

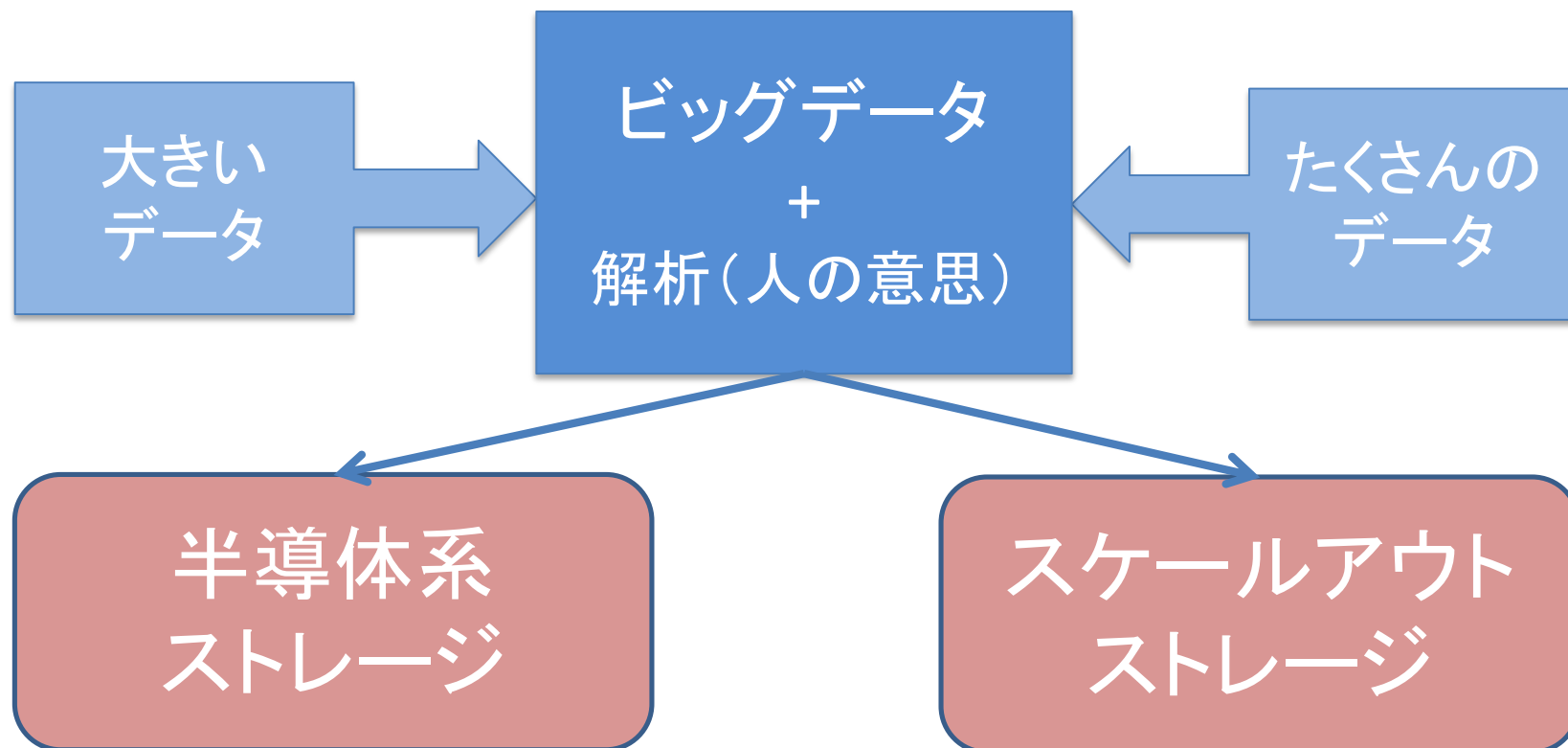


バズワード「ビッグデータ」第四步目！ AI技術は未来のストレージに応用できるのか？

ビッグデータWG

JDSF 新春セミナー
2016年 1月27日

- 昨年のTechForumでは、バズワード「ビッグデータ」第三歩目！としてビッグデータの活用事例から見たストレージの分類と未来像をご紹介しました。今回はその未来像をさらに深掘しAI技術の応用の可能性を探りました。
- 2. 本コンテンツはIT初心者の方が理解できるように、ベースの部分から解説をしております。上級者の方々には物足りない内容になっているかもしれませんが、お含みおきください。
- 3. 本コンテンツの一部は後日公開予定でございます。
- 4. 本コンテンツを転載されたい場合には、事務局までご一報ください。



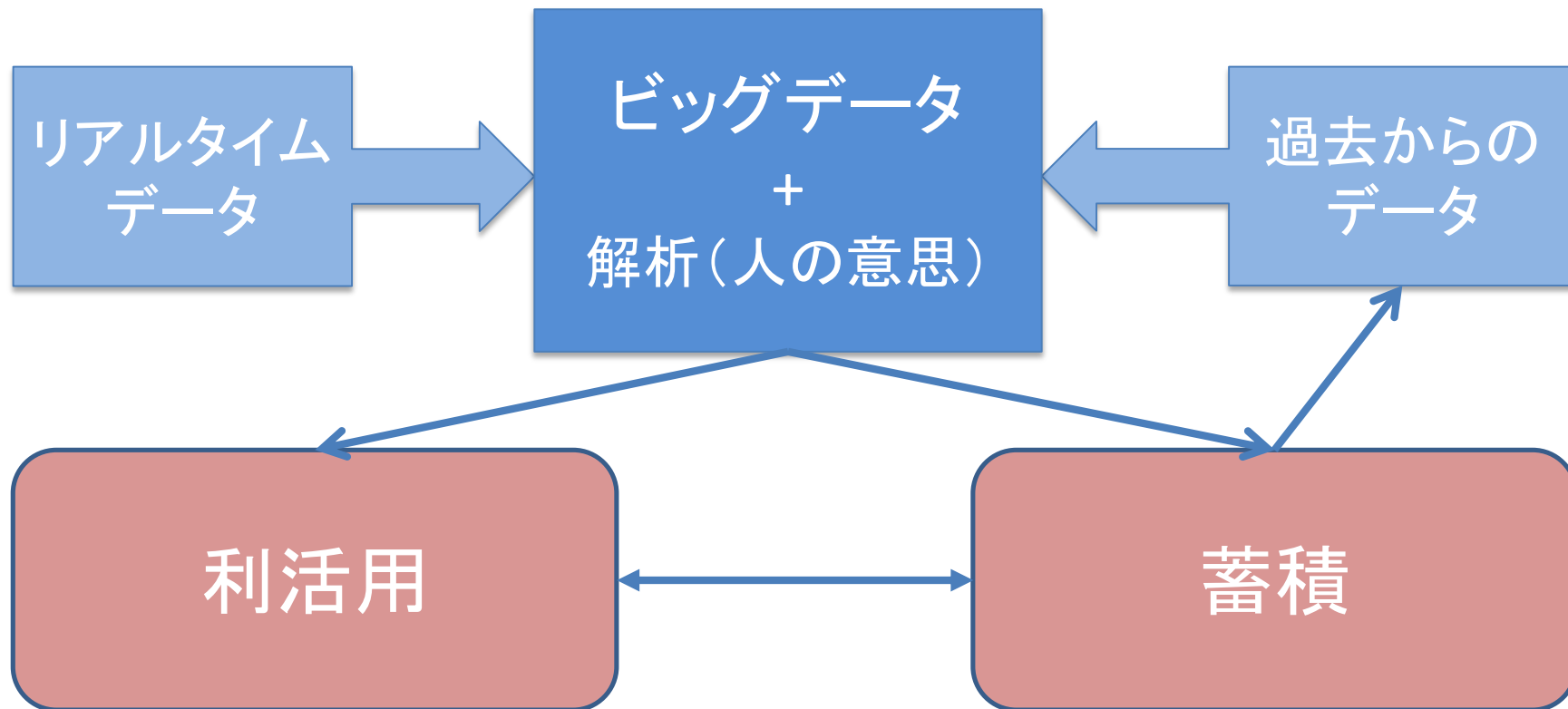
解析・分析

データ保管

- 今までのストレージテクノロジー(RAID、HDDやTAPE)と趣向を異にするストレージインフラの台頭

おさらい 2

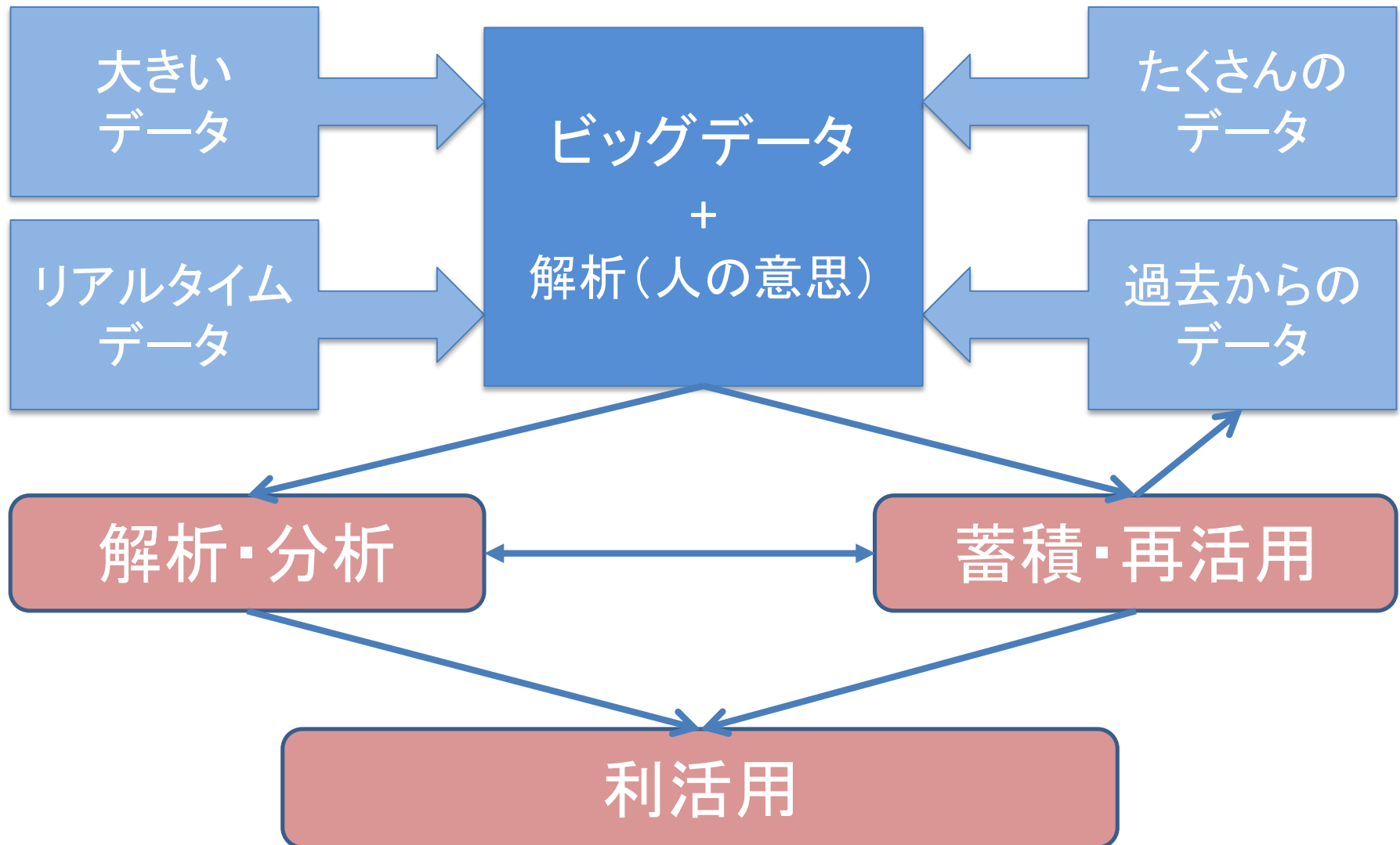
(図解)ビッグデータライフサイクルとストレージ要件



- テクノロジーは発散気味

おさらい(1+2)

(図解)ビッグデータライフサイクルとストレージ要件





ビッグデータの潮流と考察

What is the Trend of Big Data?

