



IoT時代に向けたビッグデータ解析

- インテルIoT・ビッグデータ戦略

田口栄治

インテル株式会社 データセンター事業開発部

シニア・スペシャリスト

2014年7月



議題

1. IoTのビジネス機会
2. IoTからの価値の創出への壁
3. インテルのIoTソリューション向けビルディング・ブロック
4. インテルのApache Hadoopを中心としたビッグ・データ分析基盤への取り組み
5. インテル[®] アーキテクチャー機能を使ったAWSインスタンスでのビッグデータ負荷処理の最適化

IoT: ネットワークにつながるデバイスが急増



2014:
ネットワークに接続している
デバイスはデバイス全体の**15%**

20億
デバイス
2006年

150億
デバイス
2015年

500億
デバイス
2020年¹

2020年に向けて
急速に加速

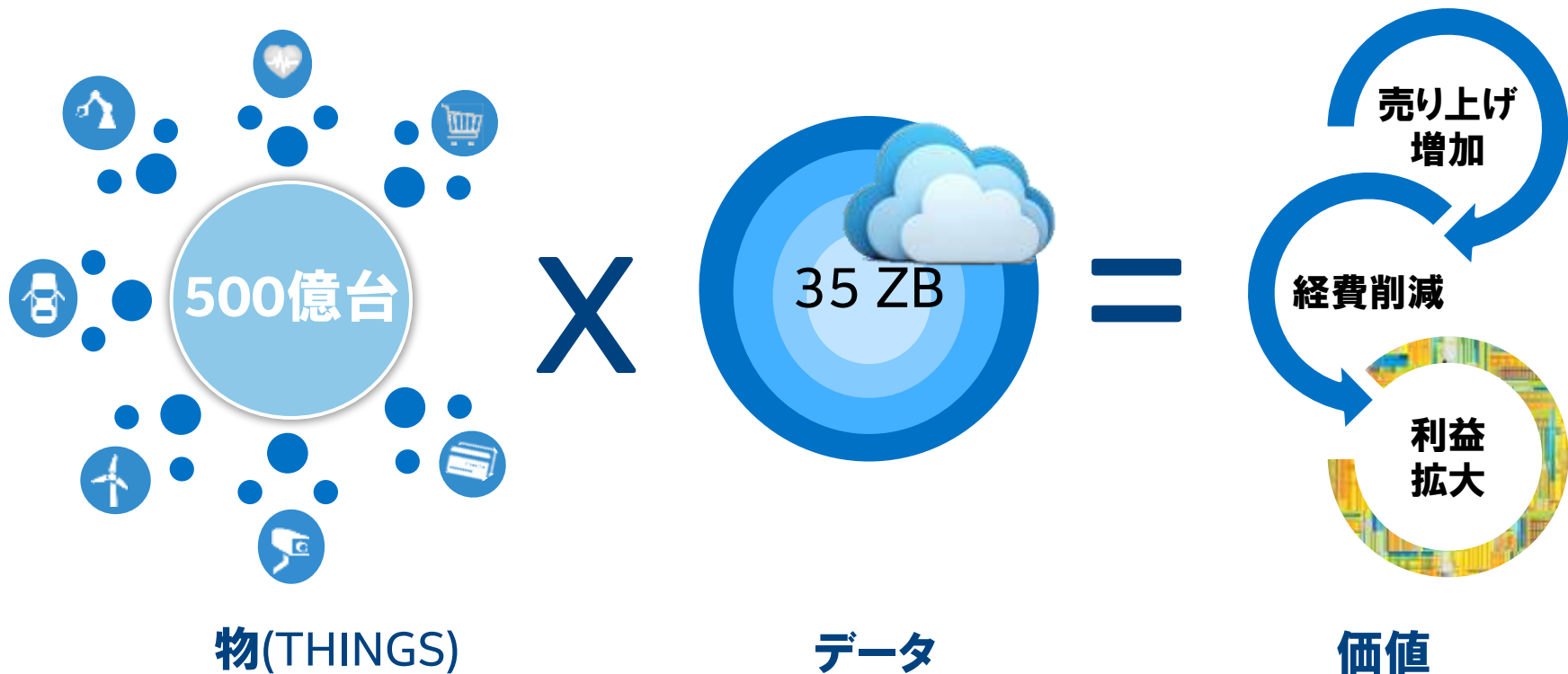
¹IDC*, Intel, United Nations

³McKinsey Global Institute*

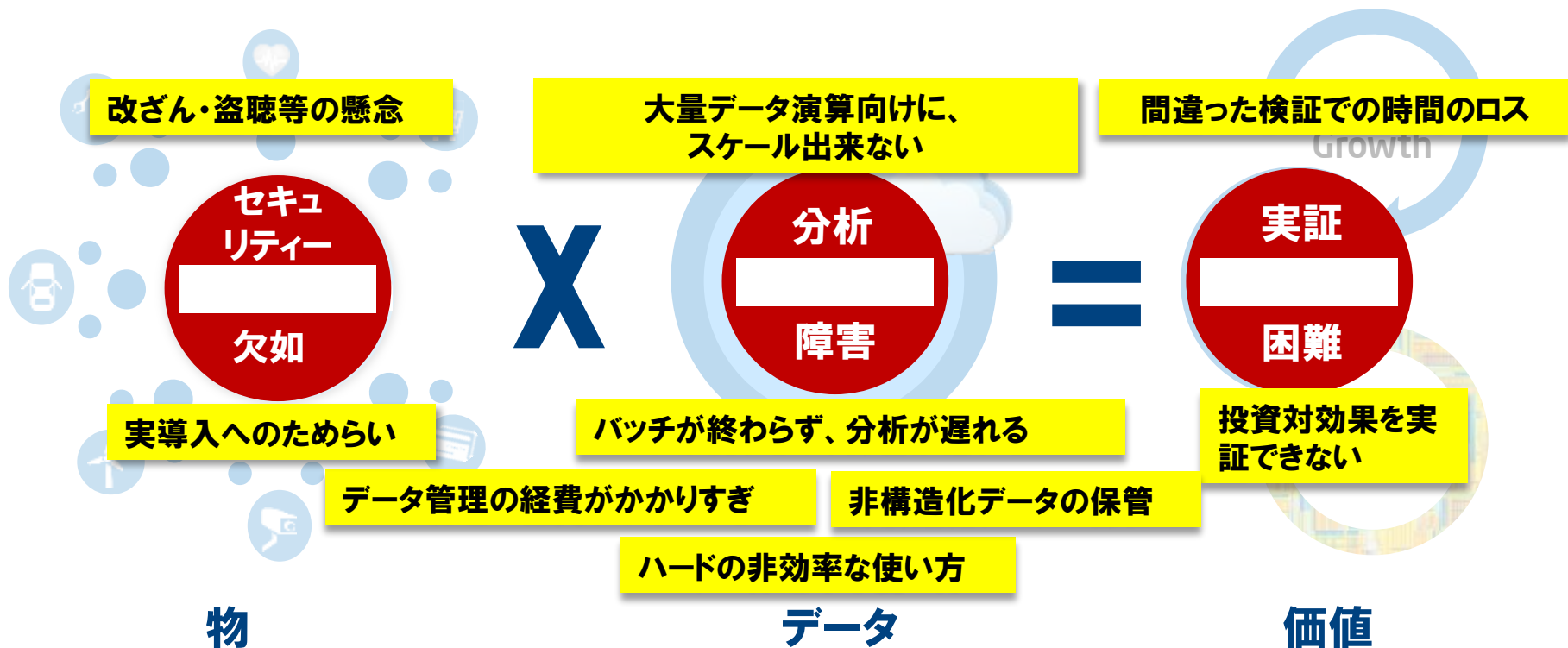
²IDC Digital Universe Study, Dec 2012

*Other names and brands may be claimed as the property of others.

