PCクラスタワークショップ in 大阪2023「ビッグデータとHPC」

Big Data Analytics on IA

～ビッグデータ技術に関するインテルの取り組みと事例のご紹介～

インテル株式会社
AIセンター・オブ・エクセレンス
AIテクニカル・ソリューション・スペシャリスト
大内山　浩

注意事項および免責条項

- 性能は、使用状況、構成、その他の要因によって異なります。詳細については、http://www.Intel.com/PerformanceIndex/ (英語) を参照してください。性能の測定結果は、システム構成に記載された日付時点のテストに基づいています。また、現在公開中のすべての更新プログラムが適用されているとは限りません。構成の詳細については、http://www.Intel.com/InnovationEventClaims/ (英語) を参照してください。絶対的なセキュリティーを提供できる製品やコンポーネントはありません。
- すべての製品計画およびロードマップは、予告なく変更されることがあります。 製品版出荷前のシステムやコンポーネントで測定された結果は、インテル・リファレンス・プラットフォーム (新しいシステムの社内サンプルモデル)、インテル社内の分析、アーキテクチャー・シミュレーション、モデリングでの推定 / シミュレーション結果を含め、情報提供のみを目的としています。システム、コンポーネント、仕様、構成に対する今後の変更によって、結果は異なる場合があります。インテルのテクノロジーを使用するには、対応したハードウェア、ソフトウェア、またはサービスの有効化が必要となる場合があります。
- 本資料は、明示されているか否かにかかわらず、また禁反言によるとよらずにかかわらず、いかなる知的財産権のライセンスも許諾するものではありません。 開発コード名は、一般向けに発表または出荷されていない製品やテクノロジー、サービスを識別するためにインテルによって使用されているものです。いずれも「商用」の名称ではなく、商標としての機能を前提としたものではありません。
- インテルは、Principled Technologies が統括する BenchmarkXPRT 開発コミュニティーを含め、さまざまなベンチマーク制定団体 / 機関へのスポンサーとしての参画、また技術的サポートの提供によって、ベンチマークの開発に貢献しています。
- 将来的な計画や予測について言及している本プレゼンテーション資料内の記述は、多数のリスクや不確定要素を伴う将来の見通しです。「想定される」、「見込まれる」、「意図する」、「目標とする」、「計画する」、「考えられる」、「求める」、「推定する」、「継続する」、「可能性がある」、「予定である」、「期待する」、「はずである」、「仮定する」などの表現やその変化形および類似表現は、いずれも将来の見通しであることを示しています。推定、予想、予測、不確定な事象、または仮定について言及している、またはこれらに基づいている記述も、今後リリースされる製品やテクノロジー、こうした製品やテクノロジーに期待される利用可能性とメリット、市場機会、インテルの事業および関連市場に見込まれるトレンドに関する記述を含め、将来の見通しであることを示しています。経営陣による現在の予測に基づくものであり、多数のリスクや不確定要素を伴う将来の見通しです。これらの要因によって、実際の結果はこれらの予測的記述に明示的または黙示的に示された結果と著しく異なる可能性があります。実際の業績をインテルの予測と大きく異ならせる重要な要因には、インテルの投資家向け IR サイト (https://www.intc.com/) または証券取引委員会 (SEC) のウェブサイト (https://www.sec.gov/) で入手可能な Form 10-K および Form 10-Q に関するインテルの最新の報告書を含め、インテルが SEC に提出 / 登録した報告書に記載されています。インテルは、法律で開示が義務付けられている場合を除き、新しい情報、新規開発、その他の結果にかかわらず、本プレゼンテーション資料に記載されたいかなる記述も、更新する義務を一切負わないことを明示的に表明します。
- インテルは、サードパーティーのデータについて管理や監査を行っていません。ほかの情報も参考にしてデータの正確さを評価してください。
-

インテルと
ビッグデータ

インテル@IT
～インテル社内のビッグデータとHPC環境～

データ容量

653PB

(2022年時点)

年間ストレージ増加率:39%

HPC環境

330万コア

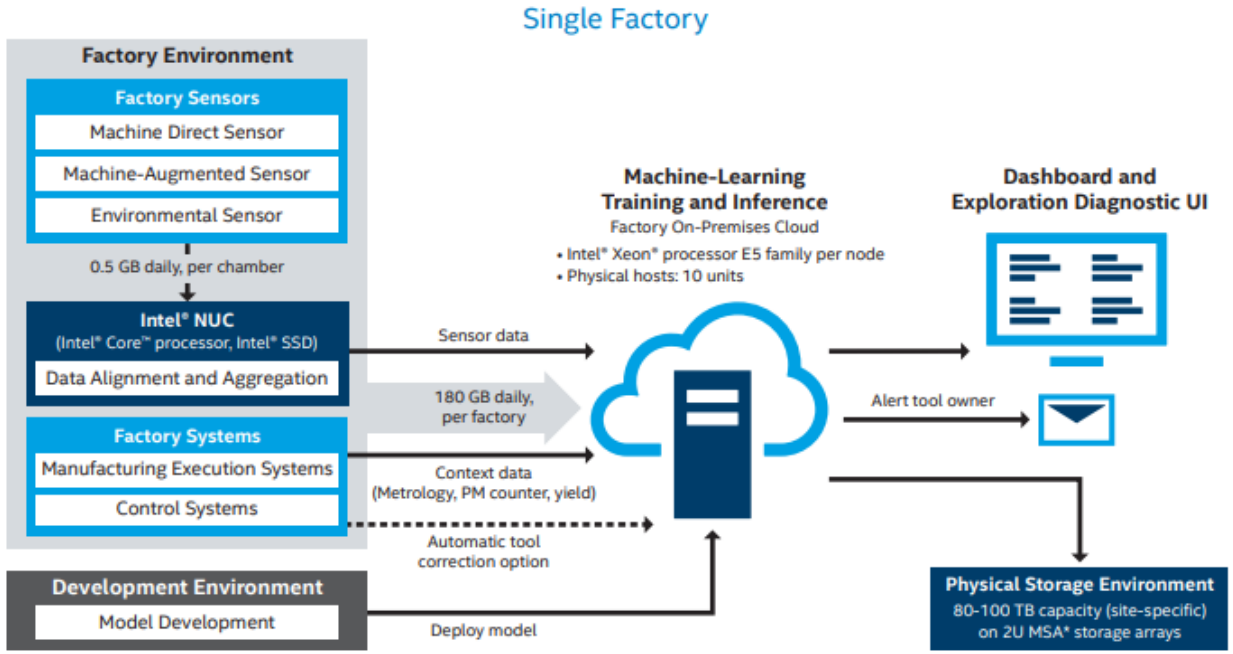
(2022年時点)

年間コンピューティング需要増加率:31%

インテル IT 2021-2022 年次業績報告書

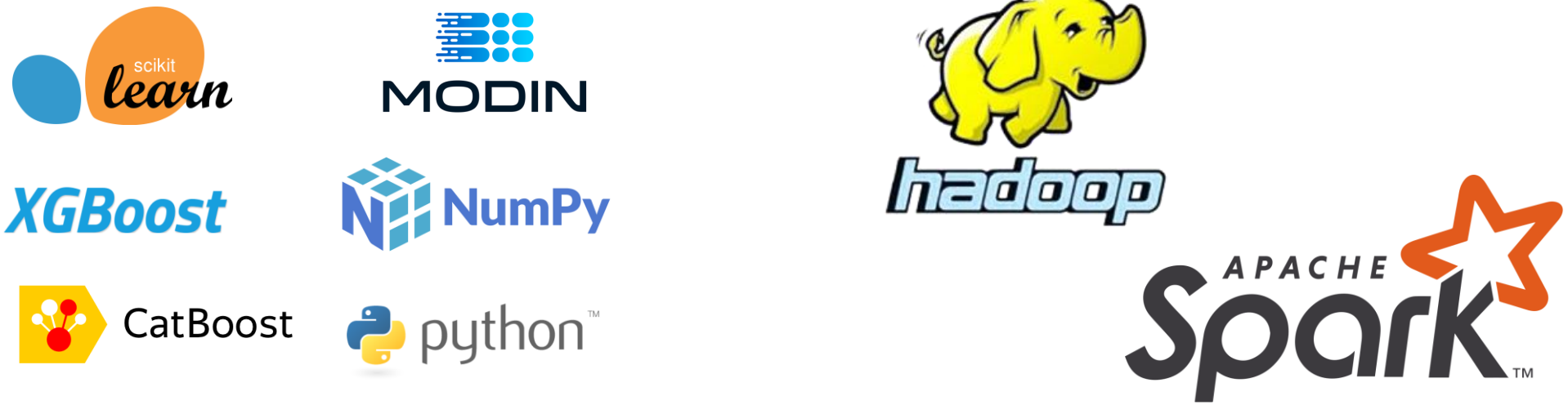
インテル@IT
～ビッグデータの製造への活用例～

品質と歩留まり向上*1	製造テスト時間を50%削減*2	スマートファクトリーオートメーション 4時間⇒30秒で情報抽出*2



*1 https://www.intel.com/content/www/us/en/it-management/intel-it-best-practices/increase-product-yield-and-quality-with-machine-learning-paper.html
*2 https://www.intel.co.jp/content/www/jp/ja/it-management/intel-it-best-practices/intel-it-annual-performance-report-2021-2022-paper.html

インテル@IT
～様々なアナリティクス・フレームワークの活用～



■Best practices for implementing Apache Hadoop at Intel
https://indico.cern.ch/event/282578/contributions/644028/attachments/520402/717933/Best-practices-for-implementing-apache-hadoop-paper.pdf

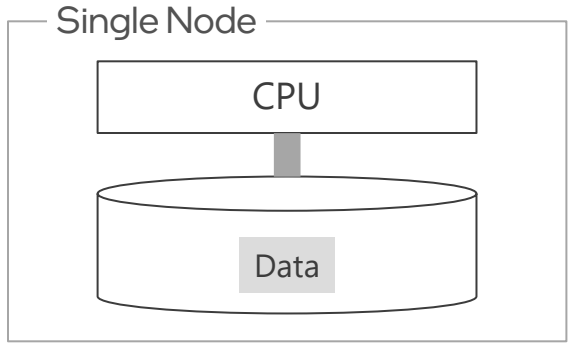
■Speed it up and Spark it up at Intel
https://www.slideshare.net/Hadoop_Summit/speed-it-up-and-spark-it-up-at-intel

ビッグデータ分析基盤の変遷

ビッグデータ分析基盤の変遷（1/2）

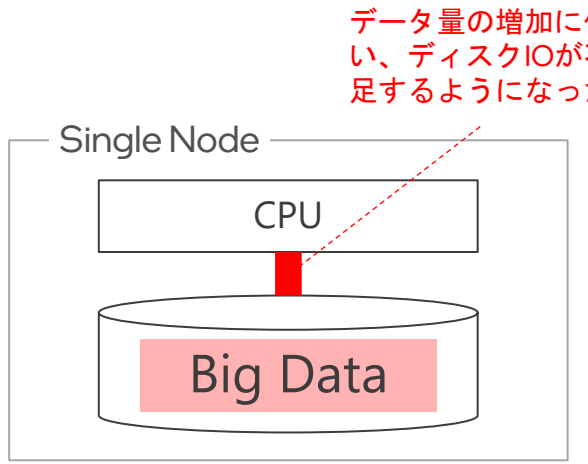
1. シングルノード

- かつては、すべてのデータを1つのノードのディスクに保存し、同じノードで分析できるほどデータ量は多くなかった。RDBはその代表的な技術である。
- より高い性能が必要な場合は、スケールアップアプローチが適用された。



