

文部科学省 科学技術人材育成費補助事業 データ関連人材育成プログラム
第5回データアントレプレナーカンファレンス
- データサイエンスと量子コンピュータ -

データコンピューティング論

Engineering of Data Computing

2019年6月19日

清洲 正勝

Masakatsu KIYOSU

目次

1. はじめに
2. 歴史
3. 概念
4. データ駆動
5. データ
6. データ量
7. データ表記法
8. データパイプライン
9. ビッグデータ
10. ビッグデータ型式
11. Neumannアーキテクチャ
12. 人工知能プロセッサ
13. マイクロコンピュータ
14. アルゴリズム
15. プログラミングライブラリ
16. ベンダーフレームワーク
17. 量子コンピュータ / 量子プロセッサ
18. 量子コンピュータソフトウェア開発キット
19. まとめ

はじめに

本講演は，データコンピューティングの

背景 を概観する

観点 を多角的に提示する

技術 を一覧する

※ 本資料の詳細は講演にて説明しています．

歴史

科学革命^[1] Paradigm Shift

第一の科学	第二の科学	第三の科学	第四の科学
経験的	理論的	計算的	データ主導
1000年前 観察及び観測による処理	数百年前 分析による理論化、法則化	数十年前 計算機の演算処理	現在 データ探索、集中処理

産業革命^[2] Industrial Revolution

第一次産業革命	第二次産業革命	第三次産業革命	第四次産業革命
機械化	電力化	自動化	最適化
18～19世紀 水力、蒸気機関、紡績機	20世紀初頭 電力、石油、大量生産	1970年代初頭 計算機、情報通信技術	2010年代 人工知能、ビッグデータ、ロボット、IoT

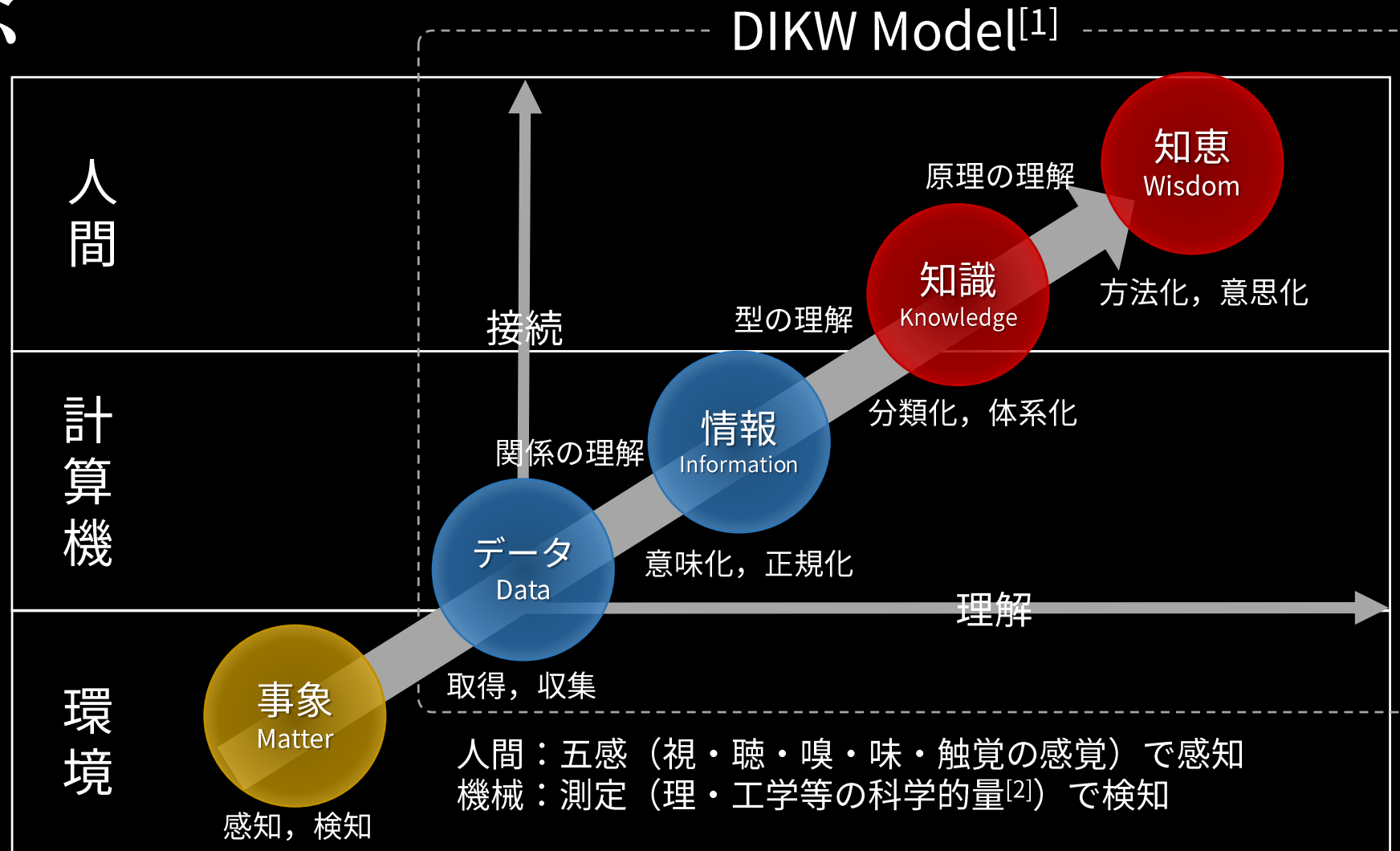
図：歴史

1. Tony Hey, "The Fourth Paradigm: Data-intensive Scientific Discovery", James Nicholas Gray, 2009, <http://www.systems-thinking.org/dikw/dikw.htm>.

