

米国から始まったオープンソースソフトウェア第3の波 IoTとBigDataの活用を支える 「次世代データプラットフォーム」

2017年11月16日

ホートンワークスジャパン株式会社
執行役員社長
廣川裕司



議題

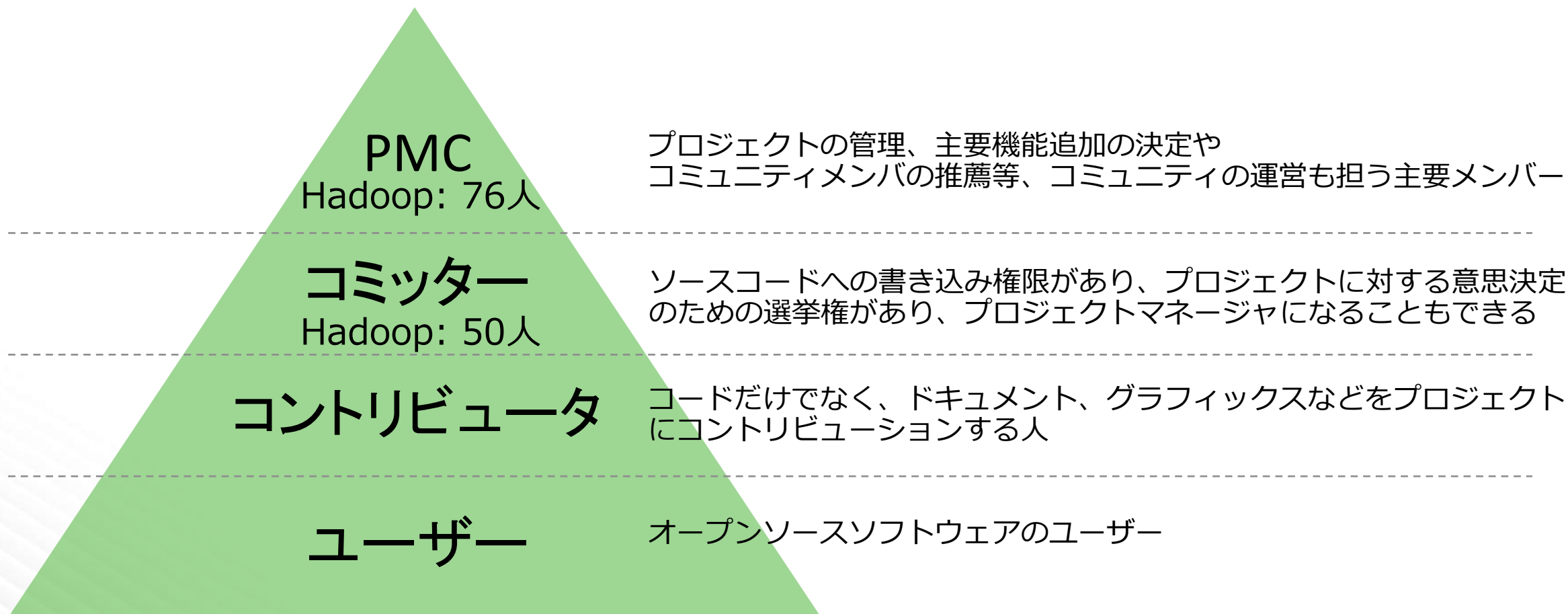
1. オープンソースソフトウェア (OSS) 動向
2. 事業成長を牽引する「データのトレンド」
3. OSS「次世代データプラットフォーム」
= Data Lake と Hortonworks の紹介 =
4. IoT と米国における IoT 活用状況

1. オープンソースソフトウェア (OSS)とは

- ソフトウェアのソースコードが無償で公開
(通常： ライセンス料は無償)
基本コンセプト「いいものはシェアしよう！ Collaboration」
- 誰でも
 - ◎ 開発に参加できる
 - ◎ 可変できる
 - ◎ 使用できる
 - ◎ 配布できる
- 有名なもの
 - 第1の波： LINUX OS (1991年 Linus Torvalds開発開始)
 - 第2の波： JBoss, PostgreSQL, MySQL (ミドルウェア・DB)
 - 第3の波： Hadoop, SPARK, NiFi, Kafka等 (ビッグデータ・IoT)

最大のオープンソースソフトウェア支援財団

Apache Software Foundation (ASF)



オープンソースソフトウェアのトレンド（革新イノベーション）

OPEN
COMMUNITY



THE
INNOVATION
ADVANTAGE

PROPRIETARY APPROACH

TIME

INNOVATION

Power of Open Source

- 1991年に Linux OSが登場して以来 OSの領域に始まり、ミドルウェアの領域まで OSS開発のパワーは独自開発を遥かに凌ぐまでに至る
- IoT・ビッグデータの領域でも先端の革新はオープンソースコミュニティで起きている

Value of Hortonworks

- ホートンワークスはIoT・ビッグデータ基盤をエンタープライズ企業でも安心安全に活用できるサポート・サービスを提供

Expanding Ecosystems

- 世界中で、日本中で官公庁・企業が活用できるエコシステムを急速に構築中
- 開発コミュニティのリードに加え、2100を超えるエコシステムパートナー、1200社を超える企業での活用実績がある

2. データのトレンド



44ZB

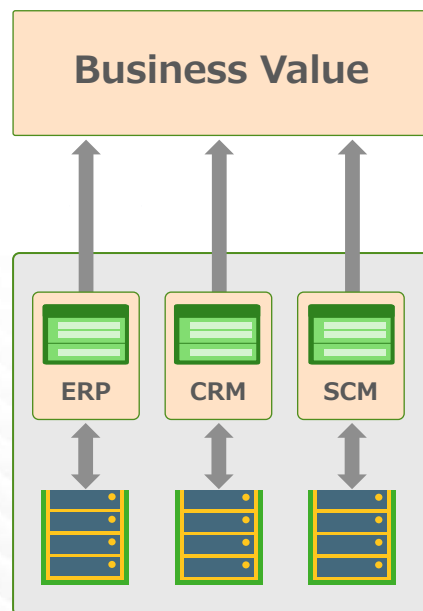
DATA

TOMORROW

次世代ビッグデータプラットフォームが直面する課題

1 Challenges

- アプリケーション毎にデータがサイロ化されている。
- 新しいデータの格納方法（非構造化・半構造化データ）
- 拡張時のコストバランス



新たに生成されるデータの 85%は
非構造型データ
(音声、写真、動画、センサデータなど)