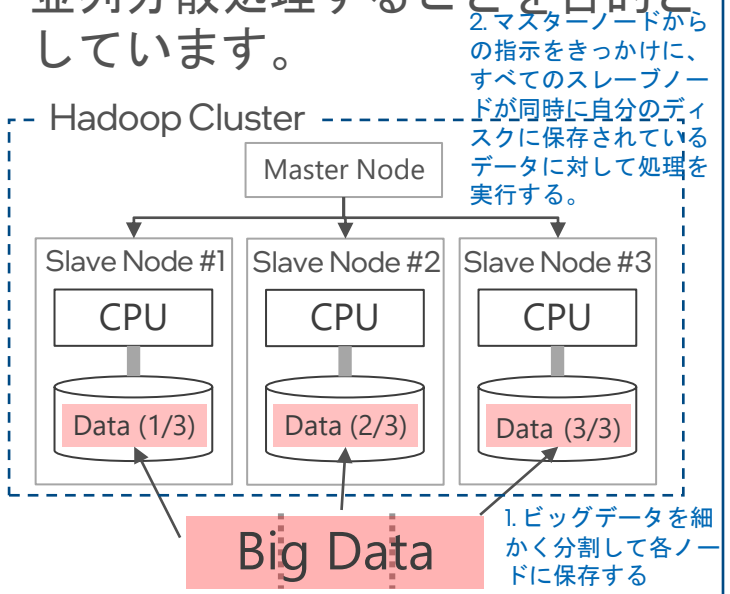
2. シングルノードの性能ボトルネック

- データ量が増えるにつれて、1つのノードでの処理効率が低下する。CPUの処理能力よりもDisk IOがボトルネックになるため。



3. Hadoopの登場

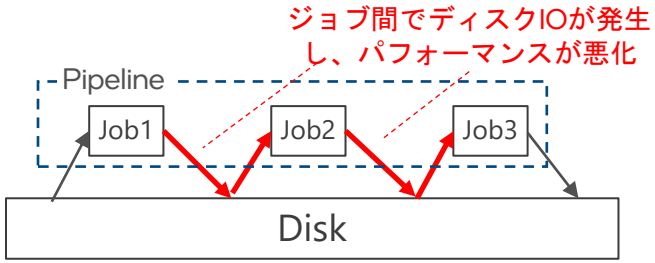
- シングルノードの限界を打破するため、2007年にGoogleがOSSとして作成したHadoopは、ビッグデータを複数のノードで並列分散処理することを目的としています。



ビッグデータ分析基盤の変遷（2/2）

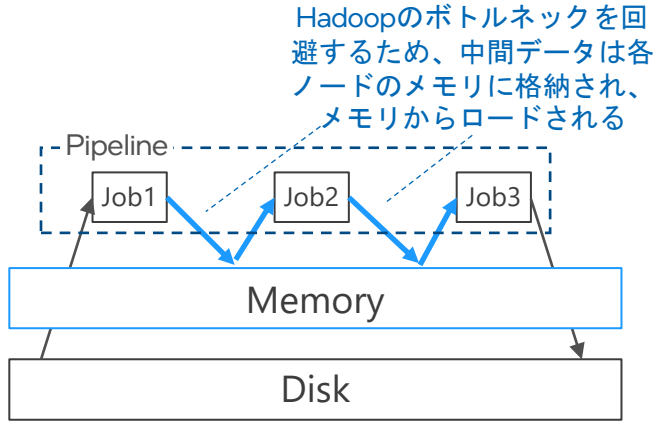
4. Hadoopのパフォーマンスに関する問題点

- Hadoopのスレーブノードでは、データ処理をジョブという単位で実行し、複数のジョブを組み合わせてパイプラインとして複雑な処理を実行します。
- 機械学習などの複雑なパイプラインを実行する場合、各ジョブの出力を1つずつディスクに書き戻すため、処理速度が遅くなる



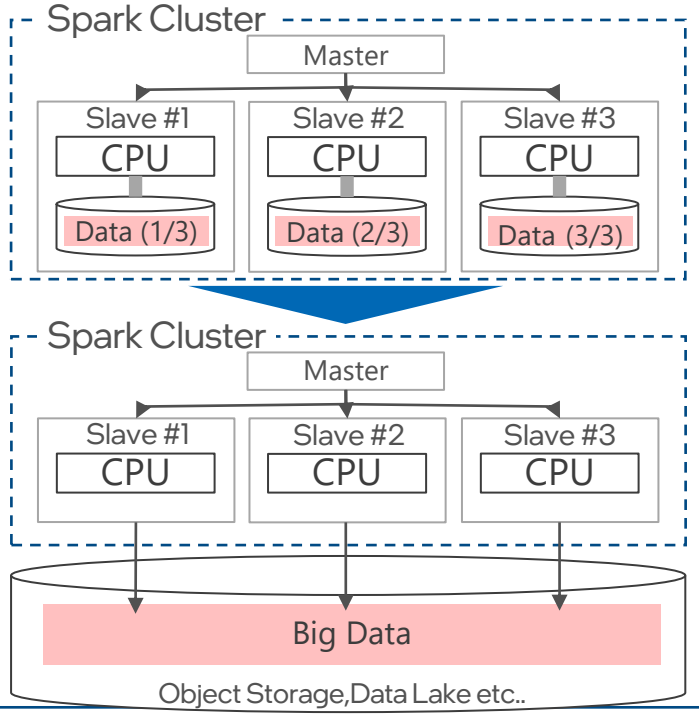
5. Sparkの登場

- Apache Sparkは、Hadoopユーザーを苦しめていた複雑なパイプラインを高速化するために2014年に誕生しました
- 最大のアップデートの1つは、各ジョブの出力をシステムメモリに保存することです。



6. Sparkの進化

- コンピュート層とストレージ層を分離し、よりクラウドネイティブな構成へと進化させる



インテル
ビッグデータ関連製品

ビッグデータ分析に向けた製品強化

ハードウェア

+

ソフトウェア

Optimized Analytics Package (OAP)

BDTK

Gluten

scikit learn

MODIN

APACHE Spark

Spark MLlib
The Machine Learning Library

XGBoost

NumPy

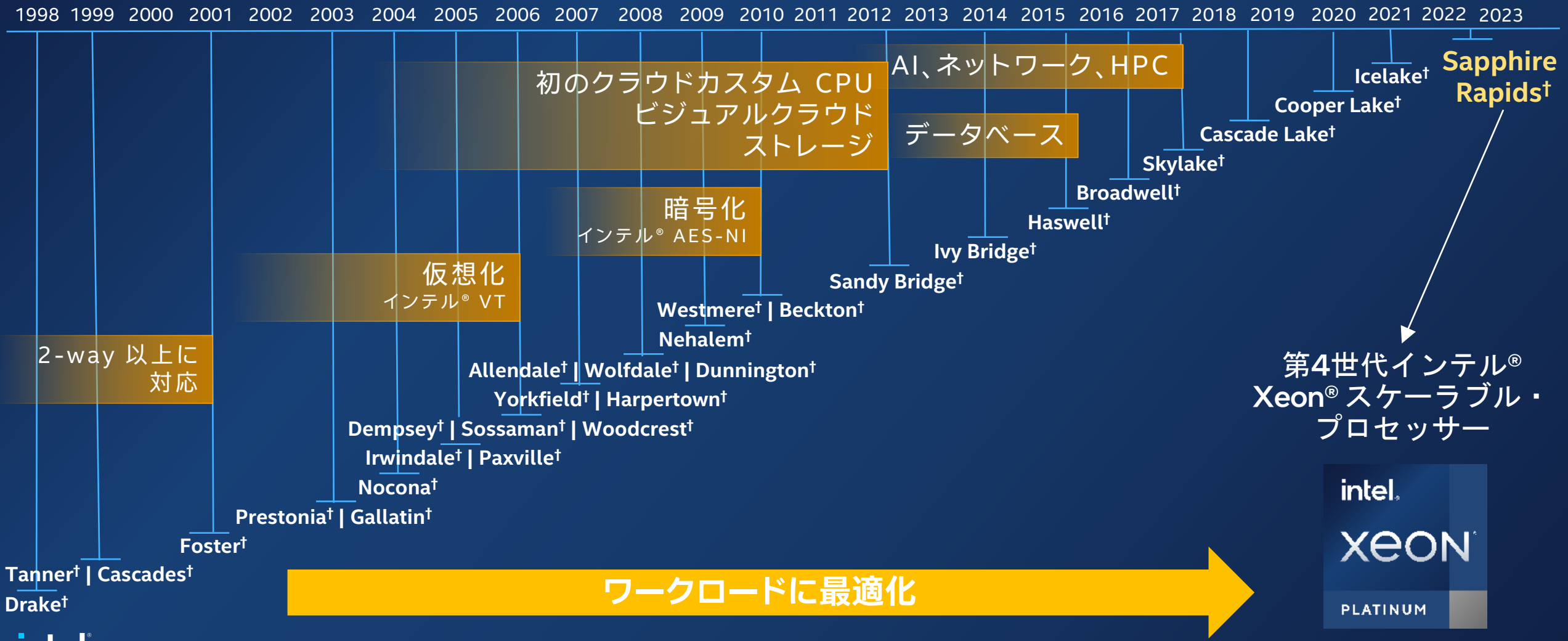
CatBoost

python

BigDL

oneAPI

20 年以上にわたるプロセッサー・イノベーション

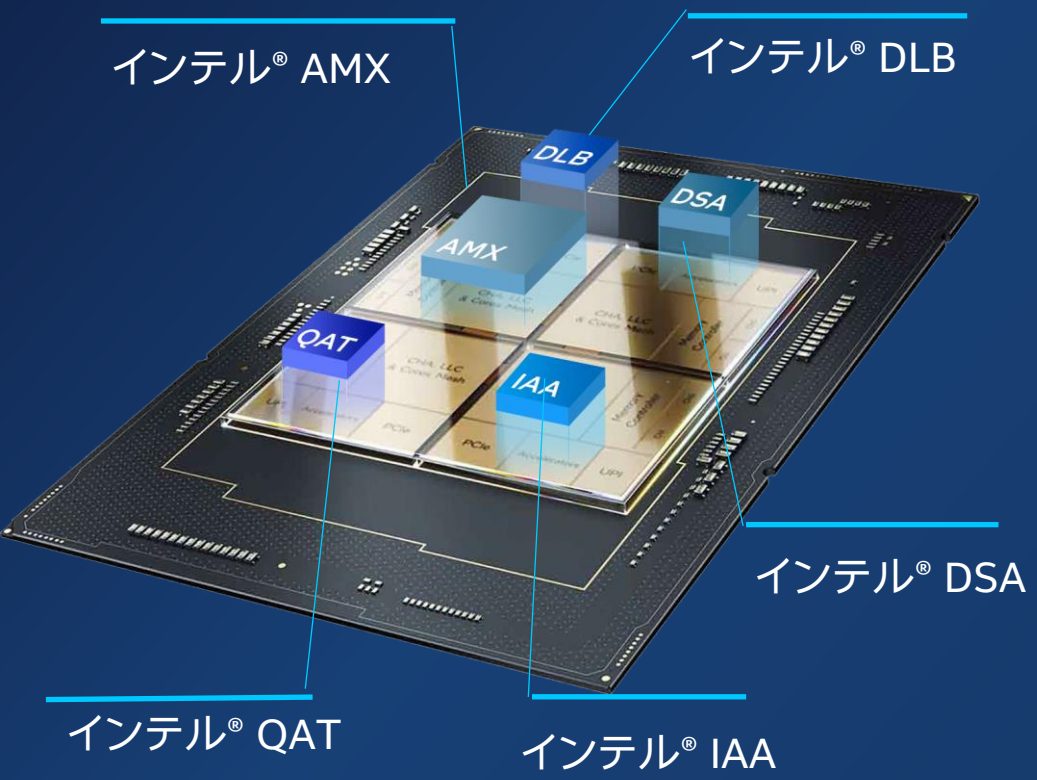


intel XEON Accelerate with Xeon

† 開発コード名

インテル® アクセラレーター・エンジン

CPUにかかる演算負荷をオフロードし、実環境でのワークロード性能向上を実現



インテル® アドバンスト・マトリクス・エクステンション (インテル® AMX)