(付録)ICT新事業創出に向けたPROJECTとACTION

■ 中核を成しているのは、ビッグデータ、オープンデータ、パーソナルデータ

【ACTION】ユーザ参加型テストベッド

- 事業創出や技術開発のコストを下げ、多くのトライアルを実現する場の提供
- ベンチャーがアセットを活用するスタイル
- 最先端のクラウド環境や情報資源も活用可能なテストベッドの整備
- 異業種のユーザも利用可能なオープンテストベッド

【ACTION】データ利活用マッチング・プラットフォーム(「場」)の構築

- 起業や創意を促す機運の醸成
- アイディアソン・ハッカソン、コンテスト等の活用
- ビッグデータやオープンデータの利活用の促進のためのマッチング・プラットフォーム(「場」)の構築

【ACTION】トライ&エラー型実証

- 小規模の社会実装と連動したプロジェクトの推進
- リスタートアップモデルの導入(ICT海兵隊)
- 国を実証フィールドに活用(先端環境の先行的導入)
- 省庁の垣根を超えたICT利活用の実現

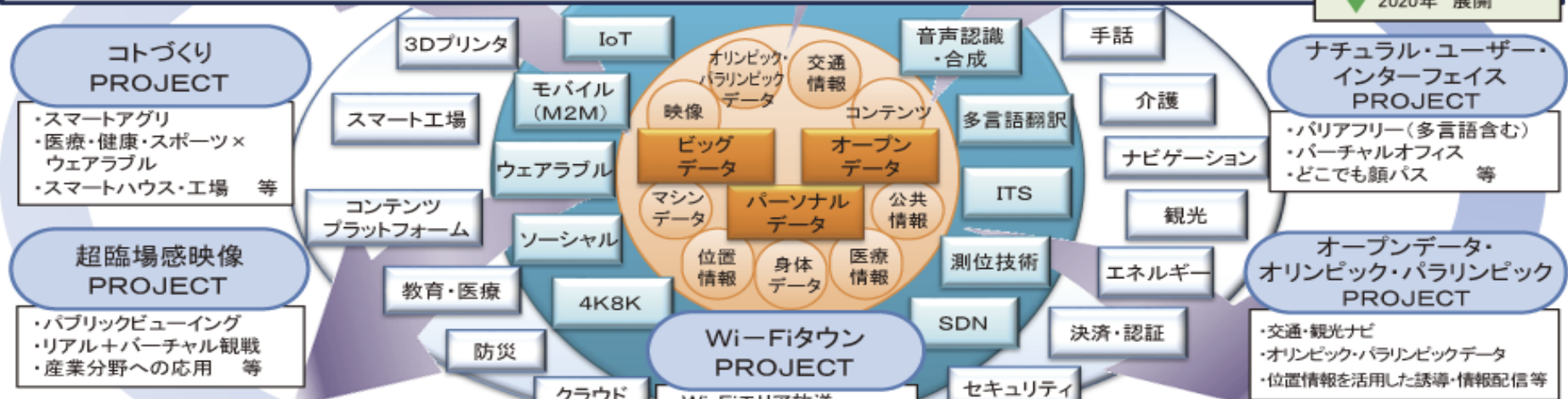
今後どのような新事業・新サービスが創出されていくか

【背景】2020年～社会・技術トレンド、震災から9年目の復興、オリンピック・パラリンピック

【方向性】「オープンデータ・ビッグデータ」、「新技術」と「ネットワーク・アプリケーション」の掛け合いによる、新事業・新サービスの創出

パイロット・プロジェクト

- 2016年 実証
- 2018年 社会実装
- 2020年 展開



【ACTION】ワークスタイル・人材

- 新事業創出を担う仕事環境や人材育成
- ICTを活用したスマートワークの実現
- プログラム人材の育成
- 各業種の優れた人材に対するICT利活用導入

ビッグデータ、ブロードバンド、クラウド等のICT技術の進展により、今まで実現しなかった新事業・新サービスが可能に

【ACTION】オープンデータ

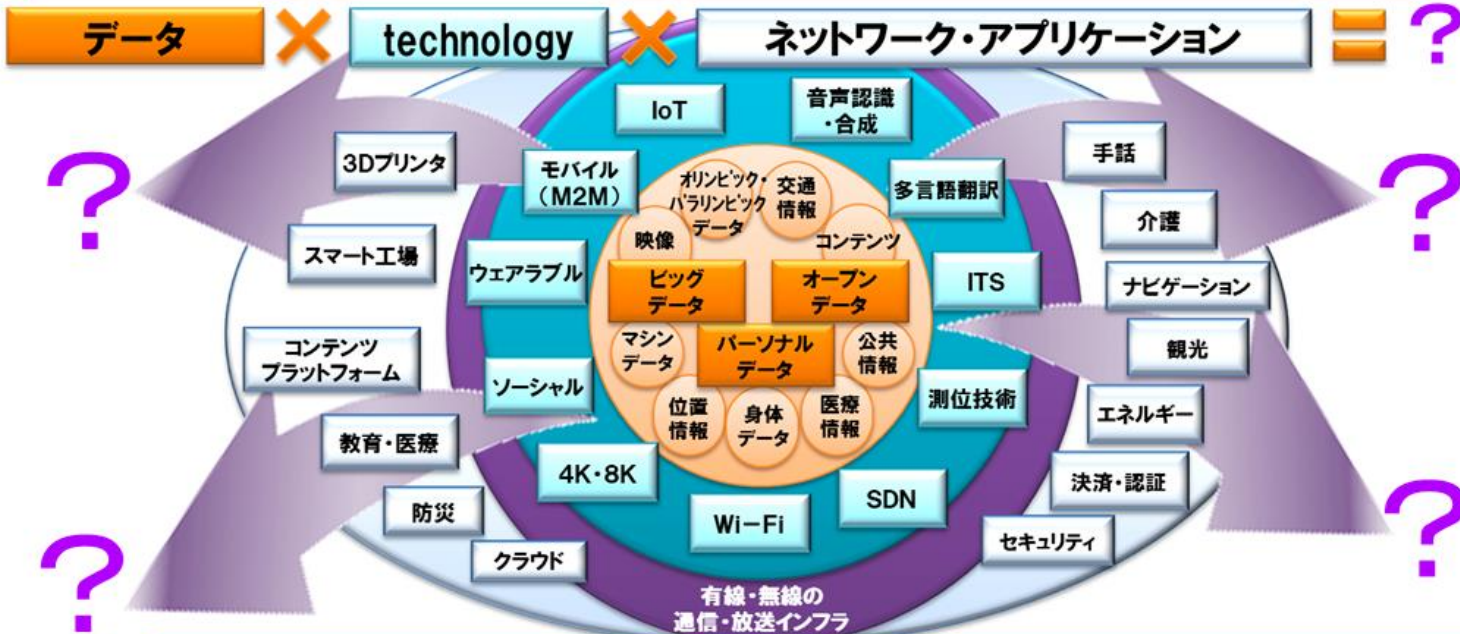
- オープンデータを核としたサービスイノベーション
- 公共オープンデータ促進によるビジネス活性化
- 民間における取組を加速するための体制強化
- ビジネスモデル検討、データサイエンティスト資格

新事業検討の方向性

新事業のイメージ 最先端ICTをどのように活用すれば新事業が創出されるのか？

◆新事業検討の方向性

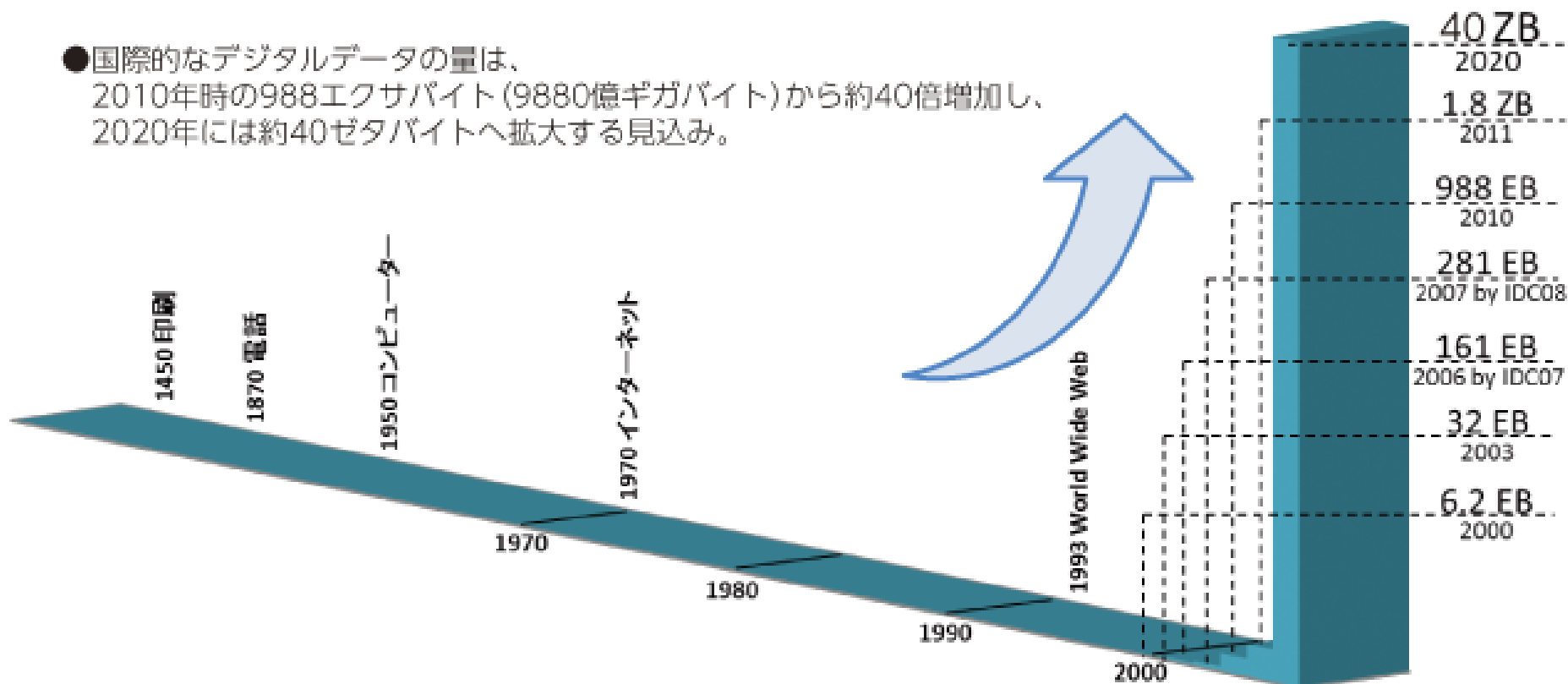
- ① 多様なデータ(ビッグデータ、オープンデータなど)の流通を促進するために必要な環境 ⇒ **データ**
- ② 最先端ICTの融合により期待される新技術 ⇒ **technology**
- ③ データを安心・安全に利活用するために必要なネットワーク ⇒ **ネットワーク・アプリケーション**
- ④ 新技術、ネットワーク、データ利活用の掛け合いにより期待される新事業・新サービス