2 IoTなどの発展によるデータの爆発的増加傾向

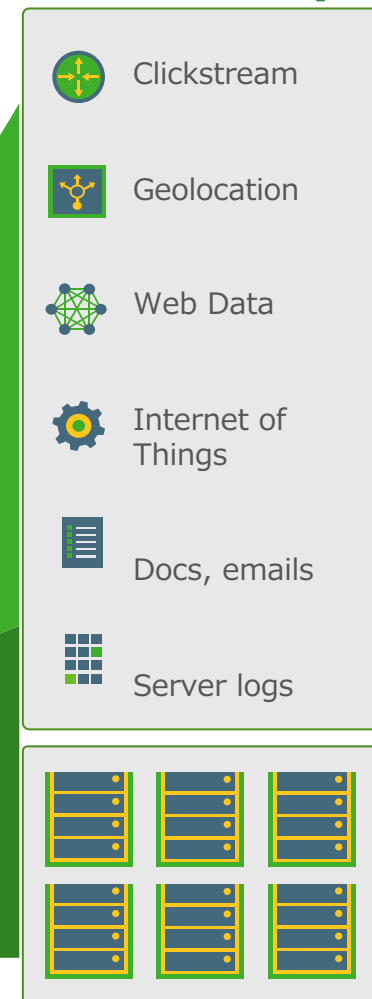
適切な技術・コストによる対応方法が求められている => OSSの活用が必須

2012
2.8 Zetta bytes

Traditional

New

2020
44 Zetta bytes



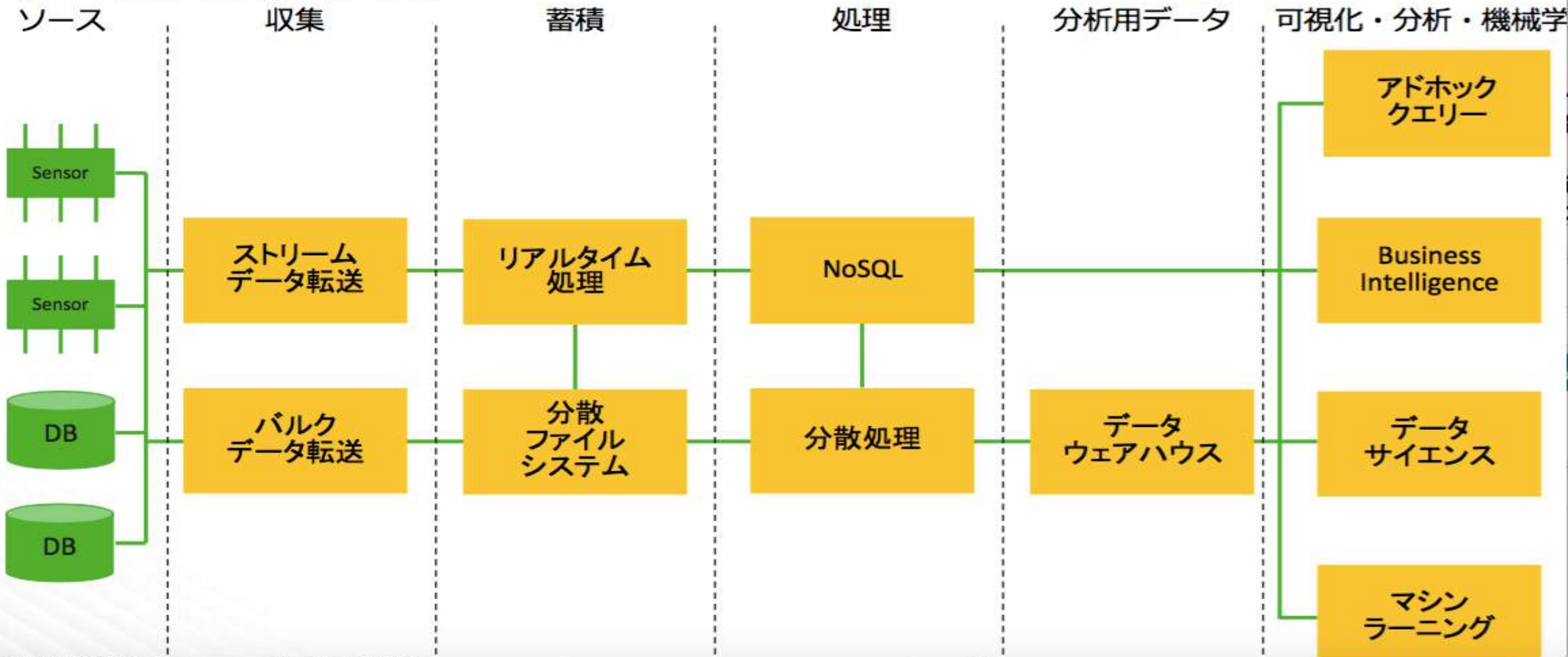
*Multiples of Bytes
Kilobyte
Megabyte
Gigabyte
Terabyte
Petabyte
Exabyte
Zettabyte
Yottabyte

1,000,000,000,000,000,000,000

IoT時代のデータ活用の流れ

データソースが多様 (IoT Device, Sensor, Connected Car...), データ量が膨大、リアルタイム性維持は重要

ビッグデータ分析のための データパイプライン



クラウド・IoT時代のデータ活用に適用される 代表的OSSコンポーネント

クラウドサービスの紹介

OSS ビッグデータ・アナリティクス

ソース

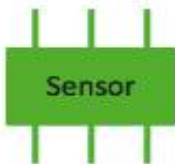
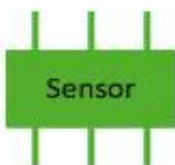
収集

蓄積

処理

データ

可視化・分析・機械学



ストリームデータ転送

NiFi, Kafka

Sqoop

バルクデータ転送

HDFS

Storm, Spark
Streaming

Hadoop /
Spark

Mongo DB

HBase,
Cassandra

Hive / Druid

Hive QL / LLAP

Superset

Zeppelin

Spark MLlibな
どの機械学習
ライブラリを
使用して開発

3. OSS「次世代データプラットフォーム」

= ホートンワークスの紹介 =

次世代データプラットフォームのデファクトスタンダードを提供する
オープンソースソフトウェア企業の世界のリーダー

会社概要

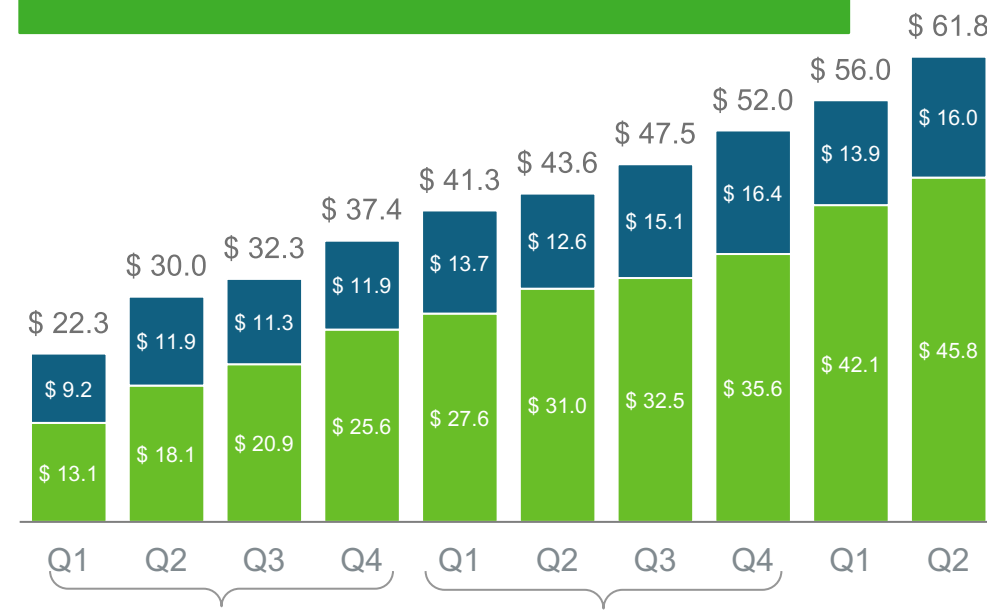
創立	2011年 Yahoo!のApache Hadoop オリジナルチームのメンバー24 人のエンジニアが設立
役員	CEO: ロブ・バーデン、COO:スコット・デイビッドソン オープンソースソフトウェアへ100%コミット Apache Hadoop プロジェクトへの貢献世界一
2011年	創業、 マイクロソフト社 (Azure HDInsight)と提携
2014年	9月 日本法人ホートンワークスジャパン株式会社設立 12月 NASDAQ上場(NASDAQ: HDP)
2015年	創業以来最速で売上\$100Mを達成 Apache NiFiのOnyara社買収し、HDFを市場投入
2016年	Billingが\$270M越す、HDC for AWSを市場投入
2016年	Dell EMC社 と提携
2017年	6月 主力製品 新版HDP 2.6、HDF 3.0を市場投入 6月 IBM社 と提携 9月 新製品・サービス HCPとDPSを市場投入

実績

Billing実績 \$ 270M (前年同期比 +62.7%)
売上実績 \$ 184.5M (前年同期比 +51.4%)

2Q17/2Q16 Revenue Growth:

42% YoY



主なパートナー様

グローバルパートナー（2100社超）

プラチナパートナー

- Accenture
- Dell EMC
- HPE
- IBM
- Microsoft
- NEC
- Teradata

ゴールドパートナー

- Fujitsu
- NEC
- Pivotal
- SAP
- Cisco Systems
- NetApp

ソリューションパートナー

- AtScale
- Attunity
- SAS
- SyncSort
- Informatica

コミュニティパートナー

- Talend
- NHN テコラス
- Tableau
- VUPICO
- VMware
- シグマシス
- Red Hat
- クリエイションライン

他多数

他

国内販売パートナー（12社）

- EMCジャパン・Dell EMC
- 日本マイクロソフト（HDInsightとして）
- Pivotalジャパン
- 日本IBM
- HPE日本
- シスコシステムズ合同会社
- 日本テラデータ
- 日本電気
- インサイトテクノロジー
- ノーチラス・テクノロジーズ
- マクニカネットワークス
- 日本情報通信

他多数

世界で最速で成長するソフトウェア企業

Hortonworks: \$121.9 million in 2015 revenues



Hortonworks is the fastest growing software company in history, reaching \$100 million in revenues in four years.

Fastest To Reach \$100mn in Revenue

Company	\$100mn Year	Founded In	Years to \$100mn	Company	\$100mn Year	Founded In	Years to \$100mn
Hortonworks*	2015	2011	4	FileEye	2013	2004	9
Salesforce	2004	1999	5	Intuit	1992	1983	9
Palo Alto	2011	2005	6	Microsoft	1984	1975	9
Workday	2011	2005	6	NetSuite	2007	1998	9
Informatica	2000	1993	7	Tableau	2012	2003	9
LogMeIn	2010	2003	7	Varonis	2014	2005	9
MobileIron	2014	2007	7	Red Hat	2003	1993	10
Citrix	1997	1989	8	Cornerstone	2012	1999	13
ServiceNow	2011	2003	8	Five9	2014	2001	13
Splunk	2011	2003	8	Paycom	2013	1998	15
Demandware	2013	2004	9	QlikTech	2008	1993	15

Source: Company Data, Barclays Research. Note: Hortonworks represents a potential future event

ホートンワークスが、ビッグデータウェアハウスのリーダーに選出



Source: Forrester: [The Forrester Wave™: Big Data Warehouse, Q2 2017](#)

- ◆ オープンソースベースのデータウェアハウスソリューション
- ◆ あらゆる種類のデータから判断・行動のための情報を提供
- ◆ 様々なエコシステムベンダーと連携可能
- ◆ 低コスト、スケーラブルなアーキテクチャ
- ◆ オンプレミス、クラウドなどあらゆるプラットフォームに対応 (MSFT Azure, AWS)
- ◆ 分散ストレージ、分散処理機能、データ収集、データガバナンス、サポートといった企業で必要とされる機能を提供

