

[第19回 PCクラスタシンポジウム]

HITACHI
Inspire the Next

日立ディープラーニング・クラウドのご紹介

2019/12/12

株式会社 日立製作所
IoT・クラウドサービス事業部
HPCソリューションセンタ
主任技師 寺本 康弘

Contents

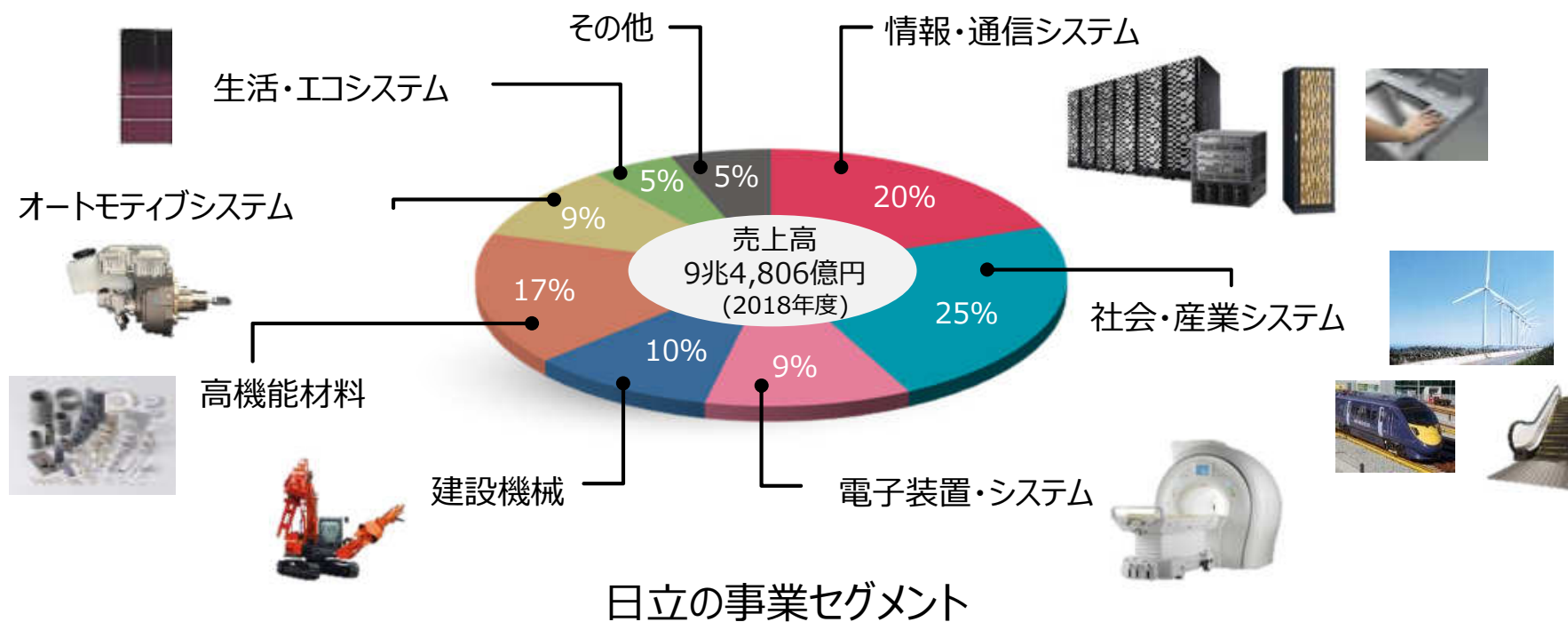
1. 日立のディープラーニング・ソリューション
2. SR24000/DL1, NVIDIA DGX
3. ディープラーニング・クラウドサービス