大きなビジネス機会：データから価値の創出

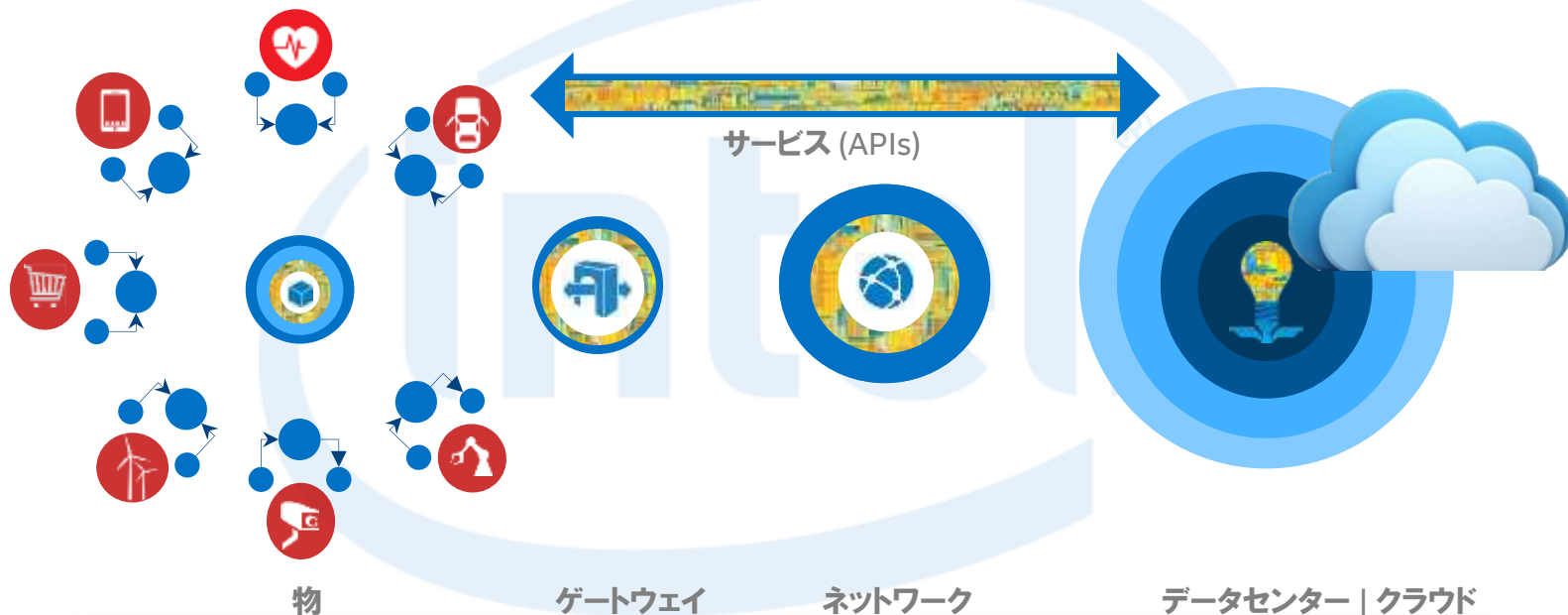


大きな課題: 実現へ向けた障害



インテルによる課題の解決: IoTに安心とインテリジェンスを提供

接続、保護、管理、多数のデバイスからのデータの分析



信頼できる基盤によるインテリジェントなソリューションで

インテルの価値: IoTソリューション展開の加速

分析: IoT向けの革新的なビッグデータ分析が可能のように、エコシステムを形成

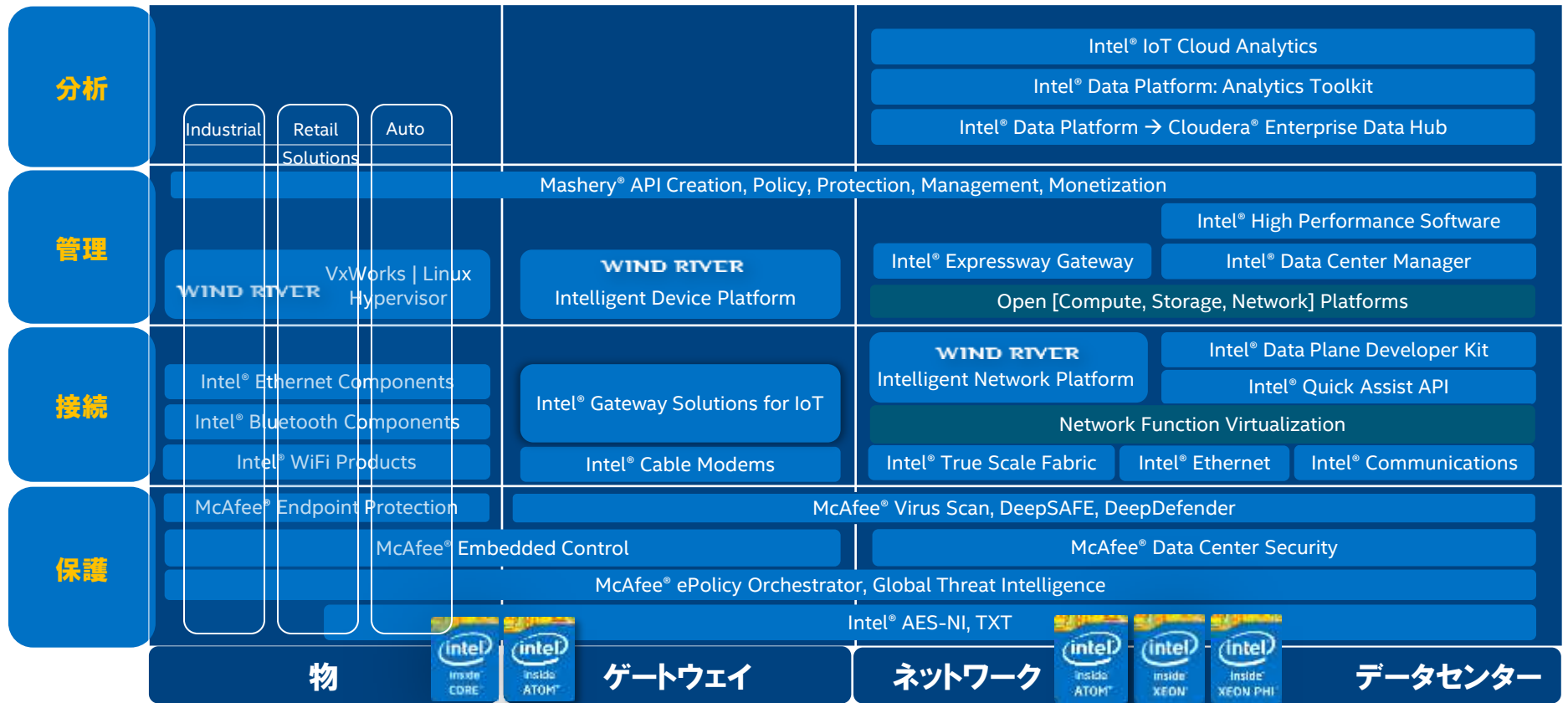
管理: 企業に対して、俊敏性を持って、デバイス、ネットワーク、システムを管理可能にする