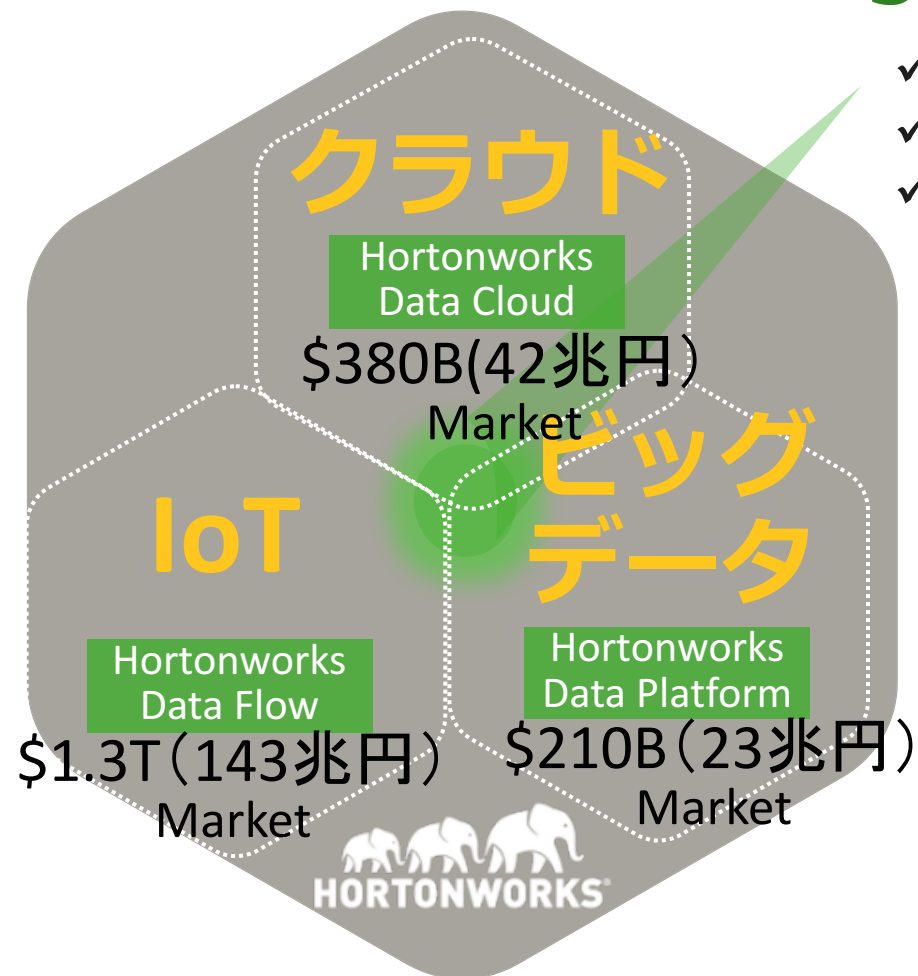
ホートンワークスの対応する市場（グローバル）

3つの市場への参入

- ✓ クラウド
- ✓ IoT
- ✓ ビッグデータ

Azure HDInsight と Hortonworks Data Cloud for AWS の提供

Onyara社買収により、IoTでのストリーミング分析用途などに Hortonworks DataFlowの提供



オープンソース技術を採用した次世代データプラットフォーム Hortonworks Data Platformの提供

Sources:

Public Cloud Services, \$204B, Gartner, <http://www.gartner.com/newsroom/id/3188817>

Big Data \$50B, IDC 2016, \$1.5T

IoT by 2020, Allied Market Research report published March 2014

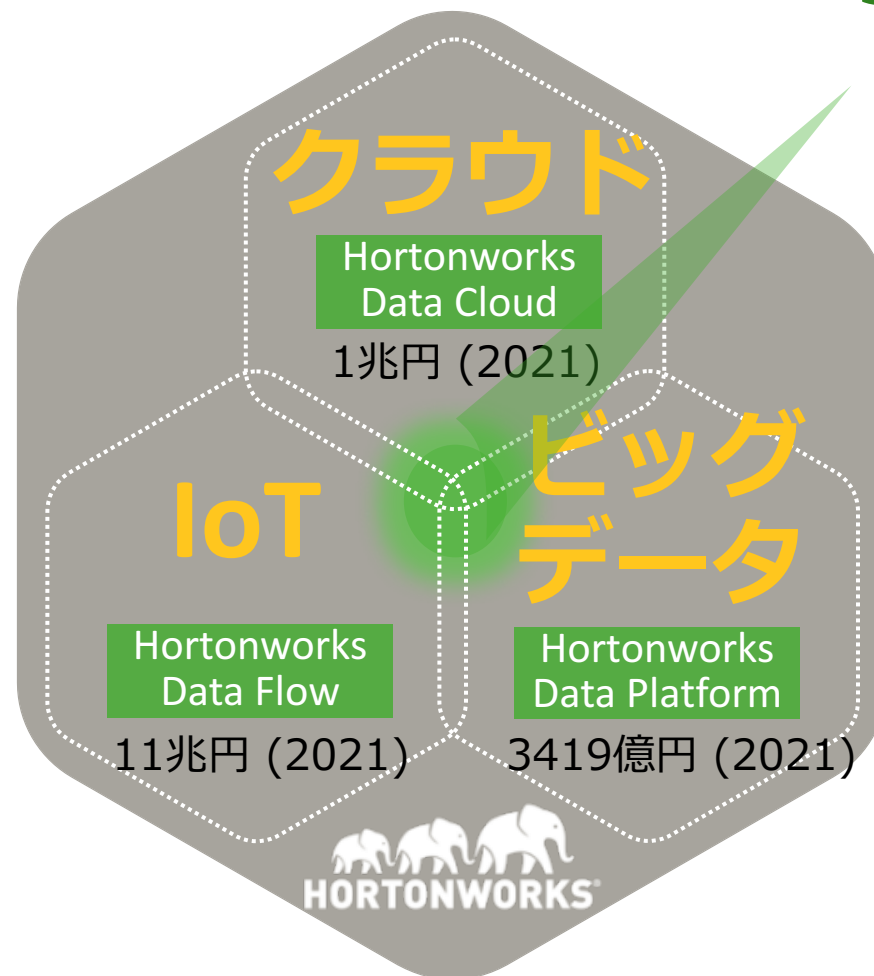
ホートンワークスの対応する市場（日本）

3つの市場への参入

- ✓ クラウド
- ✓ IoT
- ✓ ビッグデータ

Azure HDInsight と Hortonworks
Data Cloud for AWS の提供

Onyara社買収により、IoTでの
ストリーミング分析用途などに
Hortonworks DataFlowの提供



オープンソース技術を採用した
次世代データプラットフォーム
Hortonworks Data Platformの提供

出典: IDC Japan

IoT: <http://www.idcjapan.co.jp/Press/Current/201704101Apr.html>

クラウド: <http://www.idcjapan.co.jp/Press/Current/20170313Apr.html>

ビッグデータ: <http://www.idcjapan.co.jp/Press/Current/20160613Apr.html>

次世代ビッグデータプラットフォームが直面する課題

2 IoTなどの発展によるデータの爆発的増加傾向

データベース・データウェアハウス



データレイク (Data Lake)

データレイクとは 米ゼネラル・エレクトリック (GE) などが提唱するビッグデータ処理のアーキテクチャー。... 具体的には、あらゆるデータをネットワーク経由で収集し、オープンソースソフト (OSS) の分散処理基盤である「Hadoop (ハドゥープ)」に蓄積する。

2012
2.8 Zetta bytes

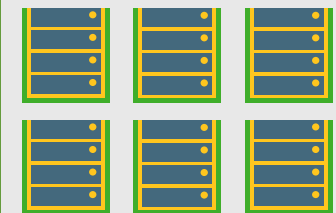
Traditional

2020

44 Zetta bytes

*Multiples of Bytes
Kilobyte
Megabyte
Gigabyte
Terabyte
Petabyte
Exabyte
Zettabyte
Yottabyte

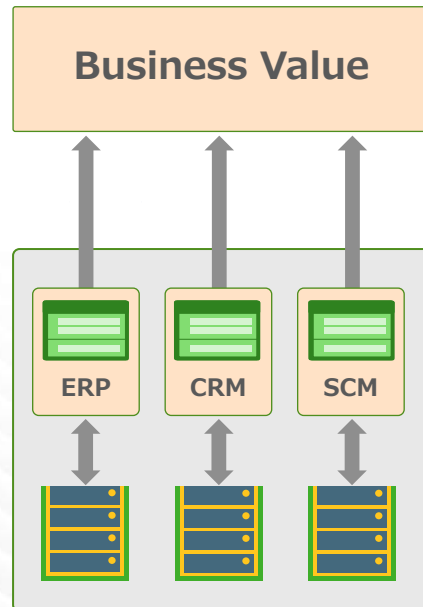
Clickstream
Geolocation
Web Data
Internet of Things
Docs, emails
Server logs



