





# 保有データと課題



# 企業の保有データの分類

## 【 非構造化データとは 】

- 一般ファイルで格納されているデータ(オフィスファイル、Eメール、画像等)
- データ量は日々増加するが、更新頻度およびアクセス頻度は低い
- J-SOXを始め法規制等で原本保証と長期アーカイブの対象
- 処理能力よりは柔軟な拡張性とスムーズなアクセス性が要求される
- コストパフォーマンスが優先され、高価なストレージまでは要求されない

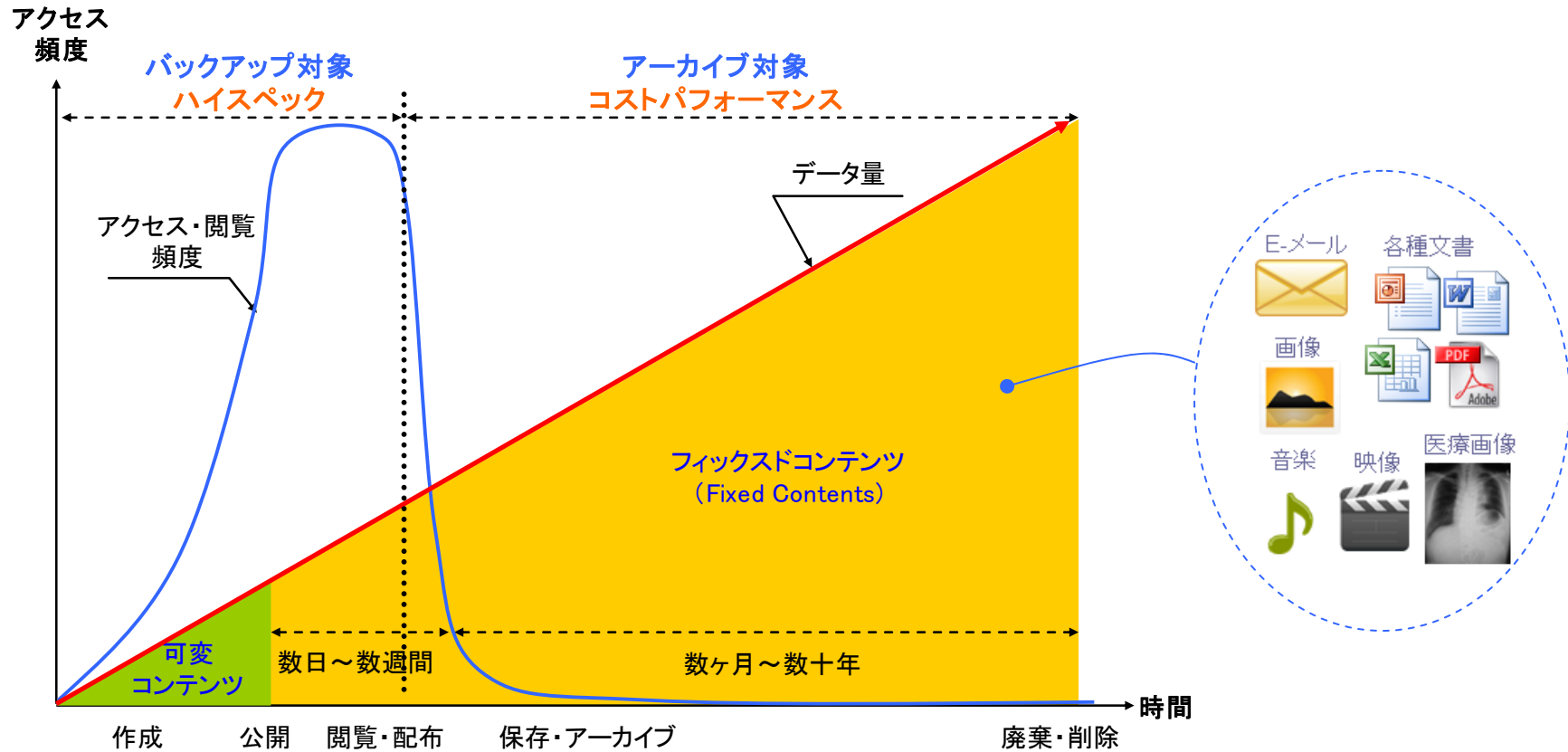


## 【 構造化データとは 】

- 主にDBなどに格納された基幹系業務で使用されるデータ(DB/ERP/CRM等)
- データ量は急増しないが、更新頻度およびアクセス頻度が高い
- 事業継続性の保証のため、バックアップ/リストアの対象
- CPUやディスクI/Oの高い処理能力が求められる
- 一般的には高価なハイスペックのオンライン専用ストレージで運用



# 非構造化データのライフサイクル特徴

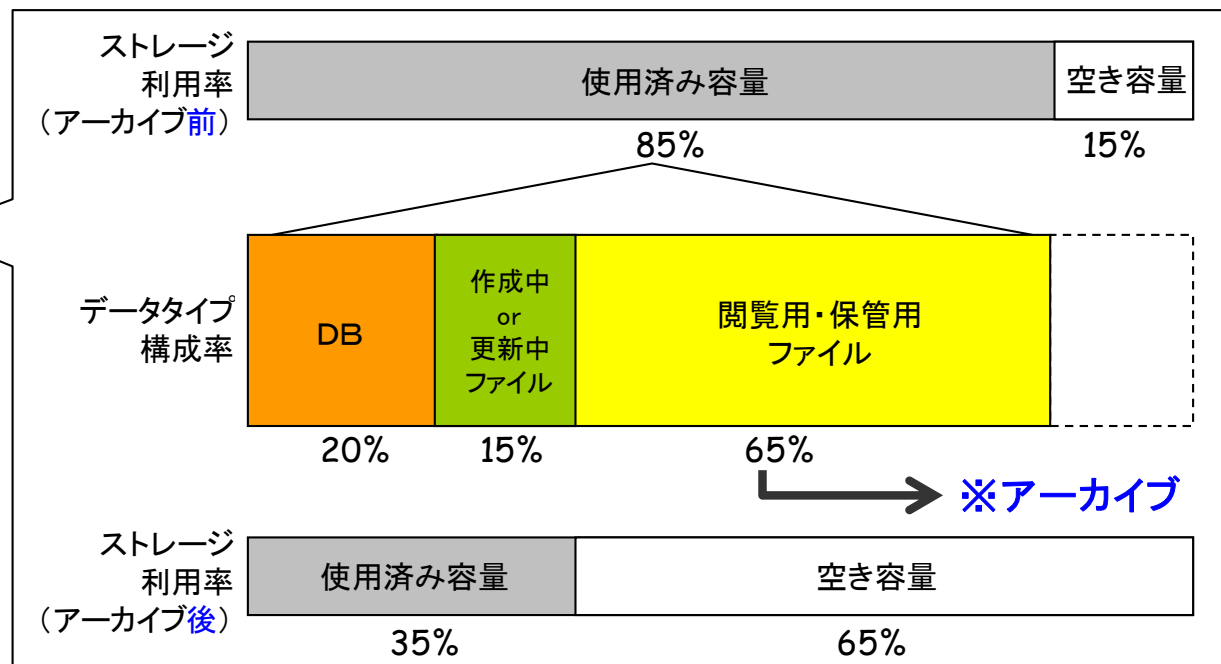
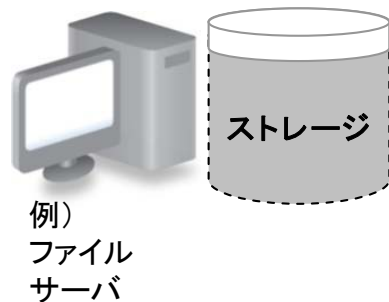