- TensorFlow
- PyTorch
- ONNX Runtime
- OpenVINO
- oneDNN (Intel oneAPI)

vRAN 向けインテル® アドバンスト・ベクトル・エクステンション (インテル® AVX)

- FlexRAN
- Data Plane dev Kit (DPDK)*

インテル® インメモリー・アナリティクス・アクセラレーター (インテル® IAA)

- Intel Query Processing Library

インテル® データ・ストリーミング・アクセラレーター (インテル® DSA)

- Storage Perf Dev Kit (SPDK)*
- Data Plane Dev Kit (DPDK)*

インテル® クイックアシスト・テクノロジー: (インテル® QAT)

- QATzip* (Intel lib)
- OpenSSL**
- Boring SSL

インテル® ダイナミック・ロードバランサー (インテル® DLB)

- VPP IPsec
- Data Plane Dev Kit (DPDK)*

intel XEON Accelerate with Xeon

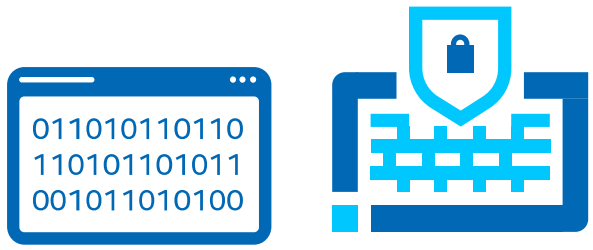
*Intel open-source library (not part of stock SW).
**Difference between Intel version and stock version.
***Intel® QPL and *Intel® DML in open-source beta, v1.0.0 coming shortly.

インテル® クイックアシスト・テクノロジー (インテル® QAT)

インテル® クイックアシスト・テクノロジーは、計算負荷の高いワークロードのハードウェア・アクセラレーションを統合。
インテル® QAT ハードウェアにオフロードすることで、一括暗号化、公開キー暗号化、圧縮を高速化。
CPU効率、データフットプリントの削減、電力使用率、およびアプリケーションのスループットの大幅な向上が可能。

インテル® クイックアシスト・テクノロジー

暗号化、ハッシュ、認証

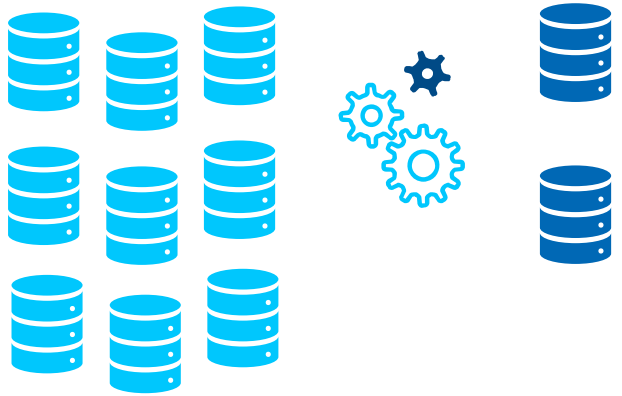


対称暗号化と認証

公開鍵暗号化

非対称暗号化とデジタル署名

圧縮 / 解凍



転送中および保存中のデータに対する
ロスレスデータ圧縮/解凍

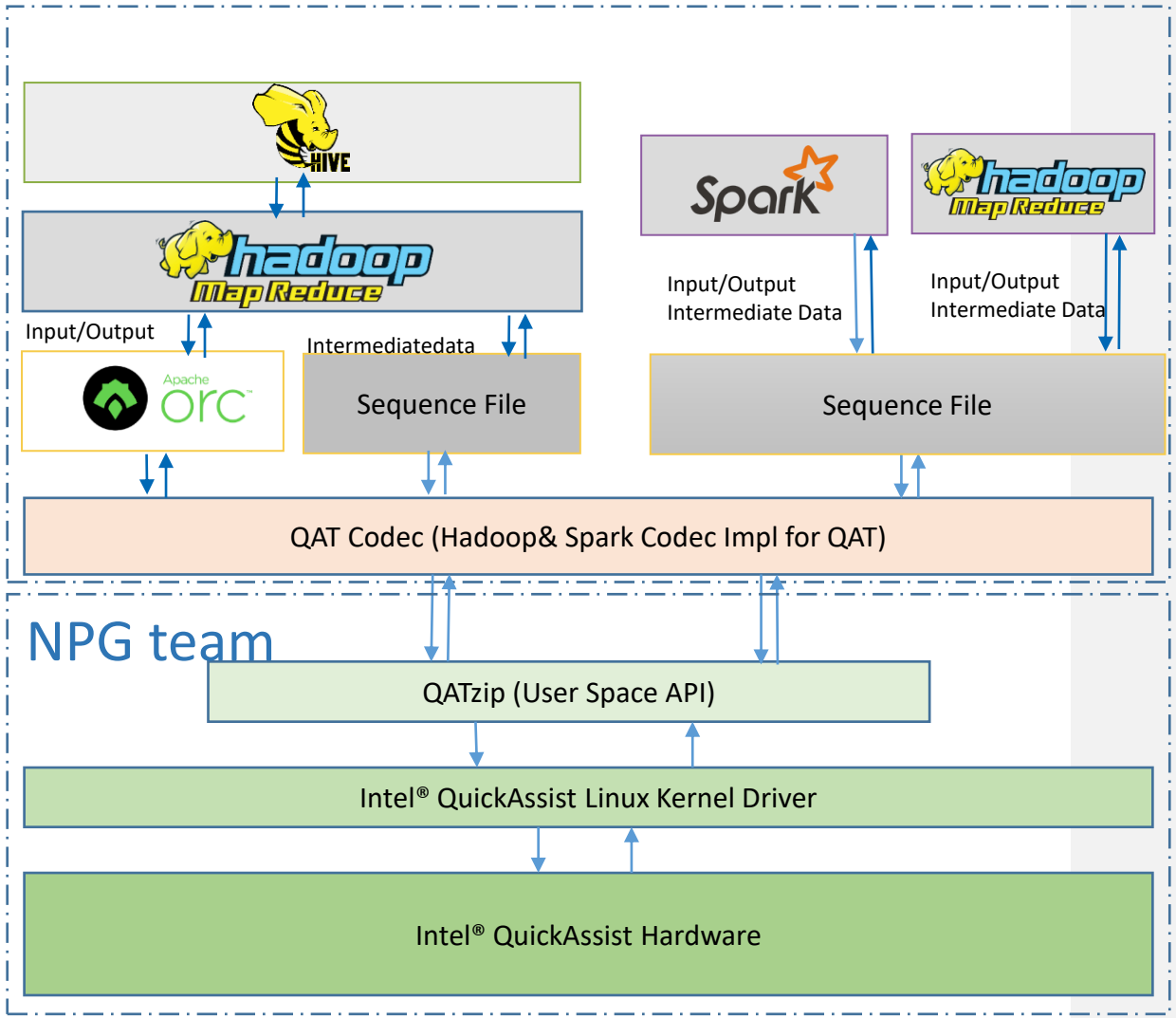
QATzip in Hadoop / Spark

https://github.com/intel-hadoop/IntelQATCodec

QATzip は QAT Codecと統合し、インテル® QuickAssist Technology を使用した Apache Hadoop/Apache Hive/Apache Spark 用の圧縮・解凍ライブラリです。

その利点は

- より良い圧縮率でより高いパフォーマンス（Snappyに対して）
  - Map-Reduceワークロードの場合： 7.29%の性能向上、7.5%の圧縮率向上
  - Sparkソートワークロード： 14.3%の性能向上、7.5%の圧縮率向上
  - Hive on MR：12.98%の性能向上、13.65%の圧縮率向上
- オープン
  - 主要なビッグデータコンポーネント（Spark、Hadoop、Hive、Parquet）で優れたサポートを提供
  - オープンソースプロジェクト



インテル® インメモリー・アナリティクス・アクセラレーター (インテル® IAA)

アナリティクス向け：IAAはクエリのスループットを向上させ、メモリフットプリントを減少

アナリティクスWLは、通常、次の2つの操作に依存

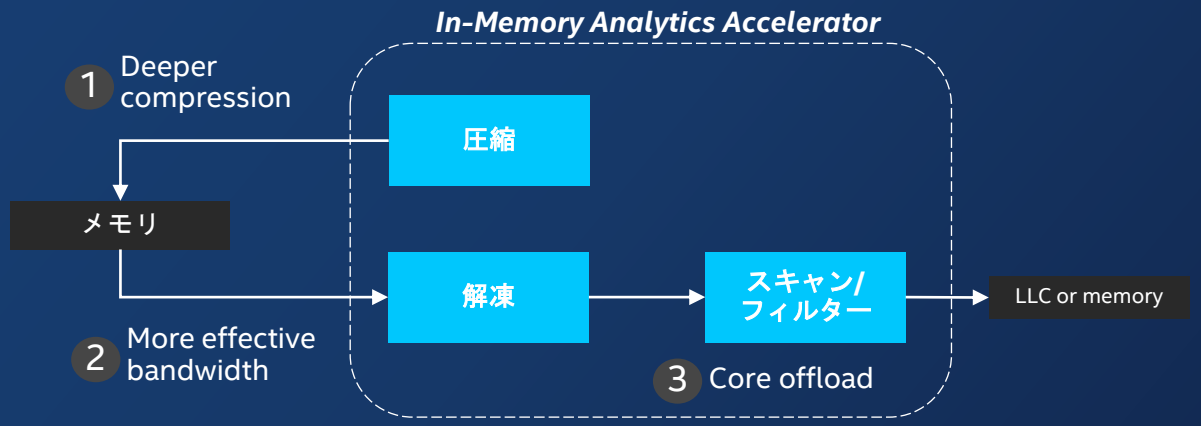
- *スキャン/フィルター：大規模なデータセットから関連するデータを特定*
- *圧縮/解凍：メモリ消費量の削減*

