

ディープラーニング、HPC/CAE/HPDA全てで 卓越したパフォーマンスを発揮するプラットフォーム

日本アイ・ビー・エム株式会社

システムズ・ハードウェア事業本部 DCC営業部
部長

笠毛 知徳 (kasamo@jp.ibm.com)

人工知能の歴史

1940s-1970s

人工知能の可能性を予見

- *概念の提唱
1956 ダートマス会議
- *研究開発投資開始
1960s 米国国防総省

1980s

第二次人工知能ブーム

- *エキスパート・システムの隆盛 (1985年 10億ドル超の投資)
- *日本の第5世代コンピュータプロジェクト

1990s

産業応用の有効性を示す

- *物流、データ・マイニング、医療診断等での部分的活用
- *1997 IBM Deep Blue がチェス王者に勝利

2010s

本格的な活用のスタート

- *ディープ・ラーニングの隆盛
- *2011 IBM Watson がクイズ王に勝利
- *2016 Alfa-GOが勝利

現代はカンブリア紀説!?

- *眼を持った三葉虫が出現！
生物進化の大転換点となったカンブリア紀(5億数千万年前)
- *コンピュータにとってのカンブリア紀は今!?!との学説もあり!



眼の誕生 草思社
—カンブリア紀大進化の謎を解く—
アンドリュー・パーカー 著
渡辺政隆、今西康子 訳



IBM Watson™

人工知能の分類

明確で固定されたルールがある問題

例：ボード・ゲーム、パズル



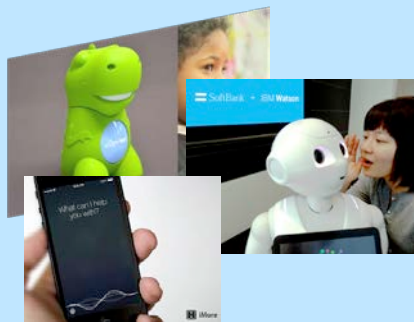
産業ロボット、マシーン自動化

例：自動運転、ロボットアーム等



会話や自由なコミュニケーション

例：コミュニケーションロボット、オープン・ドメイン質疑応答 等



専門家の高度な判断への支援

例：医療診断、株式投資判断、判例判断 等



コグニティブ・コンピューティングの中核機能

多種多様なビッグ・

データの処理

500M photos/day



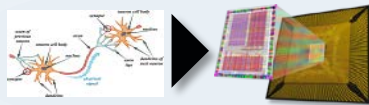
100 hours/minute



不確定で不正確さを含む
データの処理



超省電力、ハイ・パフォーマンス・コンピューティング



モバイル関連技術
(IoT/Wireless,
Security and Privacy)



人間とのコミュニケーション力



コグニティブ時代を推進するテクノロジー

自然言語処理
音声認識
画像認識

機械学習
(監督/監督なし)
サポートベクターマシン

ディープ・ラーニング
ニューラルネットワーク

遺伝的アルゴリズム
グラフ理論

アルゴリズム
(ソフトウェア)

知識
(データ)

計算能力
(インフラストラクチャー)

構造化データ
数値データ

非構造化データ
テキスト、画像
スピーチ、ビデオ

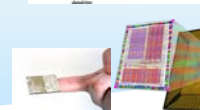
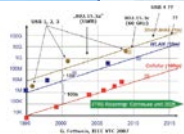
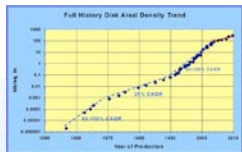
キュレーション済
データ (知識)

マシン生成データ
センサデータ



1960-2010
ストレージ密度 10^8
コンピュータ性能 10^{11}

1995-2015
ワイヤレス帯域幅 10^7



ディープ・ラーニングの“実装の流れ”と“基盤の要件”

ディープ・ラーニング実装の流れ

- やりたいことに
使えそうなデー
タを貯め込む
- 学習に使えるよ
うにデータを整
形・加工
- 大量データによる
学習の実行
- 評価フィードバッ
クと新規データをもとに反復学習
- 学習結果(モデ
ル)の活用
- 活用結果からの
フィードバック