2. Henning Kagermann et al., "Recommendations for implementing the strategic initiative INDUSTRIE 4.0", 2013, <https://www.rikanenpyo.jp/>.

概念

データ工学 データマイニング



人工知能
知識工学・機械学習

数理最適化, 統計学 応用統計学・統計科学

1. Gene Bellinger, Durval Castro, Anthony Mills, "Data, Information, Knowledge, and Wisdom", 2004, [\[URL\]](#) .

2. 自然科学研究機構 国立天文台, 丸善出版株式会社, 「理科年表」, 1925~, [\[URL\]](#) .

データ駆動

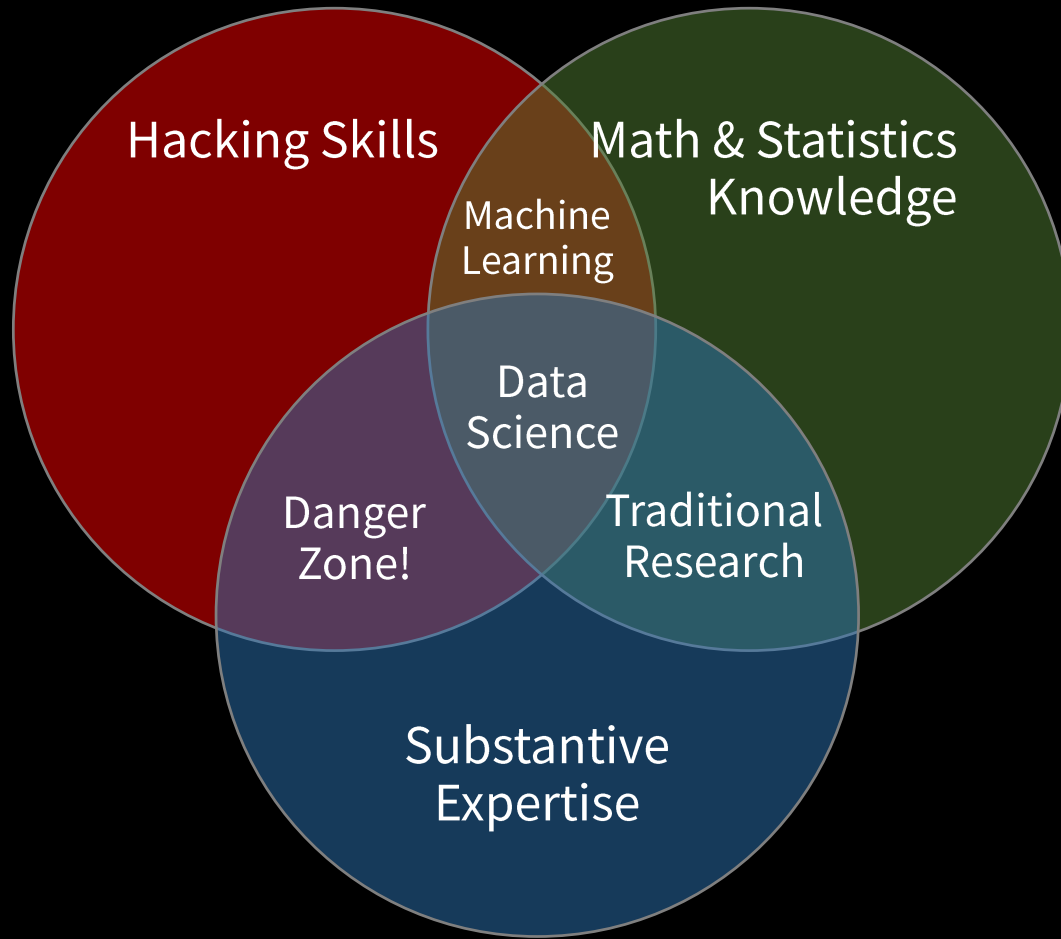


Fig. The Data Science Venn Diagram^[1]