接続: 物のインターネット(IoT)機器と既存・新規のシステムとの接続

保護: サービス、ソフトウェア、ハードアシスト保護機能によりデータの保護



安心・インテリジェントなIoTソリューションを構成する インテルのビルディング・ブロック



IoT時代のビッグデータ分析を加速



リファレンス・アーキテクチャーと
ソリューション事例を作成し、拡販の
推進

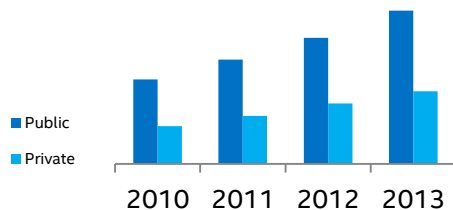
量産効果の高い基盤環境での、
先進分析基盤の活用を推進

オープンで信頼できるプラットフォームとして
のApache Hadoopの進化と適用を推進

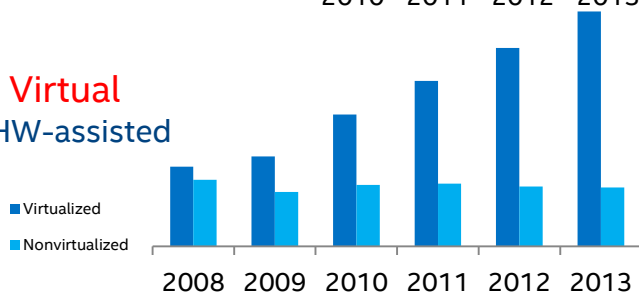
オープンで標準的なビルディング・
ブロックを使ったSoftware Defined
Infrastructure のエコシステムと
業界の発展を支援