## 【 非構造化データのライフサイクル特徴 】

- 作成中のデータへのアクセスは、ほとんど作成者に限られる
- 活用（公開、閲覧、配布）期間は比較的短期でその期間にアクセスが集中し、その後は殆ど参照されない
- ファイルサイズは数百Bから数百GBまでと、非常に幅が広い
- 処理速度よりは必要な時に目的データへのアクセス性が重要視される
- コンプライアンス対応のため、長期保管が要求されるが、いつかはデータ価値を失い廃棄／削除される

ライフサイクルに適した**効率的なアーカイブ**が必要

# 非構造化データのアーカイブメリット



日々増え続けるデータによる空き容量の不足対策として、新しいサーバを単に追加したり、ストレージを統合するだけでは障害ポイントが拡大、増える一方で根本的な解決策にはなりません。



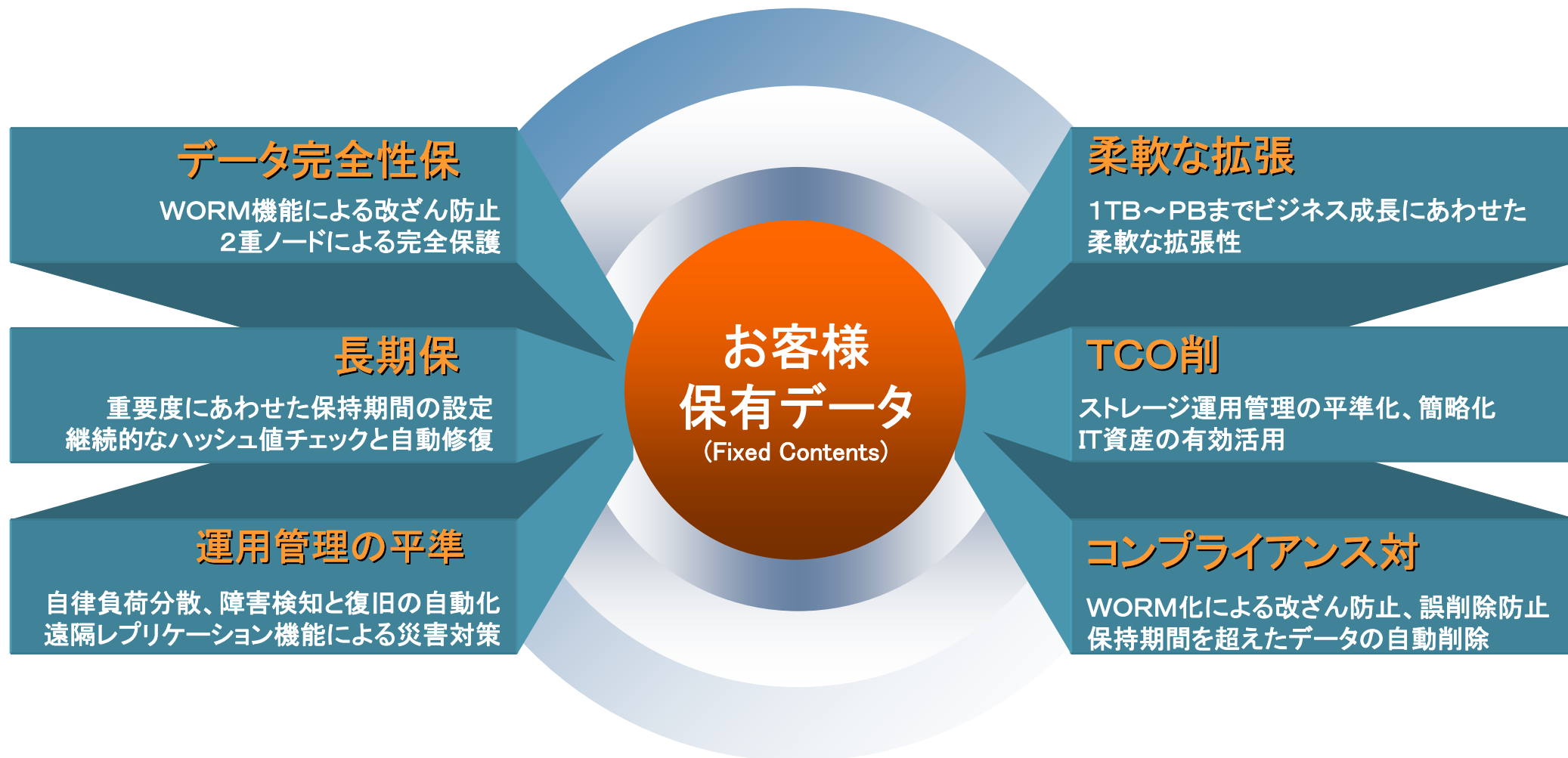
CASstor™

## 非構造化データのアーカイブメリット

- バックアップ時間の短縮
- バックアップ作業の運用管理負荷の軽減
- リストア時間の短縮による迅速な業務復旧
- 従来サーバのストレージ利用効率向上



# ストレージ運用(データ保管)に求められる要件



※フィックスドコンテンツ(Fixed Contents)とは

非構造化データのうち、既に完成された情報で、今後書き換えを行わず、長期的な価値を維持するデータ。  
その価値に沿った長期保存される情報を示します。



従来のストレージとバックアップ技術の代表であるNASやテープ装置では、急増する非構造化データ量への対応には限界があります。=>新しい仕組みが必要

### <ストレージ管理者が抱えている悩み(課題)>

- 日々急増するデータ量への対応
- 業務開始時間まで間に合わないバックアップウィンドウ(作業)時間
- リストアデータ量(時間)の増加による業務停止時間の増大
- 高いスキルと複雑さによるストレージ管理者の属人化と管理者の不足
- ストレージ運用管理の業務負荷と作業コストの増加
- 複数ベンダーの混在、サポート終了、部品調達の困難さ
- ブランチオフィスのデータ量の増大／災害対策／情報漏えい防止
- IT資産の有効活用とストレージ運用管理の平準化、シンプル化
- TCO削減
- J-SOXを始め法規制対応(コンプライアンス対応)

