
降耗10%
¥3亿元



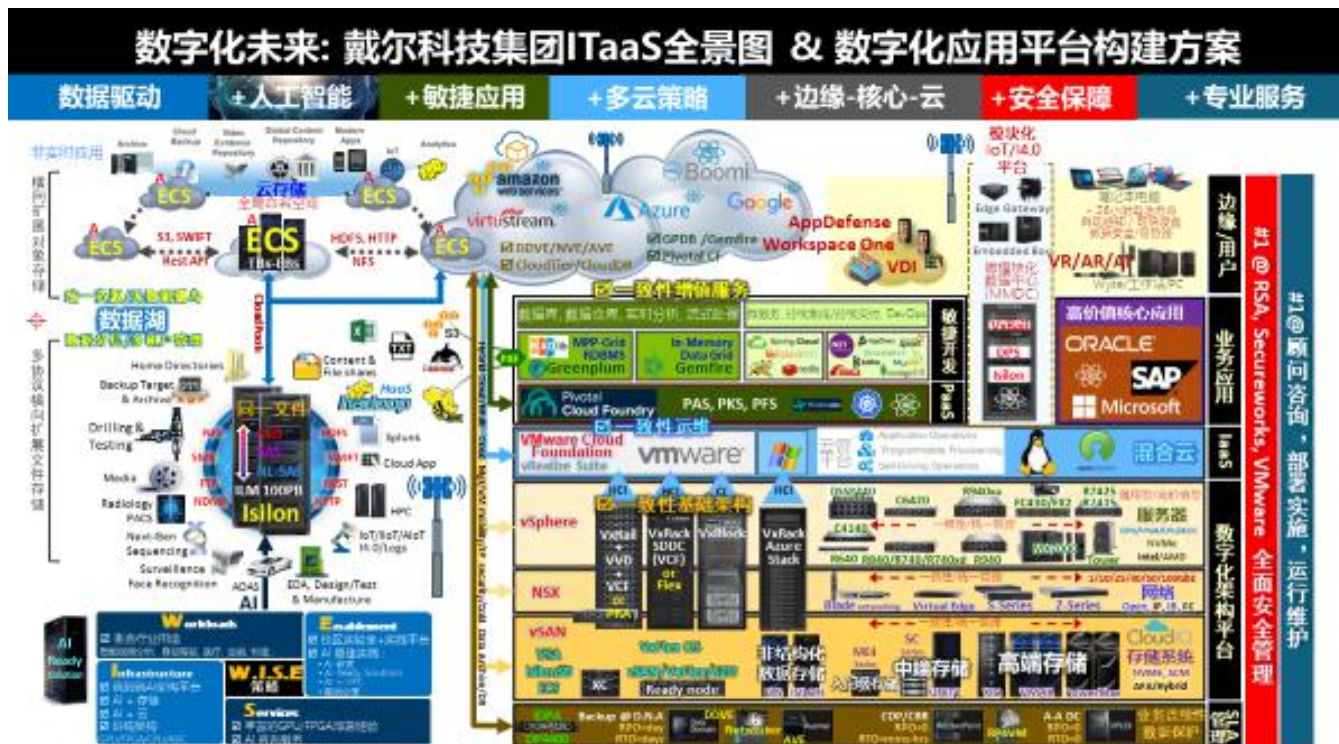
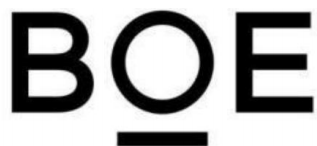
Internal Use - Confidential

19 of 17 © Copyright 2019 Dell Inc.