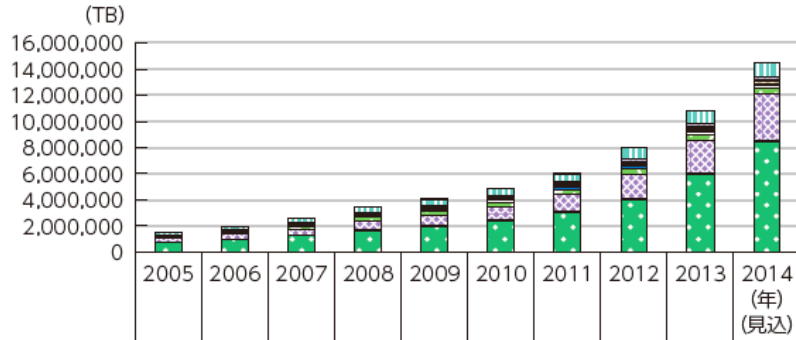
「オープンデータ・ビッグデータ」、「新技術」と「ネットワーク・アプリケーション」の掛け合いにより、さらなる新事業・新サービスが生まれるのではないかと

デジタルデータ量の増加予測

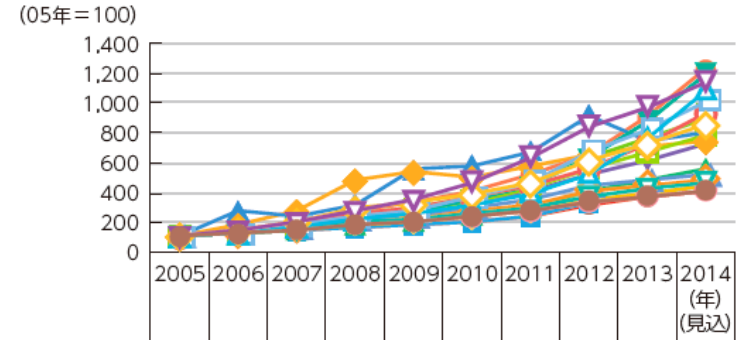
- 国際的なデジタルデータの量は、2010年時の988エクサバイト(9880億ギガバイト)から約40倍増加し、2020年には約40ゼタバイトへ拡大する見込み。



データ流通量の推移(メディア別)



	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014 (年見込)
顧客データ	4	5	6	7	8	9	10	14	16	18
経理データ	100	113	136	161	174	199	228	328	389	450
業務日誌データ	1	1	2	2	3	3	4	5	6	7
POSデータ	245,516	290,976	364,822	454,167	494,141	574,740	669,810	850,928	974,635	1,100,441
ECコマースにおける販売データ	12	15	21	28	32	40	48	61	73	85
電子メール	66,365	79,151	99,762	124,510	135,425	156,004	180,308	223,540	249,157	270,381
CTI音声データ	2,509	3,142	4,163	5,662	6,689	8,379	10,613	14,154	16,871	19,714
固定電話	38,035	47,546	62,519	80,958	89,825	108,641	131,210	167,052	183,224	196,448
携帯電話	17,765	21,883	28,339	36,068	39,441	47,117	56,232	71,289	78,822	85,388
アクセスログ	521	628	799	1,008	1,114	1,303	1,528	1,922	2,185	2,430
動画・映像視聴ログ	673	906	1,292	1,833	2,170	2,733	3,436	4,365	6,052	8,225
Blog、SNS等記事データ	155	202	278	379	447	563	709	961	1,156	1,367
GPSデータ	33,116	90,808	78,398	104,928	182,687	191,484	223,661	302,759	249,336	262,041
RFIDデータ	64,519	109,051	176,352	309,559	340,506	318,111	359,887	407,967	445,081	472,944
センサーデータ	293,287	367,601	486,795	639,348	771,822	995,800	1,284,64	1,829,23	2,568,93	3,509,038
交通量・渋滞情報データ	10,869	14,227	19,737	27,246	31,634	38,970	48,145	60,453	78,593	100,373
気象データ	877	1,102	1,472	1,958	2,401	3,120	4,067	5,826	7,223	8,789
防犯・遠隔監視カメラデータ	781,963	976,963	1,289,39	1,688,82	1,977,29	2,464,65	3,074,31	4,077,32	5,940,91	8,483,878
電子カルテデータ	227	311	448	636	714	862	1,034	1,363	1,617	1,917
画像診断データ	71	96	137	191	242	331	450	597	696	810
電子レセプトデータ	2	2	3	4	4	5	5	7	7	8



	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014 (年見込)
顧客データ	100	112	132	155	173	200	233	313	362	407
経理データ	100	113	135	160	174	199	228	327	388	449
業務日誌データ	100	117	144	177	202	242	291	408	483	559
POSデータ	100	119	149	185	201	234	273	347	397	448
ECコマースにおける販売データ	100	128	174	235	273	333	407	517	614	717
電子メール	100	119	150	188	204	235	272	337	375	407
CTI音声データ	100	125	166	226	267	334	423	564	672	786
固定電話	100	125	164	213	236	286	345	439	482	516
携帯電話	100	123	160	203	222	265	317	401	444	481
アクセスログ	100	120	153	193	214	250	293	369	419	466
動画・映像視聴ログ	100	135	192	272	323	406	511	649	900	1,223
Blog、SNS等記事データ	100	130	179	244	288	362	456	619	745	880
GPSデータ	100	274	237	317	552	578	675	914	753	791
RFIDデータ	100	169	273	480	528	493	558	632	690	733
センサーデータ	100	125	166	218	263	340	438	624	876	1,196
交通量・渋滞情報データ	100	131	182	251	291	359	443	556	723	923
気象データ	100	126	168	223	274	356	464	664	823	1,002
防犯・遠隔監視カメラデータ	100	125	165	216	253	315	393	521	760	1,085
電子カルテデータ	100	137	198	280	315	380	456	601	713	845
画像診断データ	100	135	192	268	339	463	630	836	975	1,135
電子レセプトデータ	100	119	148	181	199	235	276	340	373	409

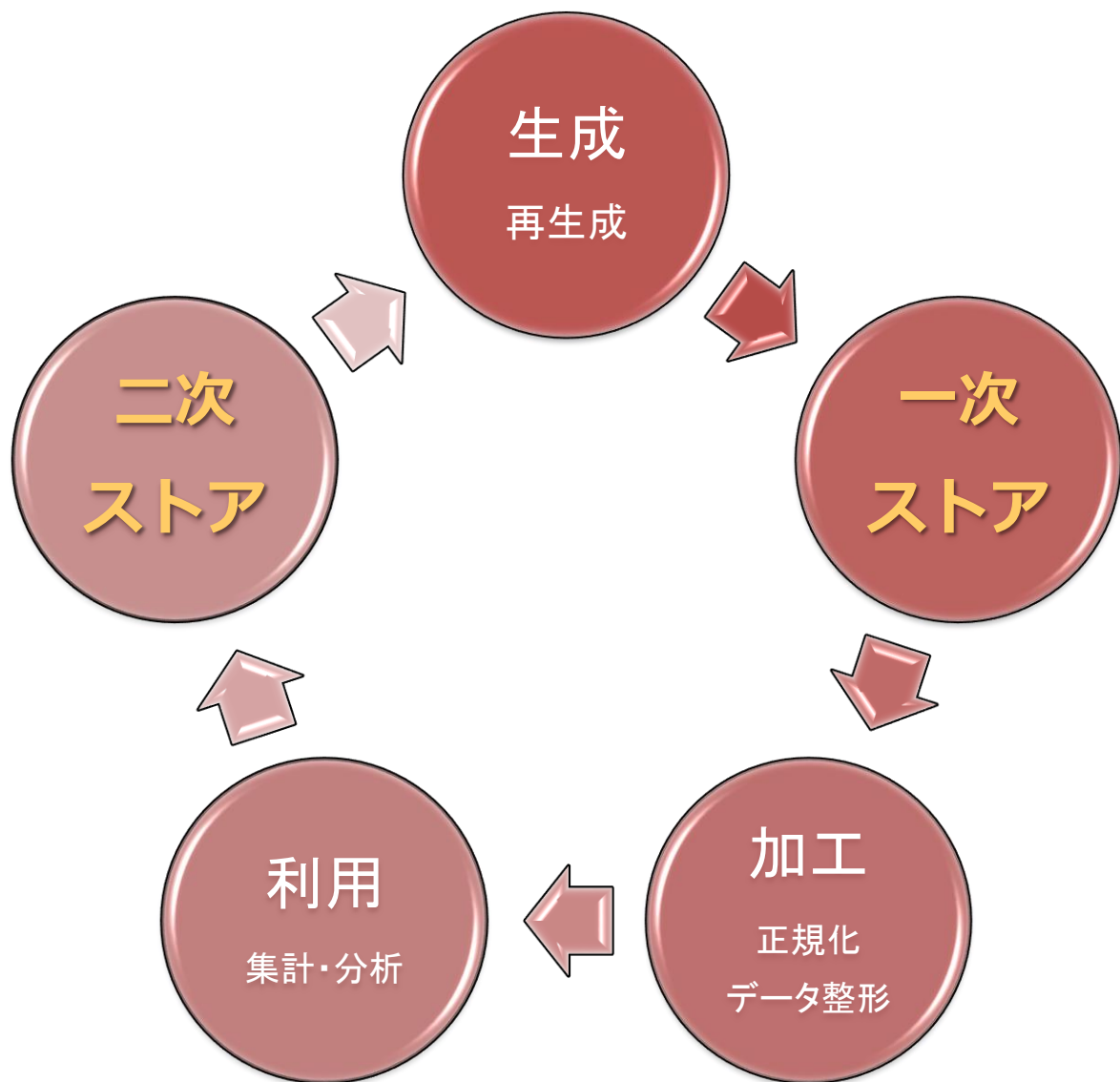
(出典)総務省「ビッグデータの流通量の推計及びビッグデータの活用実態に関する調査研究」(平成27年)



ビッグデータの活用からみる ストレージのあるべき姿

Storage should be,,,

ビッグデータのライフサイクル



- 図はライフサイクルの標準的なステップ。
- 二次ストアから加工/利用のステップにデータが供給されるケースもある。
- 二次ストアから再生成や二次加工に回らず終端するケースもある。
- 各ステップから除外・削除・消滅されるデータが存在。

ビッグデータストレージの分類と要件

■ ホットストレージ(HOT)

- ◆ 高IOPS
- ◆ 使い続けても落ちない性能が必要

■ ウォームストレージ(WARM)