データセンターの変革点概要

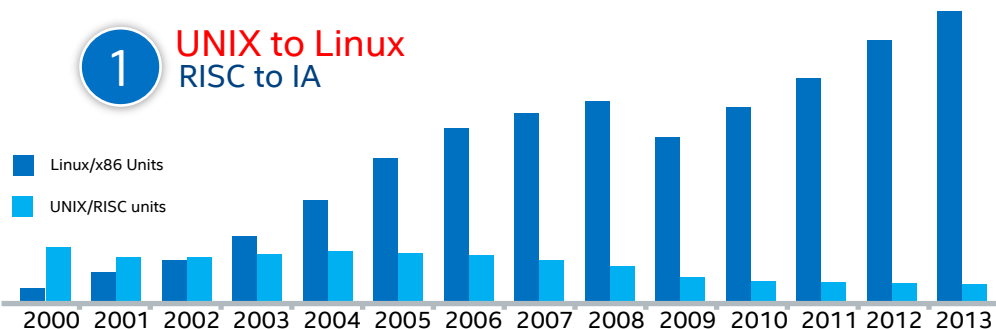
3 Cluster to Cloud ASIC to IA/Fabric



2 Physical to Virtual SW-only to HW-assisted



1 UNIX to Linux RISC to IA



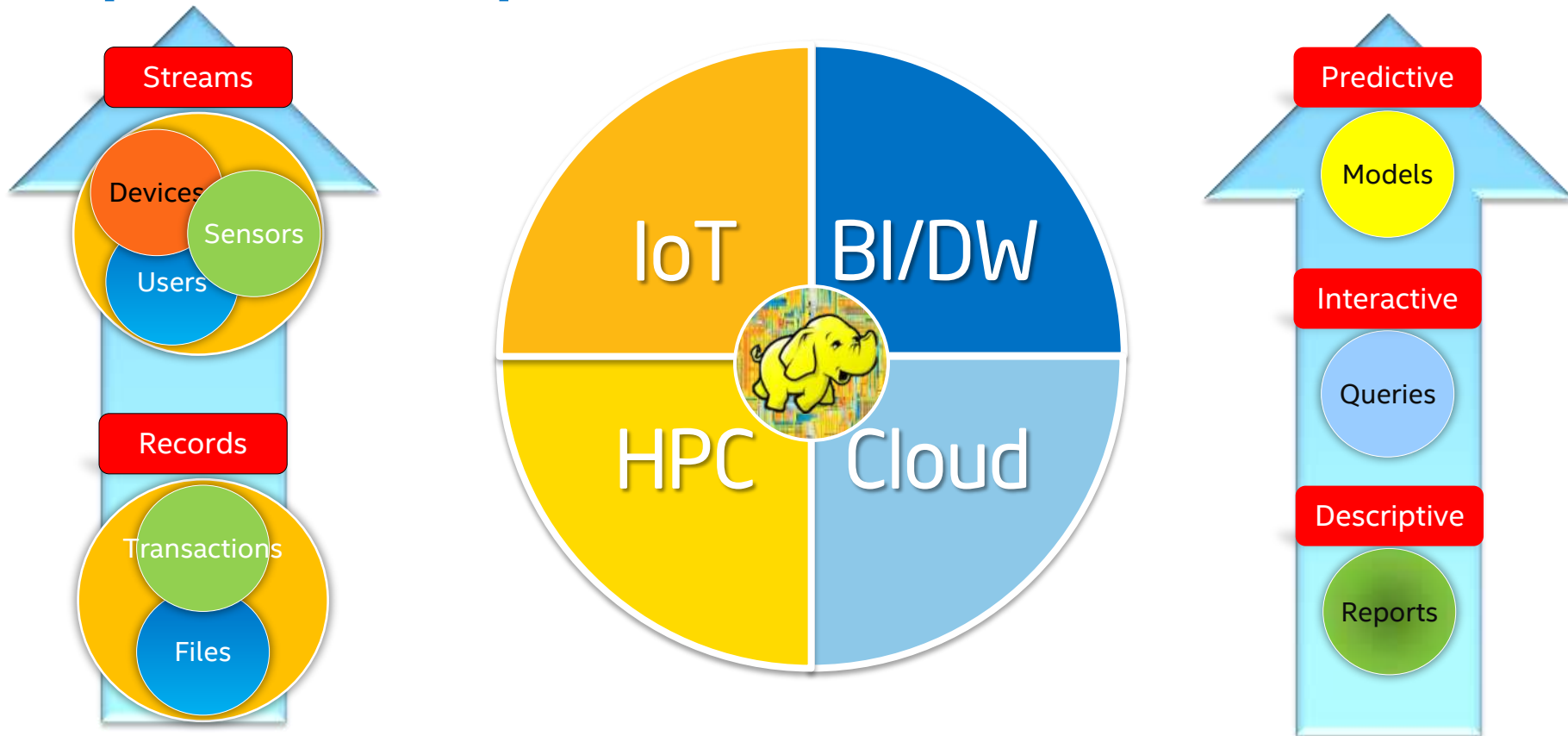
4 Big Data

2000年、インテルはLinuxの時代を予測、Red Hatに投資、2005年 仮想化時代を見据えVMWareに投資、2008年ハイパー・スケール・コンピューティングへ投資開始

ビッグデータとHadoopによる
変革はこれらすべてを凌駕する
(ほどの影響がある)

Diane Bryant, SVP & GM
Data Center Group, Intel

Apache Hadoop変革の加速



課題

IoT

マシン生成データの活用

ソースに近い場所での集約・分析

ヒストリカル情報のデータセンターへの記録・保管

パターンの認識と結果予測

デバイスへ“知識”の配布

Cloud

デバイスやユーザー生成データの投入

公共と個人のデータ・セットの結合

クラウド資源を使った分析サービスの活用

インタラクティブで、バッチ的やリアルタイム的な処理

BI & DW

ユーザー生成やトランザクショナルなデータの投入

増大するデータの従来型のDWHへの移動

増大する企業データの有意義な活用

ビッグ・データによってもたらされる価値創出のBI技術

より高速で、より経済的な分析

HPC

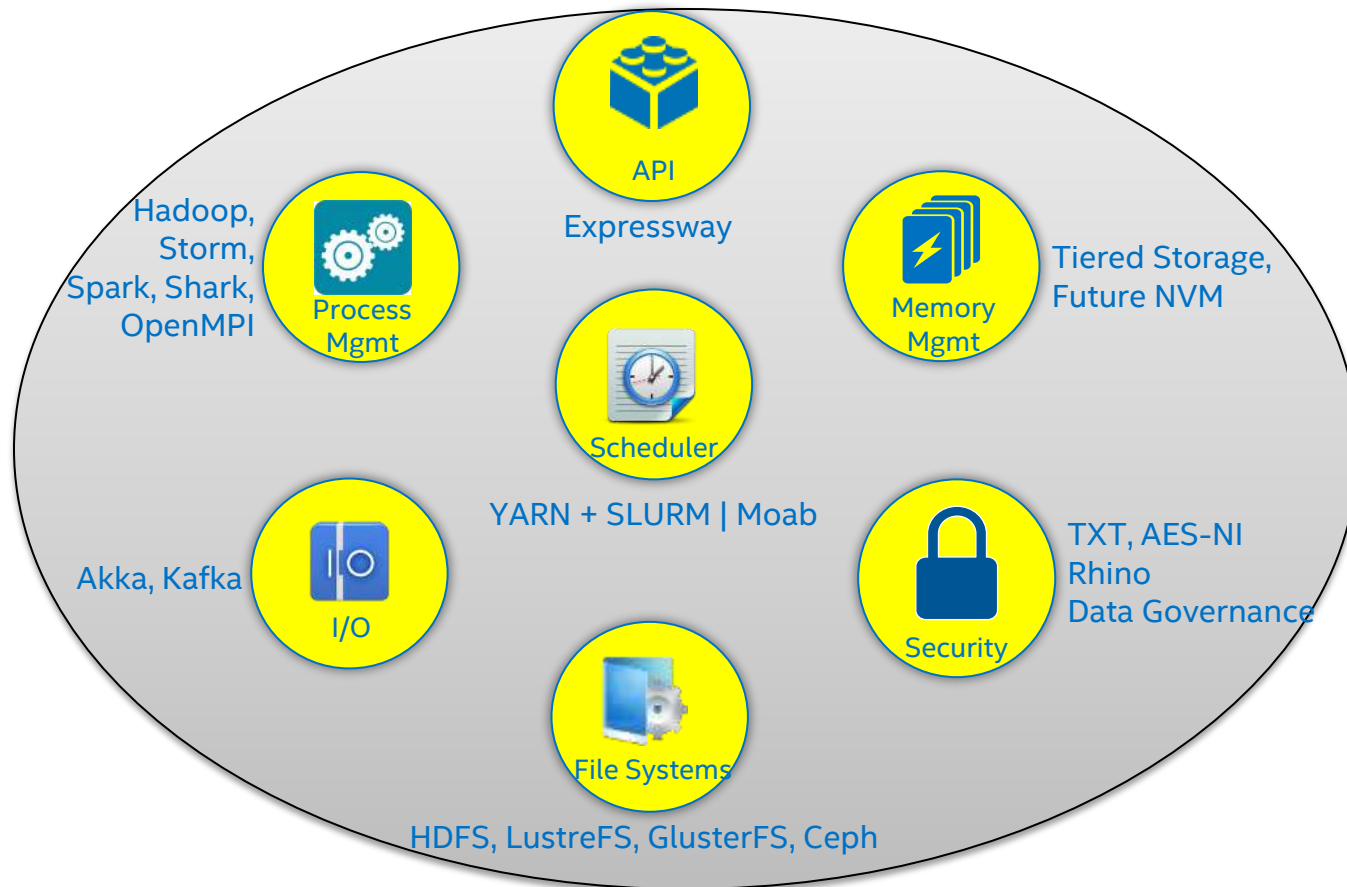
HPCクラスター上に残るデータの処理

HPCデータセットでのMapReduceやGraph-parallelコンピューティングの活用

慣れ親しんだスケジューラーでの処理の管理

ビッグ・データ分析とモデリング・シミュレーションとの統合

Hadoopをビッグ・データOS的に変革



Apache Hadoopへのインテルの貢献

Intel® Distribution for Apache Hadoop software活動を通じて



- HDFS暗号化
- AES-NI活用により暗号化を20倍高速化
- Hbaseにおける透過的な暗号化
- Hbaseでのセルレベルでの暗号化とACL
- SSLによる、ノード間通信の保護
- Zookeeperのログの暗号化



- マルチ・バッファ暗号化の最適化
- SSE 4.2での圧縮とCRC
 - iGZipのネイティブな圧縮
- MapReduce 用Hashベースシャッフル



- 構築の単純化
- Intel® Active Tunerによる構成自動化
- 管理強化とSNMP/SMTTPアラート強化
- Hadoop jobsのプロファイリングの強化
- Hbaseスキーマに単純化
- 運用管理用REST API整備