Dell Technologies

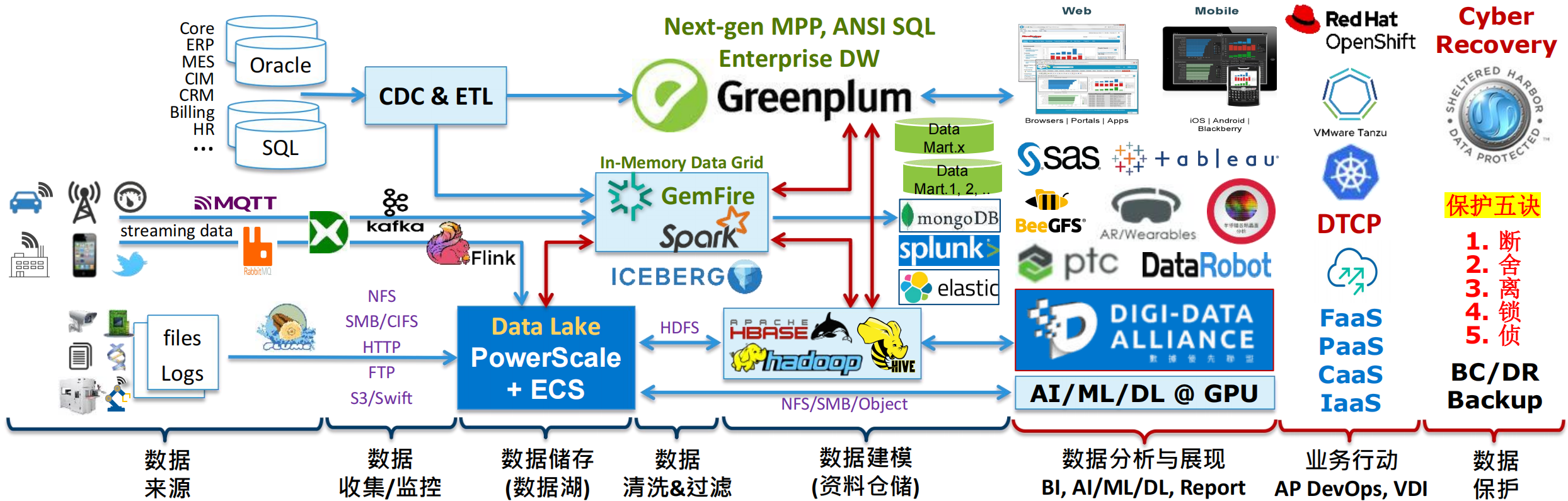


灯塔工厂建设中的数字化转型- 戴尔科技集团助力客户建设成长-落实到了增效，降本



戴尔科技智能工厂打造创意韧性: 极致大数据平台 IT 整合架构

☑ 极快效能; ☑ 极大扩充性; ☑ 极高可用性; ☑ 开放与开源; ☑ 极致性价比

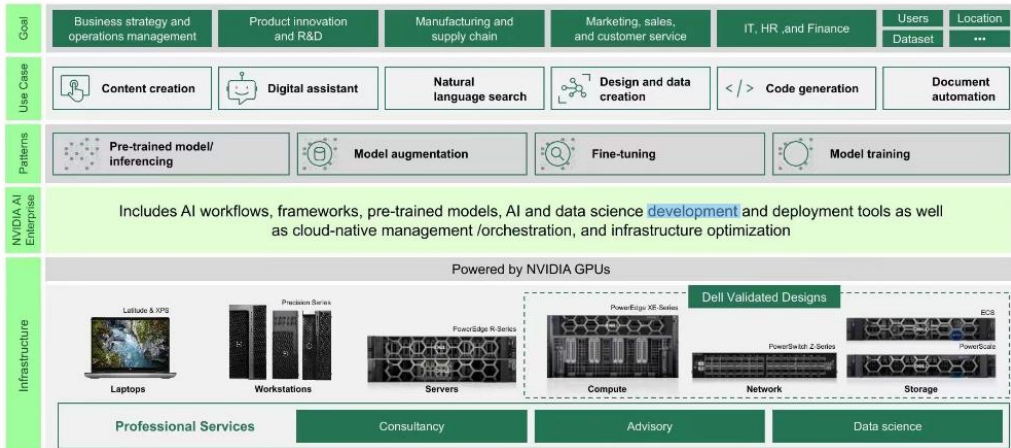


工欲善其事，必先利其器
(事 = 实现创意; 器 = 极致大数据平台 IT 整合架构)

生成式AI浪潮的降本增效

通过全新定义的产品线完成对GenAI的硬件和软件准备，并从原理上做好技术储备

Right-sizing your AI investment



Generative AI in the Enterprise

Scalable Production Infrastructure for Artificial Intelligence Large Language Models