ディープ・ラーニング基盤に必要な要件

- 多種多様な学習対
象データを貯める
ことができる柔軟
性と拡張性
- 障害時の影響範囲
の最小化・堅牢性
- データ整形・加工
を簡単にできる使
い勝手の良さ
- 時間短縮に必須な
インメモリー処理
と広いメモリ帯域
- 多量の技術情報を
活用できる人気学
習ツールへの対応
- 学習時間を最短に
できる計算能力

ディープ・ラーニング基盤を支えるIBMソリューション

IBMは、最高のディープ・ラーニング環境を簡単かつ短期間で整備・構築するためのオープン・テクノロジー・ベースの総合ソリューションをご提供しています。



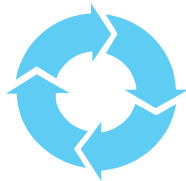
貯める



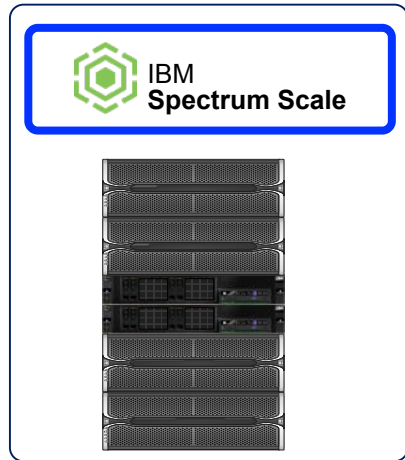
整える



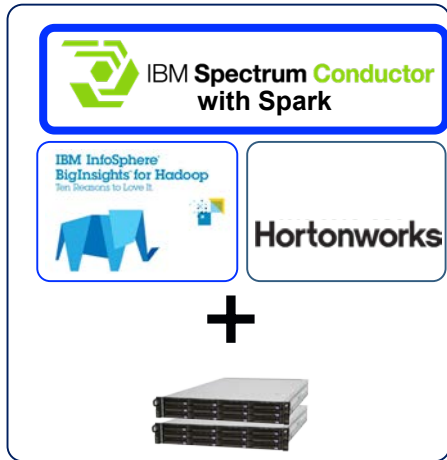
学習する



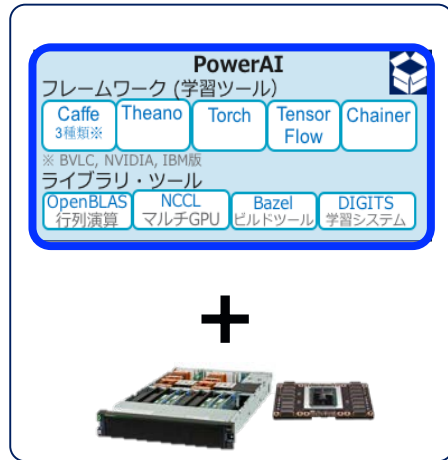
データ
実行



ESS (Elastic
Storage Server)