- LustreFSのサポート
- GlusterFSのサポート
- Graph Analytics



- HDFS HA v2の改善
- Hbaseスナップショット

大ニュース: Cloudera + インテル アライアンス

Clouderaへインテルが\$740Mの投資

- インテルとして最大のデータセンタ関連企業への投資、インテルのビッグデータへの括弧足す意思の現れ
- Clouderaの独立性を担保

インテルとClouderaによるオープン・ソース革新の推進

- 基礎技術への注力により、Hadoopの成長を加速
- Hadoop基盤とそのエコシステムにおける開発者の技術革新を推進

CDHのインテル® アーキテクチャーへの最適化

- Clouderaによるインテルのデータセンター技術の最適な活用を支援
- データセンター基盤の提供により、拡張性を担保した開発を支援

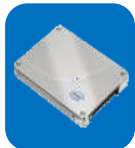
1つのオープン・ソース・プラットフォームへの統合



- 最も安定した、汎用性の高い、成熟したHadoopディストリビューション
- SQL機能と性能のリーダー (Impala)
- 豊かな運用機能と管理機能
- 150のHadoop開発者
- 100のオープンソース貢献者
- シリコンレベルから拡張されたセキュリティーと性能を引き出す唯一のHadoop
- 監査、アクセス制限、暗号化などに対応した最先端小セキュリティー機能
- 50のHadoop開発者、12の貢献者
- Linux, KVM, Xen, Java, OpenStack, Hadoop 等オープンソース活動への1000人を越える貢献者と長年の実績

複雑性を回避、強固なセキュリティー、高速分析

cloudera®
Ask Bigger Questions



オープンソースによる改革の加速

- ビッグデータ用水平分業プラットフォームのオープン化の維
- Apache Hadoopと関連プロジェクトの加速

CDHのIAへの最適化

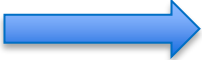
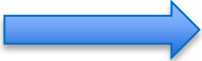
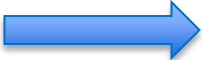
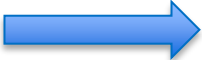
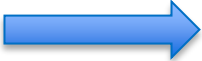
- サーバー、ストレージ、ネットワーク基盤を俯瞰した最適化
- 拡張されたハード機能による基盤保護の強化

ビッグデータのエコシステムの成長支援

- 業界標準のベンチマークや活用方法の提供
- リファレンス・アーキテクチャーや業界全体にわたるソリューションの開発

オープンソースの革新の加速

Apache Hadoop基盤の進化を、活動の分散化を避けながら推進

SQL	Project Gryphon Impala		Impala
ストリーミング	Apache Storm Apache Spark Streaming		Spark Streaming
性能	Apache Tez Apache Spark		Spark
セキュリティー	Project Rhino Apache Sentry		Project Rhino
ストレージ	Apache HDFS Apache HBase		Accelerated investment in both

CDHをインテル® アーキテクチャーへ最適化

シリコンとソフトウェアの進化による劇的な効果

1

演算主体の処理負荷をシリコン(HW)へ

2

最大20倍のメインメモリー活用の増加

3

ラック・スケール・アーキテクチャーを見据えた設計

暗号化
圧縮
計算

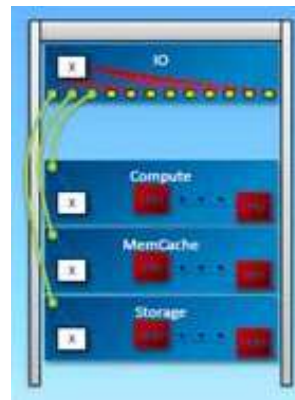
(AES-NI)
(SSE 4.2)
(MKL)

Disk : Memory
レシオ改善

200:1



10:1



インテル® アーキテクチャー上でのApache Hadoopの性能向上

最大
50%高速

Compute



最大
80%高速

Storage & Memory



最大
50%高速

Network



10ノードクラスターにおける1TBのソートでの性能比較結果

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Source: Intel Internal testing

For more information go to intel.com/performance

インテル® アーキテクチャーによるApache Hadoopのセキュリティー向上

AES-NI

- AES暗号化をプロセッサ支援で高速化
- 暗号化ソフトの高速化・堅牢化を可能に

Data in Motion

Secure transactions used pervasively in ecommerce, banking, etc.

Data at Rest

Full disk encryption software protects data while saving to disk

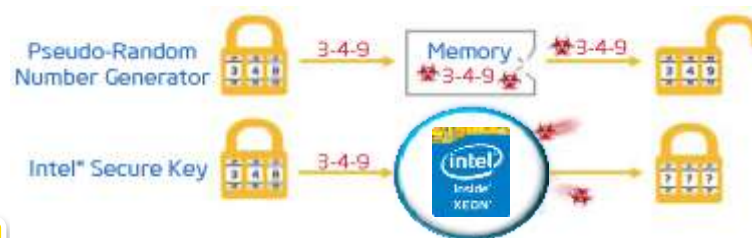
Data in Process

Most enterprise and cloud applications offer encryption options to secure information and protect confidentiality

AES-NI - Advanced Encryption Standard New Instructions

セキュアキー(DRNG)

- プロセッサによる真の乱数発生
- より堅牢、コンプライアンス対応、より高性能



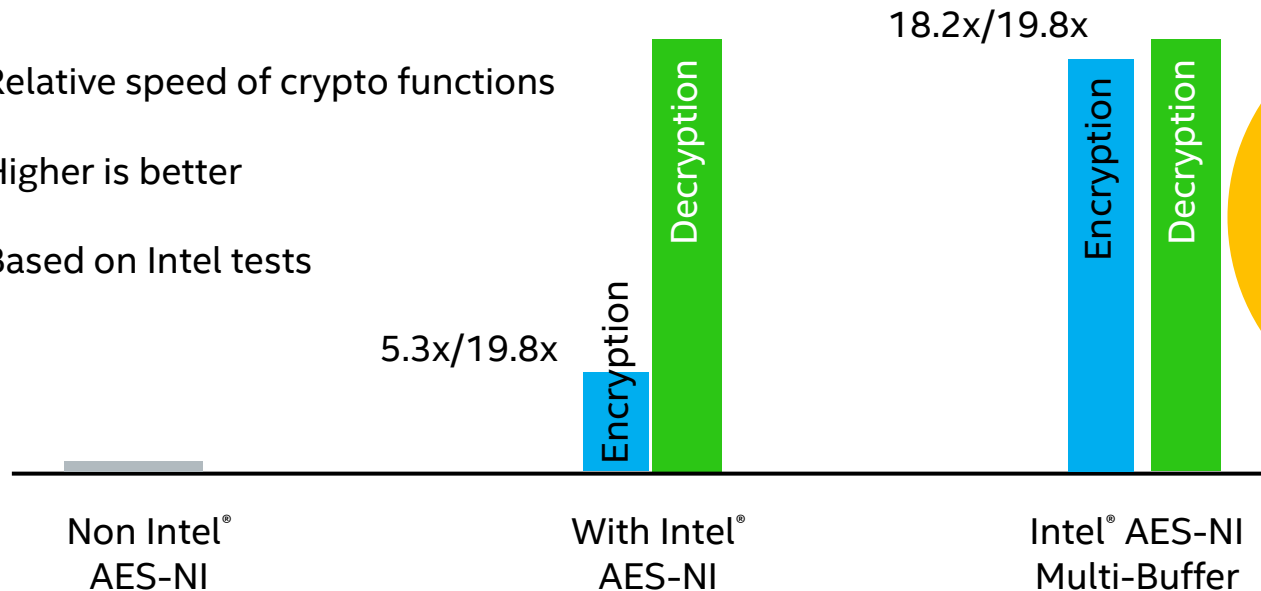
Secure Key - previously known as Intel Digital Random Number Generator (DRNG)