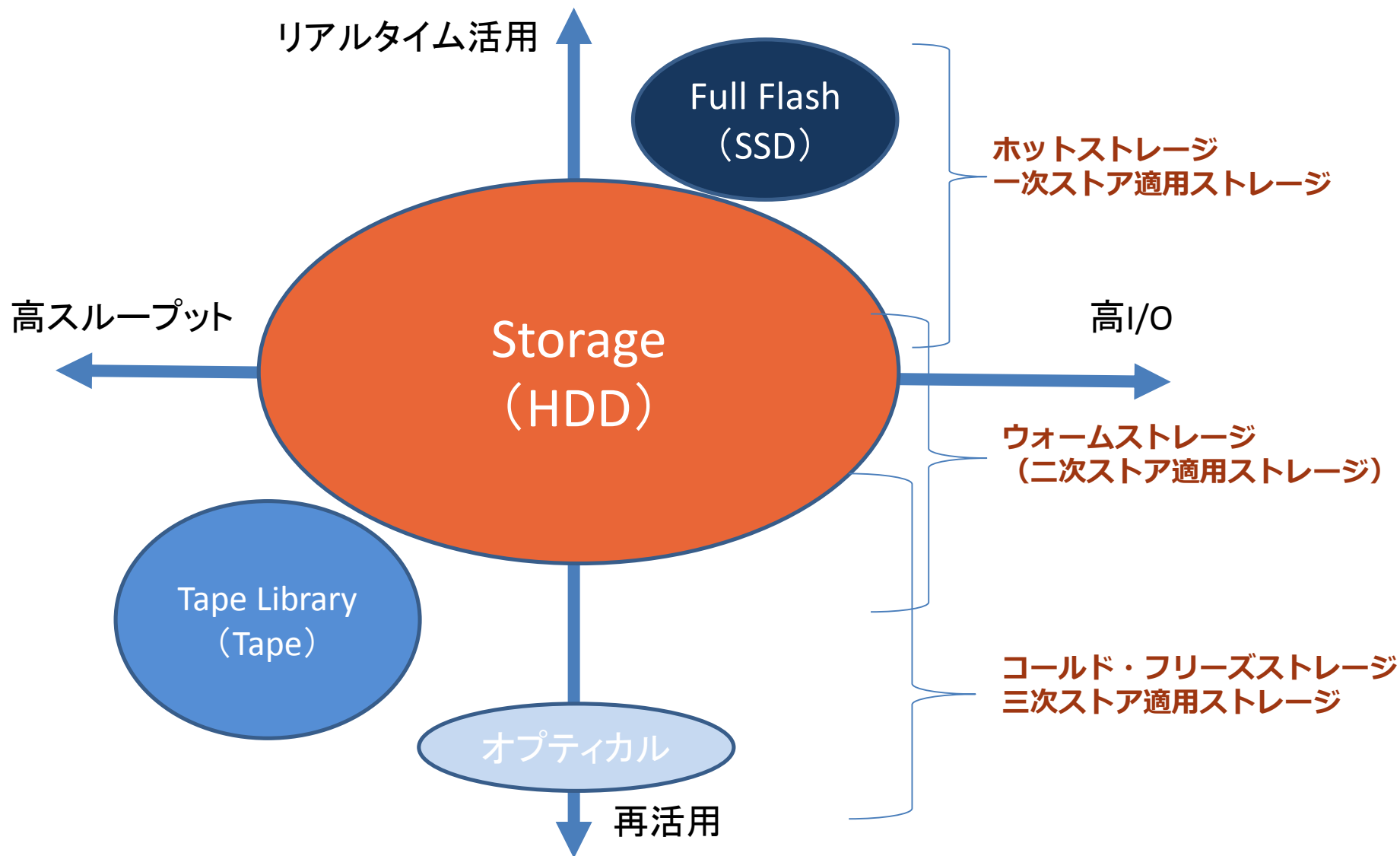
- ◆ スケールアウト(ホットにもコールドにも必要)
- ◆ 高スループット
- ◆ 可用性の確保
- ◆ データの完全性を確保
- ◆ セキュリティ
 - ▶ 消去の完全性・機密性

■ コールド・フリーズストレージ(COLD)

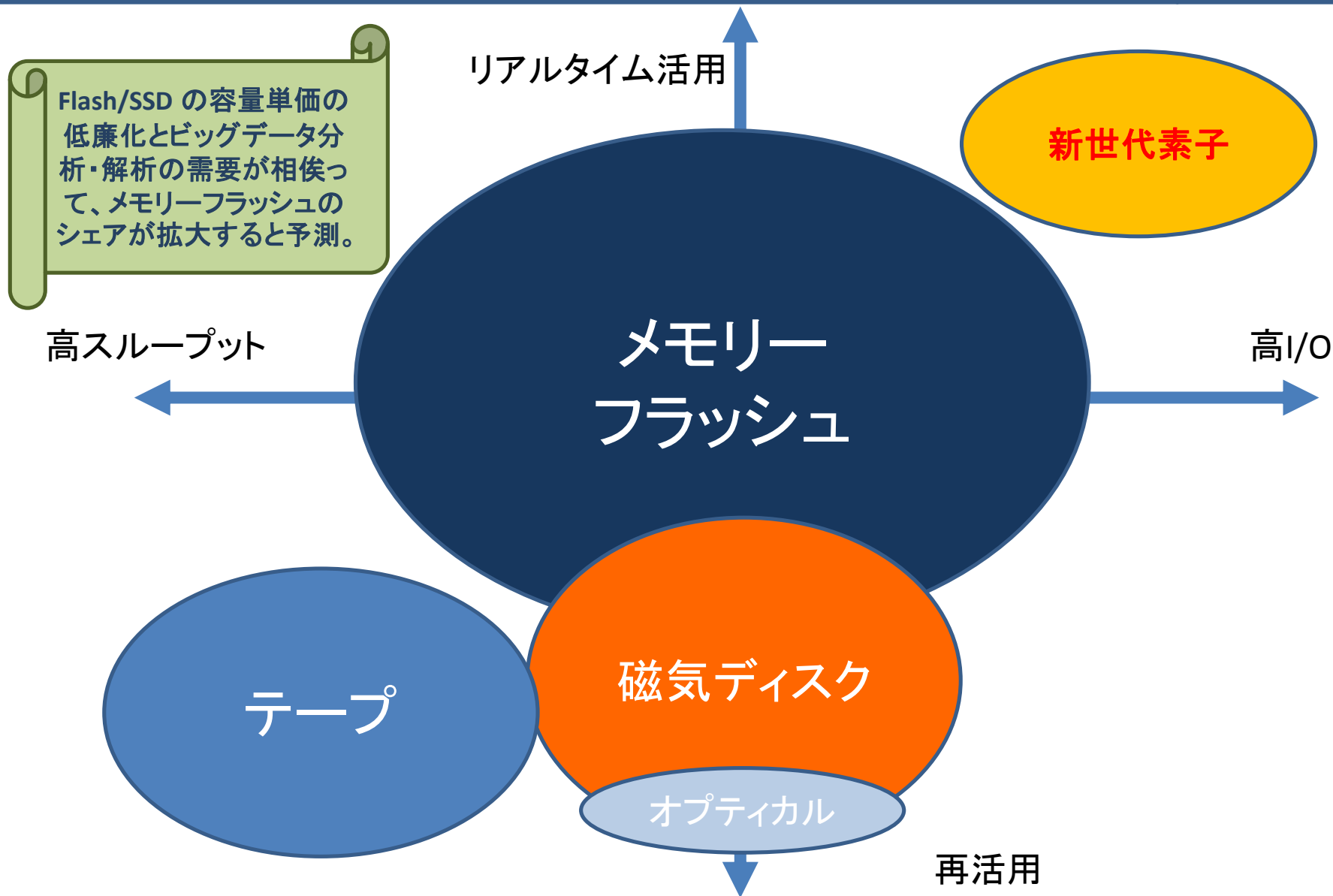
- ◆ データの完全性を確保
- ◆ セキュリティ
 - ▶ 消去の完全性・機密性

・インフラへ簡単アドオン
・壊れても即復旧できる
・考慮点:セキュリティ
(消去の完全性・機密性)


媒体とストレージの位置付け・・・現在



媒体から見た分類とその特性・・・数年後こうなる？



- データが壊れない(壊れても意識しなくて良い)
 - ◆ データロストは絶対なくしてほしい
 - ◆ 止まらないでほしい(いつでもアクセス、自動復旧)
- 設定が簡単
 - ◆ インストールが簡単
 - ◆ 特殊な技術が必要ない
- 運用が容易
 - ◆ 煩わしい運用を簡潔にしたい
 - ▶ バックアップ→できればやりたくない
 - ▶ 災害対策→できれば考えたくない
 - ▶ 容量追加→LUNとか嫌
 - ▶ 設定変更→手順が面倒
 - ▶ データ移行→できればやりたくない
 - ▶ ファイルシステム→色々あって大変、互換性も無い
 - ◆ 何かの性能指標を入れると自動的に設定してくれる
 - ▶ SDS注目の理由
 - ▶ フラッシュが人気の理由



オブジェクト
ストレージにより
解決しつつある

- ますますビッグデータ活用は広まっていく
- データ管理**は今でも大変だが、ビッグデータ活用により、ますます大変になっていく

新しいテクノロジーを使った**自動化**や
昔からの課題克服がビッグデータを支える
ストレージ基盤になっていくと考えられる

ビッグデータ分析から見たストレージ

■ 適材適所

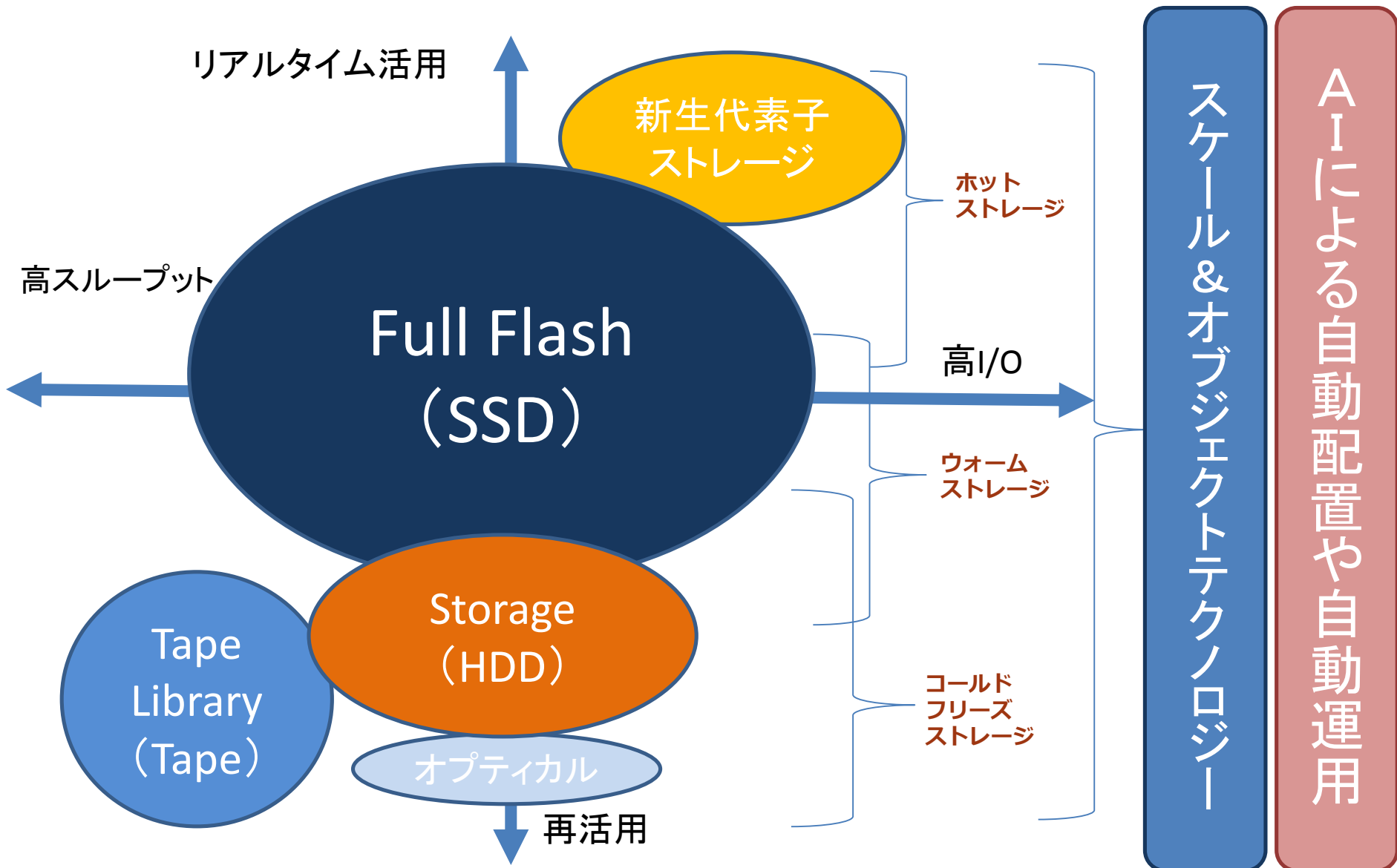
- ◆ 分析 → 一定の性能が必要
- ◆ データ保管 → ストレージの階層化が必要