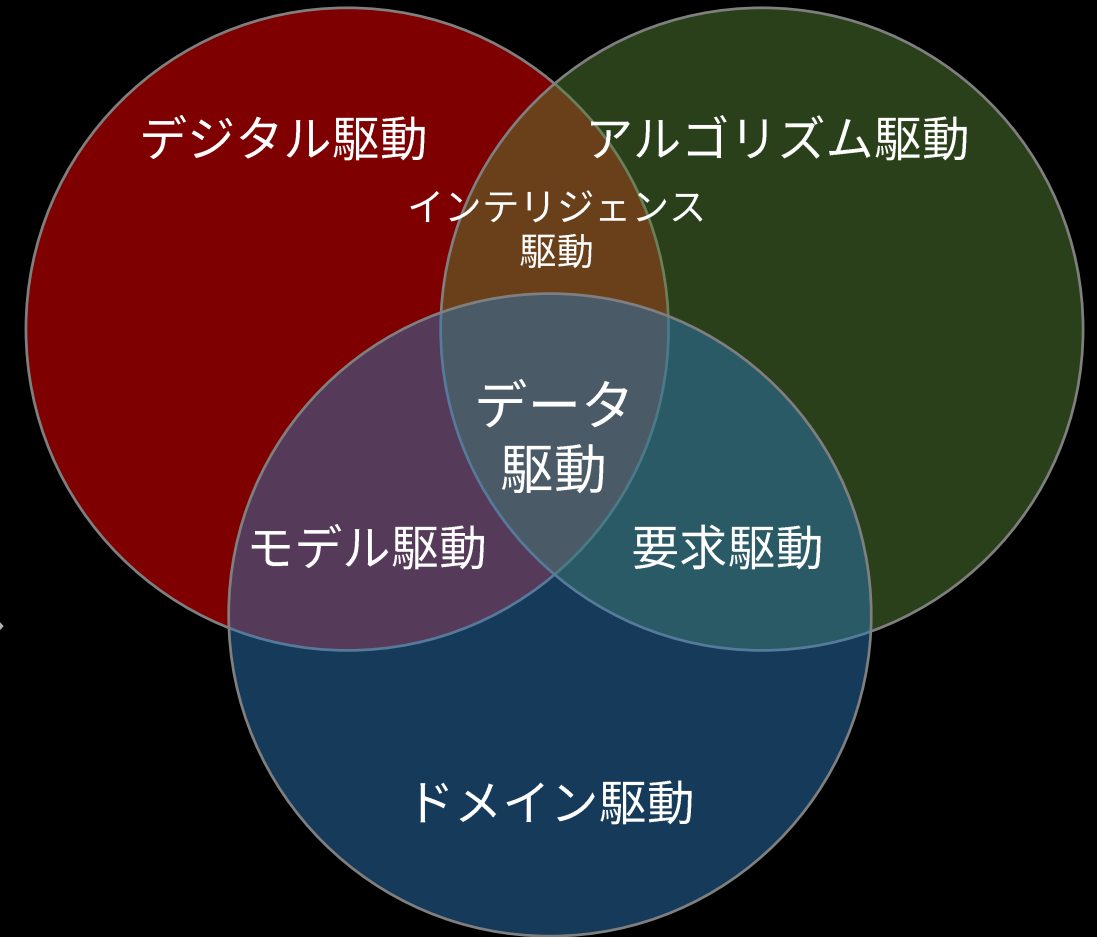


図. 駆動アプローチ^[2]

- 駆動アプローチの図は，左図に駆動（xD）アプローチを近しい部分にマッピングしたもので，厳密に写像ではない。

1. Drew Conway, "THE DATA SCIENCE VENN DIAGRAM", 2010, <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>.

2. 清洲 正勝, 「データコンピューティング論」, 2019.

データ

表：データ^[1]

OSI参照モデル	TCP/IP4層モデル	データ	
		TCP	UDP
アプリケーション	アプリケーション	ストリーム	メッセージ
プレゼンテーション			
セッション			
トランスポート	トランスポート	セグメント	パケット
ネットワーク	インターネット	データグラム	データグラム
データリンク	ネットワークアクセス	フレーム	フレーム
物理			

- TCP/IPのモデルは層数や各層の名称の違いを含めて幾つか存在している。
- プログラミングのデータ構造でもリスト型の「データフレーム」がある。
- ネットワーク管理は、SDN（Software Defined Network）のOpenFlowやNETCONFなどの標準でコントローラを使って制御し、NFV（Network Functions Virtualization）のOPNFVやONAPなどの標準でハードウェアから仮想環境上のソフトウェアに置き換えることができる。

1. Craig Hunt, 「TCP/IPネットワーク管理 第3版」, p.14, 2003. を参考に作成

データ量

情報理論において、発生確率が1/2の事象が持つ情報量1Sh (shanon) は、データ量1Cbit (classical bit | bit) として扱うことができる.

$$1 \text{ Sh} \leftrightarrow 1 \text{ Cbit}$$

量子情報理論^[1]において、Booleanの状態0と1をブラ・ケット記法のうちケットベクトル“ $|\rangle$ ”を使い $|0\rangle$, $|1\rangle$ と表したとき、量子ビット (quantum bit: qubit | Qbit) は、

$$|0\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |1\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

となり、1量子の量子重ね合せ状態 (quantum superposition) の量子ビット $|\psi\rangle$ は、

$$|\psi\rangle \rightarrow \psi_0|0\rangle + \psi_1|1\rangle$$

で表し、2量子間で一方の状態が決まるともう一方の状態が確定する量子もつれ (quantum entanglement) は、

$$\sqrt{\frac{1}{2}}(|00\rangle + |11\rangle)$$

で表される.

- 上記の ψ は波動関数で、ブラケット記法は、理論物理学者のPaul Adrien Maurice Diracが定義し、量子ビットは、理論物理学者のBenjamin Schumacherが定義した.