Conductor with Spark
+ Spark on Power



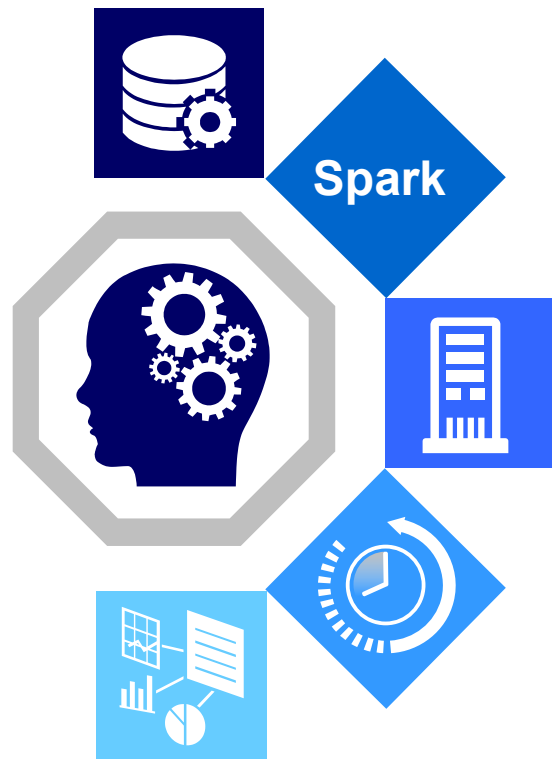
PowerAI
+ S822LC for HPC

分散データ処理基盤 学習基盤



IBMディープ・ラーニング基盤ソリューションの特長

オープン・テクノロジ・ベースのIBMディープ・ラーニング基盤には5つの特長があります。



1

Elastic Storage Server (ESS) で
多種多様な大容量データを安全に保管

2

Spark on Powerと優れたツールで
データの整形・加工を簡単に短時間に実施

3

S822LC for HPC(コードネーム：Minsky)だけが
搭載する最新技術で学習時間の大幅短縮を実現

4

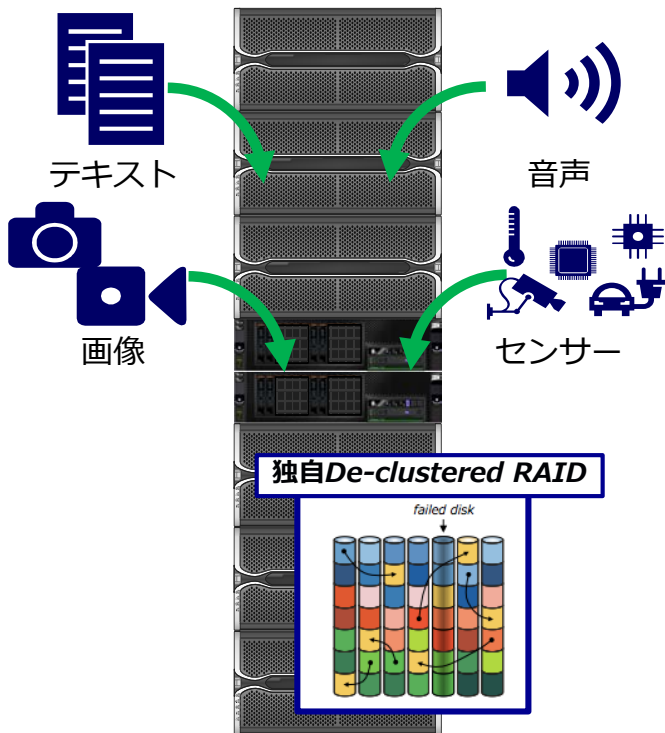
人気学習ツールをMinsky上で最適化して、簡単
導入パッケージにした“Power AI”を無償で提供

5

データ蓄積→整形→学習のディープ・ラーニング
学習の全てに対応できる統合基盤を構築可能

特長① ESS(Elastic Storage Server)で多種多様な大容量データを安全に保管

Elastic Storage Server



1台で多種多様なデータ管理

- 様々な標準プロトコルのアクセスおよび認証をサポート
NFS,SMB,GPFS,HTTP(ObjectStorage),iSCSI,HDFS...
- それぞれのデータ領域を**単一画面**から管理
- 必要な性能、容量を**Building-Block方式**で拡張

ディスク障害の影響を最小化する独自のRAID機能

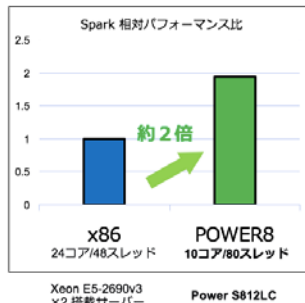
- **最短のRAID再構築時間**
ディスク障害時の数時間以内でリビルドが完了
(他のHDDでは8TB HDDで数日必要)
- **サイレントエラーの検知**
クライアント-サーバー間、サーバー-ディスク間のそれぞれにおいてデータが正しくRead/Writeされるかを確認