1,000,000,000,000,000,000,000

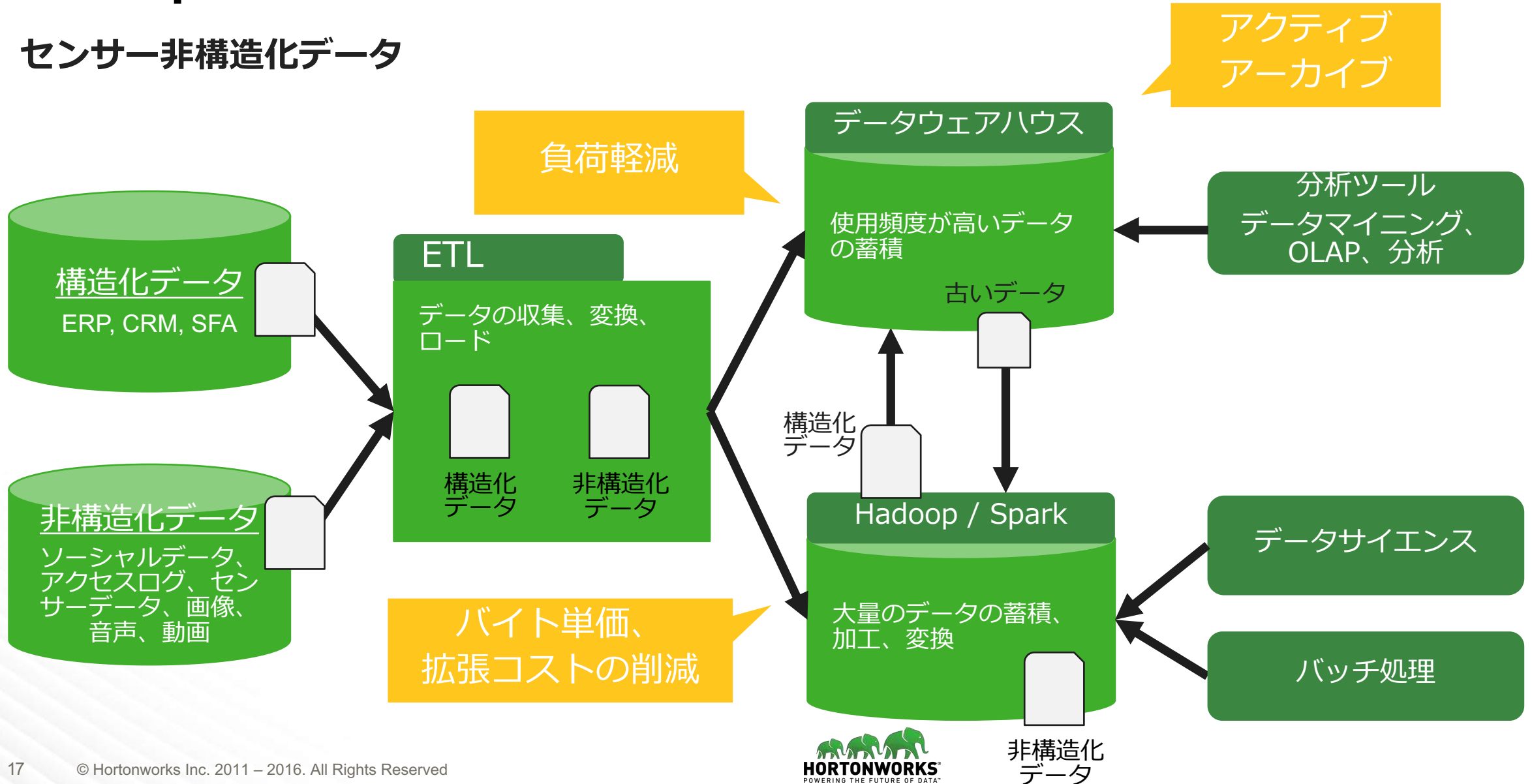
1 Challenges

- アプリケーション毎にデータがサイロ化されている。
- 新しいデータの格納方法 (非構造化・半構造化データ)
- 拡張時のコストバランス



Hadoopを利用したデータウェアハウス最適化

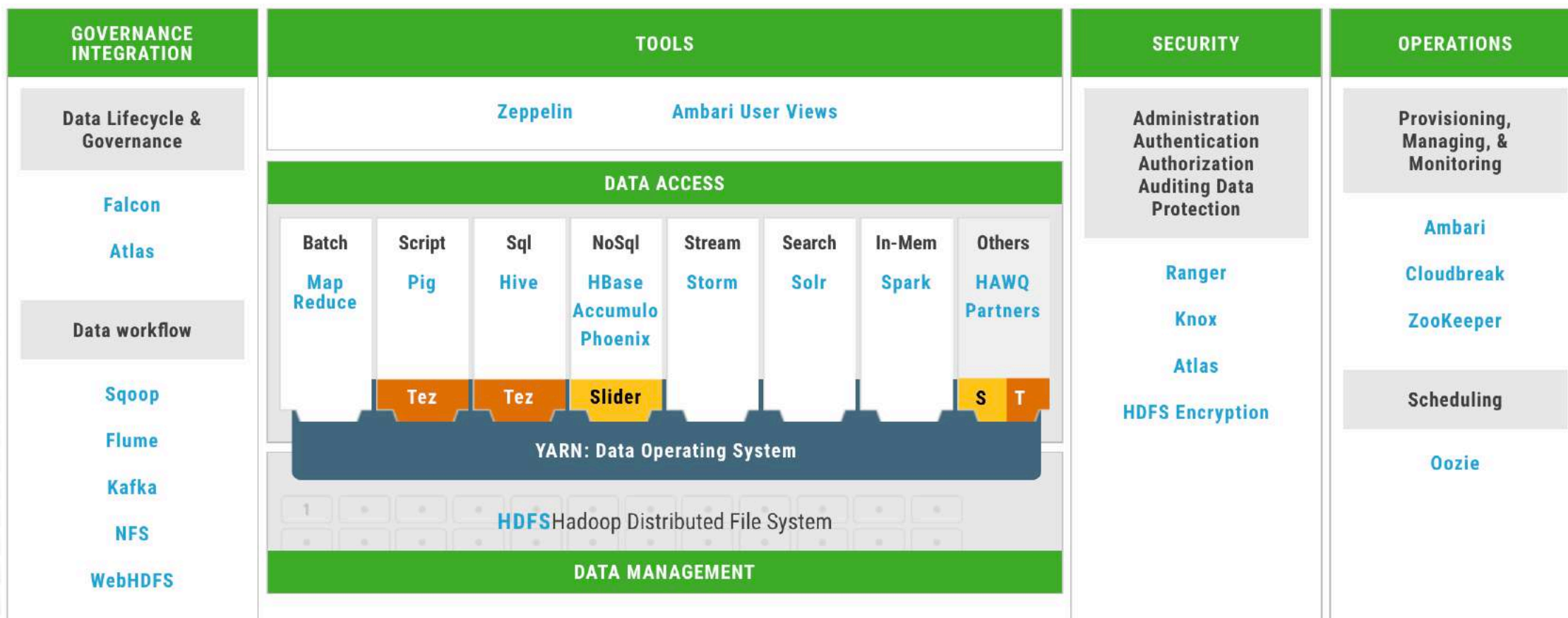
センサー非構造化データ



データレイクの基盤 Hortonworks Data Platform (HDP)



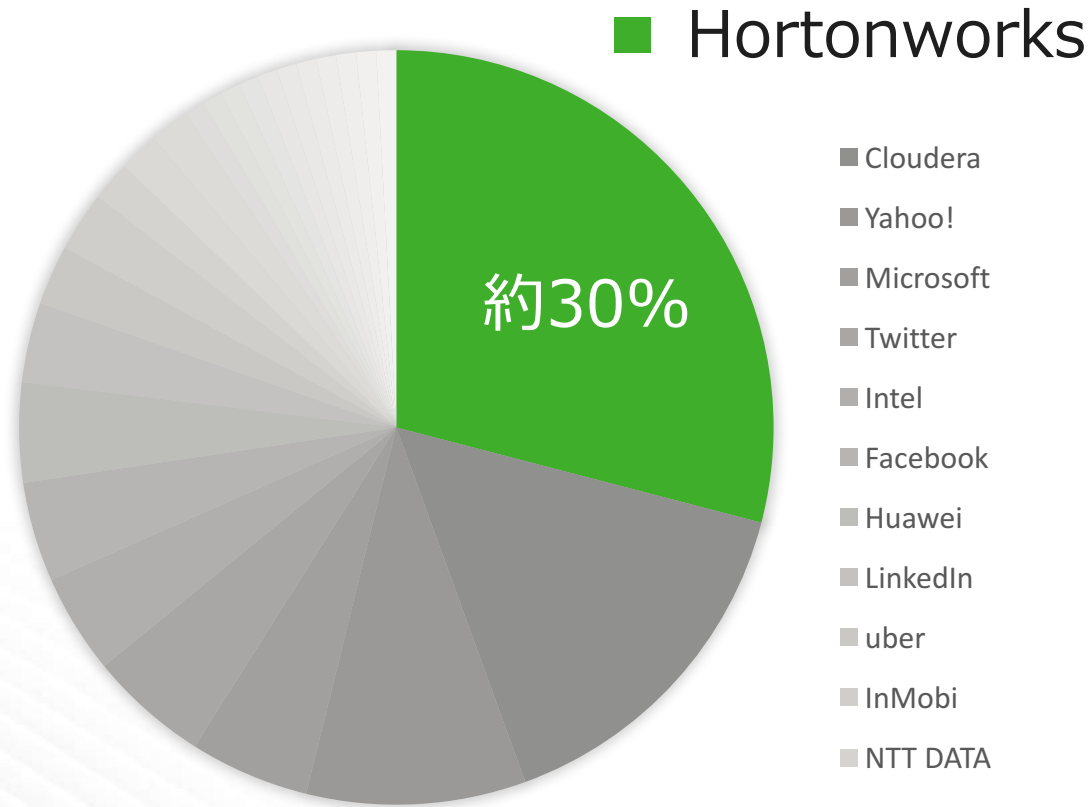
データ分散処理ソフトウェアの「Apache™ Hadoop®」や「Apache Spark™」等のオープンソースソフトウェアをベースとし、セキュリティやガバナンスを強化した次世代データプラットフォーム