May 2023

White Paper

Dell Technologies Solutions

Dell Technologies

Launching & Running Large Language Models on a single Dell server produces outstanding results

Launching & Running Large Language Models on a single Dell server produces outstanding results

Author: Ben Fisher, Ph.D., Senior AI Research Scientist and Distinguished Technical Staff, Microsoft, Redmond, WA, Senior Distinguished Engineer

Introduction
Large language models (LLMs) are an exciting new class of AI models that can process vast amounts of natural language text and generate human-like text. They are used in a wide variety of applications, from content creation to customer service. However, these models are computationally intensive and require significant infrastructure to run. This white paper discusses the challenges of running LLMs and provides a blueprint for a scalable, secure, and cost-effective infrastructure solution.

Large Language Models (LLMs)
LLMs are trained on vast amounts of text data and are capable of understanding and generating human-like text. They are used in a wide variety of applications, from content creation to customer service. However, these models are computationally intensive and require significant infrastructure to run.

Challenges of Running LLMs
Running LLMs on a single Dell server produces outstanding results. The Dell PowerEdge XE980 is a high-performance server designed and optimized to enable uncompromising performance for artificial intelligence, machine learning, and high-performance computing workloads. Dell PowerEdge is launching our innovative 8-way GPU platform with advanced features and capabilities.

Dell Technologies PowerEdge XE980 Server
The Dell PowerEdge XE980 is a high-performance server designed and optimized to enable uncompromising performance for artificial intelligence, machine learning, and high-performance computing workloads. Dell PowerEdge is launching our innovative 8-way GPU platform with advanced features and capabilities.

Key Features:

- 8x NVIDIA H100 80GB 700W SXM GPUs or 8x NVIDIA A100 80GB 500W SXM GPUs
- 32x Fourth Generation Intel® Xeon® Scalable Processors
- 32x DDR5 DIMMs at 4800MT/s
- 12x PCIe Gen 5 1.6 ETH Slots
- 8x SAS/NVMe SSD Slots (U.2) and BOSS-N1 with NVMe RAID

Dell PowerEdge XE980

2-socket Config:

- Up to 64 GB Dual In-Line Memory Modules (DIMMs)
- Up to 32 DDR5 DIMMs
- Up to 4800 MT/s (DDR5)
- All are cooled up to 10°C ambient
- 100W max TDP

Support for high speed and memory capacity:

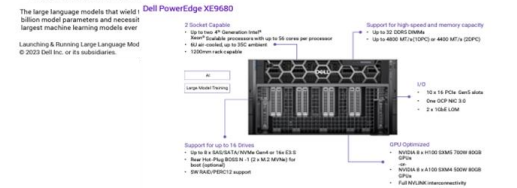
- Up to 32 DDR5 DIMMs
- Up to 4800 MT/s (DDR5)
- 1x 100W max TDP
- 1x 100W max TDP

Support for up to 14 drives:

- 8x SAS/NVMe SSD Slots (U.2)
- New 14.1" 8-Bay BOSS-N-1 (2x M.2 NVMe) for boot
- 2x 100W max TDP
- 2x 100W max TDP

GPU Options:

- 8x NVIDIA H100 80GB 700W SXM GPUs
- 8x NVIDIA A100 80GB 500W SXM GPUs
- 8x NVIDIA A100 80GB 500W SXM GPUs
- 8x NVIDIA A100 80GB 500W SXM GPUs



1 Launching & Running Large Language Models on a single Dell server produces outstanding results

Harnessing the Power of Specialized Language Models like ChatGPT for Business Success

Harnessing the Power of Specialized Language Models like ChatGPT for Business Success

Author: Ben Fisher, Ph.D., Distinguished Member Technical Staff, Senior AI Research Scientist, Dell Technologies

Introduction
Large language models (LLMs) are an exciting new class of AI models that can process vast amounts of natural language text and generate human-like text. They are used in a wide variety of applications, from content creation to customer service. However, these models are computationally intensive and require significant infrastructure to run. This white paper discusses the challenges of running LLMs and provides a blueprint for a scalable, secure, and cost-effective infrastructure solution.