インテル® AES-NIによる暗号化の加速

Relative speed of crypto functions

Higher is better

Based on Intel tests



約20倍
の
高速化

AES-NI - Advanced Encryption Standard New Instructions

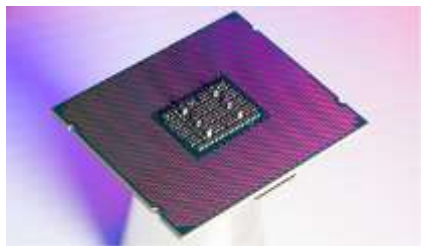
ビッグデータ向けのより安全なクラウド基盤

AES-NI搭載の インテル® Xeon® プロセッサ

インテル® Xeon® プロセッサ E7 ファミリー

インテル® Xeon® プロセッサ E5 ファミリー

インテル® Xeon® プロセッサ E3 ファミリー



AES-NIをサポートする アマゾン EC2インスタンス

EC2ハードウェア専用インスタンス等

アマゾン ウェブ サービス

Amazon EC2

ご利用開始にあたって

製品の詳細

インスタンスタイプ

料金表

旧世代のインスタンス

購入オプション

Amazon EC2 スポットインスタンス

Amazon EC2 リザーブドインスタンス

Amazon EC2 ハードウェア専用インスタンス

開発者用リソース

よくある質問

ハードウェア専用インスタンスとは、単一のカスタマー専用のシングルテナントハードウェア上で実行される Amazon EC2 インスタンスです。ハードウェア専用インスタンスは、企業の方針や規制傾向により、その他のカスタマーに属するインスタンスからホストハードウェアレベルで物理的に分離された EC2 インスタンスを必要とする作業負荷に適しています。ハードウェア専用インスタンスを使用すれば、Amazon EC2 コンピューティングインスタンスをハードウェアレベルで物理的に分離しながら、オンデマンドの伸縮自在なプロビジョニング、従量課金制といった、AWS クラウドの利点を十分に活用いただけます。

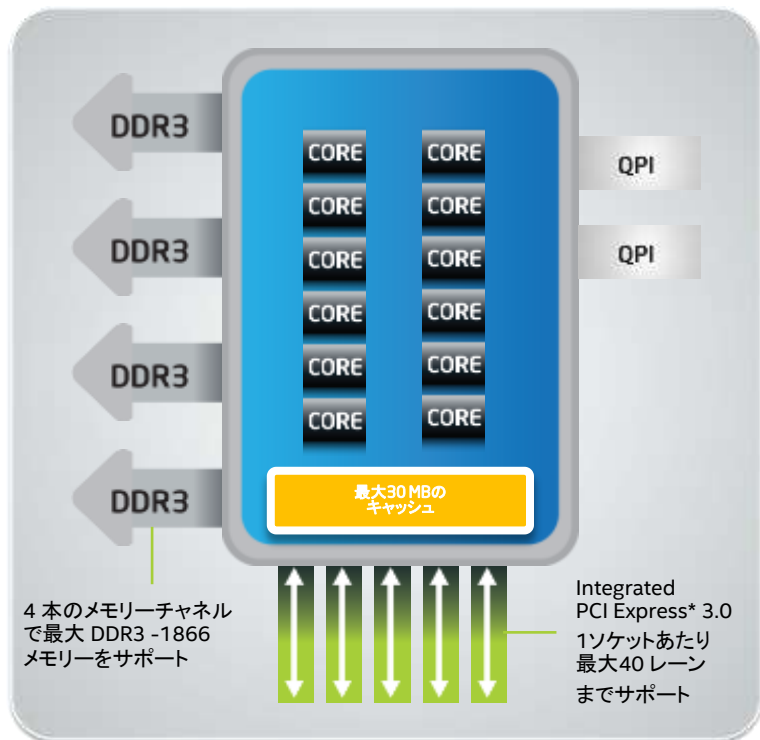
Intel プロセッサの機能

Amazon EC2 インスタンスには、下記を含む Intel プロセッサの機能が搭載されています。

- **Intel® AES New Instructions (AES-NI)** : Intel AES-NI は、Advanced Encryption Standard (AES) アルゴリズムを改良し、より高速のデータ処理、より優れたセキュリティを提供する専用ハードウェア機能です。
- **Intel® Advanced Vector Extensions (Intel® AVX)** : Intel AVX は、256 ビットの命令セット拡張で、浮動小数点 (FP) を多倍するアプリケーションで使用します。画像およびオーディオ/ビデオ処理、科学シミュレーション、財務分析、3D モデリングや分析などのアプリケーションのパフォーマンスを向上します。
- **Intel® Turbo Boost Technology**: Intel Turbo Boost Technology は、必要に応じてより高いパフォーマンスを提供します。些小的に定率の動作周波数よりも高速でコアを動作させ、処理速度を高速化します。



インテル® Xeon® プロセッサー E5-2600 v2 製品ファミリー



エネルギー効率の向上

- 同じ消費電力で最大**40%**の性能向上を実現¹

更なる性能向上

- コア数とキャッシュサイズを**50%**増加 (最大12コア、30MB)
- 高速DDR3メモリーをサポート(最大**1866 MTS**)

データセンター・クラスの仕様

- 最大**1.5 TB**のメモリー容量をサポート(2ソケットシステム)
- iPCIe 3.0の改善で10GbEを最大限活用
- 仮想化とセキュリティの技術革新