Apache Hadoop開発コミュニティへの貢献 (世界最大の Hadoopコミッタ数を有する OSS企業)

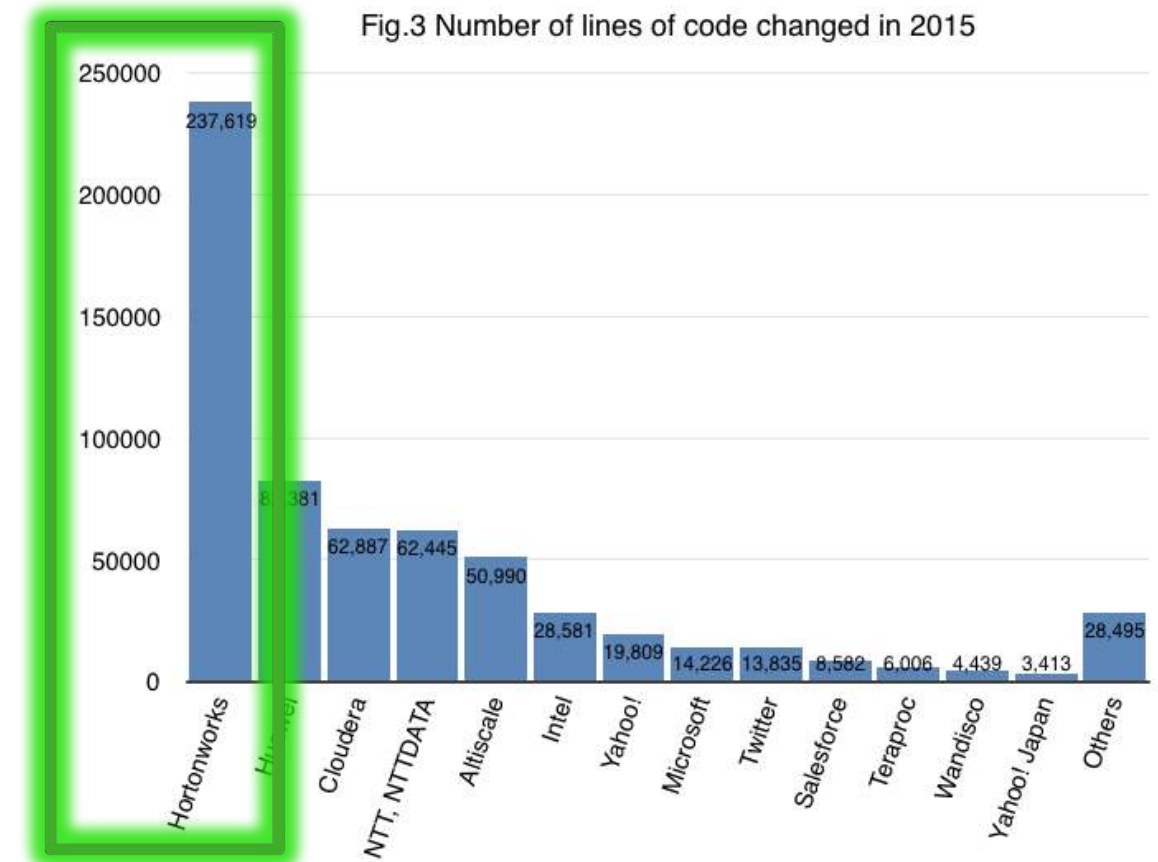
(開発者の数)

コミッター数



(貢献したコード数)

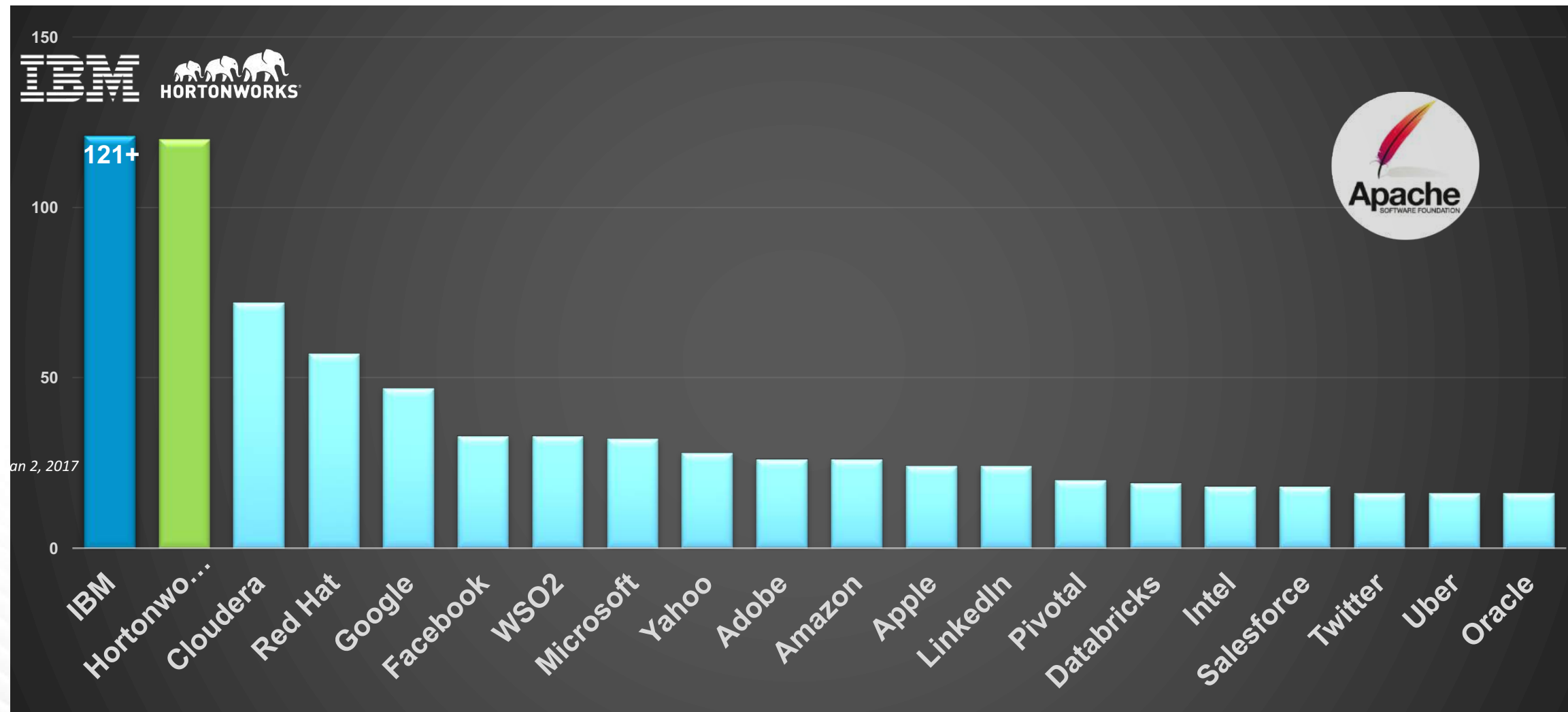
コントリビュートしたコード数



<http://hadoop.apache.org/who.html> を元に作成 (2017/03/02)

出典: <http://ajisakaa.blogspot.jp/2016/01/the-activities-of-apache-hadoop.html>

上位20社におけるApache OSS コミッターの数



次世代データアプリケーションを支えるプラットフォーム

- 最適な保健料金の選定システム
- 故障予測サービス
- サイバーセキュリティ
- スマートメータ連携システム
- レコメンドシステム
- その他、インダストリーアプリケーション

次世代データアプリケーション



セキュリティ脅威の監視

HCP

Hortonworks
Cybersecurity
Platform



流れているデータ

HDF

Hortonworks
DataFlow



蓄積されたデータ

HDP

Hortonworks
Data Platform

Spark

hadoop

4. IoTと米国における IoT活用状況

1. IoTとは？ (1999年: MIT AutoIDセンサー創始者ケビン・アシュトン氏が唱える。実用は その10年後から)
Internet of Things (モノのインターネット)とは、従来は主にパソコンやサーバー、プリンタ等のIT関連機器が接続されていたインターネットにそれ以外の様々なモノを接続することを意味します。