## ノードクラスタリングと単一ボリューム化技術を実装した次世代アーカイブソリューション

### <CAS:CAStor™導入によって得られるメリット>

- アーカイブによるバックアップ容量の縮小でバックアップ時間の短縮と作業負荷の低減
- 保存データを複数ノードにより多重化保護、障害時にも事業継続性向上
- 非構造化データの統合一元管理によるストレージ運用・管理の平準化
- ストレージ管理者のスキル不足や人員不足を自律管理機能と修復機能により解消
- 必要な時に必要な分だけ段階的にストレージを拡張(TB～PBまで)
- 既存IT資産やリソースの有効活用
- データごとの改ざん防止機能と保持期間設定によるコンプライアンス対応
- IA-64ベースの汎用HW使用で容量単価の削減とストレージ利用効率向上
- 遠隔レプリケーション機能による災害対策

#### ※CAS(Content Addressed Storage)とは

CASはブロックアクセス型(SAN)やファイルアクセス型(NAS)とは異なる、コンテンツオブジェクトアクセス型。

データ(例えば、画像データ)とメタデータ(画像データの属性:撮影日時、フォーマット、画像サイズ)などをパッケージ化したもので、ファイル名の代わりに、

固有のコンテンツID(UUID)を付与することで、コンテンツの物理的な格納場所をOSやアプリケーションから管理する必要がない。

コンテンツはコンテンツIDからアクセスされることで、シンプルな管理による、フィクスト・コンテンツの長期間に渡る安全な保存、容易な参照を実現。



クラウド・ストレージ  
carīngo **CAStor**<sup>TM</sup> シリーズのご紹介



- 2005. 3 Caringo, Inc創業(米国、テキサス州オースティン)
- Paul Carpentier(CTO), Jonathan Ring(Pres.), Mark Goros(CEO)
  - 各30年を超える業界経験、17社以上の創業に関わる  
Oracle, Siebel Systems, Sybase, TechGnosis, Wave Research, FilePool 等
- CAS(Content Addressed Storage)の巨匠
  - CTOのPaul Carpentierと、VP Advanced TechnologyのJan Van Rielは、ストレージ業界ではCASの父母と呼ばれているCAS概念と技術を確立、実装してきたメンバー。Jan Van Rielは、8年間EMC社Centera製品のDirector of Technologyで去年Caringo合流。

○ 1999	FilePool社時代に CAS基礎となる4つの特許申請
○ 2001	EMC社 FilePool買収
○ 2002	FilePool技術基盤にEMC社CenteraCAS(HW+SW)製品化
○ 2006	Caringo社 SoftwareのCAS製品CASTor™を製品化
○ 2009	CASTor™CSS、CFS、CRの製品ライナップ



# Caringo™ ストレージ専用ソフトウェア製品群

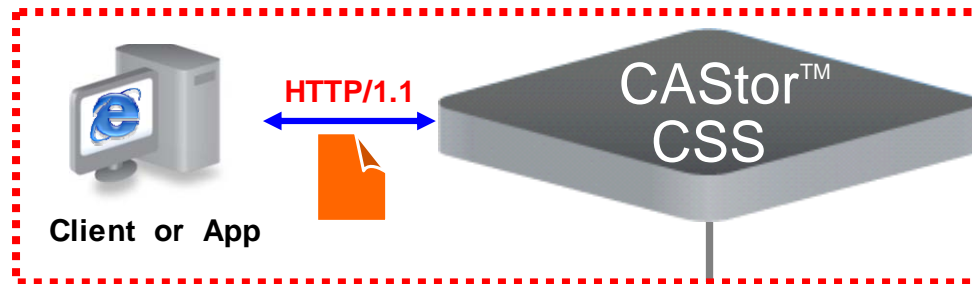
フィックスドコンテンツへの接続(CFS)、保管(CSS)、配布(CR)の3パッケージ



## Access

### CFS (Content File Server)

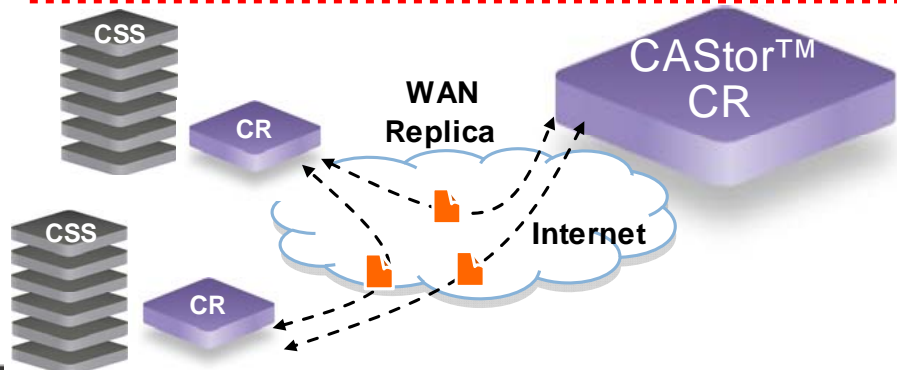
従来のアクセスプロトコル(CIFS/NFS等)を利用し、フィックスドコンテンツのRead/Writeを提供するアクセスゲートウェイソフトウェア



## Store

### CSS (Content Storage Software)

ノードクラスタリングと単一ボリューム化技術を利用し、フィックスドコンテンツの保管・保護に最適化したアーカイブ専用ソフトウェア



## Distribute

### CR (Content Router)

ルールベースでディザスタリカバリやミラーサイト構築を可能にした遠隔自動レプリケーションソフトウェア  
(1⇔1, 1⇔多, 多⇔多)

## ユーザ事例



### Vodafone360

ケータイとPC・Mac間でアドレス帳・SNSの連絡先・写真・音楽等を共有することができるポータブルの新しいWebサービスVodafone360を支えるプラットフォームとしてCAStorが採用されました。Vodafone360については<http://www.vodafone360.com>をご覧ください。



### DocuShare

ゼロックスECM(企業文書管理)製品DocuShareのストレージとしてCAStorが採用されました。大量の紙データをデジタルアーカイブとして保存するストレージプールとして、顧客のビジネスの成長に合わせて、柔軟にかつ経済的に容量を拡張できるアーキテクチャが評価されました。



### ジョンホプキンス大学

ジョンホプキンス大学の遺伝子研究データセンターで研究データのアーカイブを行うためにCAStorが採用されました。1日に生成される1~5TB(テラバイト)の遺伝子情報をアーカイブするためのストレージとしてスケーラビリティ・パフォーマンス・経済性が評価され、既に数100テラバイトのCAStorクラスターが運用されています。



### サンディエゴ・スーパーコンピュータセンター(SDSC)

カリフォルニア大学サンディエゴ校の1930-1940年代の歴史的重要な文献をアーカイブするプロジェクトを進めています。カリフォルニア大学のから集められた重要な文献はSDSCのCAStorクラスターにアーカイブされ、研究者がいつでもアクセスできるように管理されています。