IAAは、クエリのスループットを向上させ、メモリフットプリントを減少：

1. ソフトウェアのみの圧縮技術よりも、より深く圧縮
2. 深く圧縮されたデータは帯域幅を消費しないため、より効果的な帯域幅が得られる
3. コアの代わりにIAAが計算負荷の高いスキャンとフィルター処理を行うため、コアのオフロードが可能

IAAは、SPR CPUに内蔵されたアクセラレーターで、解析プリミティブ（スキャン、フィルターなど）、CRC計算、圧縮、解凍などを高速化

- すべてのEGS互換メモリで動作
- ごくわずかな電力フットプリント



Intel Query Processing Library（Intel® QPL）

3 Facebookは、PrestoDB分析ソリューションのプロファイリングを行った際、次のことを確認した。

“***全世界のPrestoのCPU時間の60%近くがテーブルスキャンに起因している...***”

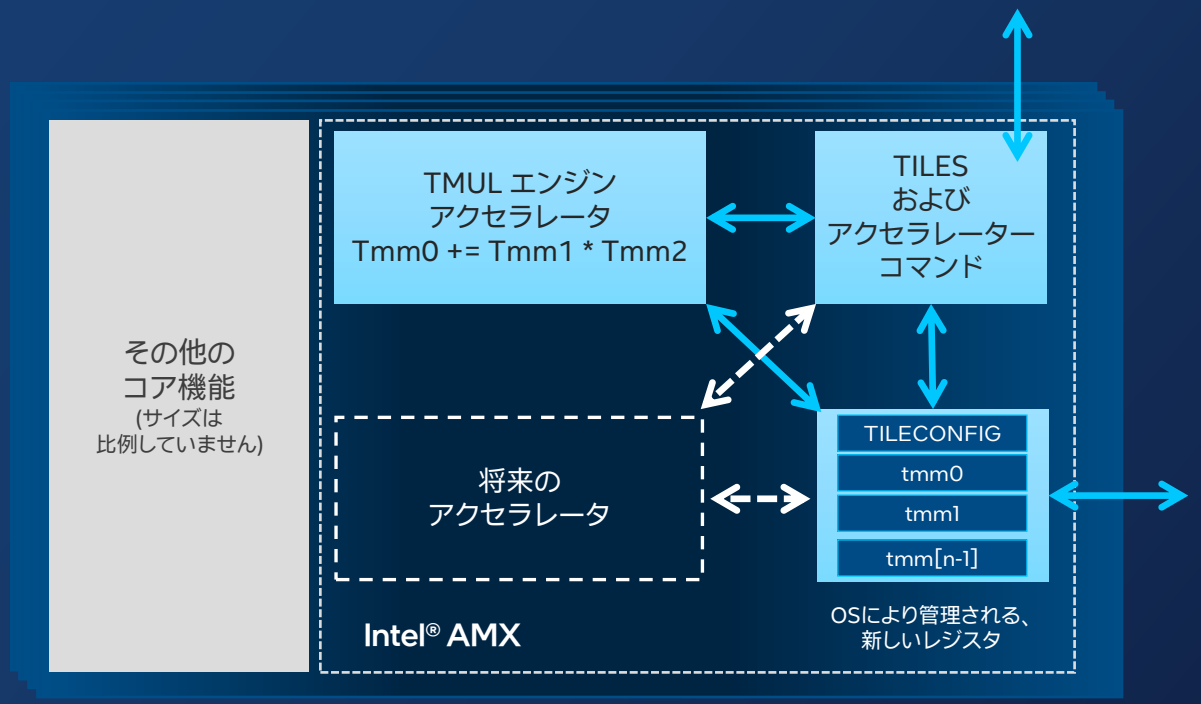
Aria Presto: テーブルスキャンをより効率的に, Facebook Engineering (link)

インテル® アドバンスド マトリックス エクステンションズ (インテル® AMX)

第4世代 インテル® スケーラブル・プロセッサー各コアに内蔵されたDeep Learningアクセラレータ

データを
大きな塊で保存

大きな行列を**一つの処理**で計算可能な命令



サポートされるデータタイプ

BF16: 学習と推論

INT8: 推論

intel XEON Accelerate with Xeon

より高性能なサーバー・アーキテクチャー

インテル® アクセラレーター・エンジンがもたらすメリット

インテル® アドバンスト・マトリクス・エクステンション (インテル® AMX)

最大

8.6 倍

内蔵のインテル® AMX BF16 により、音声認識の推論パフォーマンスが向上 (FP32 との比較)

インテル® ダイナミック・ロード・バランサー (インテル® DLB)

最大

96%

インテル® DLB により、同等のスループットを維持しながら Istio-Envoy Ingress のレイテンシーを低減 (Istio Ingress ゲートウェイ用ソフトウェアとの比較)

インテル® データ・ストリーミング・アクセラレーター (インテル® DSA)

最大

1.7 倍

内蔵のインテル® DSA により、SPDK-NVMe の IOPS が向上 (ISA-L ソフトウェアとの比較)

インテル® インメモリー・アナリティクス・アクセラレーター (インテル® IAA)

最大

2.1 倍

インテル® IAA により、RocksDB のパフォーマンスが向上 (Ztsd ソフトウェアとの比較)

インテル® クイックアシスト・テクノロジー (インテル® QAT)

最大

84%

内蔵のインテル® QAT により、少ないコア数で NGINX の毎秒接続数を同等に維持 (既成ソフトウェアとの比較)

アクセラレーター群が基本アーキテクチャーを超えて 1 段階上のパフォーマンスを実現

intel XEON Accelerate with Xeon

ワークロードと構成については、補足資料を参照してください。結果は異なる場合があります。

ビッグデータ分析に向けた製品強化

ハードウェア

+

ソフトウェア

Optimized Analytics Package (OAP)

BDTK

Gluten

scikit learn

MODIN

APACHE Spark

Spark MLlib
The Machine Learning Library

XGBoost

NumPy

CatBoost

python

BigDL

oneAPI

Modin –データ操作用ライブラリ-
1行のコード変更で無限のスケーラビリティを実現

```
import pandas as pd
```

MODIN

0.9のバージョンでは、ModinはPandas APIを100%サポートしています。

その他のPython*エコシステムとの統合（NumPy、XGBoost、Scikit-Learn*などを通常通り使用可能）

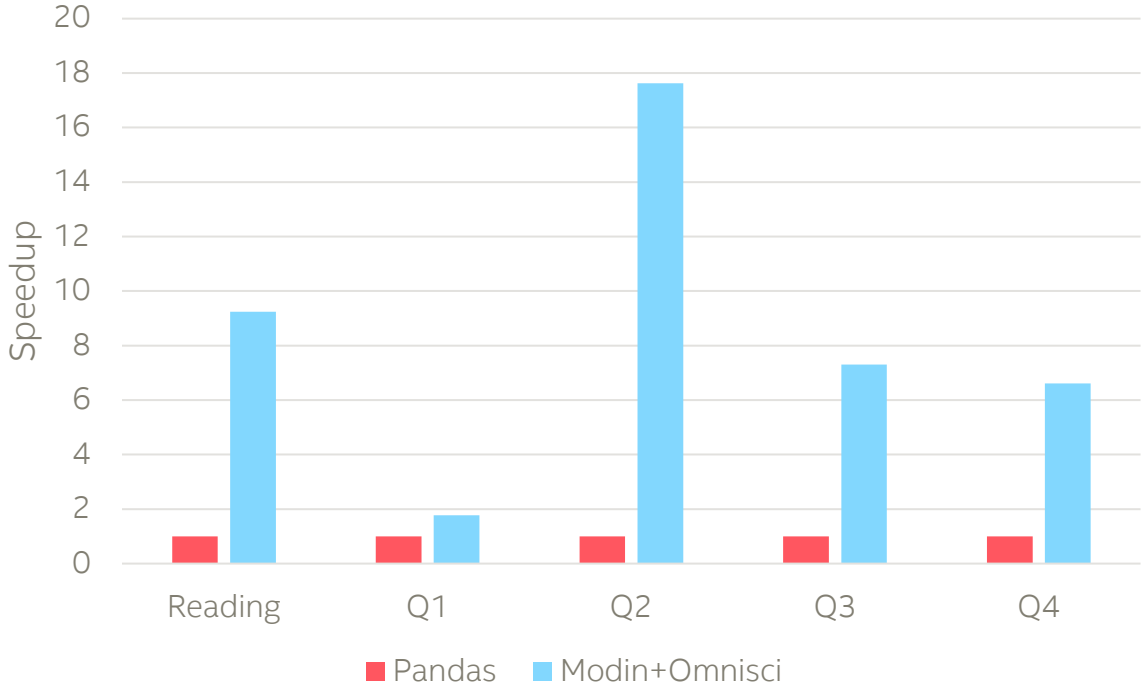
インテル・ディストリビューションのModinは、バックエンドでは既存および新規のインテル®ハードウェアのコンピューティング・パワーを活用するために最適化されたエンドツーエンド・アナリティクスのための高性能フレームワークであるOmniSci*をサポートしています。

omnisci

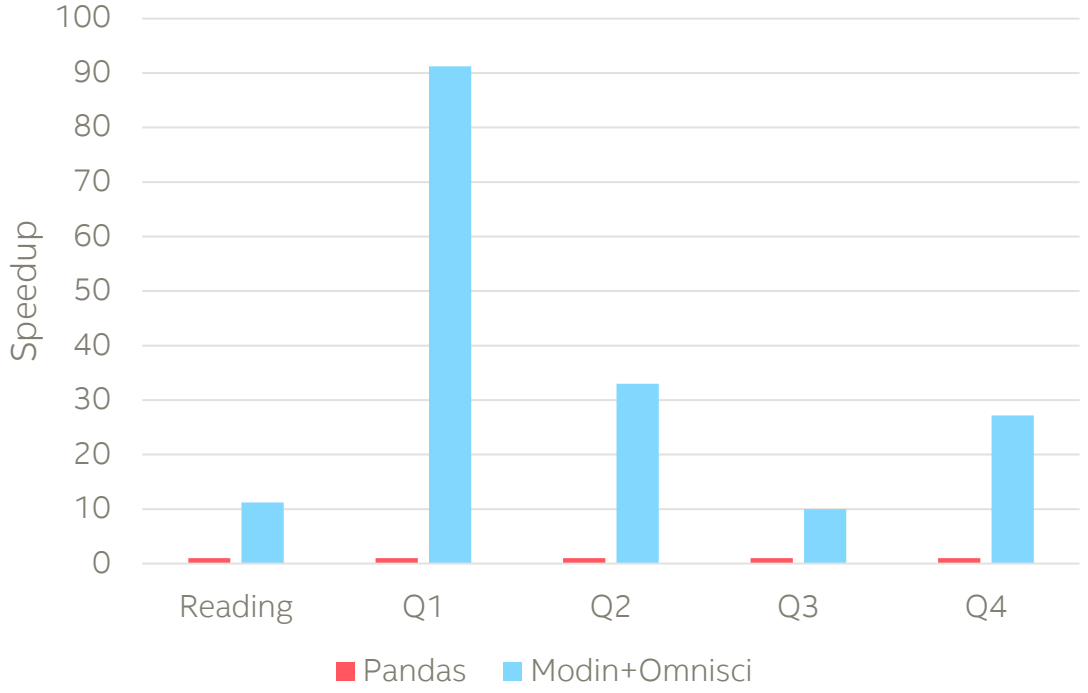
NYCTaxi ワークロード性能

Pandas vs Modin – Higher is Better

NYCTaxi (20 Million rows)
Performance improvement with Modin+Omnisci



NYCTaxi (1 Billion rows = 1.6 TB in mem)
Performance improvement with Modin+Omnisci – using 3TB Optane