特長② Spark on Powerと優れたツールでデータの整形・加工を簡単に短時間に実施

Sparkは
インメモリーで速い

Sparkは
アプリ開発が楽

さらに、POWER上のSparkは約2倍速い



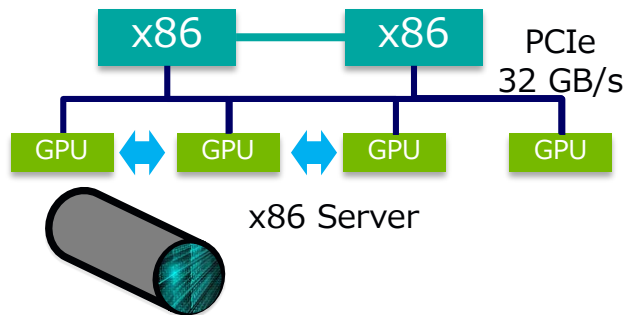
並行処理を最適化

1. Spark taskの並行処理を実現するマルチスレッド機能 (SMT8)
1コアを8スレッドで並列処理 – vs x86で4倍
2. メモリーへの高速アクセスを実現する広いメモリーバンド幅
最大192GB/sのメモリーバンド幅 – vs x86で 約4倍
3. メモリースケジュールを最適化するL4キャッシュ
16MBx4のL4キャッシュによりメモリーへの書き込み回数の削減

特長③ S822LC for HPC(コードネーム:Minsky)だけが 搭載する最新技術で学習時間の大幅短縮を実現

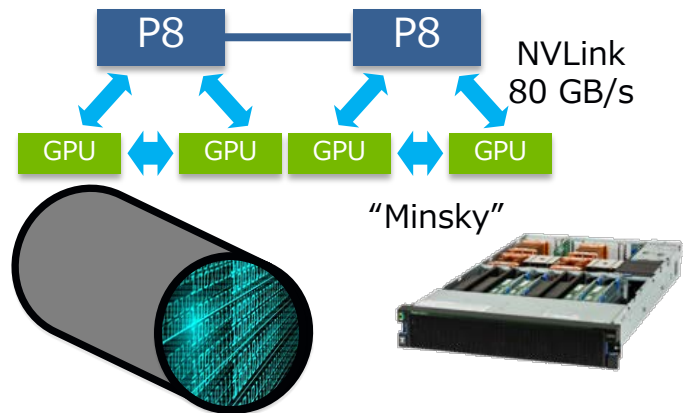
x86基盤とGPU

CPUとGPU間でNVLinkなし
PCIe 接続のためボトルネック



S822LC for HPC ("Minsky")

CPUとGPU間でNVLinkあり
2.5倍の接続速度を実現



特長④ 人気学習ツールをMinsky上で最適化して、簡単導入パッケージにした“Power AI”を無償で提供

PowerAI Software Distribution

Deep Learning Frameworks

Caffe

NVCaffe

IBMCaffe

Torch

TensorFlow

DL4J

Theano

Chainer

Supporting Libraries

OpenBLAS

Bazel

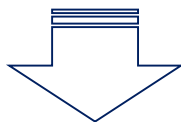
Distributed Frameworks

NCCL

DIGITS

ソフトウェア：
“**PowerAI**”として
人気ツールをMinsky用
に最適化・パッケージ化
して無償でご提供

ハードウェア：
Minsky &



簡単に導入・環境構築できます！



人気のツールをPOWERのUbuntu環境向けにコンパイルし提供するため、オープンソースのビルドの煩わしさがありません

HW+SWで統合的に最適化



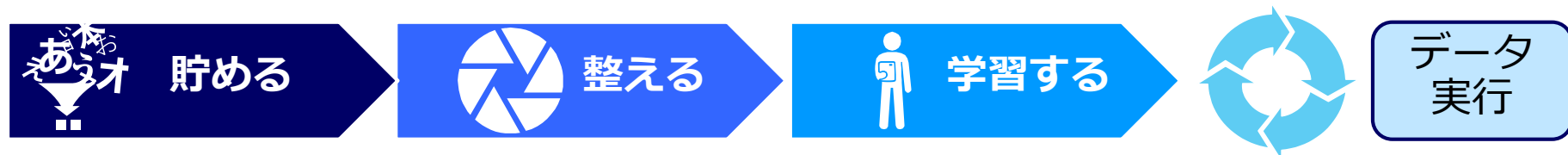
Minsky(S822LC製品群)向けに特別に最適化されたツール群が含まれます
(例:IBM CaffeはTokyo Research製)