## パートナソリューション



### SharePoint2007統合

SharePointのコンテンツをCAStorにアーカイブするSTEALTH Software製品  
<http://www.stealth-software.eu/>



### Alfresco (ECM)統合

AlfrescoのコンテンツをCAStorにアーカイブするコンテンツリポジトリインタフェイス(オープンソース)  
<http://forge.alfresco.com/>



### E-Mailアーカイブ

Messaging Architects M+SecutreStoreのE-mailアーカイブ製品  
<http://www.messagingarchitect.com>



### 総合ログアーカイブ管理システム

独自の圧縮、暗号化によりあらゆる機器のログの非改ざん性保証。リアルタイムSyslog監視と長期アーカイブ機能を統合したLogStare Tetraアーカイブ版  
<http://www.secuavail.com/>  
<http://www.logstare.com/>



### オンラインバックアップ

デスクトップから遠隔地のCAStorへのバックアップが可能なVembu TechnologyのバックアップソフトStoreGrid  
<http://www.vembu.com/storegrid/>



### 医療業界向けアプリケーション統合

医療情報管理アプリケーションEvercoreのデータをCAStorにアーカイブするソリューションを提供  
<http://www.teramedica.com/>



## ■ 特徴

1. ソフトウェアベースのフィックスドコンテンツ専用アーカイブソフトウェア
2. HTTPプロトコルによるファイルアクセス
3. IA-64ベースの汎用HWを利用したアーカイブストレージ構築
4. 複数ノードによる自動的なデータ多重化保護

## ■ 主な機能

### 1. インテリジェントな自己管理機能

自動障害検知、自動複製、自動修復、自律負荷分散を提供。

### 2. 優れた拡張性

IA-64ベースの汎用HWを使い、1TBからPBまで段階的な拡張性を提供。  
1ノード追加に約60秒の容易な増設。

### 3. セキュアな長期保管・コンプライアンス

データごとに保持期間／複製数／削除可否の設定が可能。

UUID (universally unique identifier)による効率的なデータ管理。

保持期間を過ぎたデータの自動削除、WORM機能による改ざん防止、誤削除防止。

### 4. 機密性

データに付与した識別ID (UUID、128ビット)を知ったユーザのみアクセス可能。

### 5. 完全性

内部ヘルスプロセス (Health Processor)による保管データの完全性を常時チェック。

### 6. 可用性

ノードクラスタリング構造による常時データアクセス保証、多重データ保管、無停止HW増設を提供。

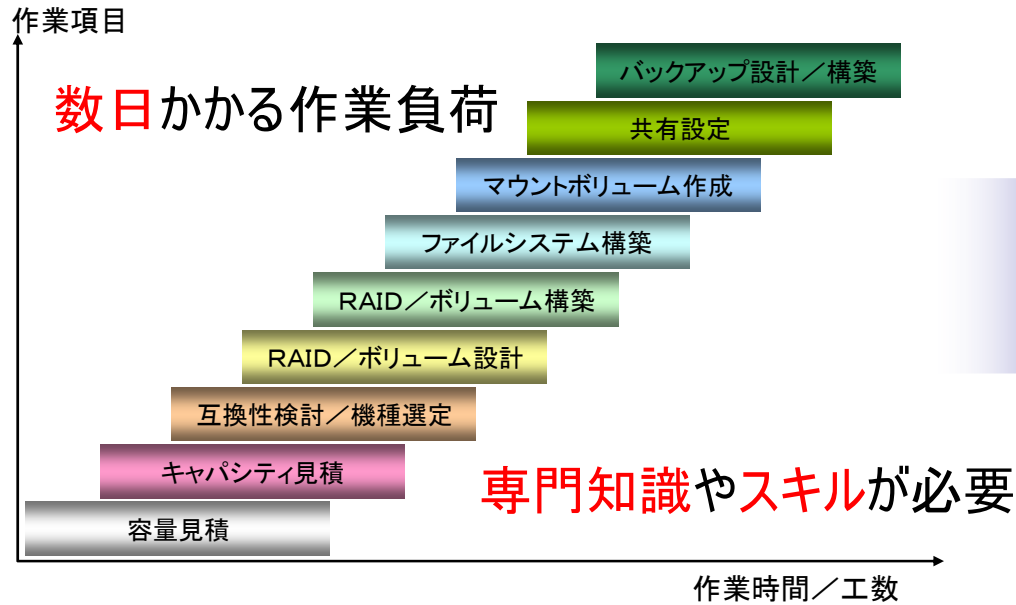
### 7. インプリメントの容易さ

業界標準プロトコルHTTP/1.1に準拠し、データの書き込み (Write)、読み込み (Read)等が可能。



# ストレージ管理に専門スキル不要 : CAStor™ CSS

従来ストレージ増設時の作業項目 (現状)



CAStor™ CSSストレージ増設時の作業項目 (導入後)



ストレージ運用管理工数を大幅に軽減

強力な自己管理機能とシンプルな仕様により専門要員を必要としない



# ストレージ運用管理負荷の軽減 : CAStor™ CSS

- バックアップ作業
- リカバリー作業
- ファイルシステム構築
- ストレージ見積、機器選定
- ボリューム設計・構築
- システムごとの仕様管理
- システムごとの保守／部品管理
- システム／製品ごとのVerUP
- システムごとの容量管理
- システム／製品ごとのスキル習得
- ストレージパフォーマンス管理
- データ完全性確認
- 不要データ削除
- 情報漏えい有無チェック
- 改ざん有無チェック
- テープの入れ替え／保管管理
- 各拠点のデータバックアップ
- .....等々

- ・ 自律的な負荷分散
- ・ 自動修復機能
- ・ データ二重化保護管理機能
- ・ 保持期間／データ削除機能
- ・ 継続的なヘルスチェック機能
- ・ 単一ボリューム／オンデマンド拡張

CAStor™

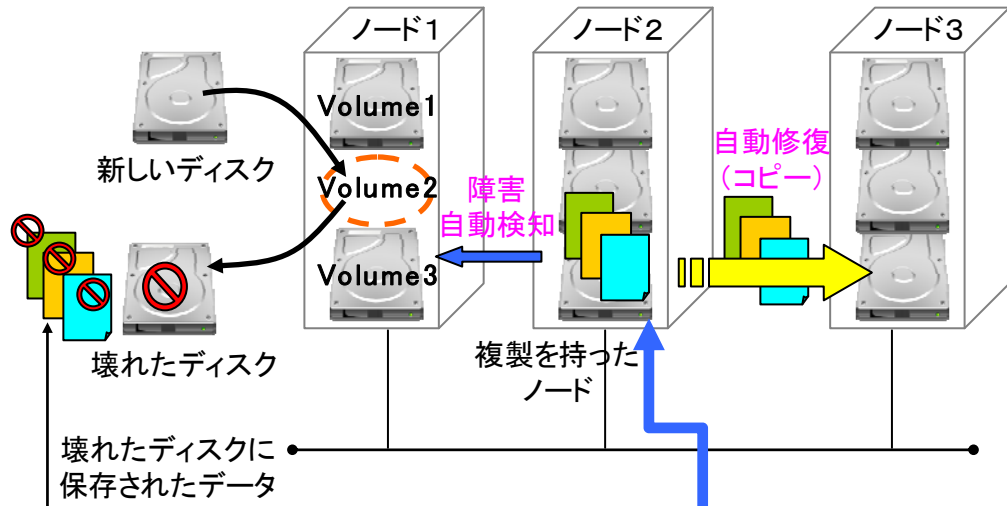
- CAStor™ CSSバージョンUp
- メモリ増設／ディスク交換
- ノードの追加／撤去
- ストレージ容量チェック