1. Charles H. Bennett, Peter W. Shor, “Quantum Information Theory”, IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 44, NO. 6, OCTOBER 1998. より抜粋
2019/6/19

データ表記法

表形式

CSV: Comma-Separated Values
RFC 4180

例 country.csv

```
name,capital,currency,code,district,gdprank  
"Japan","Tokyo","JPY",81,47,3
```

意味論

XML: Extensible Markup Language
RFC 3023

例 country.xml

```
<?xml version="1.0" encoding="UTF-8"?>  
<country>  
  <name>Japan</name>  
  <capital>Tokyo</capital>  
  <currency>JPY</currency>  
  <status>  
    <code zone="8">81</code>  
    <district title="prefecture">47</district>  
    <gdprank year="2018">3</gdprank>  
  </status>  
</country>
```

交換性

JSON: JavaScript Object Notation
RFC 4627

例 country.json

```
{  
  "name": "Japan",  
  "capital": "Tokyo",  
  "currency": "JPY",  
  "status": {  
    "code": 81,  
    "district": 47,  
    "gdprank": 3  
  }  
}
```

- データ技術にサポートされる様々なデータ表記法（単に記法とも言う）があり、格納法も技術によって様々。データ表記法はファイルフォーマットとも関係している。

1. RFC 4180, “Common Format and MIME Type for Comma-Separated Values (CSV) Files”, IETF, 2005, <https://www.ietf.org/rfc/rfc4180.txt>
2. RFC 3023, “XML Media Types”, IETF, 2001, <https://www.ietf.org/rfc/rfc3023.txt>
3. RFC 4627, “The application/json Media Type for JavaScript Object Notation (JSON)”, IETF, 2006, <https://www.ietf.org/rfc/rfc4627.txt>