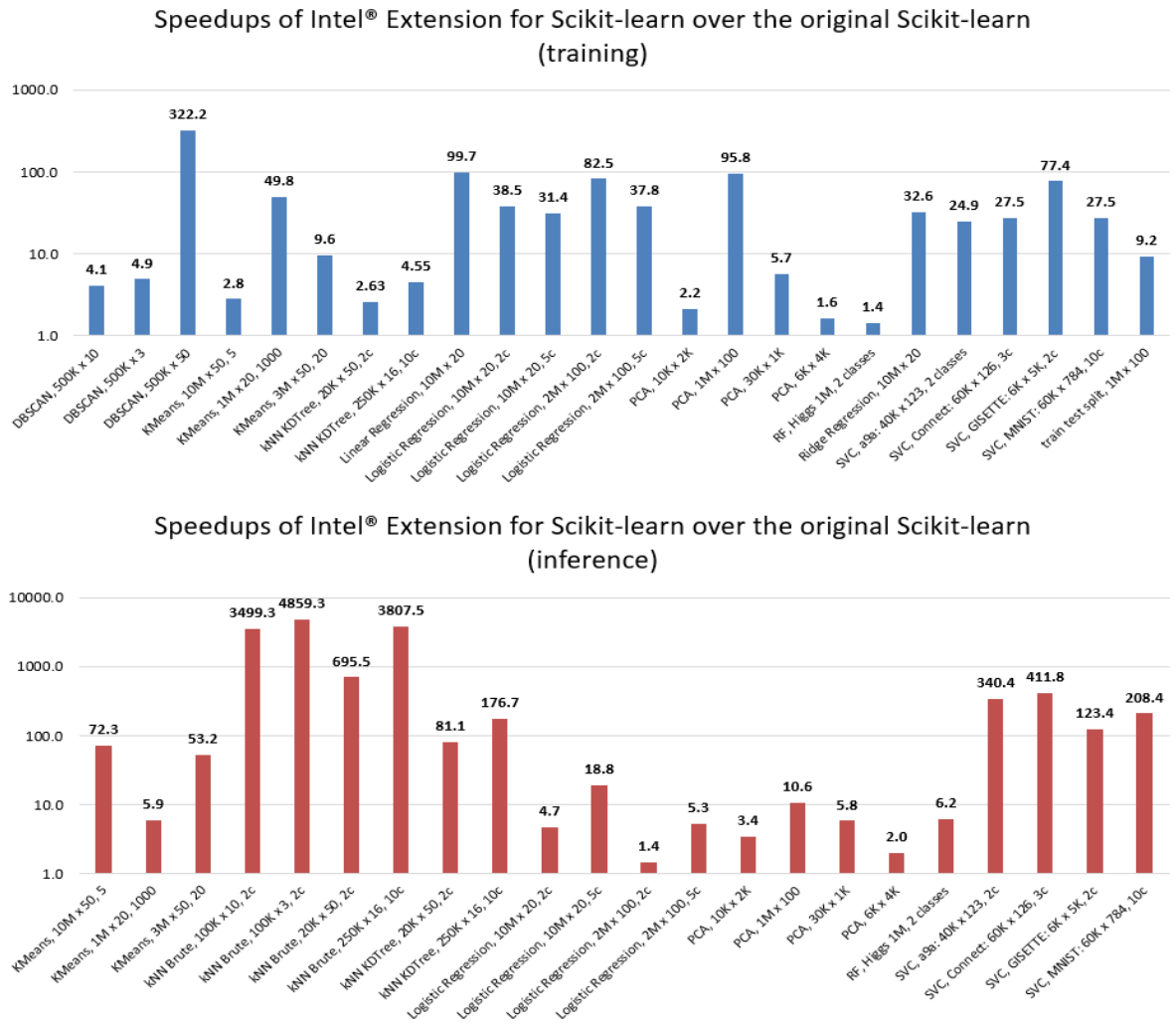


Dataset source: https://github.com/toddwschneider/nyc-taxi-data

Configurations: For 20 million rows: Dual socket Intel(R) Xeon(R) Platinum 8280L CPUs (S2600WFT platform), 28 cores per socket, hyperthreading enabled, turbo mode enabled, NUMA nodes per socket=2, BIOS: SE5C620.86B.02.01.0013.121520200651, kernel: 5.4.0-65-generic, microcode: 0x4003003, OS: Ubuntu 20.04.1 LTS, CPU governor: performance, transparent huge pages: enabled, System DDR Mem Config: slots / cap / speed: 12 slots / 32GB / 2933MHz, total memory per node: 384 GB DDR RAM, boot drive: INTEL SSDSC2BB800G7. For 1 billion rows: Dual socket Intel Xeon Platinum 8260M CPU, 24 cores per socket, 2.40GHz base frequency, DRAM memory: 384 GB 12x32GB DDR4 Samsung @ 2666 MT/s 1.2V, Optane memory: 3TB 12x256GB Intel Optane @ 2666MT/s, kernel: 4.15.0-91-generic, OS: Ubuntu 20.04.4

Intel® Extension for Scikit-learn

https://github.com/intel/scikit-learn-intelex



Same Code, Same Behavior

PASSED

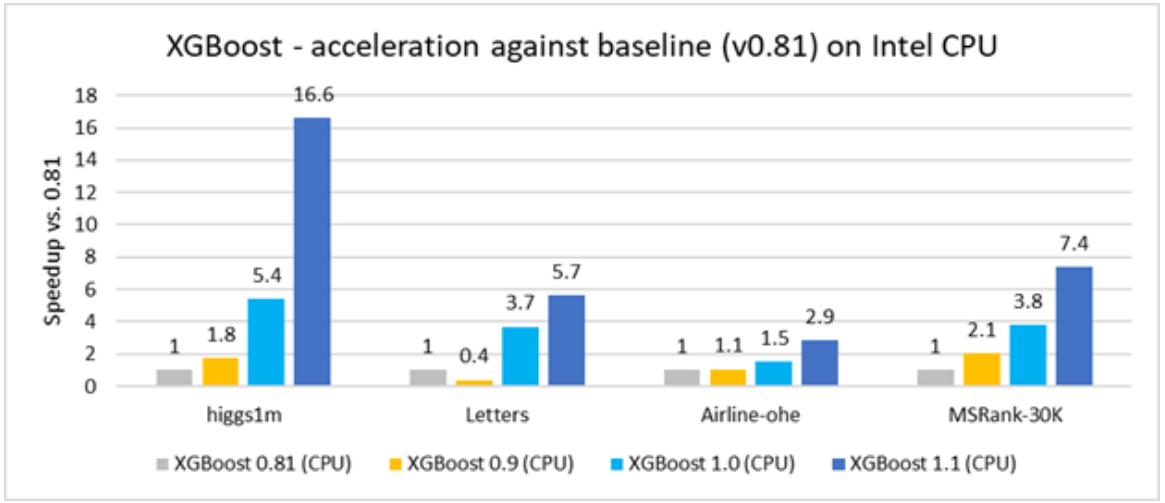
Scikit-learn, **not** scikit-learn-*like*

Scikit-learn conformance (mathematical equivalence) defined by Scikit-learn Consortium, continuously vetted by public CI

Technical details: float type: float64; HW: c5.24xlarge AWS EC2 Instance using an Intel Xeon Platinum 8275CL with 2 sockets and 24 cores per socket; SW: scikit-learn version 0.24.2, scikit-learn-intelex version 2021.2.3, Python 3.8, benchmark code

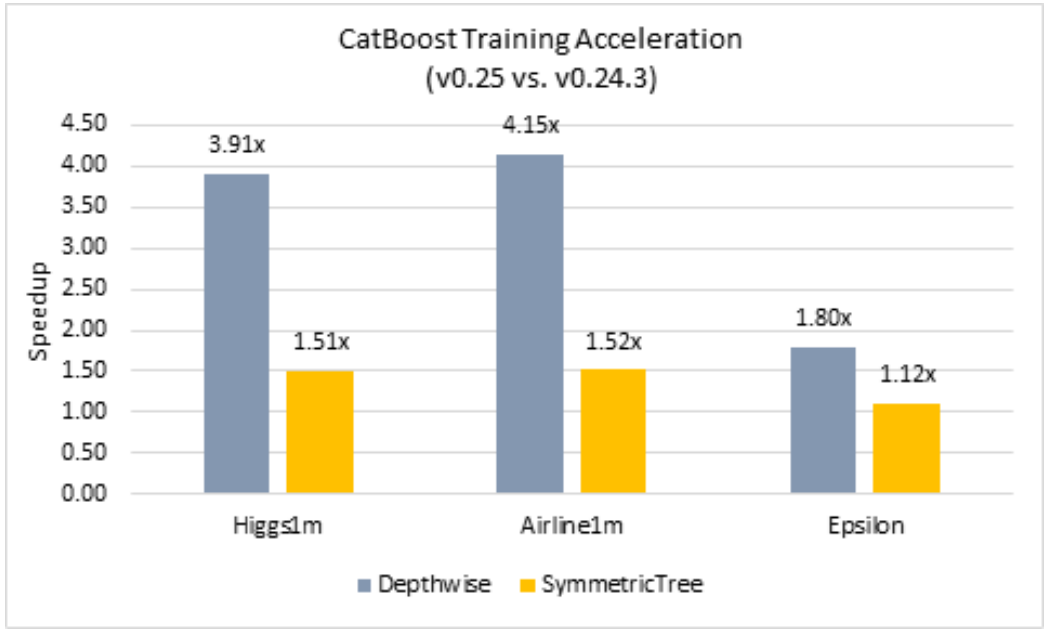
XGBoost / CatBoost

XGBoost



https://medium.com/intel-analytics-software/new-optimizations-for-cpu-in-xgboost-1-1-81144ea21115

CatBoost



https://medium.com/intel-analytics-software/optimizing-catboost-performance-4f73f0593071