■ 互換性の担保

- ◆ ファイルシステムの障壁
 - ▶ ビッグデータ解析の場合には中間データの部分が重要
- ◆ APIはあるが、色々な思惑がありストレージ間の互換性が生まれにくいケースが多い
 - ▶ ストレージベンダー、アプリベンダー等

自動化や課題克服がビッグデータを支える
ストレージ基盤になっていくと考えられる

媒体とストレージと未来に必要な技術



Artificial Intelligence

- AIによる半自動化を使ったITシステムの中のストレージ
 - ◆ **ストレージを意識しなくてもデータ管理・分析ができる**
 - ▶ 例) 問題点を音声でヒアリングし調査結果を出してくれる
 - ▶ 例) 必要になる性能指標を出してくれる
 - ▶ 例) アプリケーションに最適化してくれる
 - ▶ 例) 地理情報や気象データから災害対策のリスクを数値化してくれる
 - ▶ 例) 新旧技術の互換性を吸収
 - ◆ **適材適所にストレージリソースを自動分配してくれる**
 - ▶ 例) データを最適配置しコストバランスを鑑みた構成を推奨してくれる
 - ▶ 例) 必要なデータを必要な時に、必要なストレージリソース上へ移動させてくれる

現状では？…Nimble Storage InfoSightの例

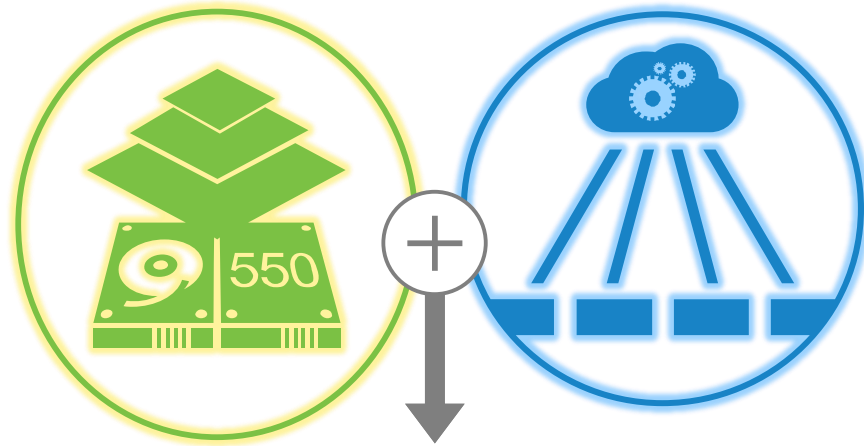
ハイブリッドを前提に作られた1つのアプローチ（予測し、キャッシュに配置）

CASL

Cache Accelerated Sequential Layout

新鋭システム機構

Flash-Optimized System Software



InfoSight

稼働しているNimbleの挙動を全て蓄積・分析し、性能をアプリ毎に最適化例) DB, VM, MAIL等

Nimbleが提供する4つのバリュー

デバイス(SSD, HDD)の性能を最大限に引き出す効率性



圧倒的に優れたコスト当たりのパフォーマンスと容量

CPU オフロード型
スピンドル・バウンド型

必要な分だけ追加出来る優れた拡張性



小さな規模から大きな規模までシステムを止めずに柔軟に追加可能な拡張性

Scale to Fit
必要な領域のみ拡張

追加ライセンス無しで使える強力なデータ保護



迅速なバックアップとリカバリ
VMware認定
VAAI対応/SRM対応

極小スナップショットブロック
VMware フレンドリ

簡単なオペレーション



障害を未然に防ぐプロアクティブなサポートと容易な運用性
(予兆検知/標準保守)