ワークロードの最適化

- インテル® AVX
- インテル® Xeon Phi™ にI/Oを最適化
- ビッグデータ処理に最適

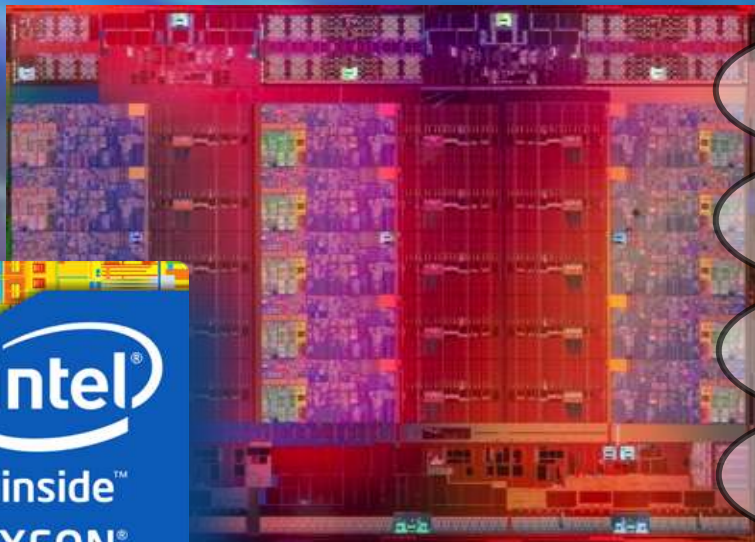
インテル® アーキテクチャー ソフトウェアとの互換性

- 既存のSWエコシステムの活用

¹パフォーマンス・テストに使用したソフトウェアおよび負荷は、インテルのマイクロプロセッサーのみに対してパフォーマンスが最適化されていた可能性があります。SYSmark、MobileMarkのようなパフォーマンス・テストは、特定のコンピューター・システム、コンポーネント、ソフトウェア、オペレーション、および機能を用いて測定されます。これらの要素のいずれかに何らかの変更を行うと、結果が変わる可能性もあります。ご購入を検討される際には、他の製品と組み合わせた場合の製品パフォーマンスなどを含め、他の情報やパフォーマンス・テストも参考にしてください。詳細については、次のページをご覧ください。<http://www.intel.com/performance>

インテル® Xeon® プロセッサー

E7-8800/4800/2800 v2 製品ファミリー



2x

データプロセッシング
性能

3x

リアルタイム・アナリティクスを
支える 大容量メモリー

Up to 6TB
in 4-Socket

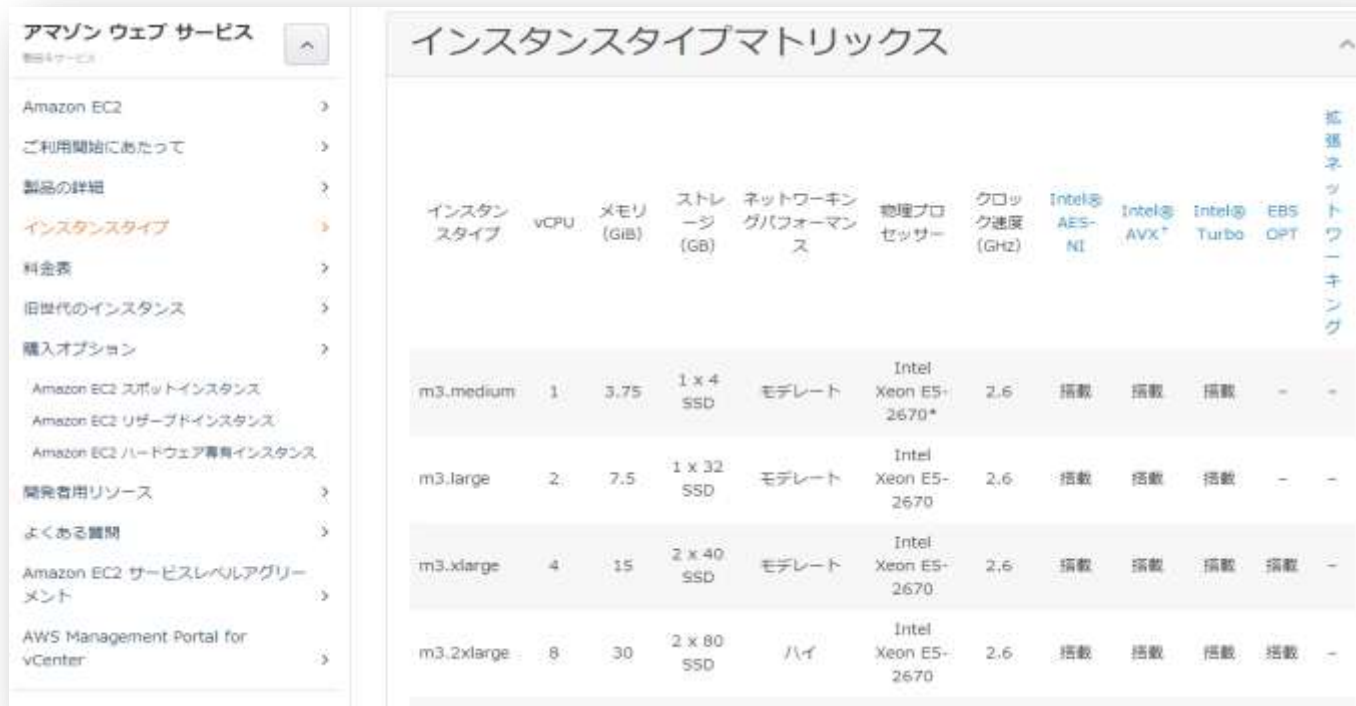
4x

応答性と拡張性を支える
革新的なI/O

5

高可用性(99.999%)を支える
インテル® Run Sure テクノロジー

AES-NI 活用可能なAmazon EC2インスタンス例



The screenshot shows the Amazon EC2 Instance Type Matrix. The left sidebar contains navigation links for Amazon EC2, including 'インスタンスタイプ' (Instance Types). The main content area is titled 'インスタンスタイプマトリックス' (Instance Type Matrix) and displays a table of instance types with their specifications.

インスタンスタイプ	vCPU	メモリ (GiB)	ストレージ (GB)	ネットワークパフォーマンス	物理プロセッサ	クロック速度 (GHz)	Intel® AES-NI	Intel® AVX*	Intel® Turbo	EBS OPT	拡張ネットワーク
m3.medium	1	3.75	1 x 4 SSD	モデレート	Intel Xeon E5-2670*	2.6	搭載	搭載	搭載	-	-
m3.large	2	7.5	1 x 32 SSD	モデレート	Intel Xeon E5-2670	2.6	搭載	搭載	搭載	-	-
m3.xlarge	4	15	2 x 40 SSD	モデレート	Intel Xeon E5-2670	2.6	搭載	搭載	搭載	搭載	-
m3.2xlarge	8	30	2 x 80 SSD	ハイ	Intel Xeon E5-2670	2.6	搭載	搭載	搭載	搭載	-

多くのインスタンスが選択可能 : aws.amazon.com/jp/ec2/instance-types

まとめ

1. IoTは、今後のビジネス成長に対して、予測不可能なほどの大きな可能性を提供
2. しかしながら、IoTからの価値の創出には、その膨大なデータの分析と保全に関して、大きな課題がある
3. インテルは、IoTソリューションを実現するために、インテル® アーキテクチャーに最適化された様々なビルディング・ブロックを提供
4. ビッグ・データ分析では、インテルはClouderaと共に、Apache Hadoopをより高性能に、よりセキュアに、処理使いやすくなるための変革を推進
5. 開発者・利用者が、ビッグ・データ分析を、より簡便でより効率よく活用できるよう、AWSと共に、インテル® アーキテクチャー機能を使ったAWSインスタンスでのビッグデータ負荷処理の最適化を推進