ビッグデータ

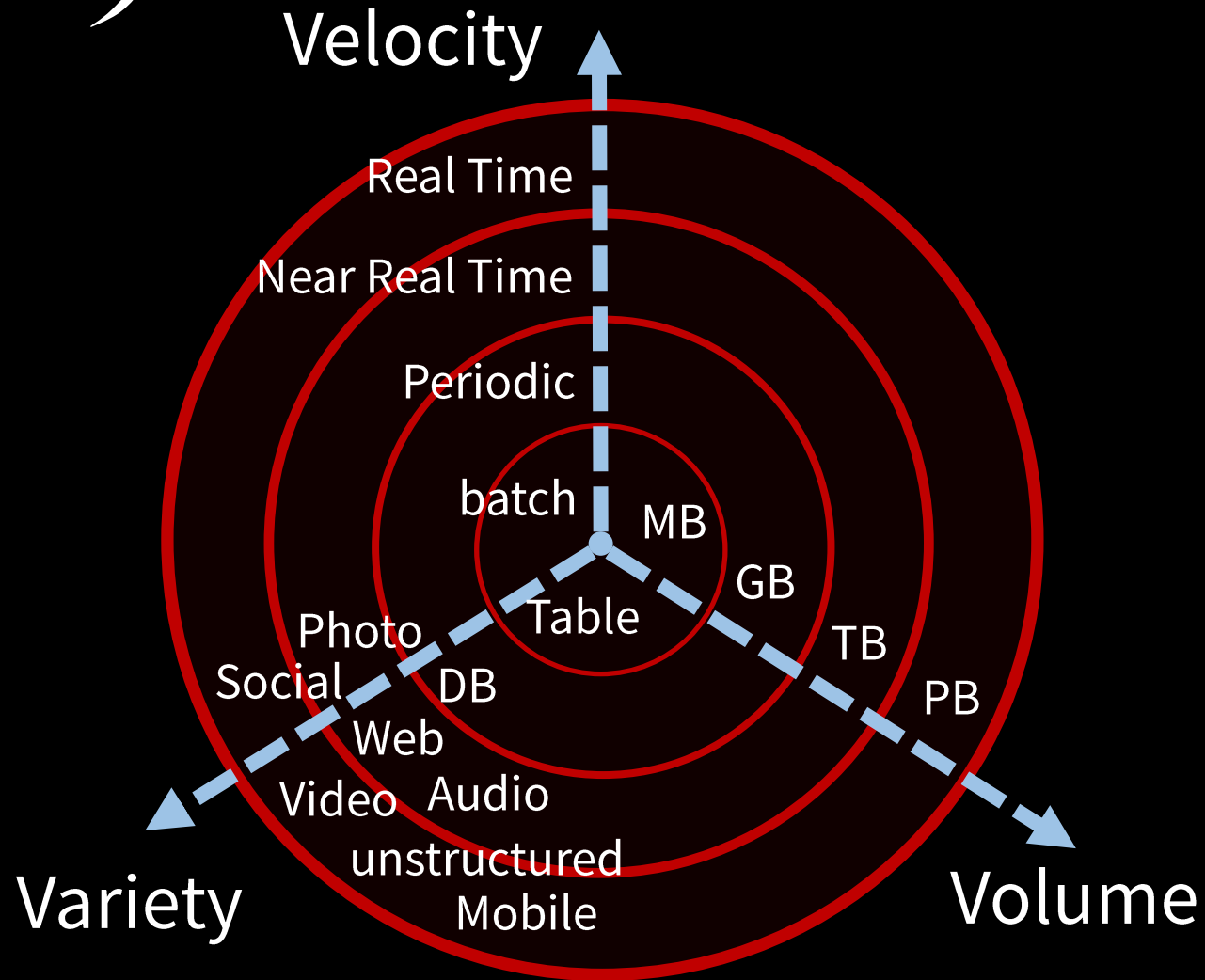
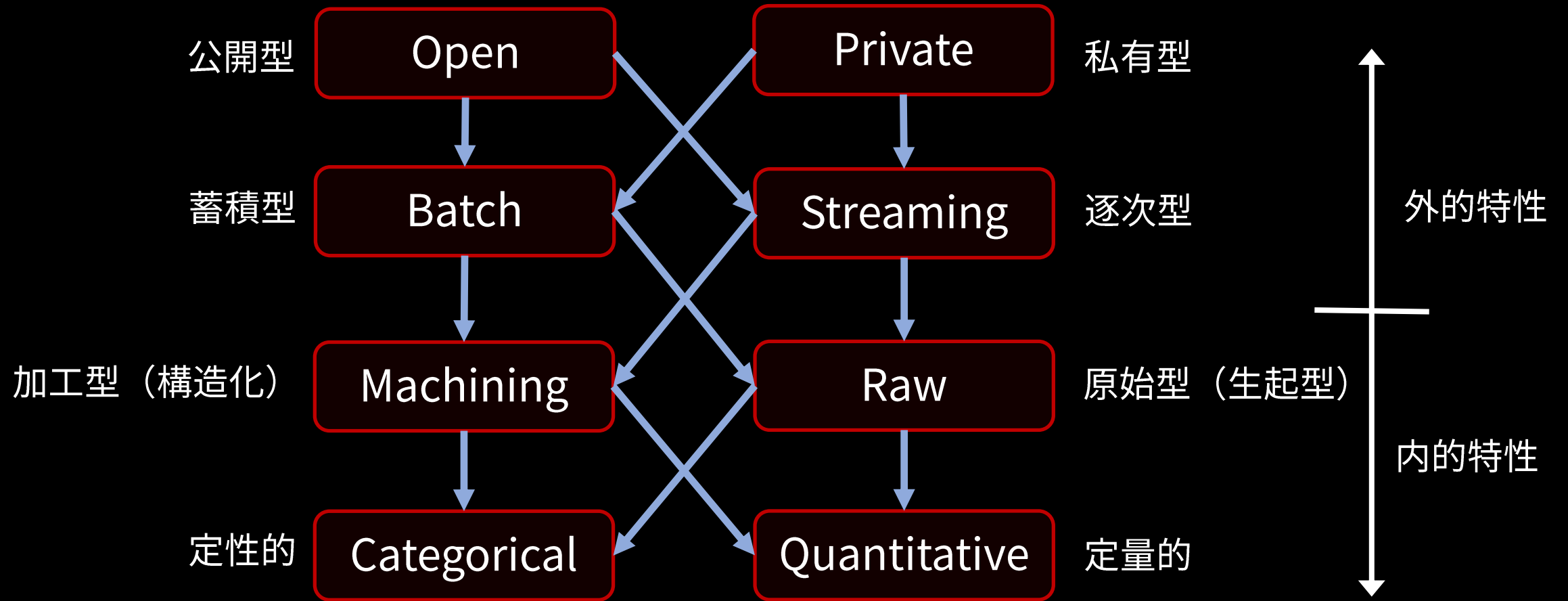


Fig. The 3Vs that define Big Data^[2]

1. Douglas Laney, "Application Delivery Strategies File: 949", META Group, 6 February 2001.

2. Diya Soubra, "The 3Vs that define Big Data", Data Science Central, 2012, <https://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data>.

ビッグデータ型式



図：ビッグデータ型式図^[1]