自律型・傾向解析・予兆検知
ストレージ SNS

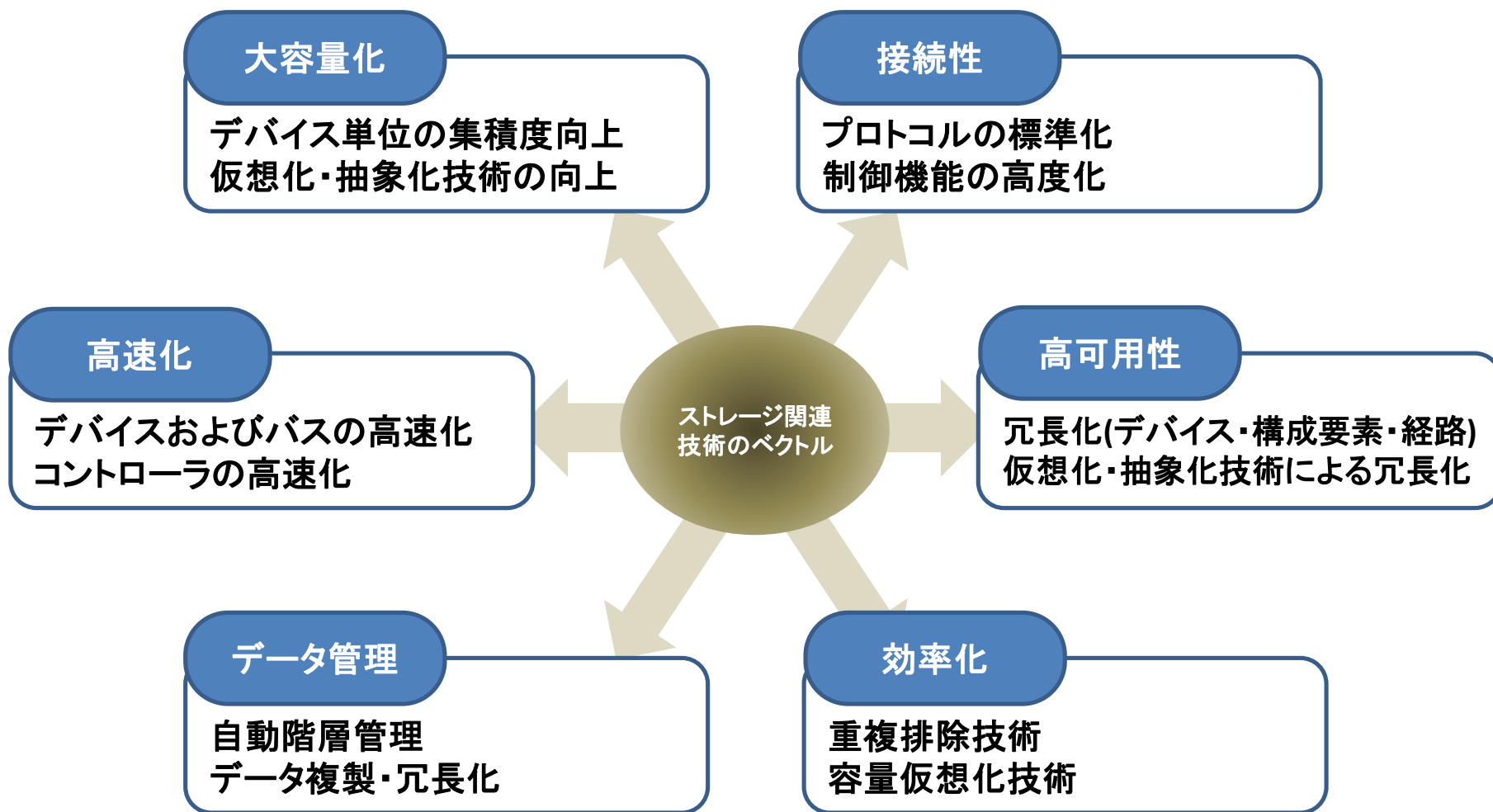


これからの新技術を予測

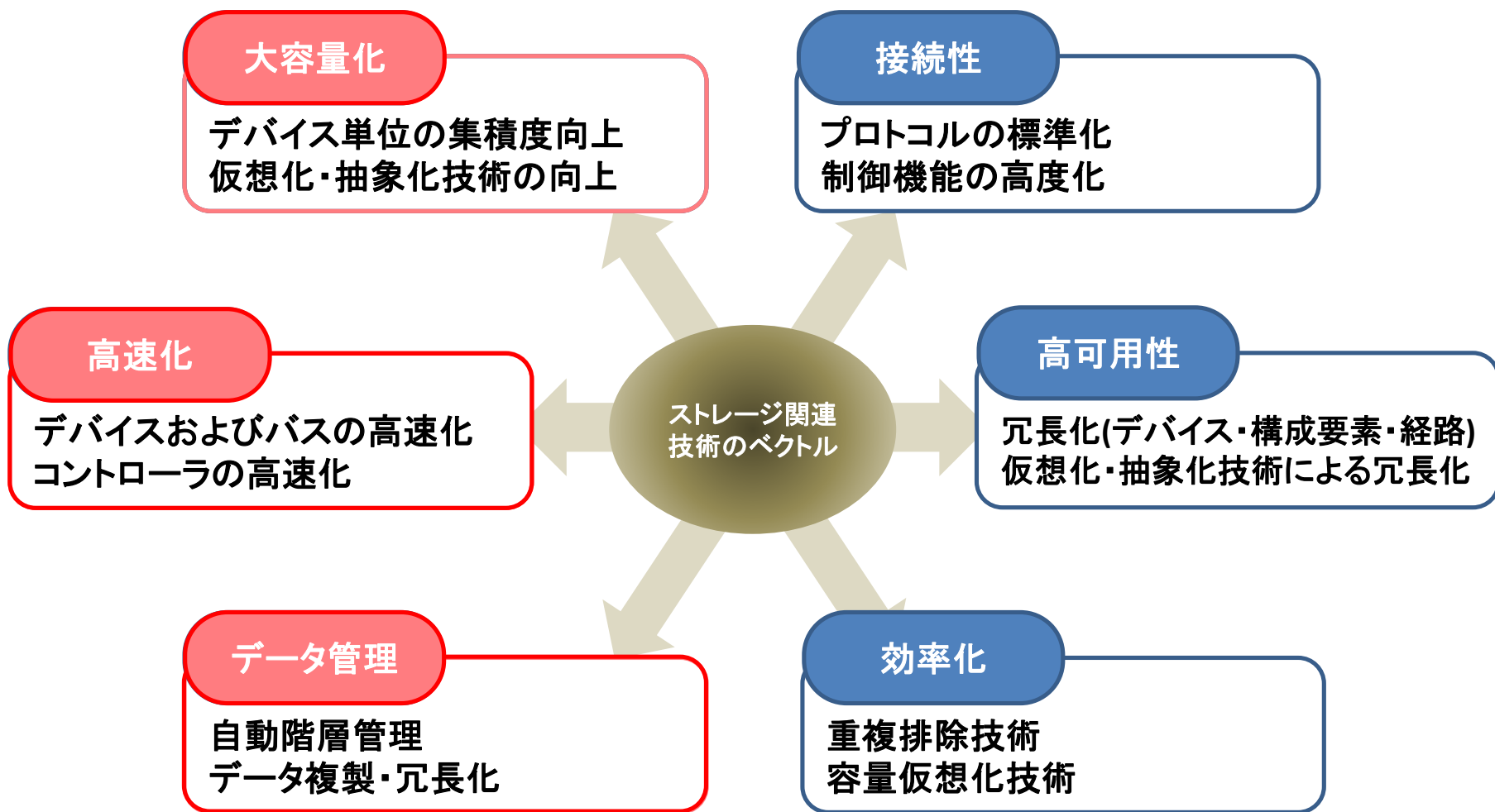
Predicting the new technology

JDSF 新春セミナー
2016年 1月27日

これまでのストレージを取り巻く 技術進化のベクトル



ビッグデータに関連する技術



ストレージ利用の効率化と運用性向上

大容量化

デバイス単位の集積度向上
仮想化・抽象化技術の向上

高速化

デバイスおよびバス的高速化
コントローラ的高速化

データ管理

自動階層管理
データ複製・冗長化

もっと使いやすく！

BigDataの活用効率化

- ✓ 大量に蓄積されたデータをいかに効率よくストレージ内で自動配置するか？
- ✓ 大量に蓄積されたデータの解析負荷軽減のため、前処理をどこまでオフロードできるか？

ストレージの運用性向上

- ✓ 最も効率的な運用をストレージが自律的に考えて自動化することができるか？
- ✓ デバイスや構成モジュールの状態をモニタリングし傾向分析することで障害予測をできないか？



機械学習 (AI) による ビッグデータの検知・分類・推奨

Machine Learning
for BigData Storage

JDSF 新春セミナー
2016年 1月27日

データの爆発的増加で発生する 業務課題

- 多種多様なデータが増加する中、
- 量的にも、質的にも人手作業では対応困難な業務が出てきています。

大量データ処理の課題が顕在化

“ヒト”では量が多くて
対応しきれない

- 手作業では処理しきれない
- 全てのデータに目を通せない

人の判断で
は限界



多様化しており、必要な
データの判断が難しい

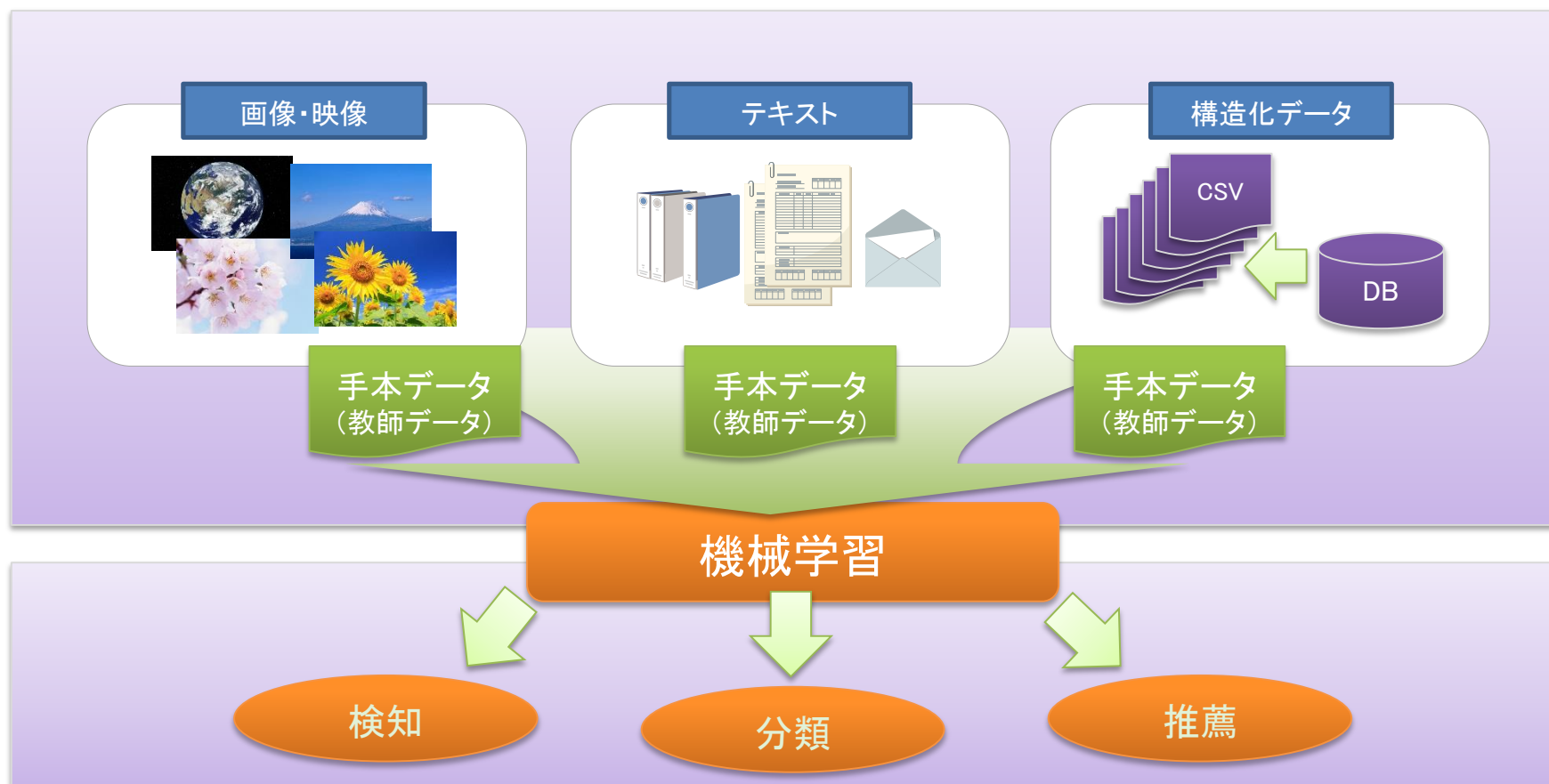
- 見るべき項目が不明
- 全ルールを精査できない

機械学習(ディープラーニング)

人に代わる「人工知能技術」を活用することで、**業務効率化を実現!**

機械学習とは

人間が自然に行っている学習能力と同様の機能を
コンピュータで実現しようとする技術・手法
お手本データを学習することで、分類/検知/推薦などの判断を実現





画像解析・データ分析への応用

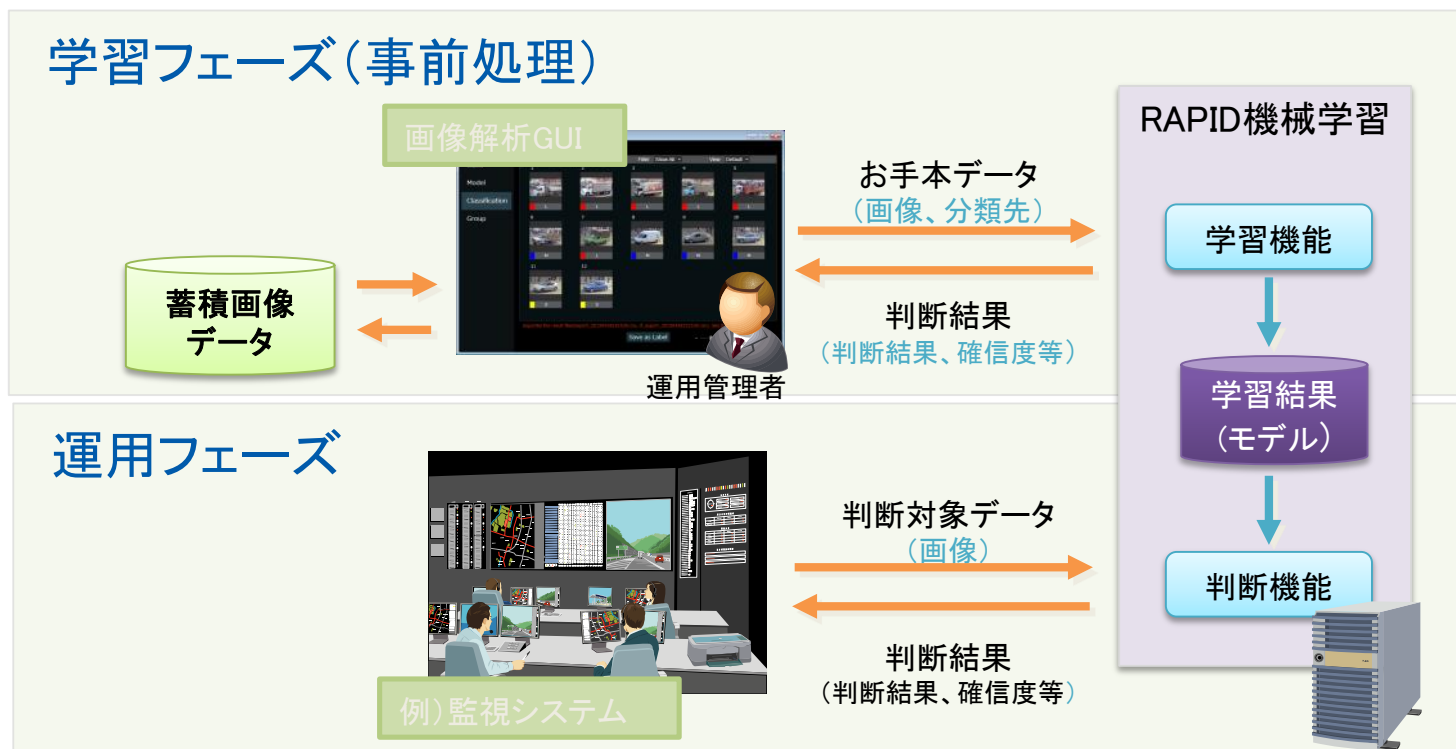
Application to Data Analytics

JDSF 新春セミナー
2016年 1月27日

NEC Advanced Analytics による RAPID機械学習 (画像解析)