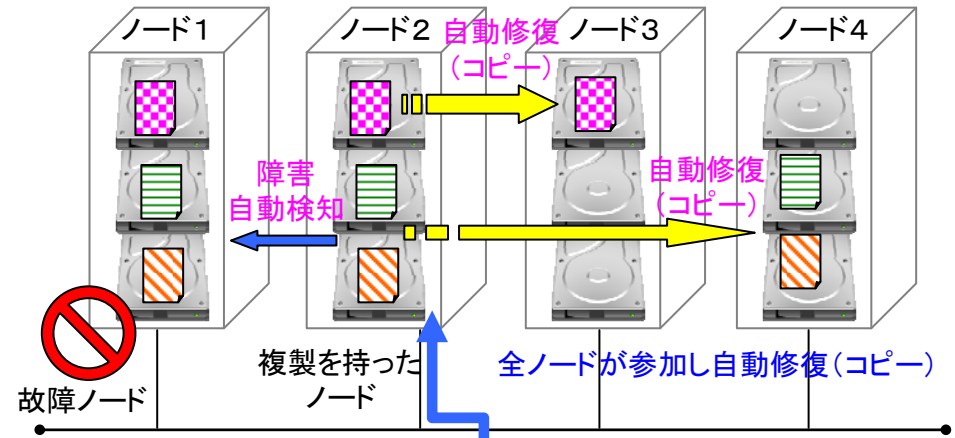
## ディスク障害の自動検知と修復

データは常に2重ノードによって自動保護されます



## ノード障害の自動検知と修復

データは常に2重ノードによって自動保護されます。障害検知した場合、全ノードがデータ復旧に参加。修復時間も短くサービス継続性を維持します。



ディスクやノード障害時でも常にデータアクセスが可能

ユーザ or アプリケーション

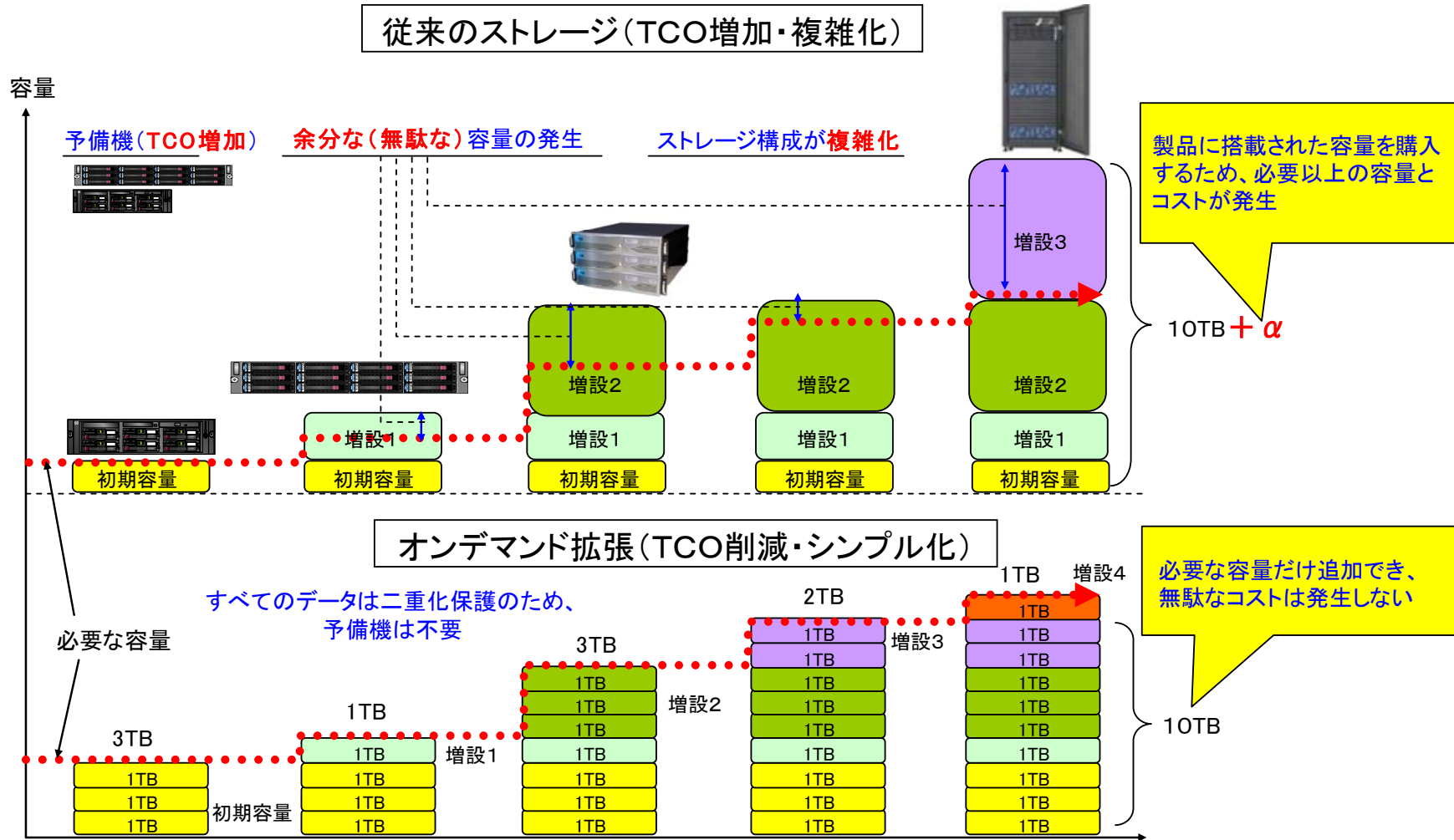
ユーザ or アプリケーション

- 保管データは、複数ノードにより常に多重化で保護
- 内部ヘルスチェックプロセス(HP)によりディスクやノード障害を自動検出
- 障害検知時は、正常なディスク/ノード上のデータをもとに他のノードによって自動コピー(修復)
- 障害時でもノードクラスタ構造により常にデータアクセスが可能