1. 清洲正勝，沼尾雅之，「データサイエンスのプロセスと業務評価モデル」，電気通信大学，情報処理学会，2015，[\[URL\]](#)。

データパイプライン

収集

蓄積

処理

保存

分析

可視化



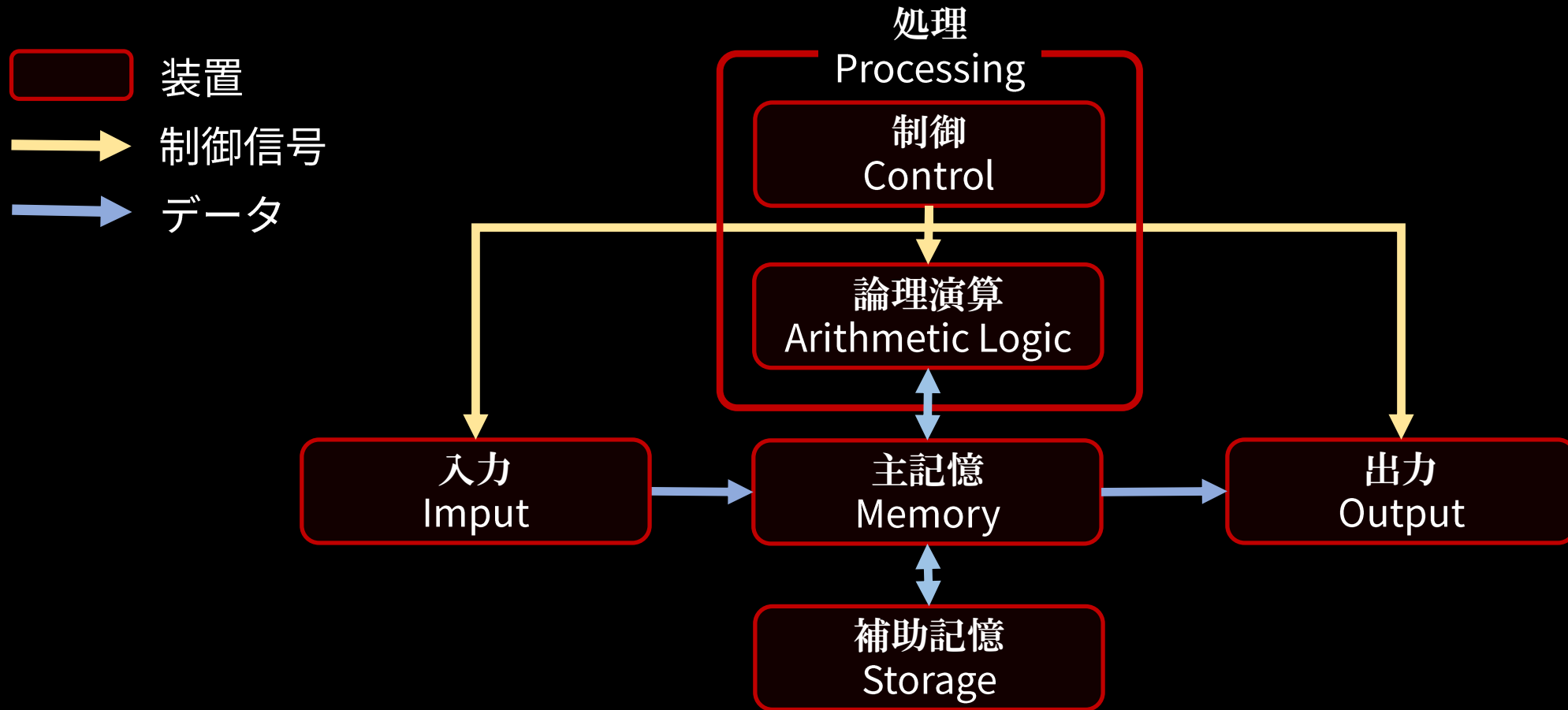
図：データパイプライン概要^[1]

データに着目したデータの流をサーバの構成とともに表現している。構成はトレンドがあり用途や目的で様々。

サーバには構成管理，仮想化，分散処理フレームワーク，コンテナオーケストレーション，監視などの機能を持たせることができる。

1. 清洲 正勝，「データコンピューティング論」，2019。

Neumannアーキテクチャ



図：Neumannアーキテクチャ

- 計算機モデルのTuringマシンを実現するハードウェアの構成概念であり、実際の設計は複雑。
- メーカーはそれぞれの設計思想で性能を向上させている。膨大なデータ処理には多大な時間と巨大な消費電力、発熱が伴う。

人工知能プロセッサ

Neuromorphic chip

2008年, DARPAはIBMとニューロモフィックチップ (Neuromorphic chip) SyNAPSE Project^[1]を開始. 2014年8月SyNAPSEの商用チップ TrueNorthを発表.

FPGA

2014年8月, MicrosoftのProject Catapult^[2]は, FPGA (Field-Programmable Gate Array) によって, データセンターにあるサーチエンジン Bingのサーバを高速化したと発表.

GPU

2015年11月, Nvidiaは, Kepler, Maxwellアーキテクチャを採用したGPU (Graphic Processing Unit) Tesla^[3]を提供. 2016年4月, 機械学習に適したPascalアーキテクチャを発表. 2017年5月VoltaアーキテクチャベースのV100を発表. 2018年8月, Turingアーキテクチャを発表.

TPU

2016年5月, Googleは, 深層学習ASIC (Application Specific Integrated Circuit) プロセッサ TPU (Tensor Processing Unit) ^[4]を発表. 2017年5月, 第2世代 Cloud TPU を発表.

NNP

2016年8月, Intelは, 機械学習ASICプロセッサのNervanaを買収. Nervana Engineの技術がCPU Xeon Phiコプロセッサ (Co-processor) へ実装. 2017年10月, ニューラルネットワークプロセッサ Nervana NNPを発表.^[5]

1. IBM, "Introducing a Brain-inspired Computer", <http://www.research.ibm.com/articles/brain-chip.shtml>.

2. Microsoft, "Project Catapult", <https://www.microsoft.com/en-us/research/project/project-catapult/>.

3. Nvidia, 「ディープラーニングと人工知能」2016, <https://www.nvidia.com/ja-jp/deep-learning-ai/>.

4. Google Cloud Platform Blog, "Google supercharges machine learning tasks with TPU custom chip", May 2016, <https://cloud.google.com/blog/products/gcp/google-supercharges-machine-learning-tasks-with-custom-chip>.