詳細

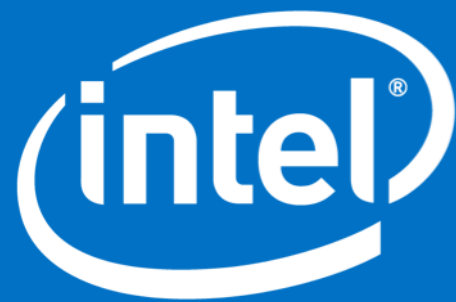
www.intel.co.jp/content/www/jp/ja/homepage.html

www.intel.co.jp/content/www/jp/ja/big-data/big-data-analytics-turning-big-data-into-intelligence.html

www.intel.com/bigdata

www.github.com/intel-hadoop/project-rhino/

www.aws.amazon.com/jp/ec2/instance-types/



法律的な免責条項

本資料に掲載されている情報は、インテル製品の概要説明を目的としたものです。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らざるにかかわらず、いかなる知的財産権のライセンスを許諾するものではありません。製品に付属の売買契約書『Intel's Terms and conditions of Sales』に規定されている場合を除き、インテルはいかなる責任を負うものではなく、またインテル製品の販売や使用に関する明示または黙示の保証（特定目的への適合性、商品適格性、あらゆる特許権、著作権、その他知的財産権の非侵害性への保証を含む）に関してもいかなる責任も負いません。

“ミッション・クリティカルなアプリケーション”とは、インテル製品の欠陥や故障によって、直接的または間接的に、人身事故または死亡事故が発生する可能性があるアプリケーションです。そのようなミッション・クリティカルなアプリケーションのためにインテル製品を購入または使用する場合、ユーザーは、インテル製品またはその部品の設計、製造、または警告におけるインテルまたはその下請会社の過失の有無にかかわらず、そのようなミッション・クリティカルなアプリケーションによって生じるあらゆる製造物責任賠償の請求、人身事故、または死亡事故に直接的または間接的に起因する保険金支払費用、損害、経費、および妥当な弁護士費用のすべてを補償し、インテルとその子会社、下請会社、関連会社、およびそれぞれの役員、職員、従業員に何らの損害も与えないものとします。

インテル製品は、予告なく仕様や説明が変更されることがあります。機能または命令の一覧で「留保」または「未定義」と記されているものがありますが、その「機能が存在しない」あるいは「性質が留保付である」という状態を設計の前提にしないでください。これらの項目は、インテルが将来のために留保しているものです。インテルが将来これらの項目を定義したことにより、衝突が生じたり互換性が失われたりしても、インテルは一切責任を負いません。この情報だけに基づいて設計を最終的なものとししないでください。

本書で説明されている製品には、エラッタと呼ばれる設計上の不具合が含まれている可能性があり、公表されている仕様とは異なる動作をする場合があります。そのようなエラッタは、インテルの保証範囲外です。現在確認済みのエラッタについては、インテルまでお問い合わせください。

法律的な免責条項

性能に関するテストに使用されるソフトウェアとワークロードは、性能がインテル® マイクロプロセッサ用に最適化されていることがあります。SYSmark* や MobileMark* などの性能テストは、特定のコンピューター・システム、コンポーネント、ソフトウェア、操作、機能に基づいて行ったものです。結果はこれらの要因によって異なります。製品の購入を検討される場合は、他の製品と組み合わせた場合の本製品の性能など、ほかの情報や性能テストも参考にして、パフォーマンスを総合的に評価することをお勧めします。

インテルは、本資料で参照しているサードパーティーのベンチマークまたは Web サイトの設計や実装について管理や監査を行っていません。本資料で参照している Web サイトまたは類似の性能ベンチマーク・データが報告されているほかの Web サイトも参照して、本資料で参照しているベンチマーク・データが購入可能なシステムの性能を正確に表しているかを確認されるようお勧めします。

各ベンチマークの相対パフォーマンスは、最初にテストしたプラットフォームの実際のベンチマーク結果をベースラインとして、1.0 の値を割り当てることによって計算されます。ベースラインとなるプラットフォーム以外のテスト対象プラットフォームの相対パフォーマンスは、各プラットフォームのベンチマークの結果を、ベースラインとなるプラットフォームの実際のベンチマーク結果で割り、報告されたパフォーマンスの向上に比例する相対パフォーマンスの数値を割り当てることによって計算しています。

SPEC、SPECint、SPECfp、SPECrate、SPECpower、SPECjobb は、Standard Performance Evaluation Corporation の商標です。詳細については、<http://www.spec.org/spec/trademarks.html> (英語) を参照してください。

インテル® ハイパースレッディング・テクノロジー (インテル® HT テクノロジー): 一部のインテル® Core™ プロセッサ・ファミリーで利用できます。インテル® ハイパースレッディング・テクノロジーに対応したシステムが必要です。詳細については、各 PC メーカーにお問い合わせください。性能は、使用するハードウェアやソフトウェアによって異なります。詳細については、<http://www.intel.co.jp/content/www/jp/ja/architecture-and-technology/hyper-threading/hyper-threading-technology.html> を参照してください。

インテル® ターボ・ブースト・テクノロジー: インテル® ターボ・ブースト・テクノロジーに対応したシステムが必要です。インテル® ターボ・ブースト・テクノロジーおよびインテル® ターボ・ブースト・テクノロジー 2.0 は、一部のインテル® プロセッサでのみ利用可能です。各 PC メーカーにお問い合わせください。実際の性能はハードウェア、ソフトウェア、システム構成によって異なります。詳細については、<http://www.intel.co.jp/jp/technology/turboboost/> を参照してください。

インテル製品は、医療、救命、延命措置、重要な制御または安全システム、核施設などの目的に使用することを前提としたものではありません。インテル製品は、予告なく仕様変更されることがあります。本資料に記載されているすべての日付および製品は、計画以外の目的ではご利用になれません。

本プレゼンテーションに記載されたインテルの製品計画は、インテルの POR 製品ロードマップではありません。インテルの現在の POR 製品ロードマップをご希望の方は、インテルの担当者までお問い合わせください。

© 2014 Intel Corporation. 無断での引用、転載を禁じます。Intel、インテル、Intel ロゴ、Look Inside、ロゴ、Look Inside.、Intel Atom、Intel Atom Inside、Intel Core、Core Inside、Xeon、Xeon Inside は、アメリカ合衆国および / またはその他の国における Intel Corporation の商標です。

インテル製品は、予告なく仕様変更されることがあります。本資料に記載されているすべての日付および製品は、計画以外の目的ではご利用になれません。

* 第三者の社名、製品名などは、一般に各所有者の表示、商標または登録商標です。