Large Language Models (LLMs)
LLMs are trained on vast amounts of text data and are capable of understanding and generating human-like text. They are used in a wide variety of applications, from content creation to customer service. However, these models are computationally intensive and require significant infrastructure to run.

Challenges of Running LLMs
Running LLMs on a single Dell server produces outstanding results. The Dell PowerEdge XE980 is a high-performance server designed and optimized to enable uncompromising performance for artificial intelligence, machine learning, and high-performance computing workloads. Dell PowerEdge is launching our innovative 8-way GPU platform with advanced features and capabilities.

Dell Technologies PowerEdge XE980 Server
The Dell PowerEdge XE980 is a high-performance server designed and optimized to enable uncompromising performance for artificial intelligence, machine learning, and high-performance computing workloads. Dell PowerEdge is launching our innovative 8-way GPU platform with advanced features and capabilities.

Key Features:

- 8x NVIDIA H100 80GB 700W SXM GPUs or 8x NVIDIA A100 80GB 500W SXM GPUs
- 32x Fourth Generation Intel® Xeon® Scalable Processors
- 32x DDR5 DIMMs at 4800MT/s
- 12x PCIe Gen 5 1.6 ETH Slots
- 8x SAS/NVMe SSD Slots (U.2) and BOSS-N1 with NVMe RAID

Dell PowerEdge XE980

2-socket Config:

- Up to 64 GB Dual In-Line Memory Modules (DIMMs)
- Up to 32 DDR5 DIMMs
- Up to 4800 MT/s (DDR5)
- All are cooled up to 10°C ambient
- 100W max TDP

Support for high speed and memory capacity:

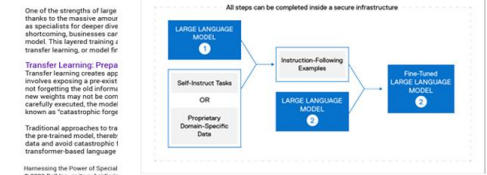
- Up to 32 DDR5 DIMMs
- Up to 4800 MT/s (DDR5)
- 1x 100W max TDP
- 1x 100W max TDP

Support for up to 14 drives:

- 8x SAS/NVMe SSD Slots (U.2)
- New 14.1" 8-Bay BOSS-N-1 (2x M.2 NVMe) for boot
- 2x 100W max TDP
- 2x 100W max TDP

GPU Options:

- 8x NVIDIA H100 80GB 700W SXM GPUs
- 8x NVIDIA A100 80GB 500W SXM GPUs
- 8x NVIDIA A100 80GB 500W SXM GPUs
- 8x NVIDIA A100 80GB 500W SXM GPUs



2 Illustration demonstrates the use of one, or multiple, large language models to either: 1) fine-tune a large language model on proprietary domain-specific data, or 2) use a language model to further refine proprietary domain-specific data for downstream language model fine-tuning. All steps can be securely performed within a customer's infrastructure with Dell Technologies private cloud and/or on-premises platforms.

Dell Technologies can help you on this journey. Recent results from a research group at Stanford University demonstrated the benefits of using a large language model to further fine-tune and refine an existing large language model, resulting in a new model known as ALPACA. This process demonstrated a new paradigm in which one large language model was used to train another large language model. This approach could further reduce the time and cost of the model training loop, such as incorporating learning with human feedback (RLHF), as previously leveraged by InstructGPT and ChatGPT.

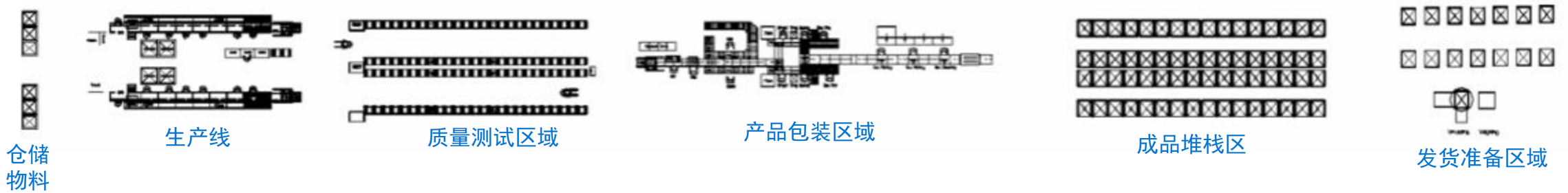
In ALPACA's case, a large generalist language model created a curated series of questions and answers on which to fine-tune another. Fine-tuning the second language model with questions generated by the first resulted in improved performance metrics of the second model. This paradigm has since been expanded to other domains outside of questions and answers to enable rapid and efficient tuning of existing large language models on new data.