自己メンテナンス機能により、メンテナンス時間は柔軟に調整可能



# オンデマンドの柔軟な拡張性：CAStor™ CSS



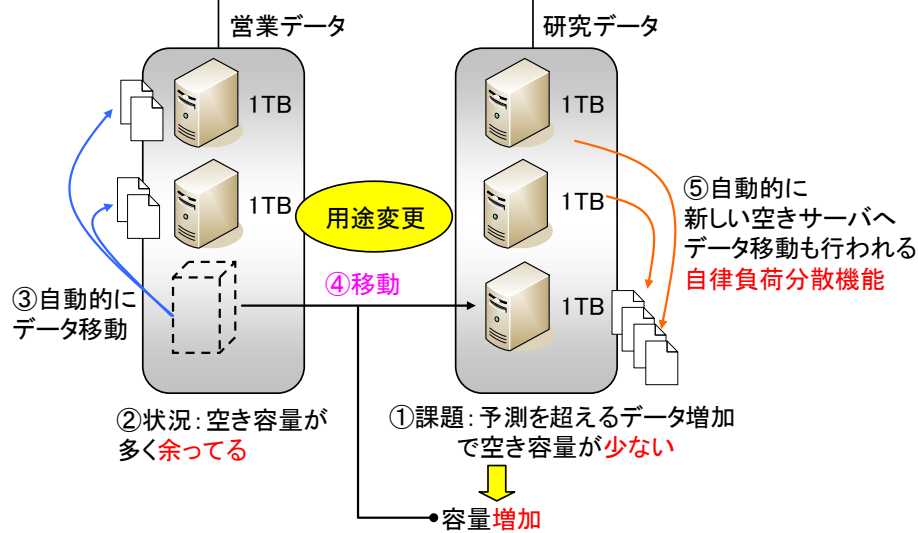
- ビジネス成長に合わせ、必要な時に、必要な容量のみ追加可能
- スモールスタート(1テラバイト)～大規模(ペタバイト)までの拡張性
- 完全二重化データ保護のため、保守用マシンは不要
- 単一ボリュームで拡張可能なため、ストレージ構成変更は不要
- 無駄な電力消費／ラックスペース／保守費が発生しない



# 既存IT資産／リソースの有効活用：CAStor™ CSS

課題 : 研究データのストレージ容量を至急増設が必要  
 状況 : 営業データのストレージ容量が多く余っている  
 解決 : 不足エリアへ余ったサーバを移動するだけ！

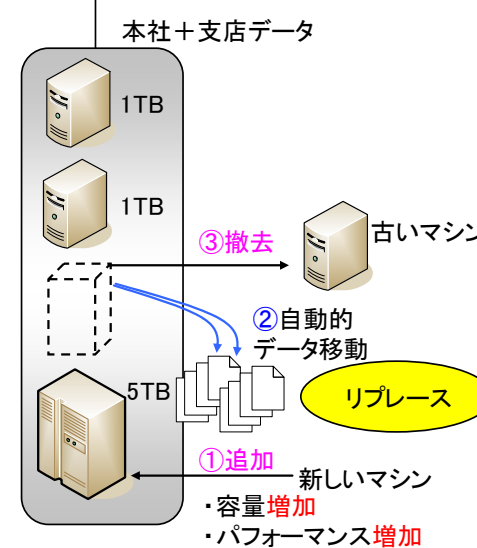
移動前の既存保管データのレプリケートやフォーマットは自動的に行われ、管理者の業務負荷は大幅に軽減できます