日本のChainerもPowerAIの仲間

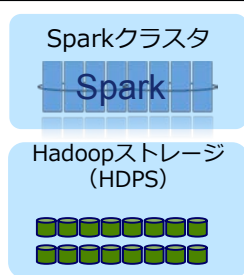


日本発のディープ・ラーニング・フレームワークChainerもメニューに加えて、国際展開拡大にIBMも貢献しています。

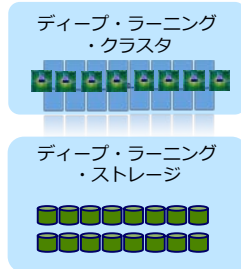
特長⑤ データ蓄積→整形→学習のディープ・ラーニング 学習の全てに対応できる統合基盤を構築可能



これまでの
ディープ・
ラーニング
基盤



前処理
タグ付け
したデータ



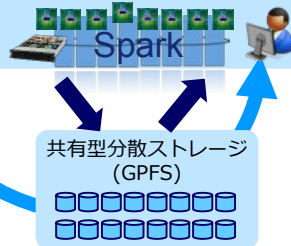
学習済みモ
デル

データの前処理とディープ・
ラーニング学習処理のインフ
ラを別々に準備するのはムダ
ではないか・・・

IBMディープ・
ラーニング基盤

データは外部のストレージ上に
必要時に共通基盤上に
前処理環境もしくはディープ・
ラーニング環境をデプロイ

前処理向けSparkクラスターや、ディープラーニング向け
Sparkクラスターを目的に合わせてオンデマンドに生成



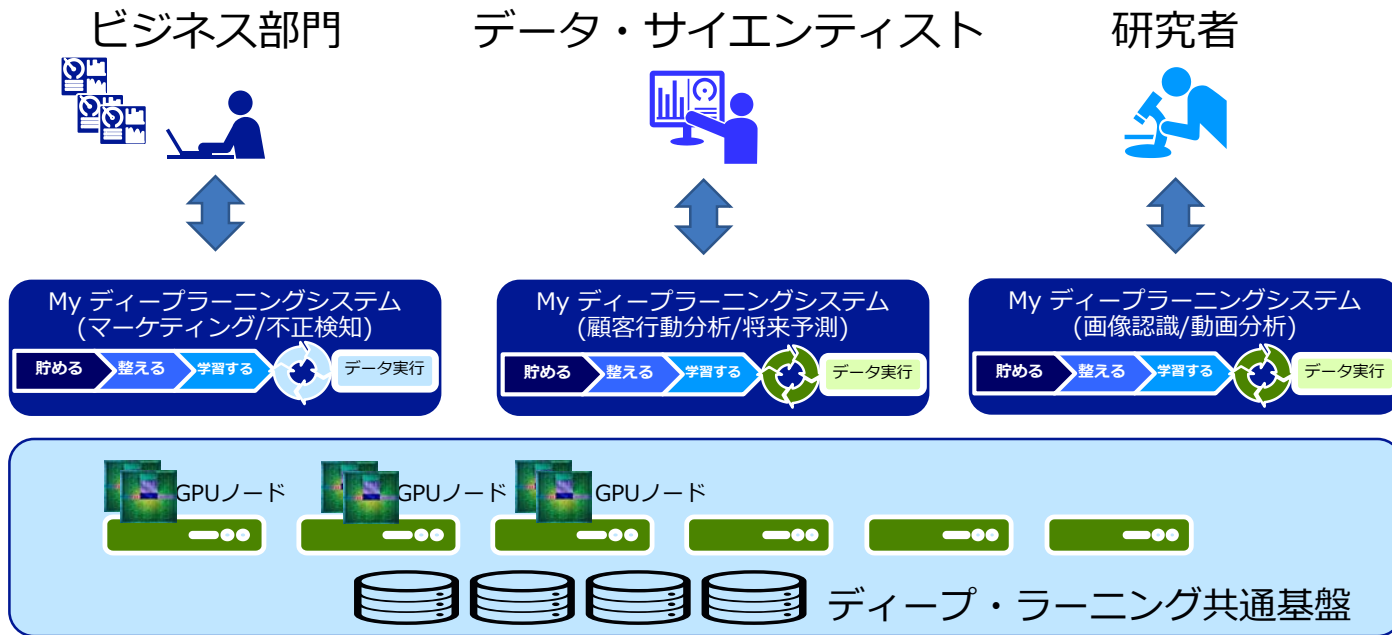
学習済みモ
デル

計算リソースの有効活用、重複投資
の削減ができた！

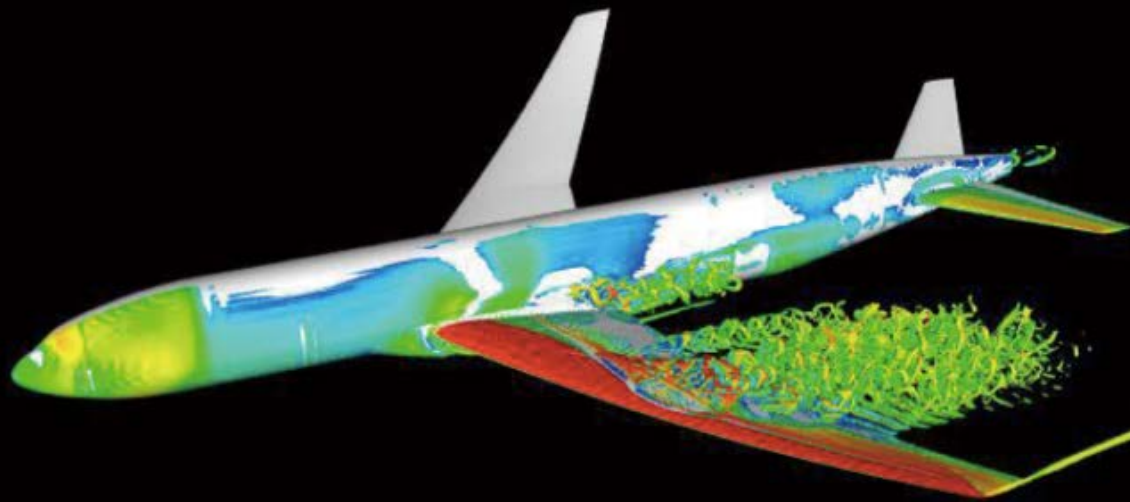
ユーザー毎に“My Spark”環境, “My
Deep Learning”環境を実現