-
- 1. 日立のディープラーニング・ソリューション**
 2. SR24000/DL1, NVIDIA DGX
 3. ディープラーニング・クラウドサービス

1-1. 株式会社 日立製作所のご紹介

会社概要

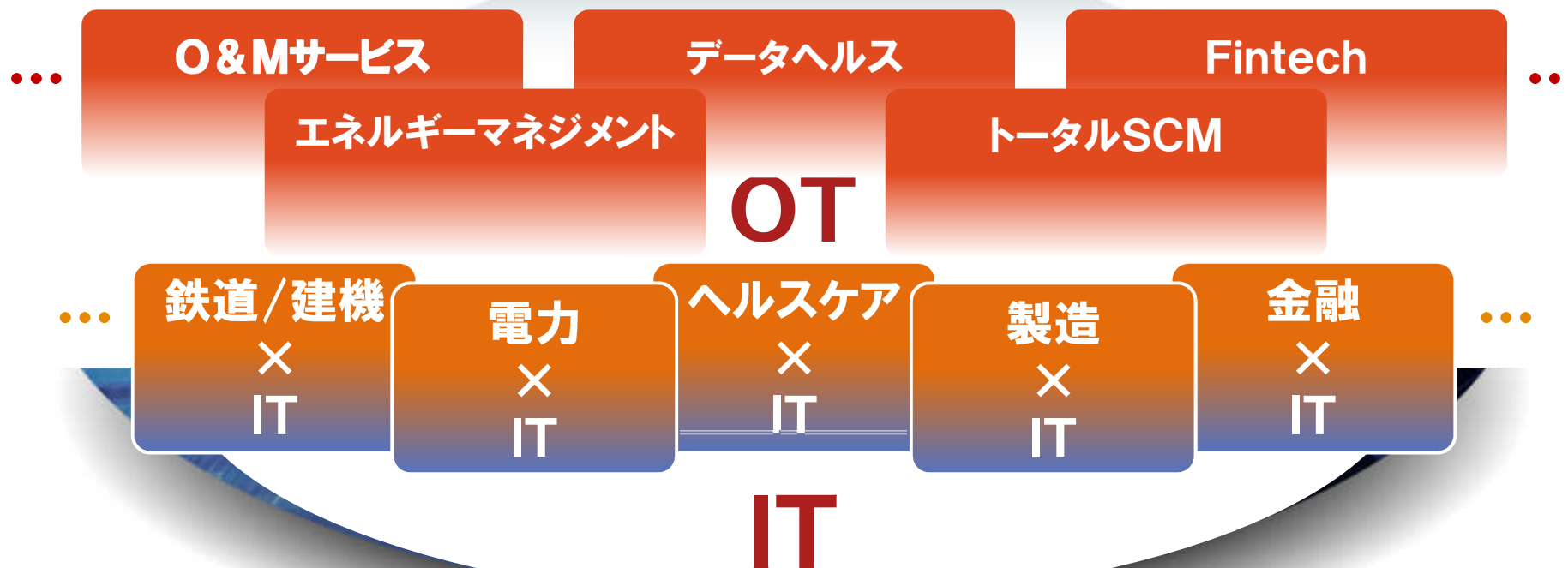
- 設立年月日 大正9年(1920年)2月1日 (1910年創業)
- 執行役社長兼CEO 東原敏昭
- 連結従業員数 295,941名 (2018年度)
- 連結売上収益 9兆4,806億円 (2018年度)