ビクトリア大学の次世代クラウド基盤が、科学者の知の広がりを支える

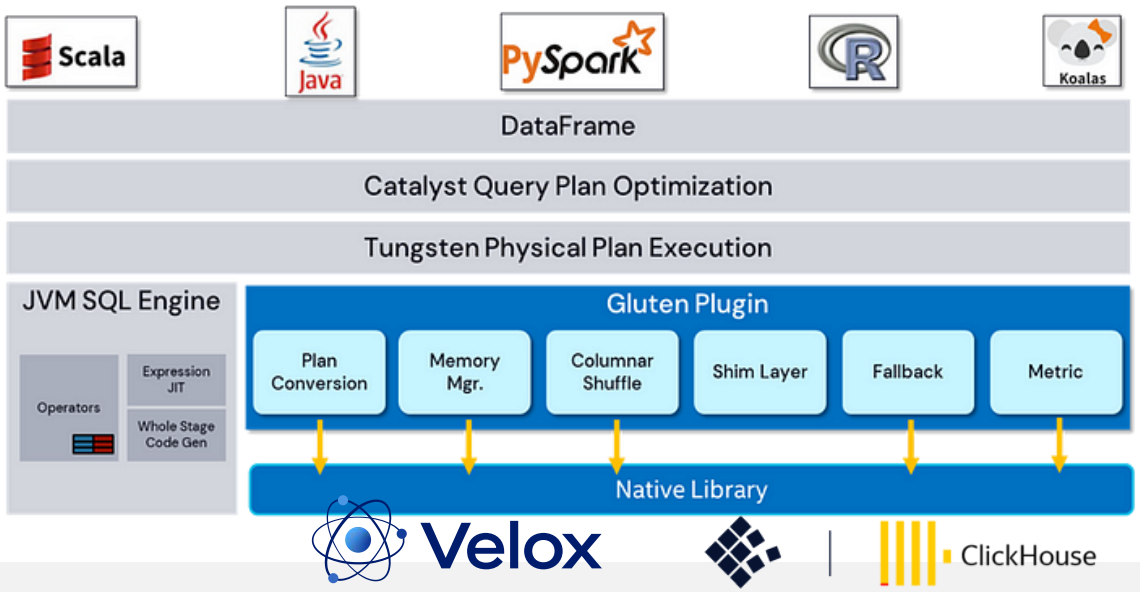
ビクトリア大学のリサーチ・コンピューティング・サービス（RCS）ユニットは、大学の研究者、国内の研究機関の科学者、そして国際的なコラボレーションを通じて、アドバンスト・リサーチ・コンピューティング（ARC）インフラとサービスを提供しています。この施設は、Compute CanadaのARCデータセンターの1つと、仮想マシンやその他のクラウドワークロードをホストすることに重点を置いたOpenStackクラウドであるArbutusクラウドをホストしています。Arbutusは、従来の大型クラスタHPCワークロードを補強し、オンライン機械学習/人工知能、ビッグデータ、共同計算など、従来のHPCクラスタとは異なる機能を必要とする研究プロジェクトをサポートするために設計されました。Arbutusは、Lenovo SR630、SR670、SD530ノードで構築され、第2世代インテル® Xeon® Scalableプロセッサーとインテル® Optane™永続メモリ、インテルSSDが使用されています。

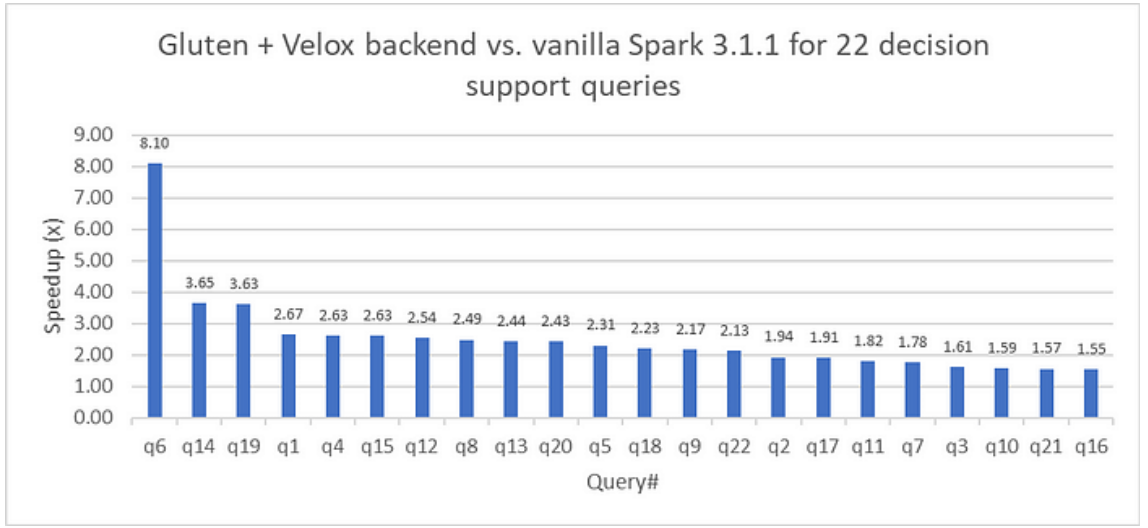
Industry	Organization Size	Country	Partners	Learn more
Higher Education	1,001-5,000	Canada	Lenovo	こちら

Gluten

https://github.com/oap-project/gluten

- Glutenプロジェクトの主な目的はSparkSQLとネイティブライブラリ（Velox, Clickhouseなど）を"繋ぐ"こと
- これによりSpark SQLのスケールアウトフレームワークとネイティブライブラリの高性能を利用することが可能

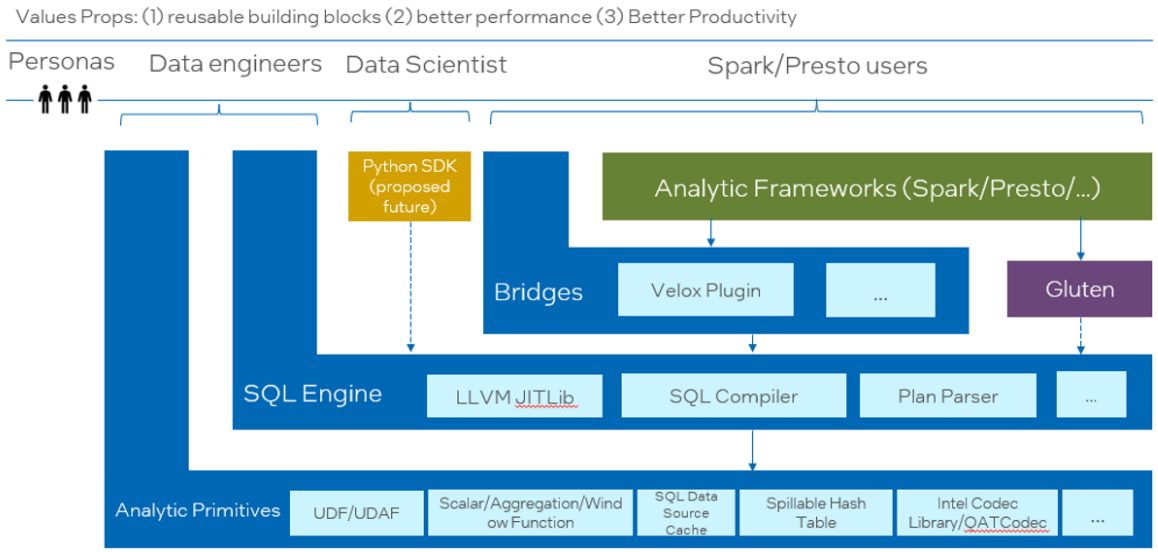


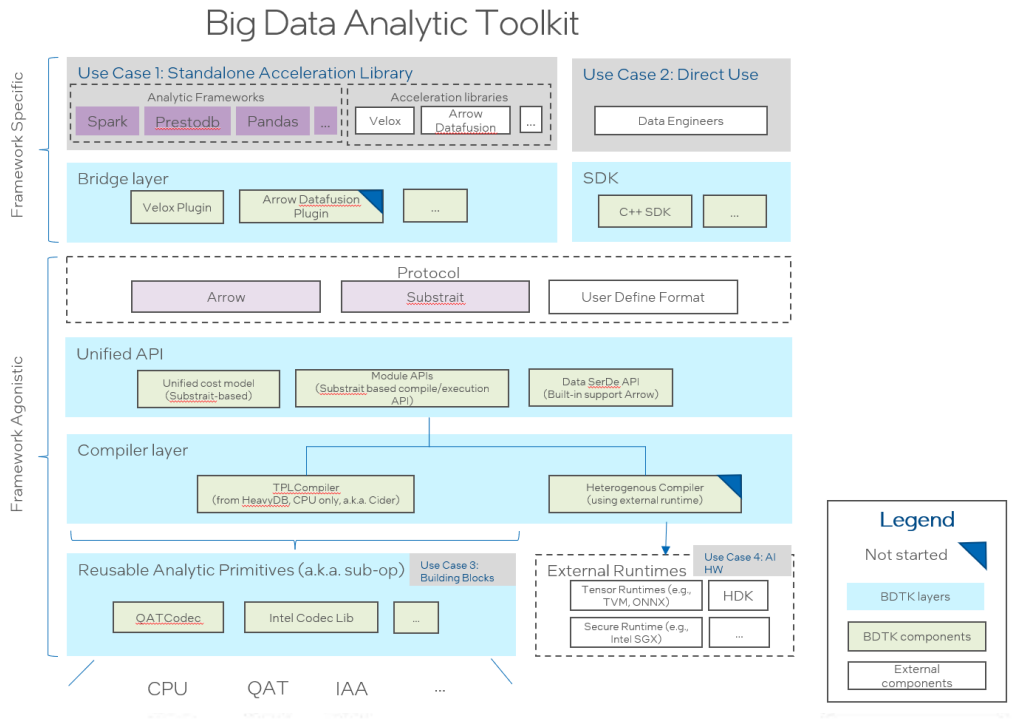


https://medium.com/intel-analytics-software/accelerate-spark-sql-queries-with-gluten-9000b65d1b4e