案例1：产线持续优化 – 自动化/ 数字化/ 智能化 厦门DELL工厂

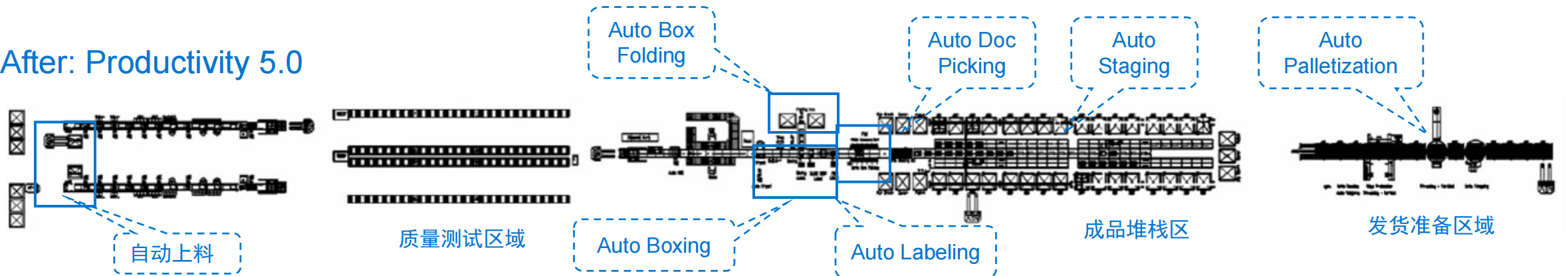
- 3年/ 5年智能制造规划
- 产线/ 车间 持续优化
- 边缘IoT/ 边缘AI计算/ AI视觉/ 后端大数据/ BI 数据深度多维度可视化

 **56%**
Improvement

Before: Productivity 3.2



After: Productivity 5.0



案例2：产品外观检测 - Dell 笔记本 Latitude A cover 组装线

[AI 机器视觉识别]

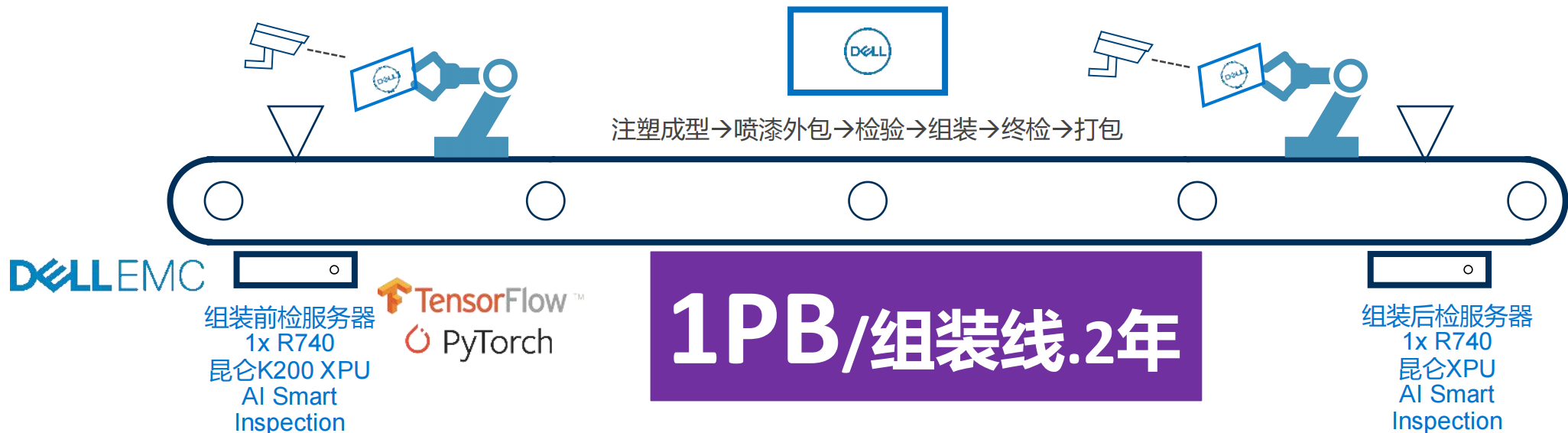
TRIVISION AI 模型，筛查十大类常见缺陷，并告知缺陷类别和位置以及可能成因,并继续学习新缺陷，其中：