1-2. 日立の社会イノベーション事業への取り組み

協創・オープンイノベーションによる社会課題解決、
社会のイノベーション

OT×IT融合サービス／異業種連携ビジネス／ビジネスエコシステム

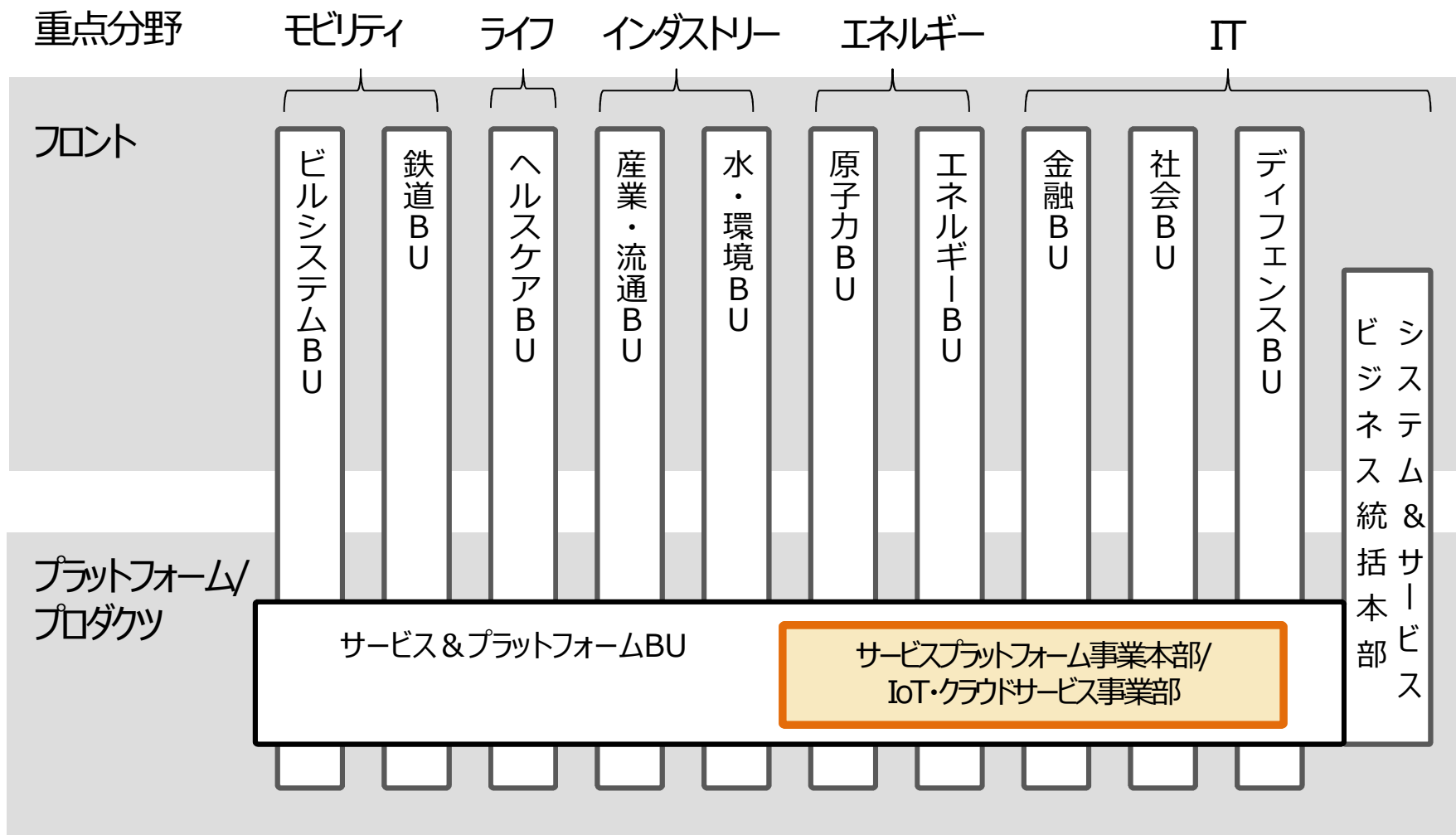


※ OT : Operational Technology
※ IT : Information Technology

※ O&M : Operation & Maintenance
※ SCM : Supply Chain Management

※ Fintech : Financial Technology

1-3. 社会イノベーション事業を加速する体制



※ BU：ビジネスユニット

日立の組織

DL 研究開発のプラットフォームから、
DL技術を活用した、幅広いソリューションを提供します。

お客さま	お客さまのニーズ	日立の商材
1 大学・研究所・民間企業の研究開発部門	研究開発向けGPUシステム	SR24000/DL1 NVIDIA DGX
2 官公庁・民間企業の開発部門・	DLモデル開発・業務用GPUリソース	日立DLクラウドサービス
3 製造業・インフラ企業の生産技術部門・保守点検部門	目視検査の代替	DL目視検査代替サービス
4 金融業・BPOの帳票の事務処理部門	AI-OCR	DL手書き帳票認識サービス

※ DL : ディープラーニング

-
1. 日立のディープラーニング・ソリューション
 - 2. SR24000/DL1, NVIDIA DGX**
 3. ディープラーニング・クラウドサービス

2-1. SR24000/DL1 (V100) のご紹介

スーパーテクニカルサーバ SR24000シリーズ モデルDL1に 最新のIBM POWER9とNVIDIA Tesla V100 SXM2を搭載



SR24000/DL1 (V100)

最新の**NVIDIA Tesla V100 GPU**を2or4基搭載。
IBM POWER9とGPUが**NVIDIA NVLink 2.0**により
直接接続する画期的なシステムです。

サーバメモリは**最大1TB** 搭載。**PCI-Exp Gen.4**のPCI
スロット (LP) を4スロット備え、外部ネットワーク (10GbE,
InfiniBand) など柔軟に構成可能。