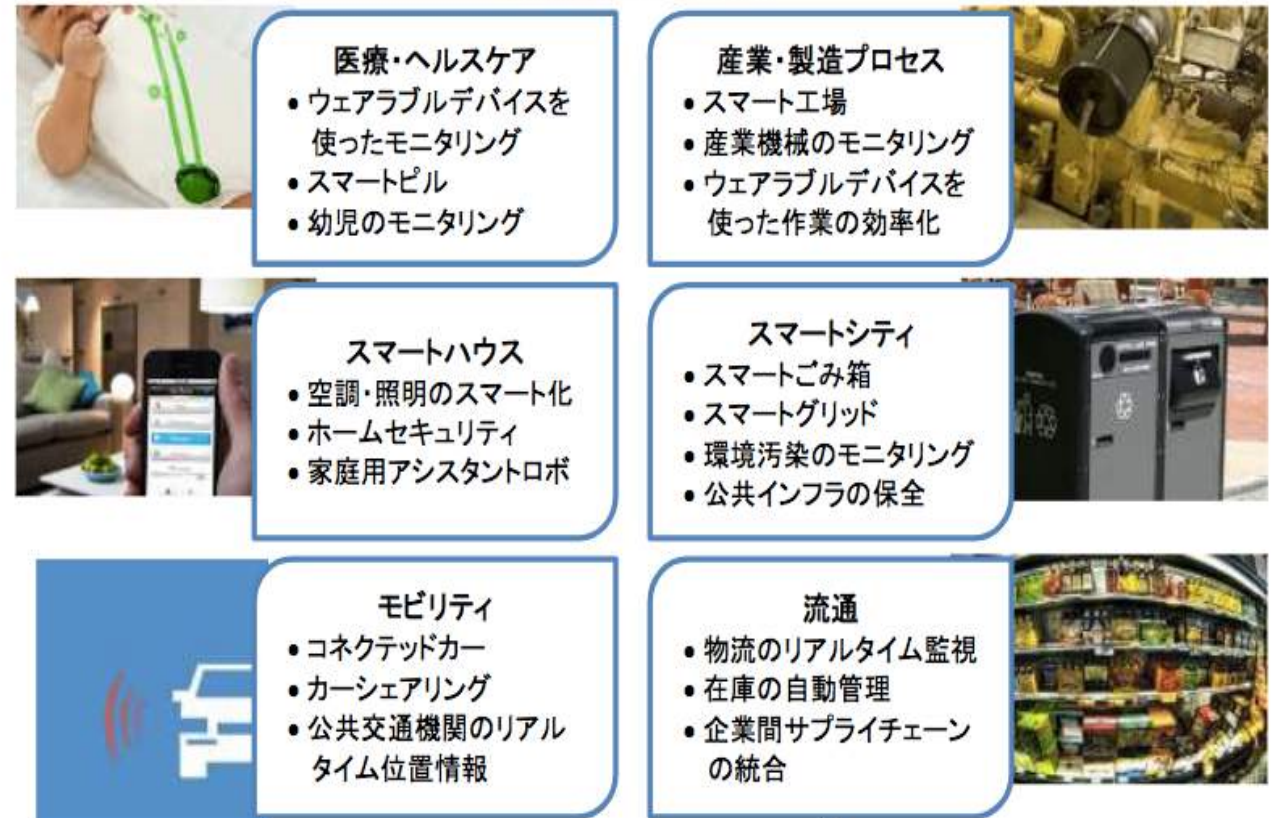
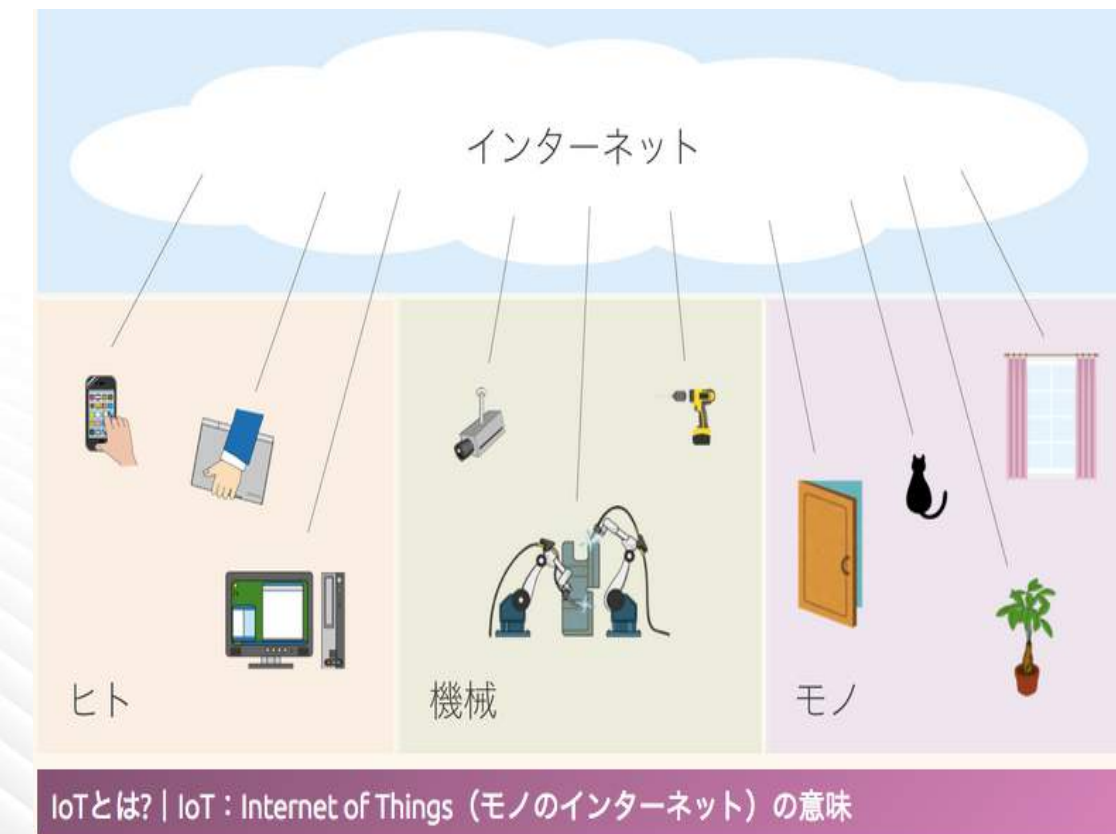
2. IoTの主構成要素: (1)モノ、(2)センサー(3)プロセッサ(スマホ、PC、ゲートウェイ) (4) Network (5) IT

3. 何を実現するか

- ・離れた場所の状態を知りたい。→離れたモノの状態を知る。
- ・離れた場所の状態を変えたい。→離れたモノを操作する。

出典: JETRO/IPA New Yorkレポート

図表 1: IoT の活用例



4. IoTと米国における IoT活用状況

1. IoTとは？：

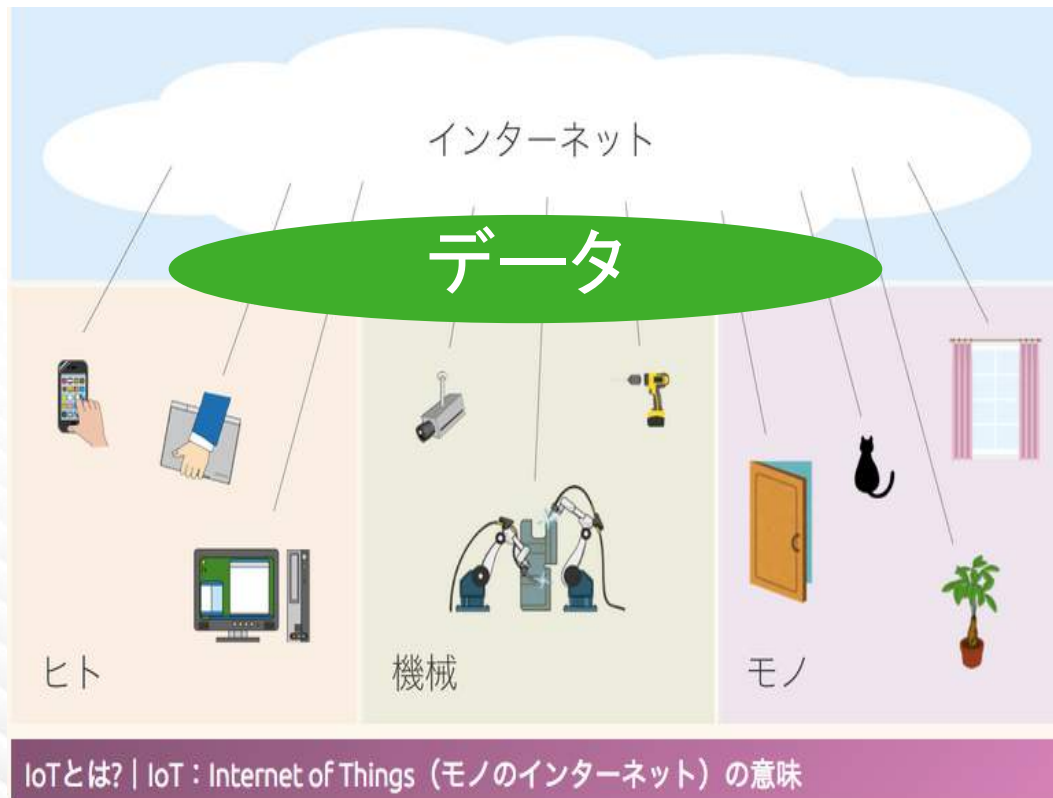
Internet of Things (モノのインターネット)とは、従来は主にパソコンやサーバー、プリンタ等のIT関連機器が接続されていたインターネットにそれ以外の様々なモノを接続することを意味します。