ストレージ利用効率向上

課題 : 1台の古いマシンをリプレースしたい  
 状況 : 新しいマシンを購入  
 解決 : 古いマシンを新しいマシンに入れ替えるだけ！

リプレース時に複雑な設定や作業は一切必要ありません。撤去マシン上にあるデータの移行は自動的に行われます

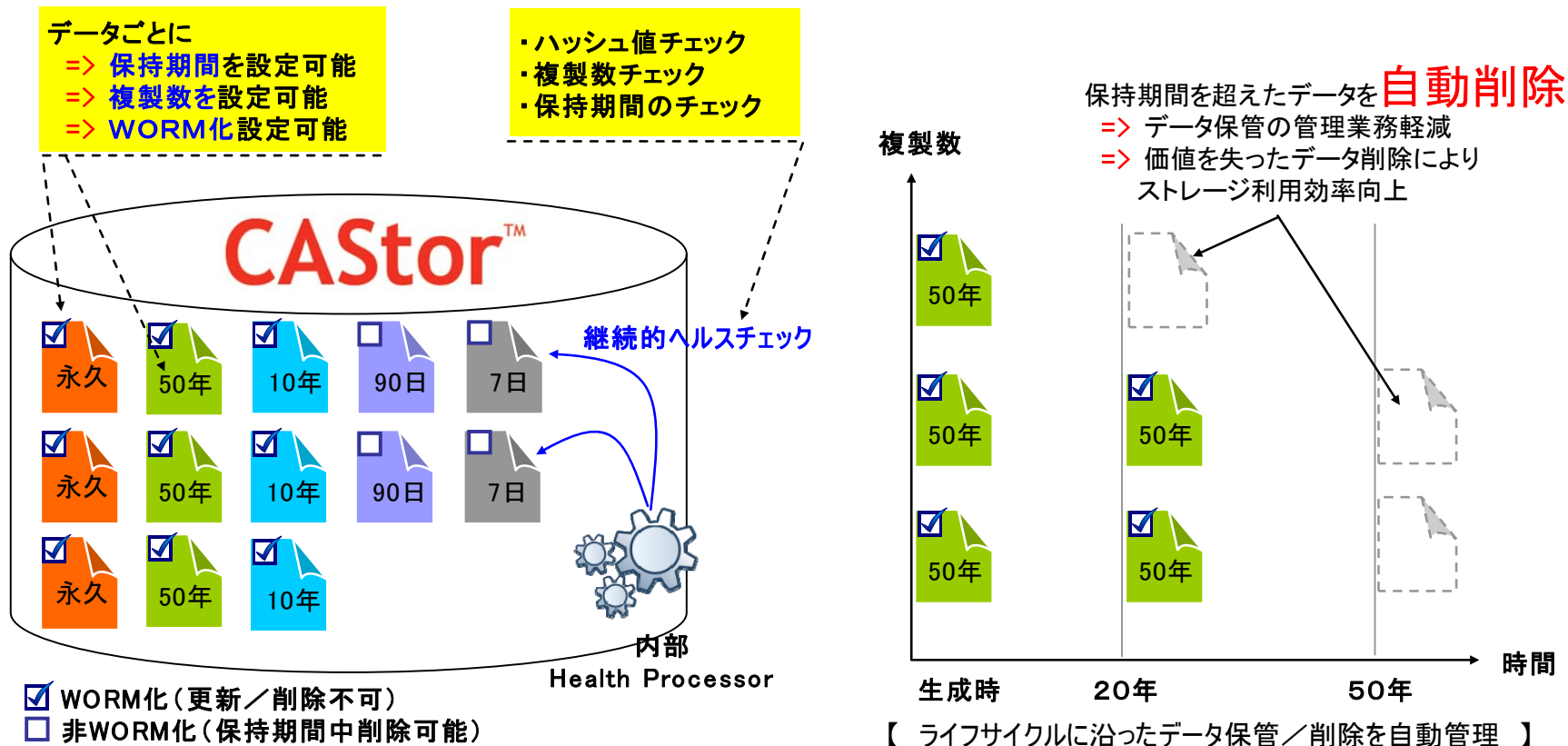


キャパシティ向上

- 用途変更 => 空き容量が少ないストレージエリアへ簡単に移動できる
  - ✓ 既存ストレージエリアからサーバをリタイアすると、該当サーバ上にあるデータは他のサーバへ自動レプリケート
  - ✓ その後、移行先ストレージエリアへ追加するだけ(自動ディスクフォーマット後、移行先ストレージへ参加)
- リプレース => 古いマシンを新しいマシンに簡単に入れ替ええる
  - ✓ 新しいマシンをストレージエリアに追加後、既存古いマシンをリタイア宣言
  - ✓ リタイアマシン上にあるデータは新しいマシンや他のサーバへ自動レプリケート

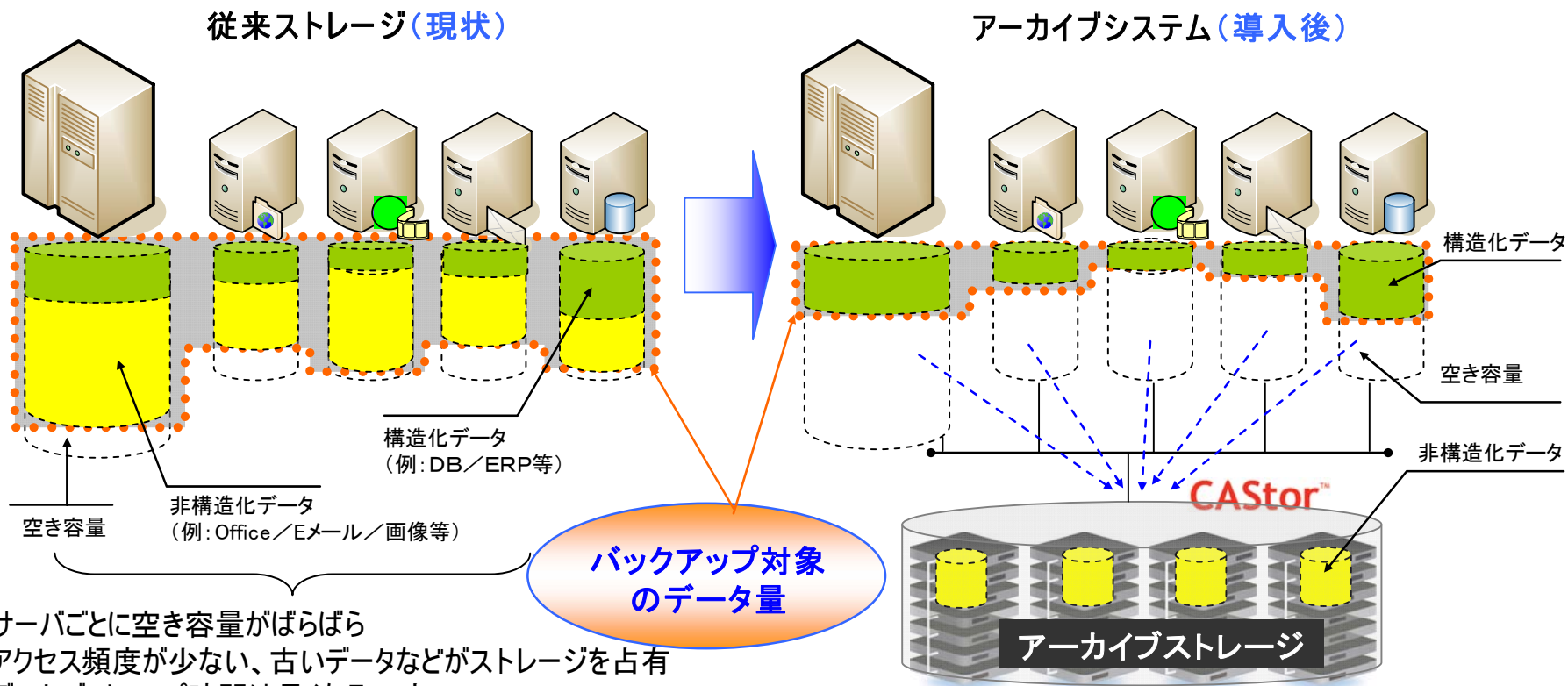


# 長期保管およびコンプライアンス対応：CAStor™ CSS



- WORM(Write Once Read Many)機能によるファイル単位での改ざん防止や削除防止
- 継続的なヘルスチェック(Health Processor)によるデータ完全性保証
  - データ単位のハッシュ値チェック → データ破損時は自動修復
  - データ単位の複製数チェック → クラスタ内複製数を自動調整／管理
  - データ単位の保持期間チェック → 保持期間を超過したデータは自動削除
- 128ビット長のデータ識別ID(UUID)により、大量のデータの長期保管に適した一意性を保証
- データごとの柔軟な保持期間の設定によりコンプライアンス対応
- データごとの複製数の設定によりデータ保護
- 設定に沿ったデータ削除の自動管理、データ保管業務の負荷軽減