■ヒトの判断を支援する高精度な画像分類機能

- ◆ 簡単に判断モデルを生成可能
 - ✓ お手本画像のラベル付け(分類情報選択)、学習、結果確認をGUIで提供
- ◆ 専門家の判断に基づく手本データを学習させることで、高精度な判別/検知が可能



- 適用先: 交通監視、店舗内監視、工場の検品など
(監視員の業務負荷軽減、人件費削減、検知精度向上、監視品質の安定化)

”イチ押し”ポイント

人工知能で ヒトの判断をサポート

お手本データを大量に学習させるだけで、複雑なルール設定は一切不要。業務専門家の判断に基づくデータを学習させることで、初心者でも高度な判別が可能になり、業務を効率化できます。

世界トップクラスの 軽量・高速性

機械学習(ディープラーニング)の研究分野で世界トップクラスであるNEC北米研の技術を製品化。軽量・高速なため、オンプレミス型のシステム構築が可能であり、「個人情報など機密性の高いデータ」「データサイズが大きく通信に制限がある画像・映像データ」も安心して活用できます。

まずは評価版(有償)で 効果を検証

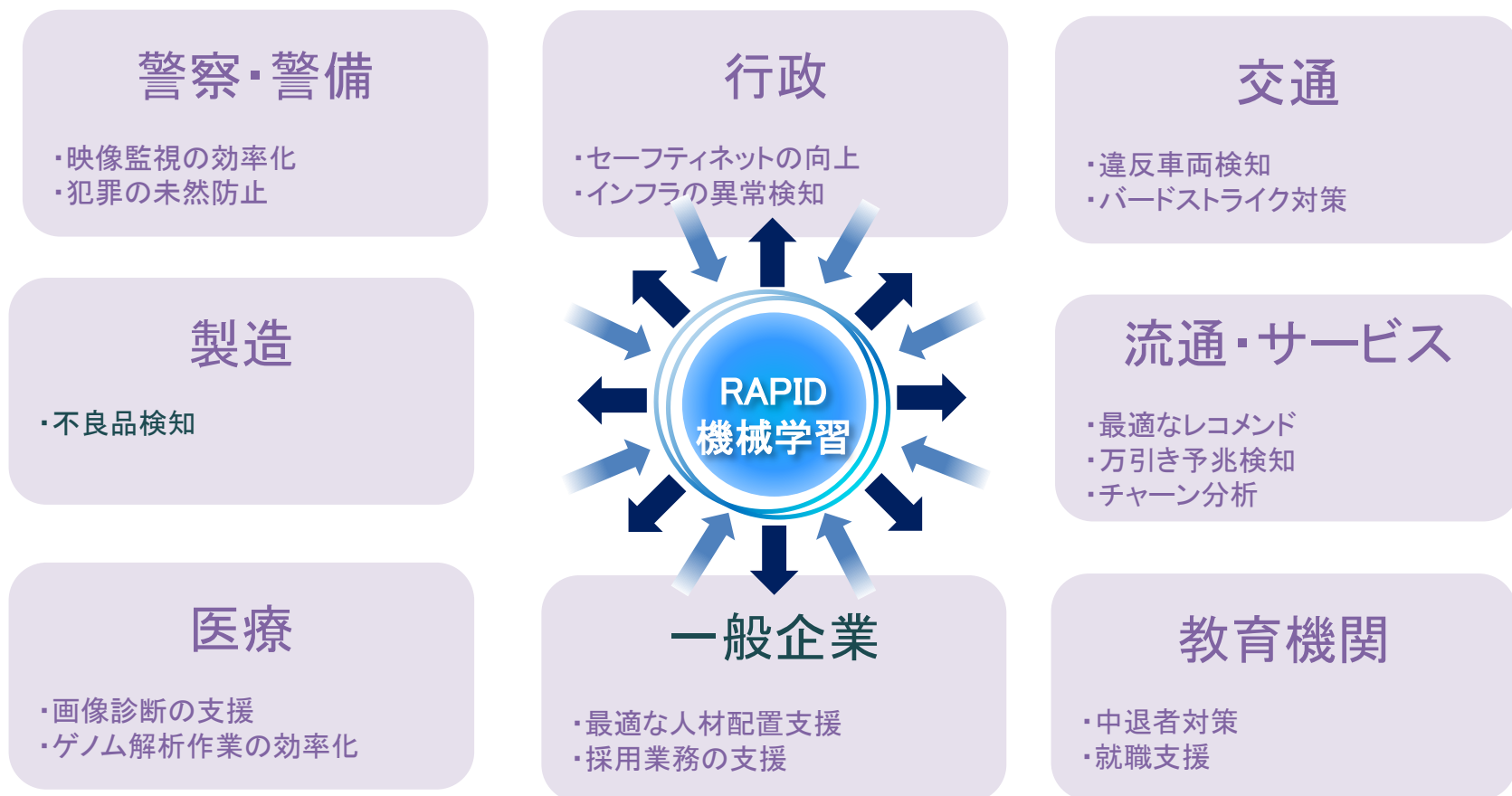
「機械学習って本当にヒトのように判断できるの!？」と不安なお客様に効果(予測精度)を検証してもらえよう、有償の評価用ライセンス(1か月)をご用意しています。

適応分野・ソリューション例

- 画像解析版: 交通監視、店舗内監視、工場の検品 など
(監視員の業務負荷軽減、人件費削減、検知精度向上、監視品質の安定化)
- 人材マッチング版: 人材仲介業、企業の採用・人事業務、大学の中退者予測 など
(仲介成約率の向上、採用・人事業務効率化、優秀人材の選考漏れ防止、中退者予防)

RAPID機械学習の適応分野ソリューション例

- 従来「人」が行っていた多様な判断作業を自動かつ高速に実現
 - ◆ 画像/映像やテキストなど非構造化データに対応し、様々な領域に適用可能





データ検索・保管管理への応用

Application to Data Search & Store

JDSF 新春セミナー
2016年 1月27日

機械学習技術を利用した社内文書検索 により高効率な検索を実行

課題

社内フリーアドレス化

社内文書の電子化

膨大な情報資産を
いかに活用するか！

解決策

エンタープライズ検索エンジン
「Neuron」を活用して社内文書
検索システムを構築

「Neuron」はブレインズテクノロジー株式会社の商標です。

欲しい情報を素早く検索

管理者は行動履歴から業務を見直し

イメージからも文書検索

導入効果

想定外の情報も入手できるようになり、通常10%程度の利用率
と言われている検索システムにおいて54%の利用率となった。

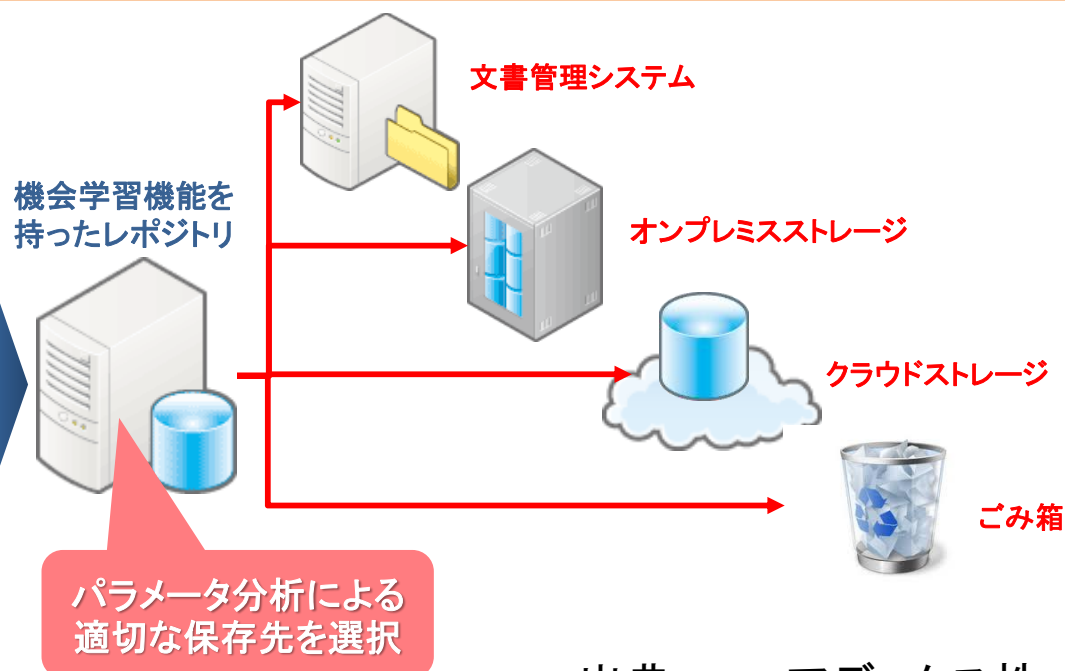
機械学習技術を利用した 社内文書分類と保管

機械学習技術を利用した検索システムの採用

- ✓ ファイルの種類、利用状態、重複有無等、様々な視点でレポジトリを見える化
- ✓ 分類した状態から対象ファイルの中身まで確認が可能
- ✓ 分析レポートを部門管理者に共有し、部門単位の整理・整頓を支援
- ✓ 特定のルールで分類したデータを用途に応じて保存先の変更(移動・削除)が可能

多様なパラメータで レポジトリの可視化・分析

- 拡張子
- ファイルサイズ
- 作成年
- 最終アクセス
- 最終更新
- オフィスファイル
- 圧縮ファイル
- メディアファイル
- 重複
- etc.





障害検知・予測への応用

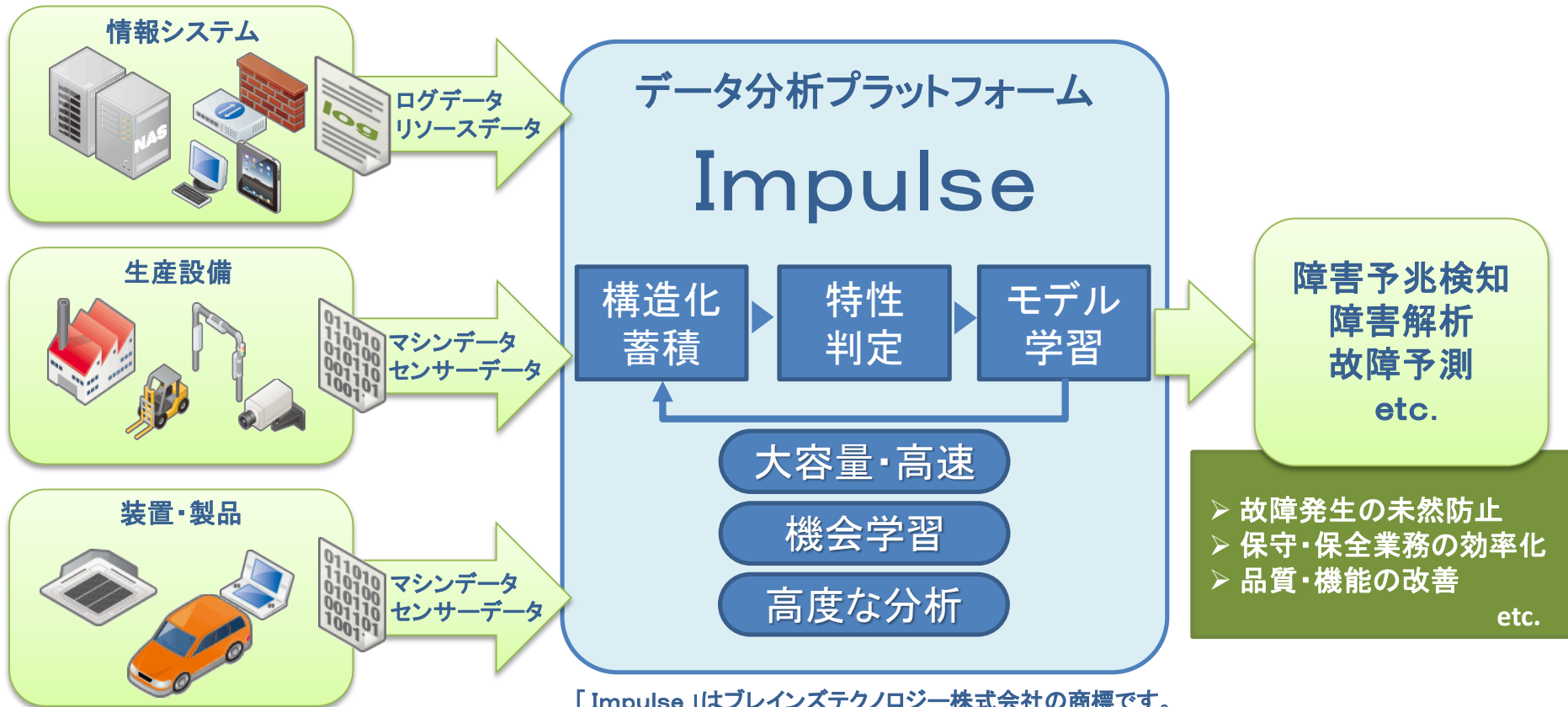
Application to Fault Detection & Prediction