Intel Big Data Analytic Toolkit (BDTK)

https://github.com/intel/BDTK



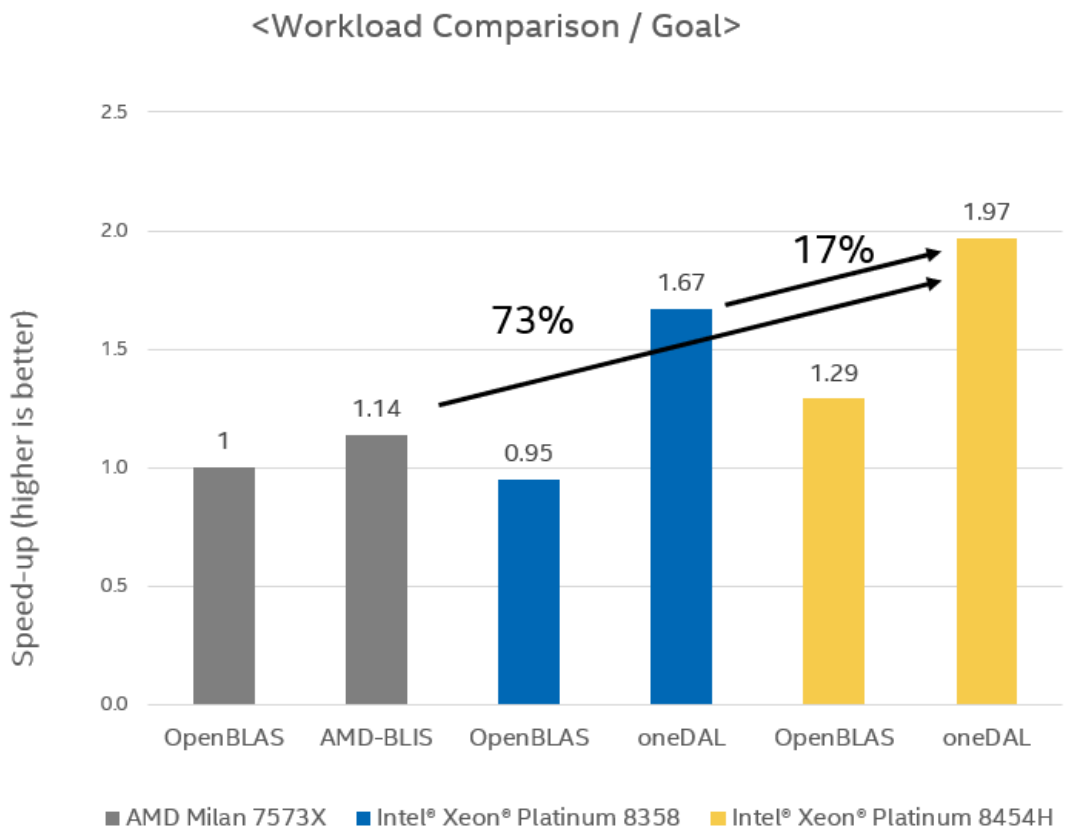


- BDTKは、データ分析フレームワークを最適化することを目的としたアクセラレーションライブラリの集合体です。このライブラリを使用することで、Prestodb/SparkのようなフロントエンドのSQLエンジンのクエリパフォーマンスが大幅に改善されます。
  - BDTKは、最新のXeonハードウェアアクセラレータ（AVX512、IAA、QAT）を活用
- BDTK は、Velox と Velox-Plugin を介して、Presto End-to-End アクセラレーション・ソリューションを提供

OAP-MLlib

https://github.com/oap-project/oap-mllib

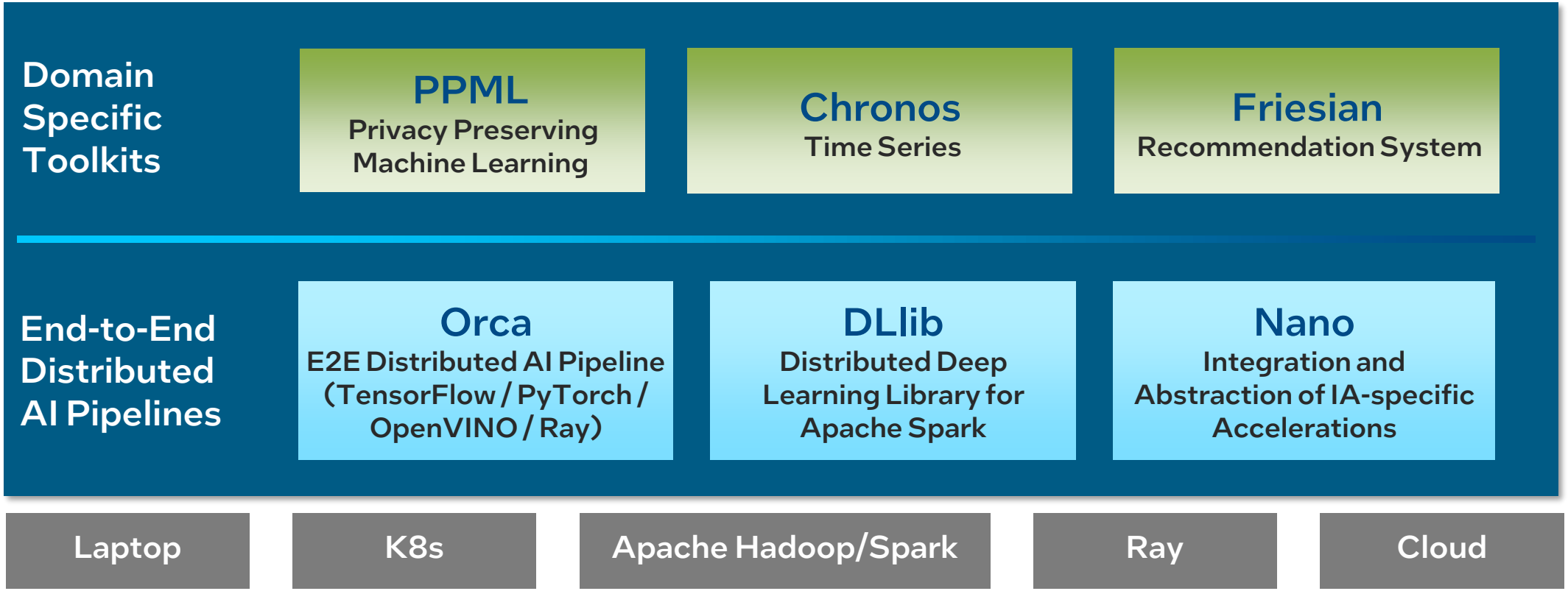
HiBench KMeans



- Application
  - KMeans is a clustering algorithm which calls BLAS routines
- Customer Impact
  - 17% performance gain on Intel Xeon 8454H when compared to Intel Xeon 8358 with Optimized MLlib
  - 73% performance gain on Intel Xeon 8454H with Intel Optimized MLlib when compared to AMD 7573X with AMD-BLIS
- Performance Drivers
  - Intel optimized MLlib relays all levels of BLAS calls to oneAPI
    - OAP-mllib calls oneDAL Kmeans implementation
    - Netlib-java relays only L1/L2 BLAS calls to native library

BigDL

Seamlessly scale end-to-end, distributed Big Data AI applications

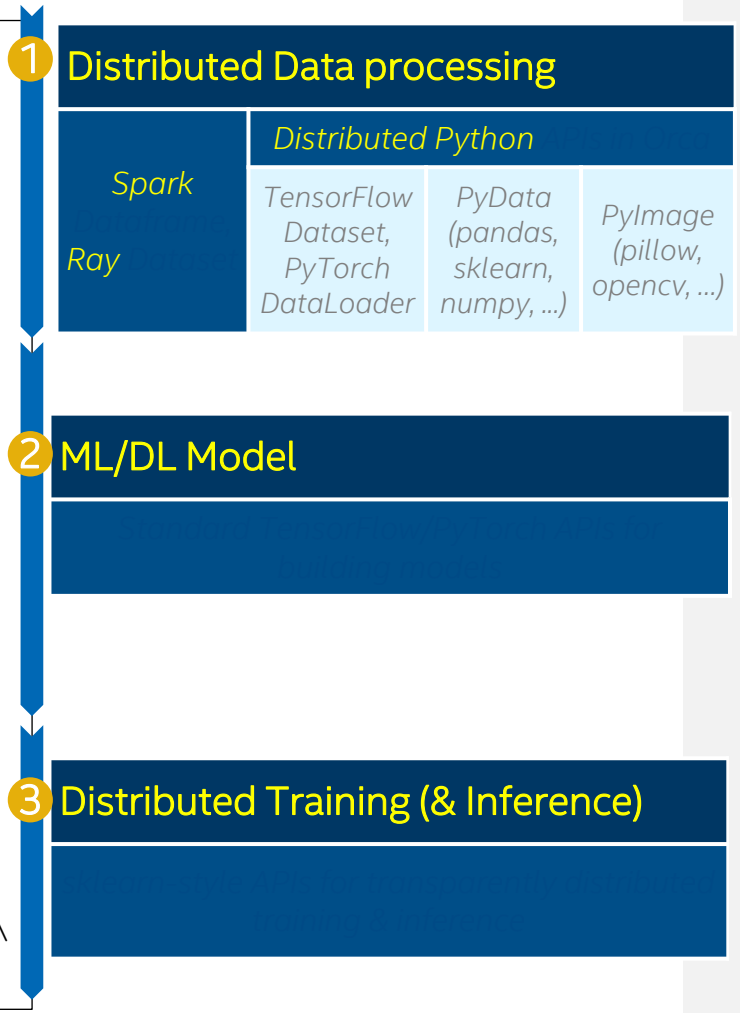


BigDL 2.0 (https://github.com/intel-analytics/BigDL/) combines the *original BigDL* and *Analytics Zoo* projects

* *"BigDL 2.0: Seamless Scaling of AI Pipelines from Laptops to Distributed Cluster", 2022 Conference on Computer Vision and Pattern Recognition (CVPR 2022)*
* *"BigDL: A Distributed Deep Learning Framework for Big Data", in Proceedings of ACM Symposium on Cloud Computing 2019 (SOCC'19)*