2. IoTの主構成要素：(1)モノ、(2)センサー(3)プロセッサ(スマホ、PC、ゲートウェイ) (4) Network (5) IT

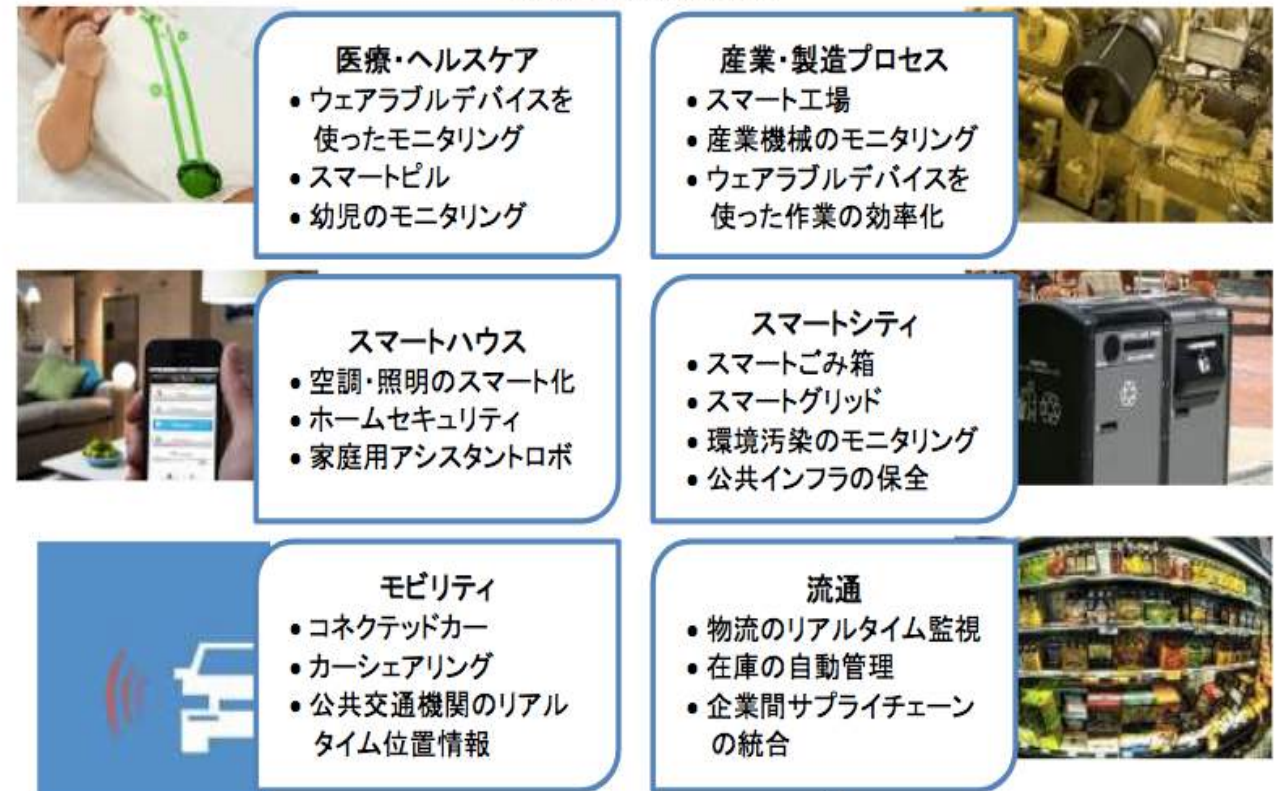
3. 何を実現するか。何が実現できるのか。

- ・離れた場所の状態を知りたい。→離れたモノの状態を知る。
- ・離れた場所の状態を変えたい。→離れたモノを操作する。

- ・ IoT・ビッグデータの基盤ソフトウェア
- ・ 分析・解析・AI・MLソフトウェア



図表 1: IoT の活用例

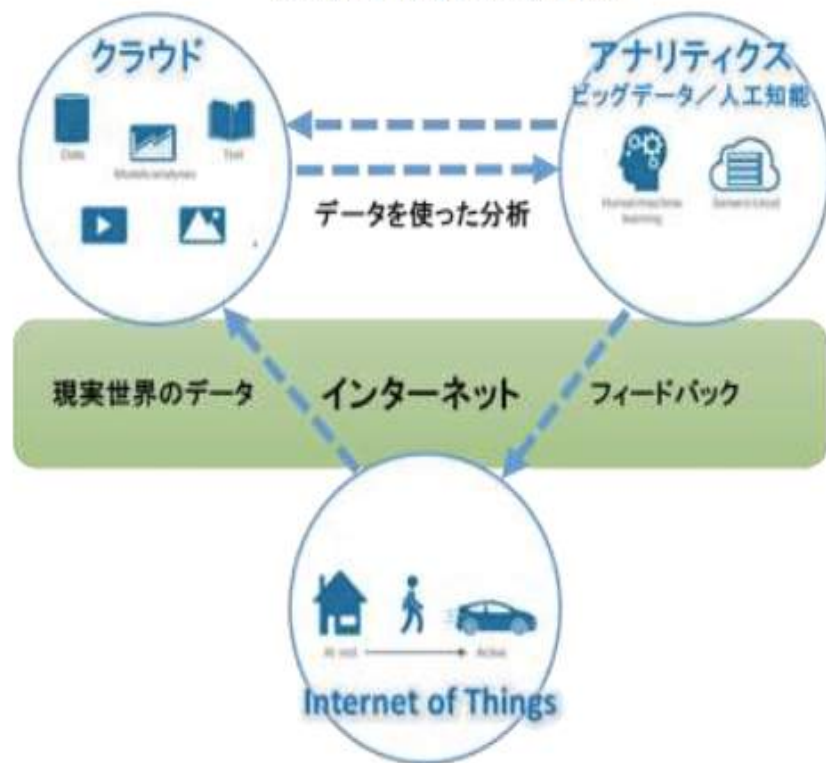


IoTの活用事例

出典： JETRO/IPA New Yorkレポート

図表 3 は GE 社の航空エンジンモニタリングサービスとなっている。

図表 2: CPS における IoT



図表 3: GE 社の航空エンジンモニタリングサービス



出典：GE Software, GE Aviation⁹

General Electric (GE) 社 Flight Efficiency Service
GEでは同社が製造した飛行機エンジンから送られてくるデータをリアルタイムで分析し、航空会社に対して飛行機の効率的な運用やエンジン操作方法を提案するサービスFlight Efficiency Servicesを提供。航空会社では年間数億円規模の燃料費セーブを実現。

CPS: Cyber Physical System
IoTの実装に於いて物理世界のデータをサイバー空間で分析し現実空間にフィードバックする、分析・計算コンポ（サイバー空間）と物理コンポ（現実世界）との連携させるコンセプト

IoTの活用事例

出典： JETRO/IPA New Yorkレポート

図表 4: Zipcar のカーシェアリング



出典: Tampa Bay Times、Zipcar¹³

図表 5: Tesla Motors 社の電気自動車(販売後もIoTでソフトウェアをアップデート可能)



出典: Tesla Motors、Extreme Tech¹⁶

Zipcar社が提供しているカーシェアリングサービス、スマートフォンで可能な車を検索し予約できる。料金は時間制であるがガソリン代や保険代込み。コネクテッド機能でスマートフォンからドアの鍵を開くことができるなど、スマートフォンで検索、予約、利用、支払いまで一括で行えるリアルタイムサービスを実現

TeslaMotors社では、ソフトウェアの更新も、インターネットで自動更新できるという仕組みを取り入れ。購入後でもソフトウェアのアップデートにより新機能を利用することができる。ユーザーにとっては常に新しい製品価値を与え、不具合があった場合でもソフトウェアでの対応が可能、メンテナンス面を自動更新できる。将来、自動運転機能Autopilotの追加・拡張してゆく予定。

Hortonworks DataFlow (IoTの OSS基盤)

Powered by Apache NiFi, Apache Kafka, Apache Storm

Ver. 3.0

GA June 12th, 2017



- ◆ データフロー管理：
 - **Apache NiFi**、**MiNiFi** により、エッジからクラウドまで、データフローを管理
- ◆ ストリームデータ処理：
 - SAM (Streaming Analytics Manager) により、アプリケーション開発者、ビジネスアナリスト、管理者は1行のコードも記述することなく、GUIからApache Kafka、Apache Stormベースの、ストリーミングアプリケーションを迅速に開発可能
- ◆ エンタープライズ向け：
 - **Apache Ranger**、**Apache Amabri** により、運用管理、エンタープライズで必要とされるセキュリティを実現
 - スキーマ用の新しい共有リポジトリを利用することで、複数のデータフロー、ストリーミングアプリケーションで利用するデータ形式を一括管理。扱うデータを明確にし、データ管理、活用を効率化