ディープラーニングおよびHPCの研究開発、業務利用に
必要な、64bit Linux OS、CUDA、主要なDLフレーム
ワークなどの開発環境をサポート。

GPUs	Tesla V100 SXM2 x 2 or 4	
FP16/FP64	31.2/7.8TFLOPS x 2 or 4	
GPU メモリ	16GB or 32GB x 2 or 4	
CPU	POWER9 20c x 2	
CPU メモリ	DDR4 256GB, 512GB,1TB	
内蔵ストレージ	1TB, 2TB 7200RPM HDD x 2 or 1.92TB SSD x 2	
ネットワーク	1GbE, 10GbE, 25GbE, IB EDR	
ソフトウェア	•Ubuntu Server Linux OS •Red Hat Enterprise Linux for POWER	
物理 諸元	サイズ	W:443 x H:86 x D:850 (mm)
	重量	30Kg
	最大電力	2,500W
	電源	200V AC
	動作環境	18 - 27 °C (推奨)

- ※ NVIDIA Tesla, NVLinkおよびCUDAは米国およびその他の国におけるNVIDIA Corporationの商標または登録商標です。
- ※ IBM, POWER9は米国およびその他の国におけるInternational Business Machines Corporationの商標または登録商標です。
- ※ Ubuntuは、Canonical Ltd.の商標または登録商標です。
- ※ Linuxは、Linus Torvalds氏の日本およびその他の国における登録商標または商標です。
- ※ Red Hat, Red Hat Enterprise LinuxはRed Hat, Inc.の米国およびその他の国における登録商標または商標です。

2-2. SR24000/DL1 のサービス

システム利用開始を支援する多様なサービス



- 専門エンジニアがシステム導入時の疑問、不安の解消をサポートします
- SR24000/DL1導入時にお客さまの**システム利用開始**を支援します
- **お客さまの実データ**を用いたディープラーニング学習の立ち上げを支援します

お客さまのディープラーニング研究や業務活用を支えるサービスを提供します

2-3. SR24000/DL1導入事例

SR24000/DL1(V100)が医療研究の最前線で活躍
従来システムでは 1.5ヶ月の解析が、1.5~3日に大幅短縮



高性能な並列演算機能を持つGPU^{*1}をディープラーニング(深層学習)に応用して、大規模シミュレーションや画像認識、医療研究などに役立てようという動きが活発化しています。国立研究開発法人 国立成育医療研究センターは、日立のスーパーテクニカルサーバ「SR24000シリーズ」のディープラーニング専用モデル「SR24000/DL1」を導入。超高速なビッグデータ解析環境を、医療研究に幅広く適用する取り組みを開始しました。 *1 Graphics Processing Unit

全ゲノム解析への対応を図るため
「SR24000/DL1」を導入

研究所は、原因不明で治療の進まない症
状を持つ小児患者さんに対し、ゲノム解

大規模シミュレーション、大量データを用
いた科学技術計算においても高い性能

2-4. NVIDIA DGXファミリー

パーソナルユースのデスクトップモデル (DGX Station) から、
最高性能のDGX-2 まで顧客ニーズに合わせて提案

NVIDIA DGX ファミリー NVIDIA製 AI スーパーコンピュータ

4x Tesla V100
500 TFLOPS
160GB/s NVLink
水冷システム
1500 W

8x Tesla V100
1 PFLOPS
160GB/s NVLink
3500 W

16x Tesla V100
2 PFLOPS
P2P 300GB/s NVSWITCH
10 kW



-
1. 日立のディープラーニング・ソリューション
 2. SR24000/DL1, NVIDIA DGX
 - 3. ディープラーニング・クラウドサービス**

3-1. DLクラウドサービスの開発の背景と特徴

社会的背景とお客さまご要望

- 働き方改革により、帳票入力など単純作業の省力化の要求
- 製造業の目視検査などを行う専門家の人材不足、高齢化の社会的課題



AI利用における課題

- AI/DL利用はカスタマイズが必要で、本格的な導入には時間がかかる
- 企業の基幹業務に利用できるセキュリティ、高信頼性のサービスではない
- 単発の導入では継続性がない