Orca: エンドツーエンドの分散型AIパイプラインの構築

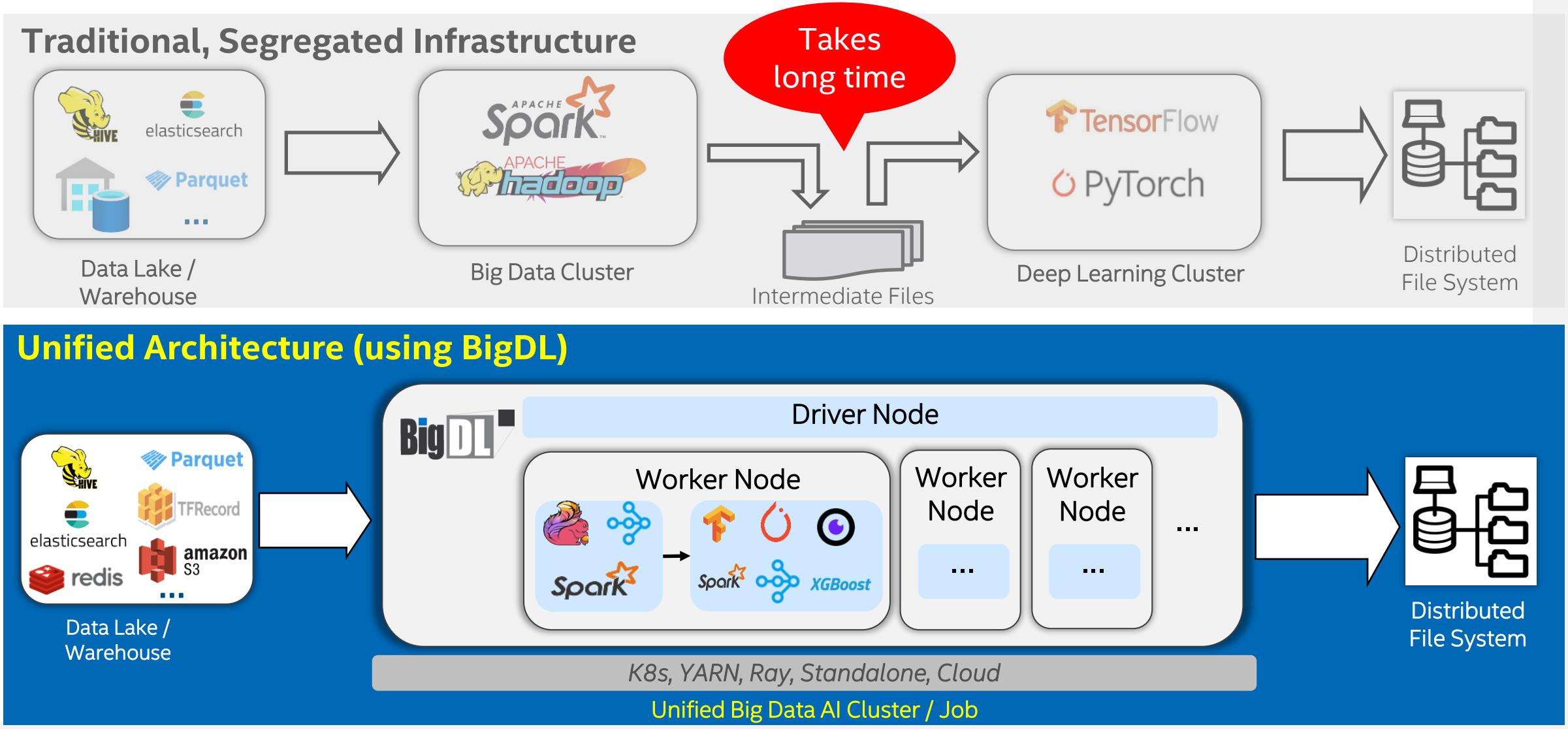
```
#1. Distributed data processing using Spark Dataframe
raw_df = spark.read.format("csv").load(data_source_path) \
  .select("Cardholder Last Name", "Cardholder First Initial", \
          "Amount", "Vendor", "Year-Month") \
  ...

#2. Building model using TensorFlow
import tensorflow as tf
...
model = tf.keras.models.Model(inputs=input, outputs=output)
model.compile(optimizer='rmsprop',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])

#3. Distributed training on Orca
from zoo.orca.learn.tf.estimator import Estimator
est = Estimator.from_keras(model, model_dir=args.log_dir)
est.fit(data=trainingDF, batch_size=batch_size, epochs=max_epoch, \
   feature_cols=['features'], label_cols=['labels'], ...)
```



https://github.com/Mastercard/udap-analytic-zoo-examples

お客様事例

事例：SK Telecom様 ~BigDLを適用してE2E処理時間を短縮~



事例：SK Telecom様

新旧アーキテクチャの比較 – 推論処理の性能

- 7,722,912 レコード = 80,447 cell towers X 8 日 X 12 レコード (5分おきの1時間分)
- 1レコードあたり8個のネットワーク品質指標

	旧（Pandas + TF）	新（Spark + Analytics Zoo）	
データExport	2.3s	N/A	
前処理	71.96s	ローカル 2.56s	3 ノード Yarn 1.43s
ディープラーニング推論	1.06s (CPU) / 0.63s (GPU)	ローカル 0.68s	3 ノード Yarn 0.18s

45 x faster

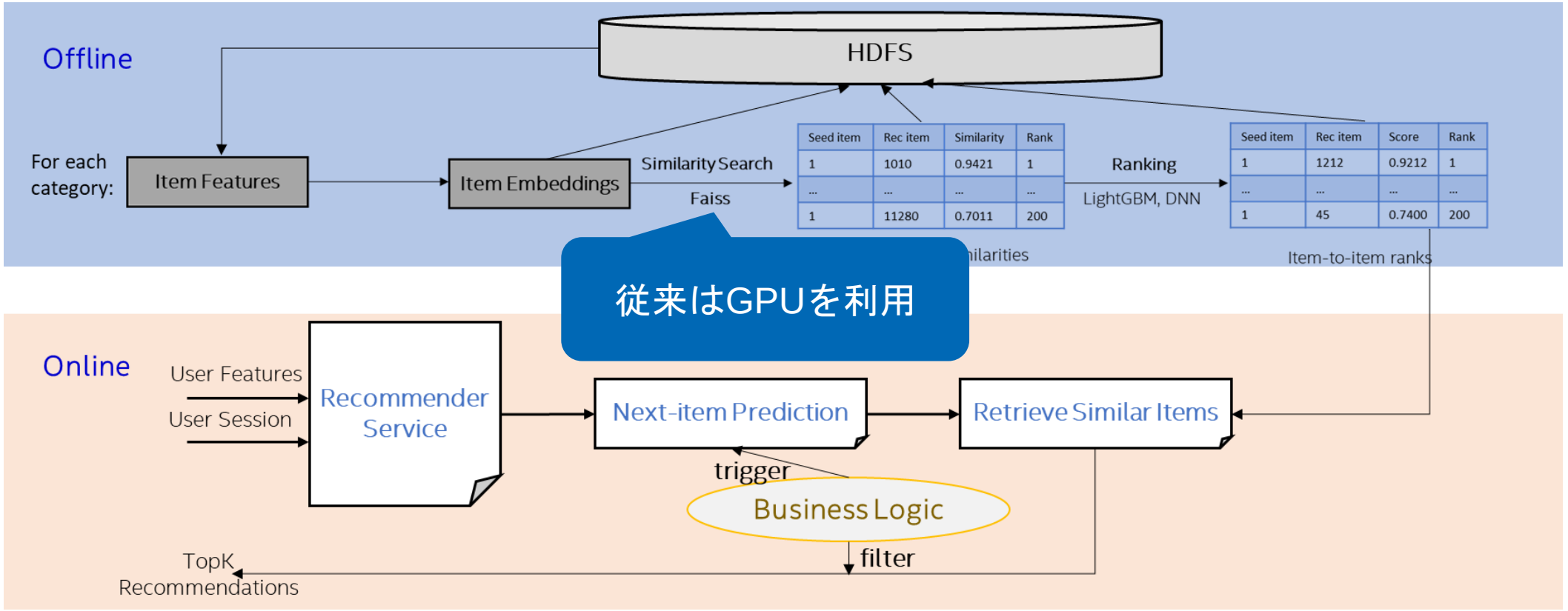
※ CPU：Intel(R) Xeon(R) Gold 6240 CPU @ 2.60GHz
※ GPU：Nvidia-K80

単一サーバーでの性能

- データと計算は50パーティションに分散
- 前処理と推論処理は単一のSparkジョブにて実行

事例：Yahoo! JAPAN様
Yahoo!ショッピングにおける商品検索性能の最適化

Yahoo!ショッピングの商品レコメンデーション・システム・アーキテクチャー

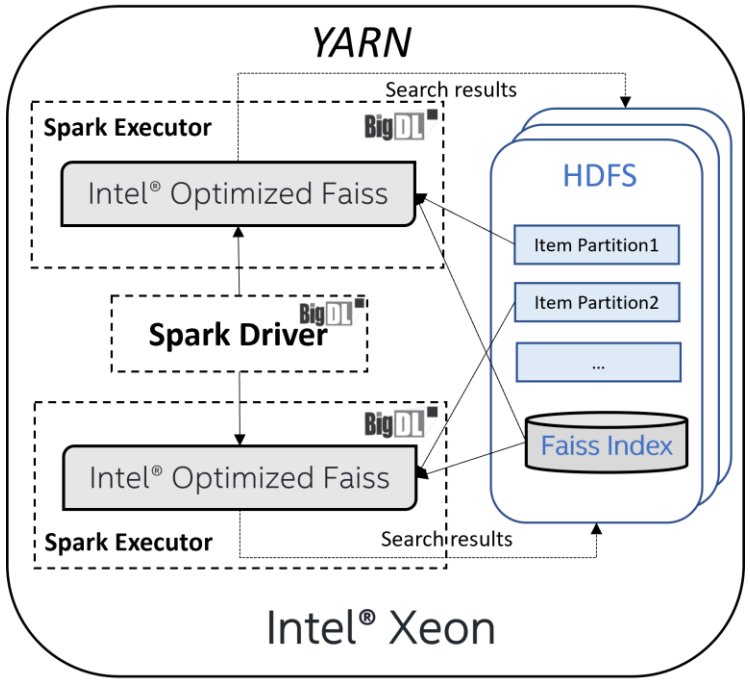


https://www.intel.com/content/www/us/en/developer/articles/technical/offline-item-search-with-bigdl-at-yahoo-japan.html

事例：Yahoo! JAPAN様
Yahoo!ショッピングにおける商品検索性能の最適化

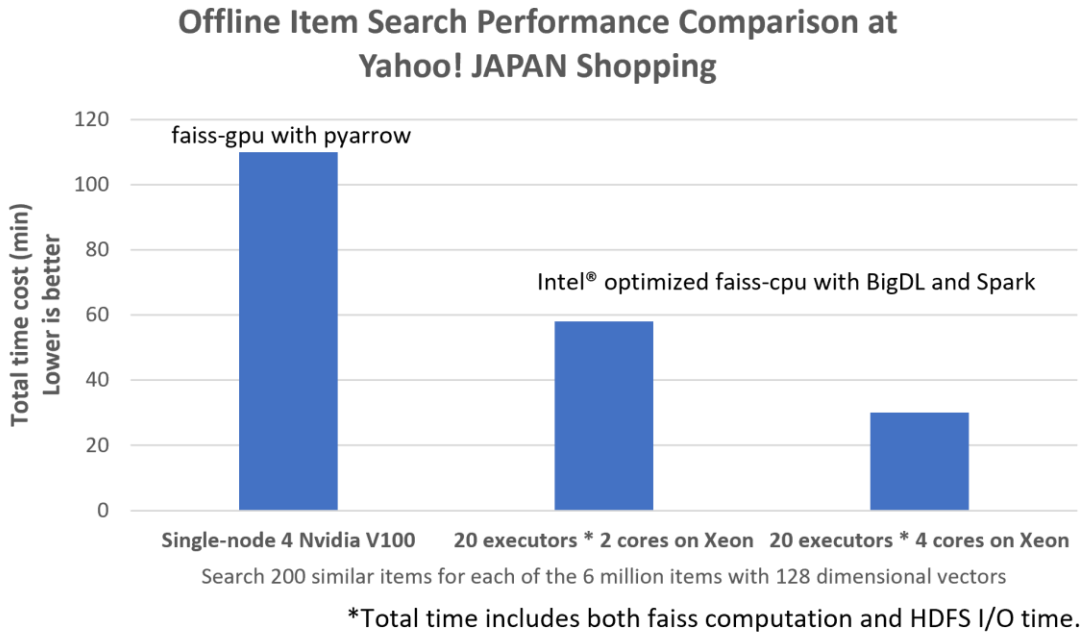
インテルによるFaissのSpark対応

商品サーチ(ベクトルサーチ)で使用される
FaissをSpark上で稼働するように最適化



結果

NVIDIA V100を用いる場合に比べて
最大で3.5倍の性能向上を実現



https://www.intel.com/content/www/us/en/developer/articles/technical/offline-item-search-with-bigdl-at-yahoo-japan.html


Try your Big Data on IA!
intel


intel