ディープ・ラーニング共通基盤

個々のユーザ各々の要望に対応する“My ディープ・ラーニング”環境を提供でき、効率的かつスピーディなデータ活用による組織の競争力強化が実現できます。



JAXA様、航空機、衛星、ロケットの設計にシミュレーションを活用。幅広い研究での活用を視野に、Minskyを評価



1億個の粒子

空力シミュレーション

NVLink高速通信

IBM POWER & NVIDIA GPU

2桁の高速化

パフォーマンス最適化

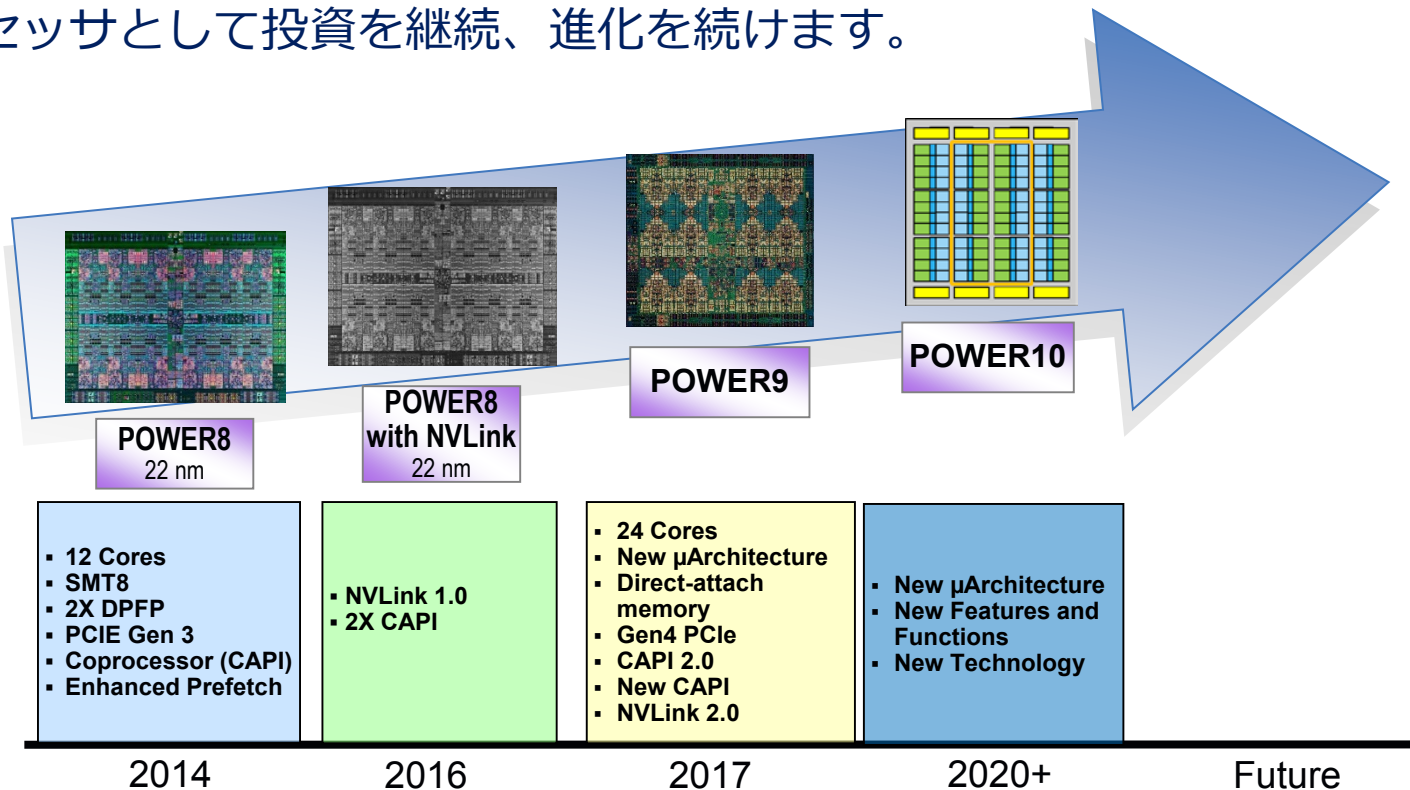
航空機 非巡航状態のBuffer分析 出典：JAXA

*こちらはイメージ図であり、Minskyによる解析結果ではありません

実際の業務に使える計算結果をより短時間で算出することが求められますので、GPUとCPU間的高速通信を実現できるシステムはとても画期的です。

コグニティブ・コンピューティングへの継続投資

POWER8からIP(仕様)を公開しオープン化したPOWERプロセッサのロードマップは、二世代先まで公開済。今後もコグニティブ・コンピューティングの根幹を支えるプロセッサとして投資を継続、進化を続けます。



深刻化するデータ処理量の増加に向けた取組み

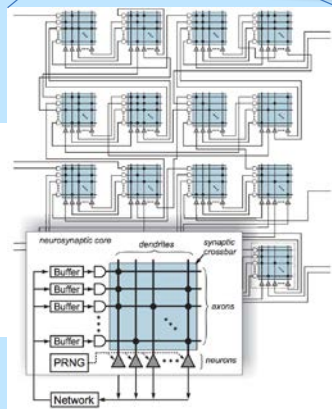
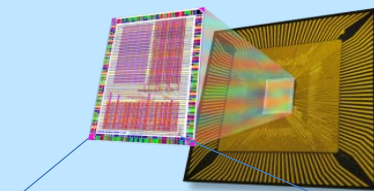
超省電力・超高性能の実現を目指した新たなテクノロジー開発に挑戦しています！