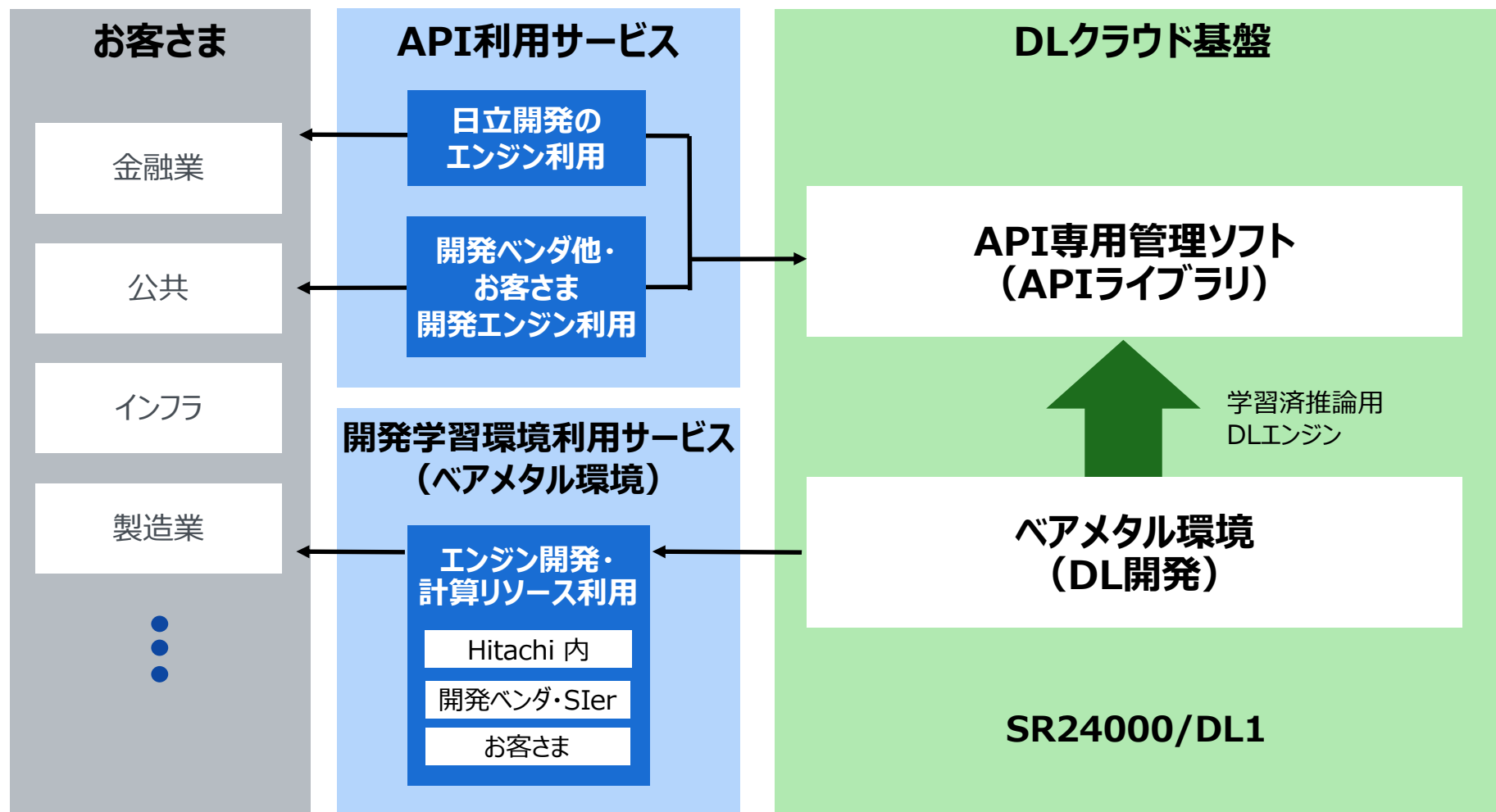


DLクラウドサービス

- AI/DLを用いた各種解析サービスをクラウドサービス提供
- 企業の基幹業務でもご利用いただける高信頼、高可用性を実現
- 解析エンジンの開発環境も提供し、サービス運用と開発を連続して実現

3-2. DLクラウドサービスの概要

DLクラウドの全体像



※ 現在開発中の内容を含みます。将来変更の可能性があります。

ディープラーニングのモデル開発、学習に必須のNVIDIA GPUをクラウド利用

基本サービス

GPUサーバ

- サーバ** SR24000/DL1
- 仕様** POWER8/9 x2基、
NVIDIA Tesla P100/V100 x4基、
512GBメモリ、SSD (1.9TB x2)



SR24000/DL1

ソフトウェア

- OS** Ubuntu 16.04/18.04 LTS for
POWER (ppc64el)