- 灰度算法，筛查严重不良；
- 比良算法，筛查未经学习的缺陷，比对未经学习过的缺陷，不良防堵以防漏失。

Data scale: **每10秒**完成一个 A cover 组装，产生**250MB**数据，**1天**产生**2TB**数据。资料需保留**2年**，**约1PB**数据储存在 Isilon 上

[具体成效]

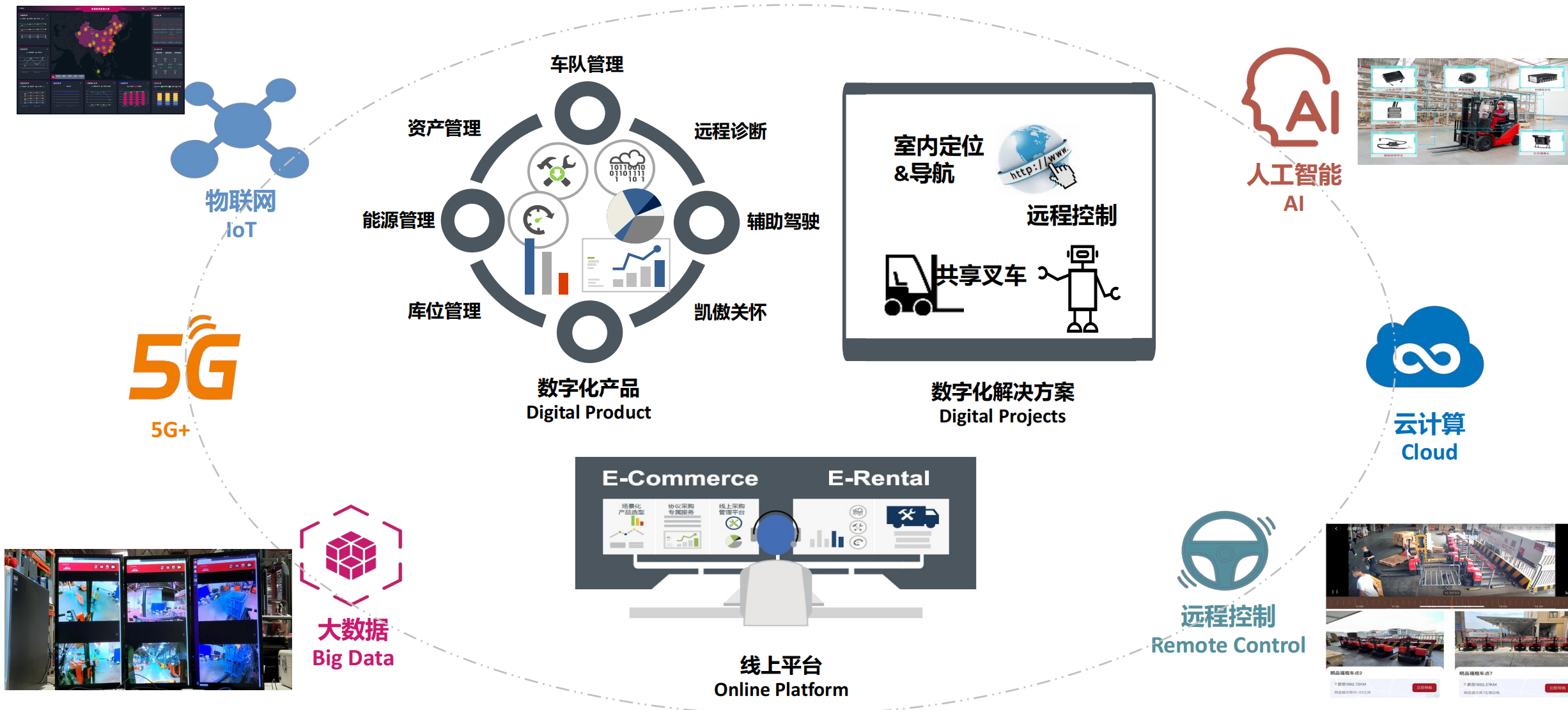
- ✓ 该环节的良品率 yield rate 提升 **10%+**
- ✓ 漏杀率 missing rate (把不良品误判为良品) **从 0.5% 到 0.3%**;
- ✓ 过杀率 overkill rate (把良品判为不良) **从 5 % 到 3%**;
- ✓ 节省组装线人力配置: 一条流水线设 **5 个人工**检测站，机构件厂还需要外派人员到ODM做入料筛查和随线筛查。