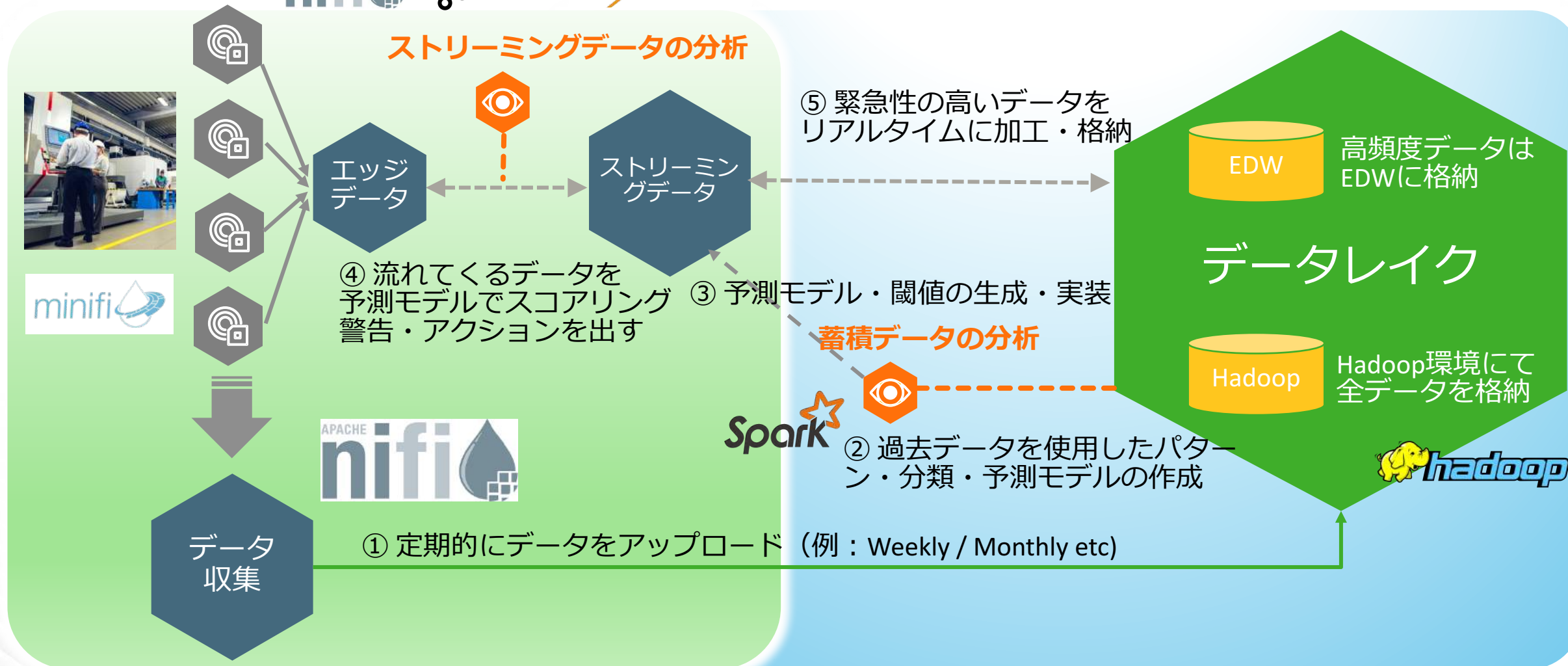
新たにIBM Power Systemsでも利用が可能。ビッグデータや認知解析のような、データを多用するワークロード向けに設計されたサーバー上の多様なストリーミングアナリティクスアプリケーションをサポート。HDFとPower Systemsの組み合わせは業界をリードするパフォーマンスと高効率のストリーミングアナリティクスを提供し、流れているデータのワークロードの管理を容易に。

IoT・ストリーミングデータ分析ソリューション

データの収集



データの蓄積、分析



オープンソースソフトウェアのトレンド（革新イノベーション）

OPEN
COMMUNITY



THE
INNOVATION
ADVANTAGE

PROPRIETARY APPROACH

TIME

INNOVATION

◆ Power of Open Source

- 1991年に Linux OSが登場して以来 OSの領域に始まり、ミドルウェアの領域まで OSS開発のパワーは独自開発を遥かに凌ぐまでに至る
- IoT・ビッグデータの領域でも先端の革新はオープンソースコミュニティで起きている

◆ Value of Hortonworks

- ホートンワークスはIoT・ビッグデータ基盤をエンタープライズ企業でも安心安全に活用できるサポート・サービスを提供

◆ Expanding Ecosystems

- 世界中で、日本中で官公庁・企業が活用できるエコシステムを急速に構築中
- 開発コミュニティのリードに加え、2100を超えるエコシステムパートナー、1200社を超える企業での活用実績がある



お問い合わせ先
info-jp@hortonworks.com

APPENDIX

IoTを活用したスマート工場の狙い

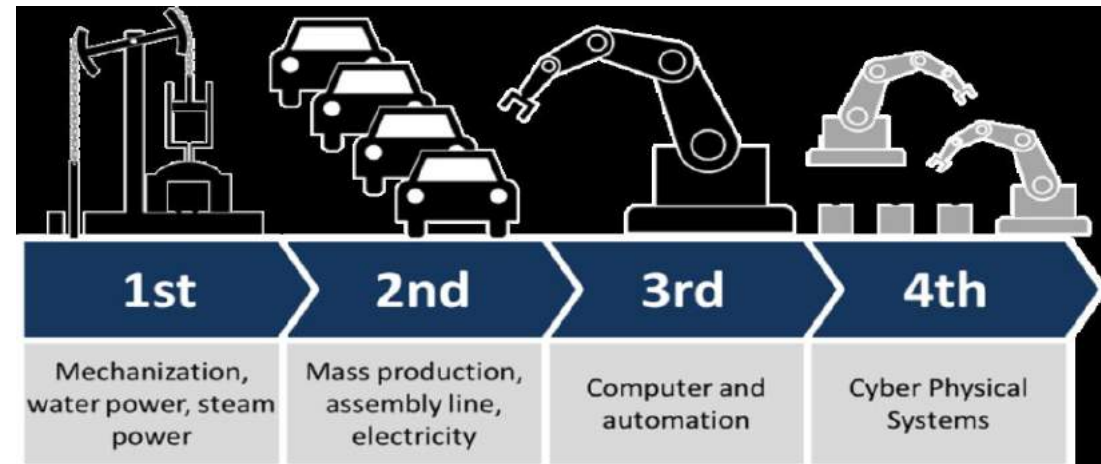
4次産業革命・ Industry 4.0時代の到来（米国、ドイツ、英国、スイス、日本中心）
老朽化する工場を活性化したり、地方や途上国等の生産の現場と
事業の本部・本社をリアルタイムで繋ぎ国際競争力のある「スマート工場」
（賢い工場・考える工場）を設立し製造業の競争優位性作り出すために
インダストリーIoTが登場し第4次産業革命を牽引し始めている。

スマート工場（賢い工場・考える工場）の特徴

1. 少量多品種生産や高付加価値製品の大規模生産を目指す
（マスカスタマイゼーション）
2. 注文と同時に生産が始まるリアルタイム性、生産・製造稼働状況の即時把握
3. 大量生産とさほど変わらない驚きの原価低減・低価格を実現

スマート工場実現の重要ポイント

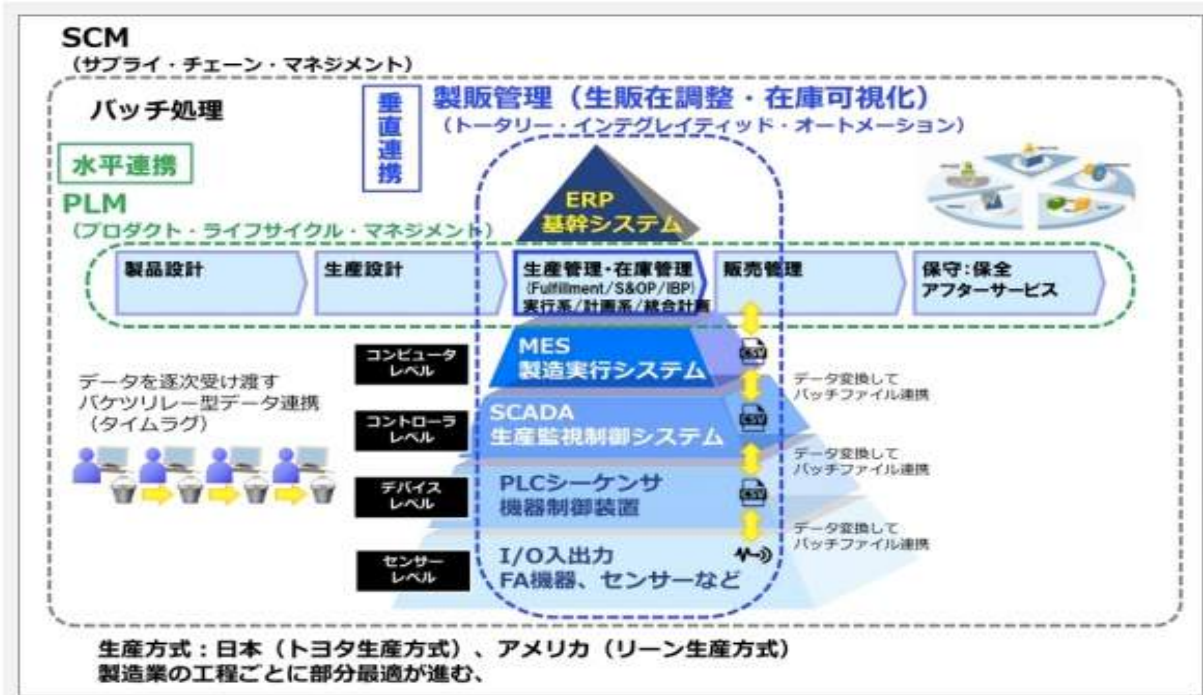
1. センサーネットワークによる
徹底した自動化
2. ビッグデータとAIを駆使
して品質向上ポイント
3. 「つながる工場」を可能に
するネットワークと
データやInterface等の標準化、
サプライチェーンとの統合



図表6：4つの産業革命（Christoph Roser作、AllAboutLean.comより）

従来の工場とスマート工場（インダストリアルIoT活用）の違い

従来の工場： 水平連携と垂直連携を別々に改善、基本バッチ
日本の「トヨタ生産方式」「改善」
ボトムアップで徹底的な生産効率を
図り短納期で安く、高品質を達成
アメリカの「リーン生産方式」
トヨタ方式にトップダウンの
考えを導入し生産性を更に改善



スマート工場：「ダイナミックセル方式」ベース
＝「ライン生産（水平連携）」（単一大量生産）
と「セル生産（垂直連携）」（一つのチーム
が全て組み立てる少量高品質生産）の折衷。
工程をセルに分けダイナミックに変更できる。
徹底した「標準化」「リアルタイム化」「デジタルデータ化」
「生産システムの接続性、効率性、
拡張性の向上」を図る。 期待できる効果
→ 大幅なコスト削減、品質向上、
生産スピード・能力向上、柔軟性 など