※ インターネット接続はプロキシ経由で可能

接続回線

SSL-VPN 1回線あたり100Mbpsベストエフォート

オプションサービス

導入支援サービス

ベアメタル環境へのインストール代行など
導入支援を提供

データ持ち込み支援サービス

ベアメタル環境への大容量の
お客さまデータ持ち込みを支援



※ ご参考価格（税別）：
75万/月（DL1/P100 4GPU, 512GB, 1回線）

3-4. API利用サービス

さまざまなDLエンジンをRestAPI で利用可能

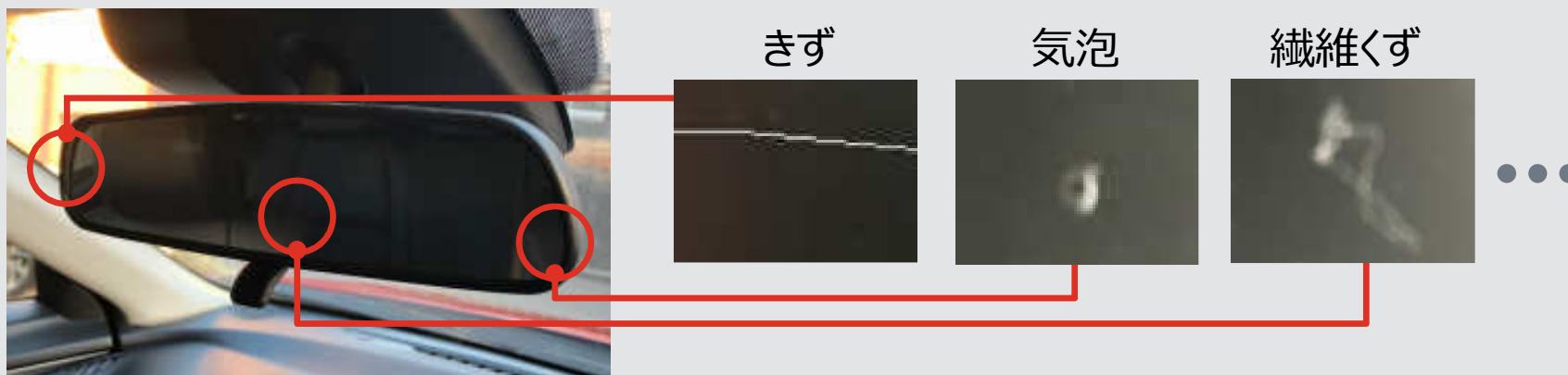
- お客様開発のDLエンジン
- ソフトウェア開発ベンダー・SIer 開発のDLエンジン
- 日立開発のDLエンジン
 - 画像解析エンジン
 - 手書き文字認識エンジン

日立API利用サービスは、独自のレイテンシ設計・スループット設計により、リアルタイム系業務への組み込みができます。

3-5. 日立DL技術の紹介 (1) 画像解析の事例

自動車部品製造メーカー様の事例

課題：外観検査装置の検出精度が不十分で**70%以下**



学習方法	きず、気泡、繊維くずなどを不良区分ごとに学習
学習データ数	約14,000枚
データ加工	回転, 反転
モデル	CNN (Convolutional Neural Network)
不良品検出率	95%以上




お客さまからのコメント

熟練工の目視・従来の画像診断と比べ実用化に十分な認識率

3-6. 日立DL技術の紹介 (2) 手書き文字認識

従来のOCR技術とディープラーニング技術の融合により、
より実用的な文字認識を実現します。

- 1 楷書の手書き文字では従来のOCR方法と同等の認識率を確認
- 2 罫線混入手書き文字など、既存OCR技術が苦手な認識の認識率向上
- 3 くせ字・崩し字なども認識（お客さま固有の個別文字にも対応）

問題	ディープラーニング学習前		ディープラーニング学習後
1 手書き文字認識 	99.0%	同等	99.1% <small>(2017年 日立内の実証実験結果)</small>
2 罫線混入手書き文字認識 	64.8%	向上	98.5% <small>(2017年 日立内の実証実験結果)</small>
3 個別文字への対応 	20%以下	向上	98%以上 <small>(2018年 日立内の実証実験結果)</small>

※ [2][3] は追加学習が必要となります。

3-7. DLクラウドサービス ～初めてご利用の方 キャンペーン～

NVIDIA Tesla P100 4基インスタンス限定

**高性能GPU環境を30日間お試し下さい！
ご利用は簡単！年内申し込み・先着5名限定**

Step-1
Web申し込み



Step-2
日立よりご連絡



Step-3
ご利用開始

【申し込み方法】

下記Webのお問い合わせフォームより「DLクラウドお試し利用申し込み」と記載の上、ご応募下さい。
ご利用に際しては日立所定の審査がございます。申し込み多数の場合など、ご希望に応えられない場合がございます。
審査後に送付する利用許諾書をご確認の上、ご利用の手引きに従いご利用下さい。

https://www.hitachi.co.jp/products/it/harmonious/cloud/service/deep_learning/index.html