5. Intel, "Nervana Neural Network Processors", 2017, <https://www.intel.ai/nervana-nnp/>.

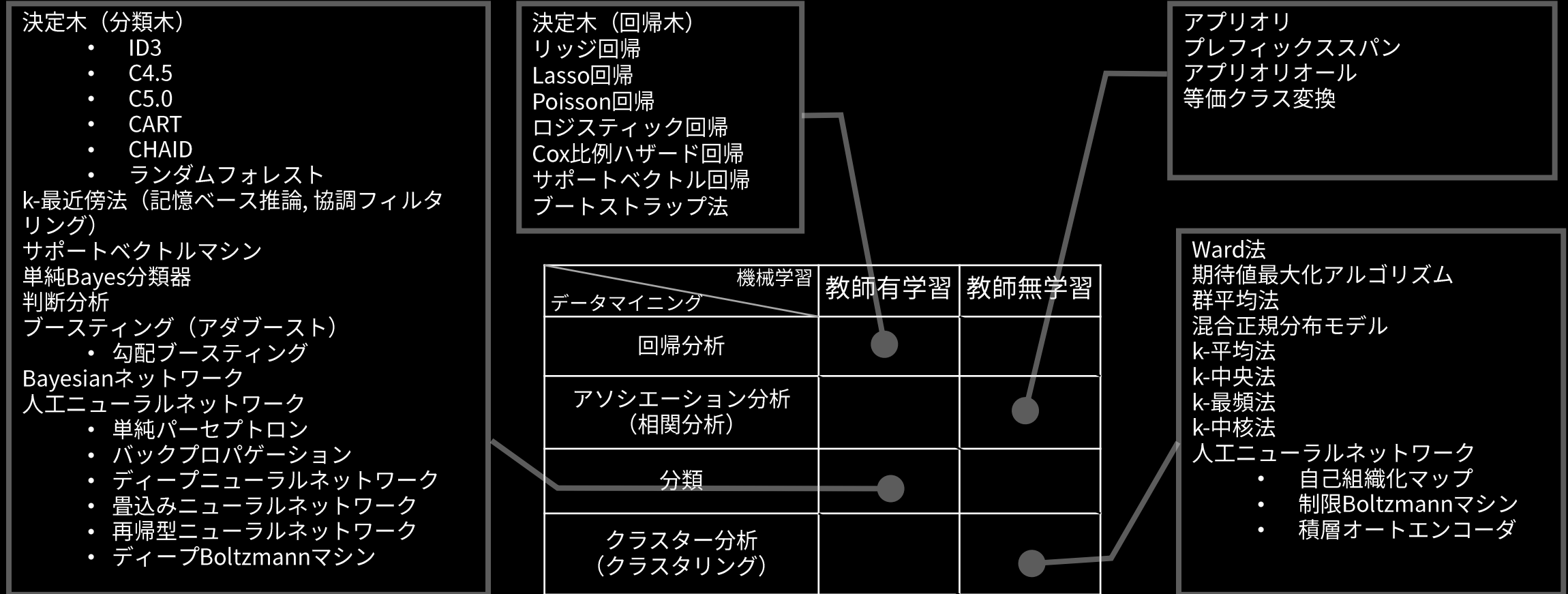
マイクロコンピュータ

表：主な開発用マイクロコンピュータ

組織	製品	プロセッサ	国	発売
Arm Holdings plc	mbed ^[1]	Cortex (ARM)	イギリス	2005
Arduino LLC	Arduino ^[2]	Cortex (ARM) , Quark (Intel)	イタリア	2005
Raspberry Pi Foundation	Raspberry Pi ^[3]	Cortex (ARM)	イギリス	2013
Intel Corporation	Intel Edison ^[4]	Atom (Intel)	アメリカ	2014
NVIDIA Corporation	Jetson ^[5]	Tegra (ARM)	アメリカ	2014
Sony Corpoaration	Spresense ^[6]	CXD5602 (Sony)	日本	2018
Google LLC	Coral ^[7]	TPU (Google)	アメリカ	2019

1. Arm Holdings plc, "mbed", <https://www.mbed.com/en/>
2. Arduino LLC, "Arduino", <https://www.arduino.cc/>
3. Raspberry Pi Foundation, "Teach, Learn, and Make with Raspberry Pi", <https://www.raspberrypi.org/>
4. Intel Corporation, "Discontinued Maker & Innovator Products", <https://software.intel.com/en-us/iot/hardware/discontinued>
5. Nvidia Corporation, 「組み込みシステムの開発者キットとモジュール」, <https://www.nvidia.com/ja-jp/autonomous-machines/embedded-systems/>
6. Sony Corporation, "Spresense", <https://developer.sony.com/ja/develop/spresense/>
7. Google LLC, "Coral Beta", <https://coral.withgoogle.com/>

アルゴリズム



図：主なアルゴリズム^[1]

- 他にも強化学習や異常検出, 時系列分析などのアルゴリズムがある.