案例3 . 凯奥集团集团林德叉车新智能工厂项目



• 生产能力提高27%、工艺质量的检测速度提高30%、能源使用量减少28%、总体效能提高12%



戴尔科技助力头部电子制造企业和物流巨头



PowerScale
大数据存储

PowerEdge
服务器

SAP HANA
综合解决方案

Siemens
PLM解决方案

运营决策 Decision

关键
决策
环节
智能化

价值流导向的
数字化

精益为基础的
柔性自动化

数据积累 Data

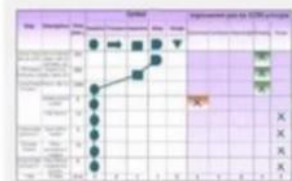
柔性自动化、数字化、智能化技术打造从“无人”到“无忧”的智能工厂



- AI加持自动智能决策;
- AI寻优, 辅助决策;



- 数字化平台提升六流活动协同性
- 基础信息化系统导入、业务流程数字化
- 设备物联, 现场透明;



- 柔性自动化产线实现少人、无人
- 精益筑基: 流程固化、现场改善、价值流分析

3W目标

- Work Reduction
- Waste Reduction
- Worry Reduction

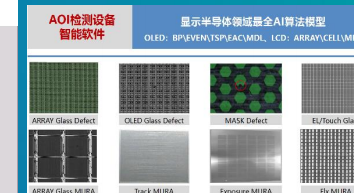
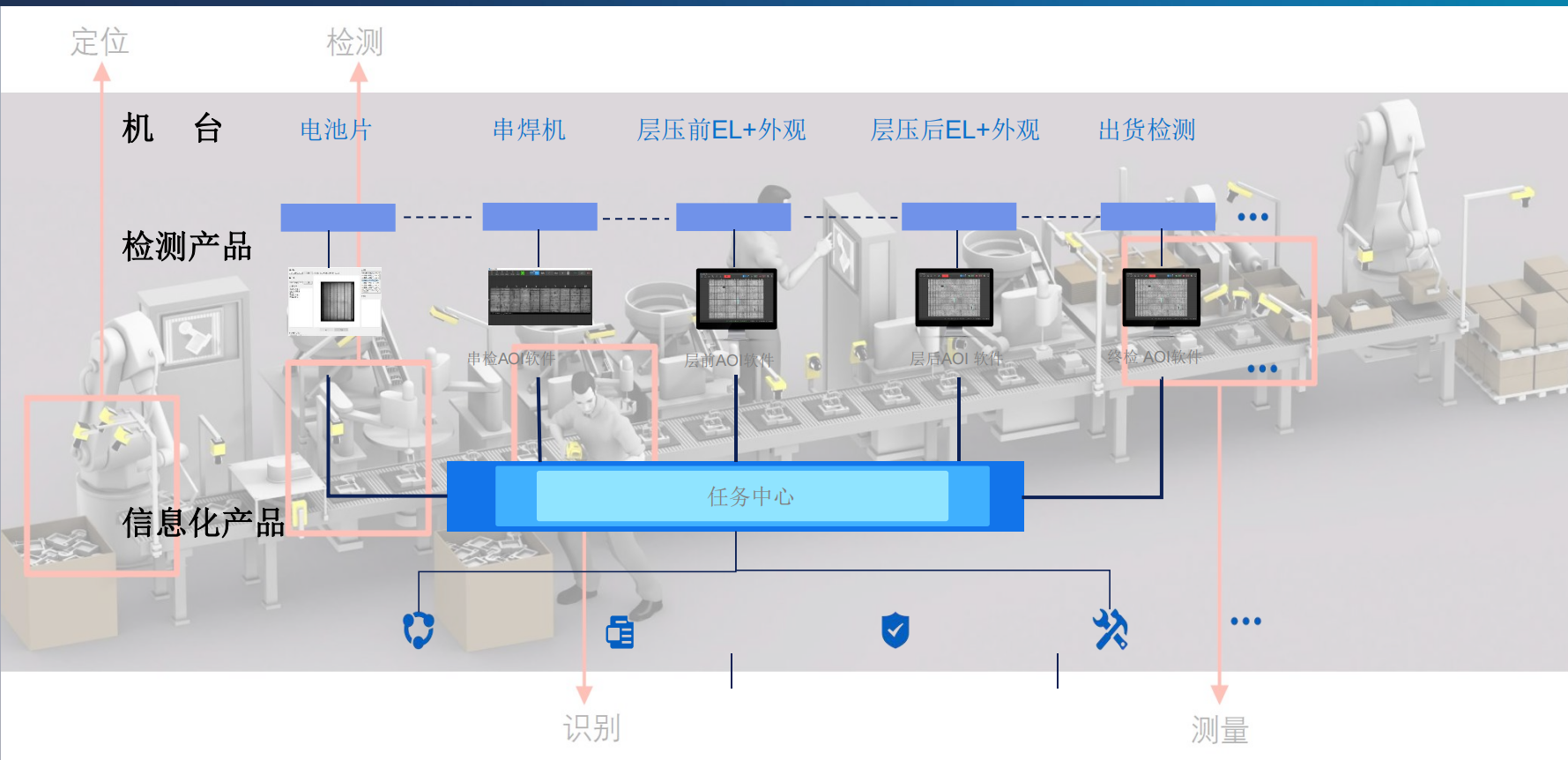
生产效率
30%