🔍 日立 ディープラーニングクラウド

1. 日立の紹介とディープラーニング・ソリューション
2. ディープラーニングに適した SR24000/DL1
3. DLクラウドサービス
 - ・ 開発学習環境利用サービス（ベアメタル環境）
 - ・ API利用サービス
 - ・ 日立DL技術のご紹介（画像解析、手書き文字認識）

END

日立ディープラーニング・クラウドのご紹介

2019/12

株式会社 日立製作所
IoT・クラウドサービス事業部
HPCソリューションセンタ
主任技師 寺本 康弘

A-1. SR24000/DL1 (V100) の特徴

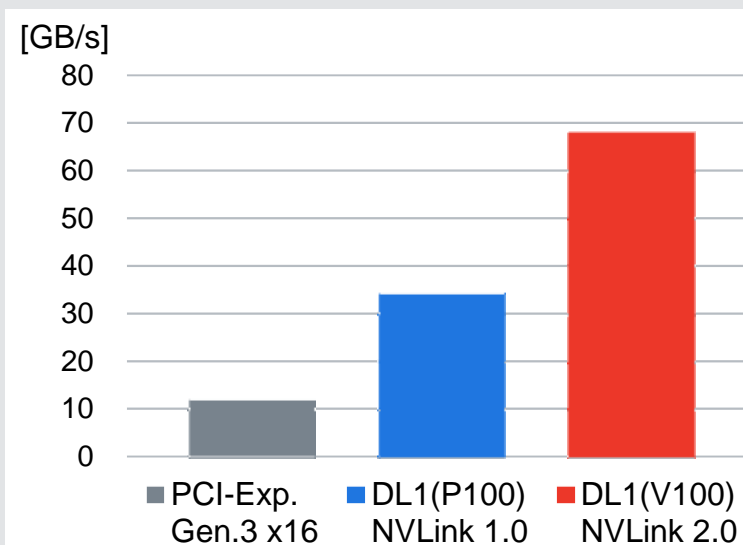
ディープラーニング向けサポートソフトウェアが充実 高性能を発揮する SR24000/DL1 (V100)

IBM PowerAI サポートによりディープラーニング開発のソフトウェアが充実。OSや開発環境はx86とも共通ですので移植性にも優れています。

ホストOS	Ubuntu18.04, Red Hat Enterprise Linux7.5 for IBM Power LE (POWER9)
開発環境	Docker-ce18.06,NVIDIA Docker2.0 NVIDIA Cuda 10.1 Anaconda 2018.12
フレームワーク・ ライブラリなど	IBM PowerAI Release 1.6.0 BVLC Caffe IBM Caffe TensorFlow PyTorch Snap ML OpenBLAS

CPU-GPU間を直結する**NVIDIA NVLink 2.0**はPCI-Exp Gen.3と比較し**5.7倍**の高速転送を実現。大規模データの計算に効果を発揮します。

CPU-GPU間 実効データ転送速度(片方向)



※ PCI-Exp. Gen.3はXeon® E5-2640v4とTesla P100間、DL1(P100)は2016年/10月発表のSR24000/DL1 P100モデル、DL1(V100)は同V100モデルにおける性能

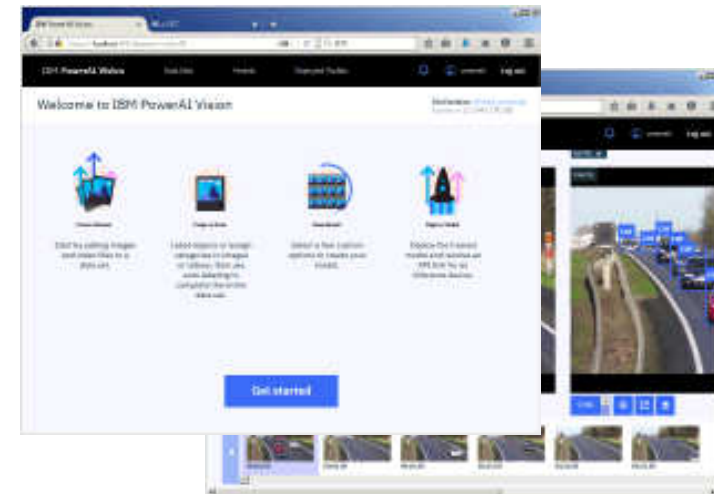
DL画像解析を簡単・迅速にオンプレミス環境に導入し、
業務改善や新事業創出に活用できます