プログラミングライブラリ

表：主なPythonライブラリ

カテゴリ	フレームワーク	Pythonライブラリ
統計		Pandas (The PyData Development Team) Statsmodels (the statsmodels development team)
機械学習		Caffe (Berkley Vision and Learning Center) MXNnet (Distributed (Deep) Machine Learning Community) PyBrain (Technische Universität München) Pylearn (LISA Lab.) PyML (Colorado State University, Asa Ben-Hur) Scikit-learn (David Cournapeau)
数値計算		NumPy (Travis Oliphant) SciPy (SciPy developers) SymPy (SymPy Development Team) Theano (University of Montreal)
自然言語処理		gensim (Radim Řehůřek) Natural Language Toolkit (NLTK Project)
描画		Bokeh (Continuum Analytics) matplotlib (the matplotlib development team) seaborn (Michael Waskom)

※ GPL, LGPL, Apache, BSD, MITなどのライセンス規約を確認。詳細は開発元の公式サイトを参照。
※ 上位ライブラリ (Wrapper) も多く開発されている。

ベンダーフレームワーク

表：主なベンダーフレームワーク

Company	Framework	Library	Cloud Platform
ABEJA		—	ABEJA Platform (2016/12)
Adobe		—	Adobe Experience Cloud (2017/3)
Amazon		—	Amazon Machine Learning (2015/4)
Airbnb		Airflow, Aerosolve (2015/6 ※2)	—
Baidu		PaddlePaddle (2016/9 ※2)	—
Facebook		Torch (2015/1 ※2)	—
Google		DistBelief (2011) , TensorFlow (2015/11 ※2)	Cloud Machine Learning (2016/3)
GRID		ReNom (2016/5)	—
HP		—	Heaven OnDemand (2016/3)
IBM		PowerAI (2017/8)	Watson Developer Cloud (2014/2, 2016/2 日本語化)
Intel		BigDL (2017/1 ※2)	—
Microsoft		Computational Network Toolkit (2016/1 ※2) Cognitive Toolkit (2017/7 ※2)	Azure Machine Learning (2014/6 β→ 2015/2)
Preferred Networks		Chainer (2015/6 ※2)	—
Sony		Neural Network Libraries (2017/6 ※2)	Neural Network Console (2017/8)
Yahoo! JAPAN		AnnexML (2017/11 ※2)	—

※ 詳細は各社の公式サイトを参照．多様な言語を利用する．赤文字は国内企業．

※ ※2 オープンソース化

量子コンピュータ / 量子プロセッサ

GQPU

2016年5月, IBMはゲート方式の5量子ビット量子コンピュータ Q System を発表. 2017年11月には20量子ビットの量子コンピュータを発表、50量子ビットの準備を発表. 2019年1月に, 商用の Q System One^[1] を発表.

AQPU

2011年5月, D-Wave Systems^[2]は, アニーリング方式の128量子ビット D-wave ONE を世界で初めて発表. 2017年11月に2048量子ビットに向上させた D-Wave 2000Q を発表.

GQPU

2018年6月, Rigetti Computingは, ゲート方式の8量子ビット Acorn を発表. 2018年11月に16量子ビットの Aspen^[3] を発表.

GQPU

2018年3月, Google は, ゲート方式の72量子ビット量子プロセッサ Bristlecone^[4] を発表.

GQPU

2018年1月, Intelは, ゲート方式の49量子ビット超伝導プロセッサ Tangle Lake^[5] を発表.

1. IBM, "IBM Makes Quantum Computing Available on IBM Cloud to Accelerate Innovation", 2019, <https://www.research.ibm.com/ibm-q/system-one/>.
2. D-Wave Systems, "The D-Wave 2000Q System", 2017, <https://www.dwavesys.com/d-wave-two-system>
3. Rigetti Computing, "QPU Specifications", 2018, <https://www.rigetti.com/qpu>.
4. Google, "A Preview of Bristlecone, Google's New Quantum Processor", 2018, <https://ai.googleblog.com/2018/03/a-preview-of-bristlecone-googles-new.html>.
5. Intel, 「量子コンピューティング」, 2018, <https://www.intel.co.jp/content/www/jp/ja/research/quantum-computing.html>.

量子コンピュータ ソフトウェア開発キット

表：主な企業の量子コンピュータSDK

	SDK	量子アセンブラ	プログラミング言語
D-Wave	qbsolv	QMI	C, C++, Python, MATLAB
Google	Cirq	Cirq	Python
IBM	QISkit	OpenQASM	Python
MDR	Blueqat	Blueqat	Python
Microsoft	Quantum Development Kit	Q#	Q#, C#, Python
Rigetti	Forest	pyQuil	Python

- 量子計算をシミュレーションするもの、量子コンピュータのあるクラウドで計算させるものなど様々。
- 最新の開発状況が刻々と変わっているので、現在の実装は上記とは異なっている可能性がある。

※ 詳細は各社の公式サイトを参照。

まとめ

『データコンピューティング論』

データを活用するための情報通信技術が急速に進歩している。高度な科学技術が進歩している現代にプログラミングスキルはリテラシーとして重要視されている。人類の難題解決のためには、計算性能（Computing Performance）の圧倒的な向上が求められており、古典コンピュータから量子コンピュータなどの次世代コンピュータへの技術革新が必要。