库存周期
15%

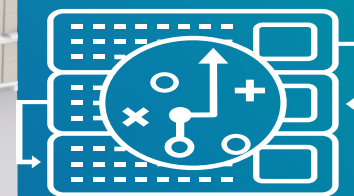
生产人力
92%

数字创新 - 智能工厂新技术新生态

SIEMENS



hongpu 洪朴



ptc
DIGITAL TRANSFORMS PHYSICAL



cogiot
native

边缘业务 拉动 数据算力/数据治理的需求

- **EDGE + AI** ---- 缩减产线时间, 实时监测产线产品质量, 提高产线检测精度和速度。
- **HPC + AI** ---- 提高专业人员工作价值, 依据算法, 投入工艺改进和实验, 提高生产效率。
- **数据管理** ---- 数据中心化处理, 生成丰富的质量信息, 为后续工艺优化提供数据基础。

intel®

DELL Technologies

戴尔边缘创新联合实验室

地址： 戴尔科技集团中国研发集团上海分公司 上海市杨浦区淞沪路252号 创智天地2号楼2楼



边缘计算应用
展示中心



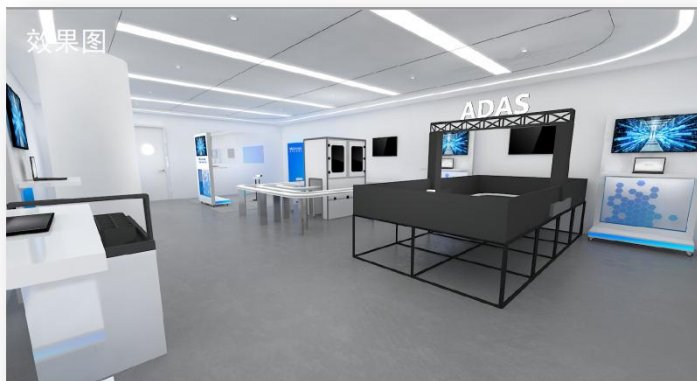
边缘计算方案
验证中心



生态合作伙伴
方案测试中心



客户及生态伙伴
赋能中心



DELLTechnologies