アプライアンスの導入メリット

- 画像分類、物体検知、行動検知などDLを駆使した画像解析を簡易に実行可能
- 学習データを自動生成するオートラベリングやデータ拡張で開発期間を短縮
- 学習データの取り込みからモデル開発、評価までGUI環境で完結します



SR24000/DL1 (V100)

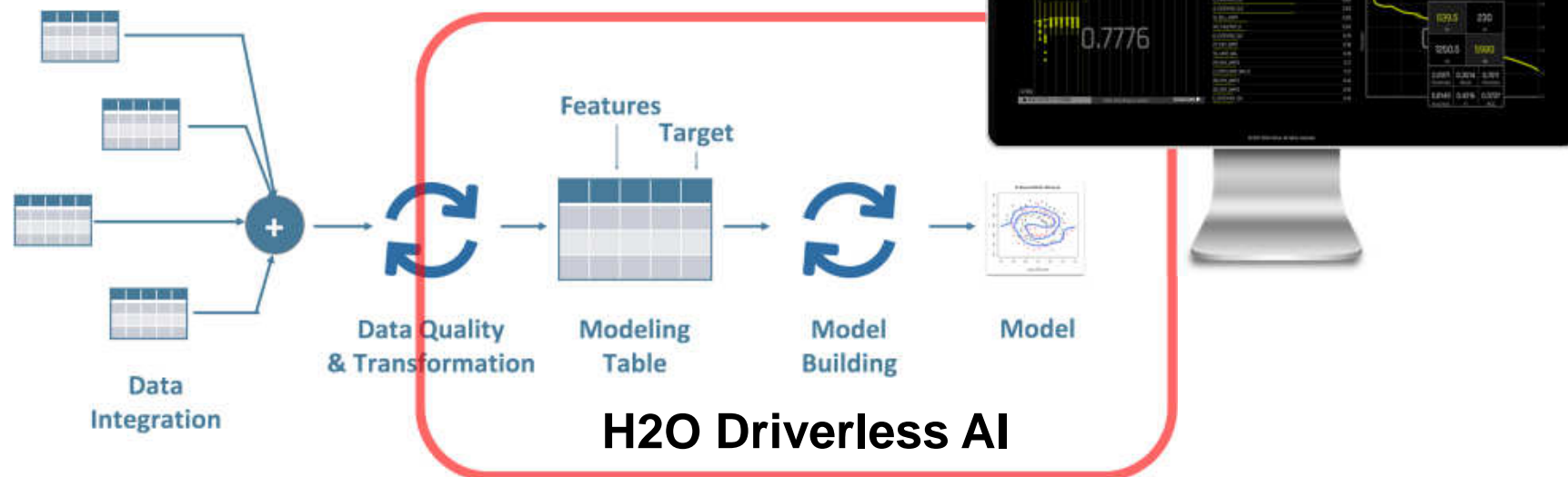


IBM PowerAI Vision

Kaggle※上位のデータサイエンティスト集団のKnowHowを 集約した機械学習の統合環境

H2O Driverless AIの導入メリット

- 複数の機械学習モデルを自動構築し最適モデルを提案
- 優れたデータの可視化と判定ロジックの説明
- GPUに最適化された高速実行



※ Kaggle : 世界中の機械学習・データサイエンスに携わっている専門家コミュニティー
<https://www.h2o.ai/products/h2o-driverless-ai/>

その他、輸出/商標などの引用に関する表示

■ 輸出時の注意

お客さまがディープラーニング・クラウドサービスを利用され輸出行為を行う場合には、外国為替および外国貿易法ならびに米国の輸出管理関連法規などの規制を確認の上、必要な手続きを実施してください。なお、ご不明な場合は、日立担当営業にお問い合わせください。

■ 商標類

- IBM, POWER9は世界の多くの国で登録されたInternational Business Machines Corporationの商標です。
- NVIDIA, Tesla, NVLinkは、米国およびその他の国におけるNVIDIA Corporationの商標または登録商標です。
- Ubuntuは、Canonical Ltd.の商標または登録商標です。
- Windows, Windows Server, Azureは、米国Microsoft Corporationの米国およびその他の国における登録商標または商標です。
- Linuxは、Linus Torvalds氏の日本およびその他の国における登録商標または商標です。
- その他記載の会社名、製品名は、それぞれの会社の商標または登録商標です。

■ 発行

2019年12月

■ 著作権

All Rights Reserved, Copyright ©2019 Hitachi, Ltd.

HITACHI
Inspire the Next