Neuromorphic Computing

人の脳にヒントを得て、開発された
「脳チップ」で超低消費電力を実現



<http://www.greencarcongress.com/2014/08/20140808-truenorth.html>



1million neurons,
256million synapses,
field-programmable on
5.4B transistor chip,
63mW

量子コンピュータ(Cloud提供開始)

IBM Quantum Experienceとして
Cloud提供開始2ヶ月でのユーザー登
録数は3万超。量子コンピュータは既
に手の届くところにきている。



30,014 registered users, in
146 countries



180,756 executions

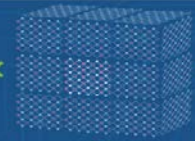
10+ professors committing to
using our system in
coursework

Building Larger
Quantum Systems:
Linear vs Square Array



Others have arranged qubits in a line to measure
only one type of quantum error at a time.
This is not enough.

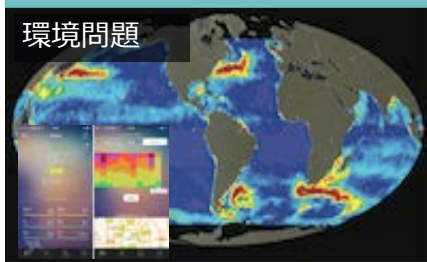
[VS]



IBM figured out a square design for a
quantum circuit that, with more qubits,
can scale to a working quantum system.

人工知能を活用した豊かな日本社会の創造に向けて

日本の発展をIBMディープ・ラーニング基盤ソリューションで支えています。



ワークショップ、セッション、および資料は、IBMまたはセッション発表者によって準備され、それぞれ独自の見解を反映したものです。それらは情報提供の目的のみで提供されており、いかなる参加者に対しても法律的またはその他の指導や助言を意図したものではなく、またそのような結果を生むものでもありません。本講演資料に含まれている情報については、完全性と正確性を期するよう努力しましたが、「現状のまま」提供され、明示または暗示にかかわらずいかなる保証も伴わないものとします。本講演資料またはその他の資料の使用によって、あるいはその他の関連によって、いかなる損害が生じた場合も、IBMは責任を負わないものとします。本講演資料に含まれている内容は、IBMまたはそのサプライヤーやライセンス交付者からいかなる保証または表明を引きだすことを意図したものでも、IBMソフトウェアの使用を規定する適用ライセンス契約の条項を変更することを意図したものでもなく、またそのような結果を生むものでもありません。

本講演資料でIBM製品、プログラム、またはサービスに言及していても、IBMが営業活動を行っているすべての国でそれらが使用可能であることを暗示するものではありません。本講演資料で言及している製品リリース日付や製品機能は、市場機会またはその他の要因に基づいてIBM独自の決定権をもっていつでも変更できるものとし、いかなる方法においても将来の製品または機能が使用可能になると確約することを意図したものではありません。本講演資料に含まれている内容は、参加者が開始する活動によって特定の販売、売上高の向上、またはその他の結果が生じると述べる、または暗示することを意図したものでも、またそのような結果を生むものでもありません。パフォーマンスは、管理された環境において標準的なIBMベンチマークを使用した測定と予測に基づいています。ユーザーが経験する実際のスループットやパフォーマンスは、ユーザーのジョブ・ストリームにおけるマルチプログラミングの量、入出力構成、ストレージ構成、および処理されるワークロードなどの考慮事項を含む、数多くの要因に応じて変化します。したがって、個々のユーザーがここで述べられているものと同様の結果を得られると確約するものではありません。

記述されているすべてのお客様事例は、それらのお客様がどのようにIBM製品を使用したか、またそれらのお客様が達成した結果の実例として示されたものです。実際の環境コストおよびパフォーマンス特性は、お客様ごとに異なる場合があります。

IBM、IBM ロゴ、ibm.com、BigInsights、IBM Spectrum Conductor、IBM Spectrum Scale、InfoSphere、POWERおよびPOWER8 は、世界の多くの国で登録されたInternational Business Machines Corporationの商標です。他の製品名およびサービス名等は、それぞれIBMまたは各社の商標である場合があります。現時点での IBM の商標リストについては、www.ibm.com/legal/copytrade.shtmlをご覧ください。