JDSF 新春セミナー
2016年 1月27日

障害検知・予測技術への応用

データ分析プラットフォーム Impulse

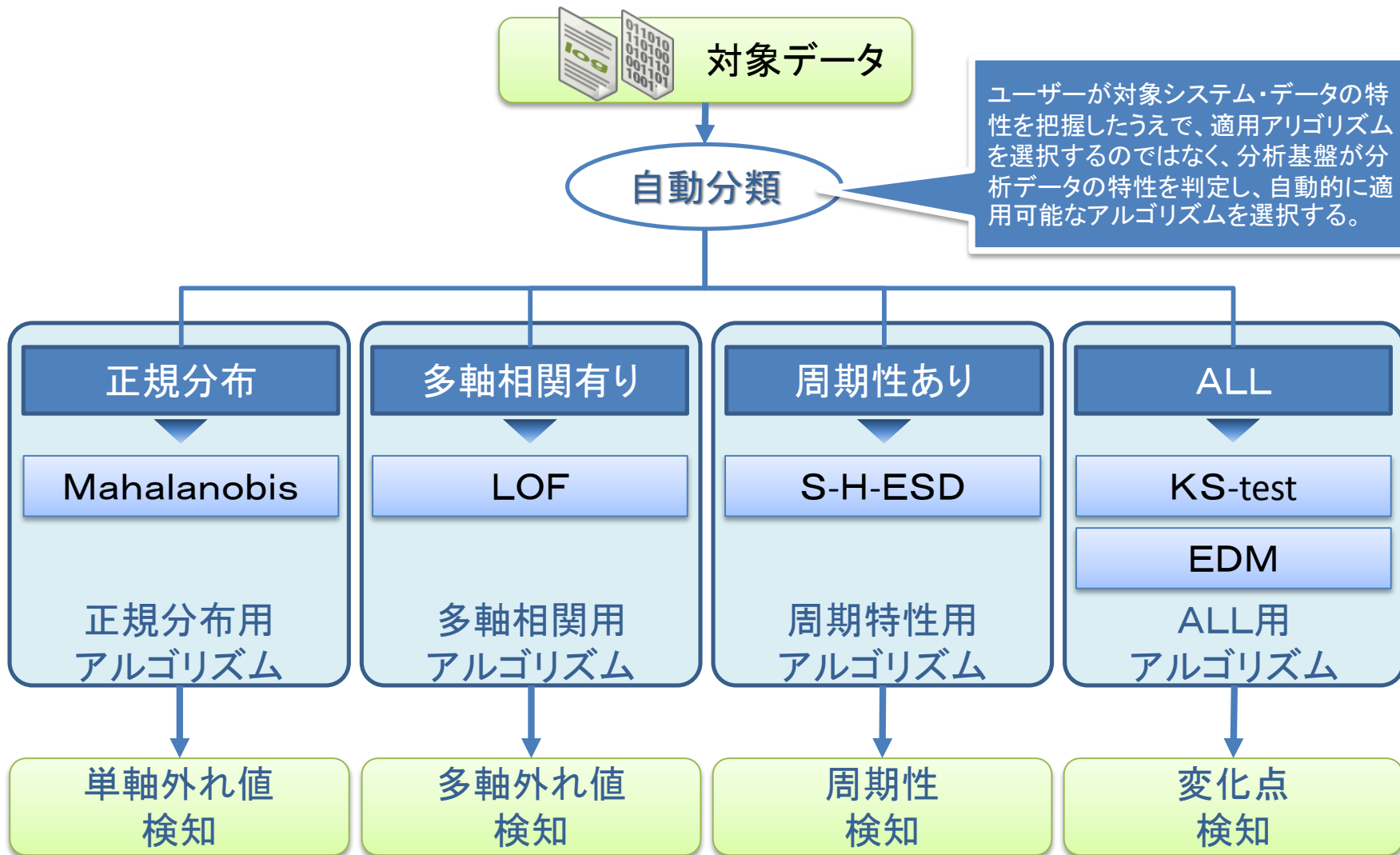
- ・企業内に蓄積されたデータを最大限に活用
- ・機会学習による深い分析と高速処理で「予兆検知を実現」



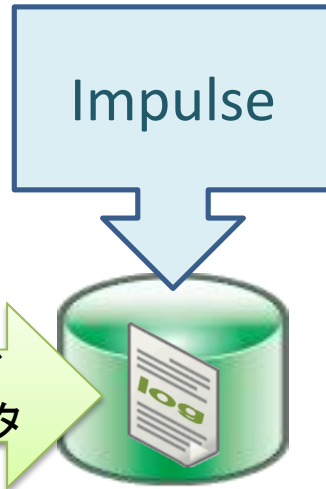
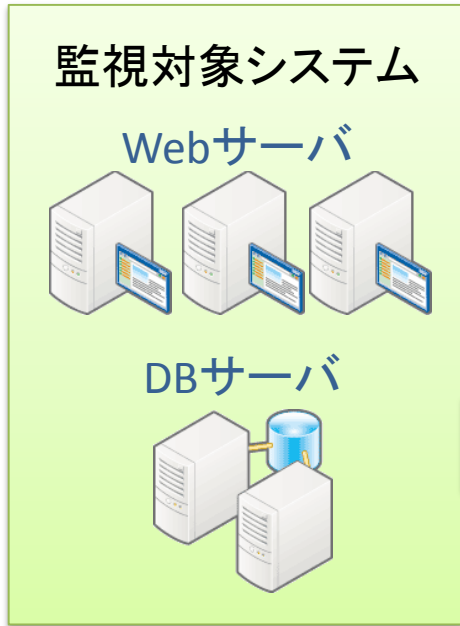
「Impulse」はブレインズテクノロジー株式会社の商標です。

出典: ユニアデックス株式会社

分析アルゴリズムの選択を自動化



実証実験で異常検知を確認



- ✓ データベースサーバーと3台のWebサーバーのログデータを基にImpulseが故障につながる異常を正しく検知できるかを調べた。
- ✓ 実際に起こった障害状況
 - ① 検索システムのデータベースサーバーのログは、正常時には周期的な特性があった。
 - ② 障害時、データベースサーバーのクエリ処理件数が落ち最終的に夜間バッチ処理が不能になった。
 - ③ 熟練技術者が後からログデータを見ると、周期性を外れる予兆が生じていた。

- データベースサーバの障害に対して、Impulseはログデータに周期性があることを自動判断し「**周期特性用アルゴリズム**」を使って異常を検知した。
⇒ 障害発生の**7日前に障害を検知**することができた。
- 3台のWebサーバーの障害に対しては、ログデータ特性から各データに相関関係があることを見抜き、「**多軸相関用アルゴリズム**」を自動選定していた。
⇒ 検証結果では、Webサーバーが故障する**12日前に異常を検知**できた。



エピローグ

Epilogue

JDSF 新春セミナー
2016年 1月27日

■ビッグデータWGメンバー

◆ パルテック

◆ シーティーシー・エスピー

◆ 日本電気

◆ アライドテレシス

◆ ユニアデックス

上原様

落合様

力石様

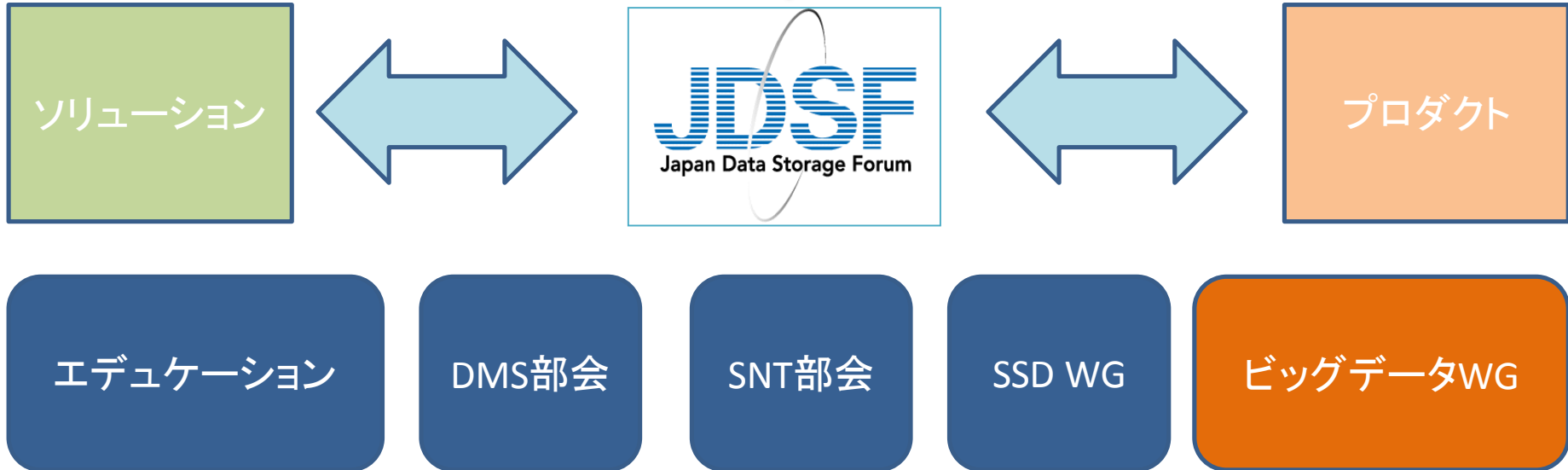
延原様

高木

ご参加お待ちしております

- 情報を取りまとめて一緒に発信しましょう
- ご参加されたい方は事務局へご一報ください

ITテクノロジー／バズワード／トレンド





ご清聴ありがとうございました。

Thank you!

JDSF 新春セミナー
2016年 1月27日