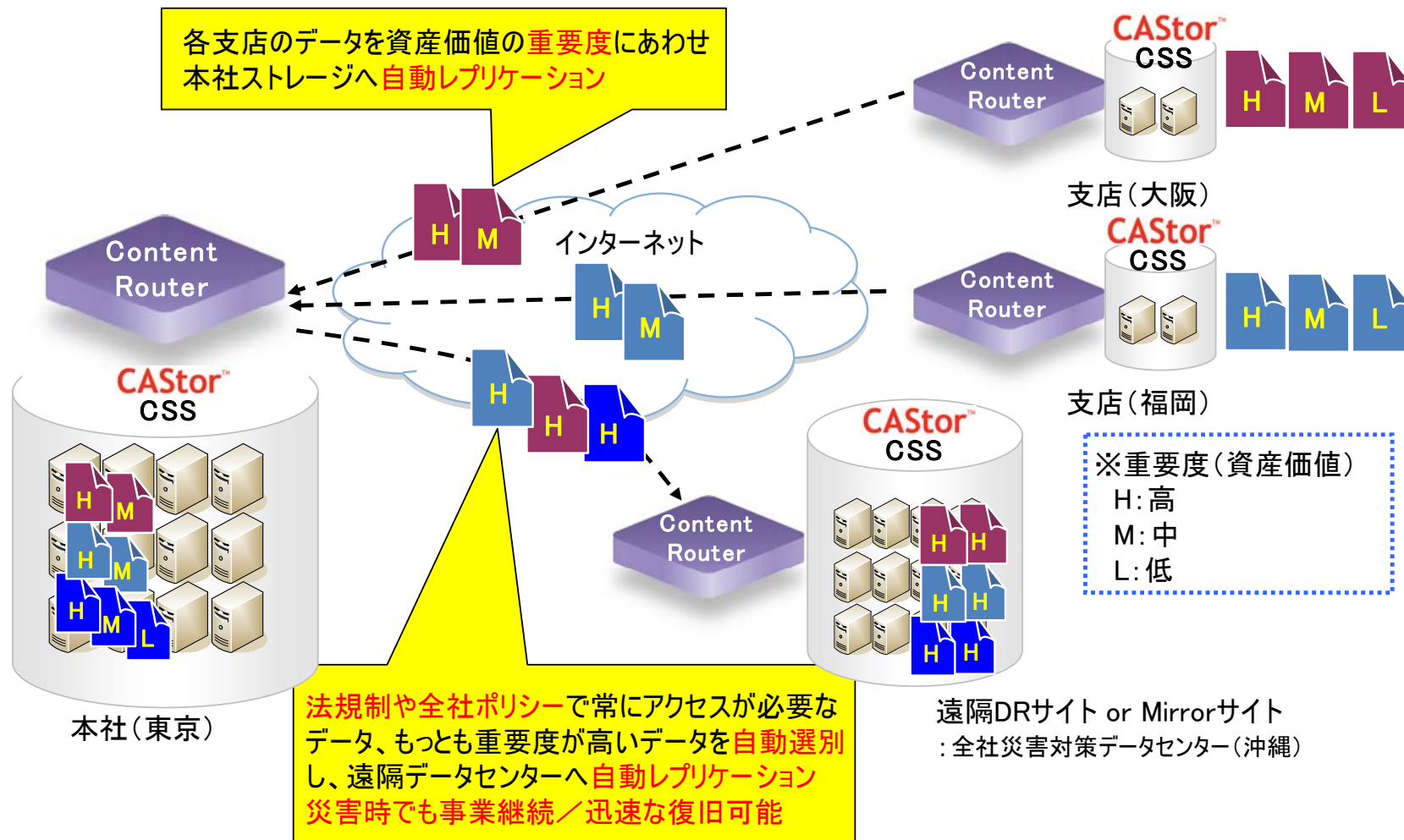
# アーカイブによるストレージ運用管理の効率化 : CAStor™ CSS



新しいサーバを単に追加するだけでは根本的な問題の解決はできず、更に悪化するのみ  
非構造化データの管理に最適化されたアーカイブシステムの導入により

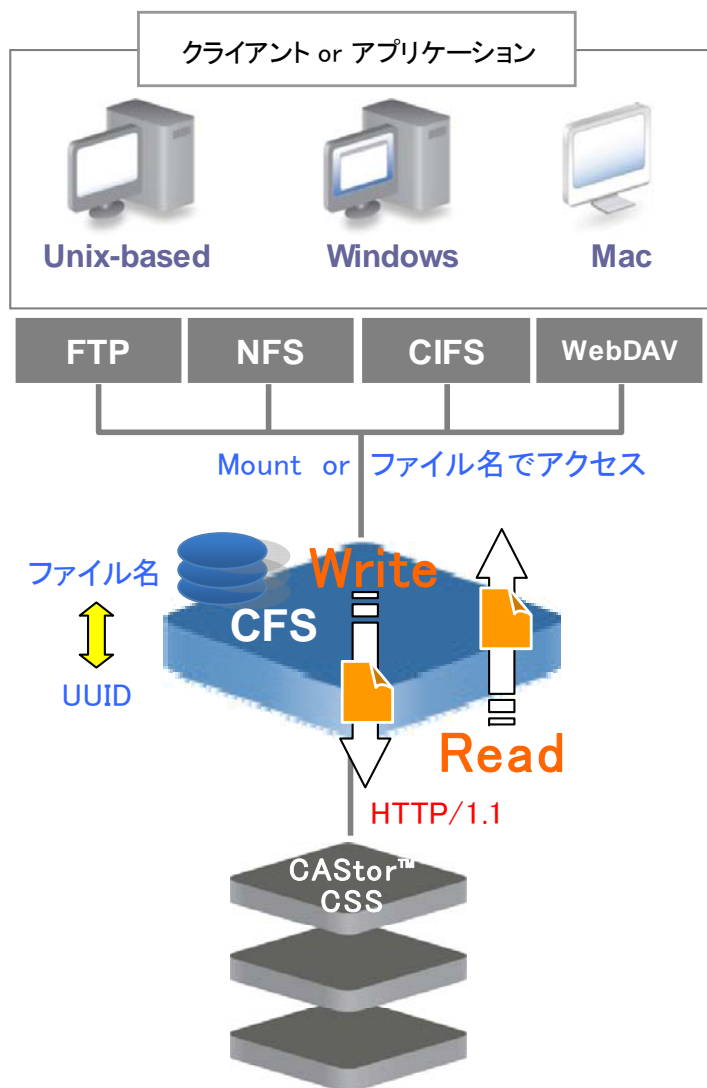
- ✓ バックアップデータ量の減少による**バックアップ時間の短縮**
- ✓ バックアップ作業のスリム化による**作業工数削減**、**運用管理負荷の軽減**
- ✓ リカバリーデータ量の減少により障害時の**迅速な業務復旧**
- ✓ 従来サーバのストレージ利用**効率向上**／**パフォーマンス向上**

# 災害対策(ローカル&遠隔地レプリケーション) : CAStor™ CR



ローカルおよび遠隔レプリケーションCAStor™ Content Router(ソフトウェア)

- データの重要度やカテゴリをMetadataに指定しCAStor™ CSSに保存します
  - Content Routerは、Metadataとポリシーに従って自動レプリケーションを行います
- ⇒ 各支店のデータをふくむ、災害対策を一元自動管理できます



CFS (Content File Server)とは、

- アーカイブストレージ (CAStor™ CSS) への専用アクセスゲートウェイ
- 既存App、ユーザ、ファイルサーバなどからは業界標準アクセスプロトコルを利用
- CFSを経由しCAStor™ CSSへフィックスドコンテンツの書き込み (Write) や読み込み (Read) が可能

## ■ サポートする業界標準アクセスプロトコル

Windows : CIFS (Common Internet File System)  
Unix/Linux/Mac : NFS (Network File System)  
その他 : FTP / WebDAV



# CASstor™ 開発検討要素

1. ファイルRead/Write(HTTPプロトコル)
2. UUIDの保管
3. データ保存先の変更

※上記は、既存Appから直接CASstor™ CSSへアクセスを想定した場合です。

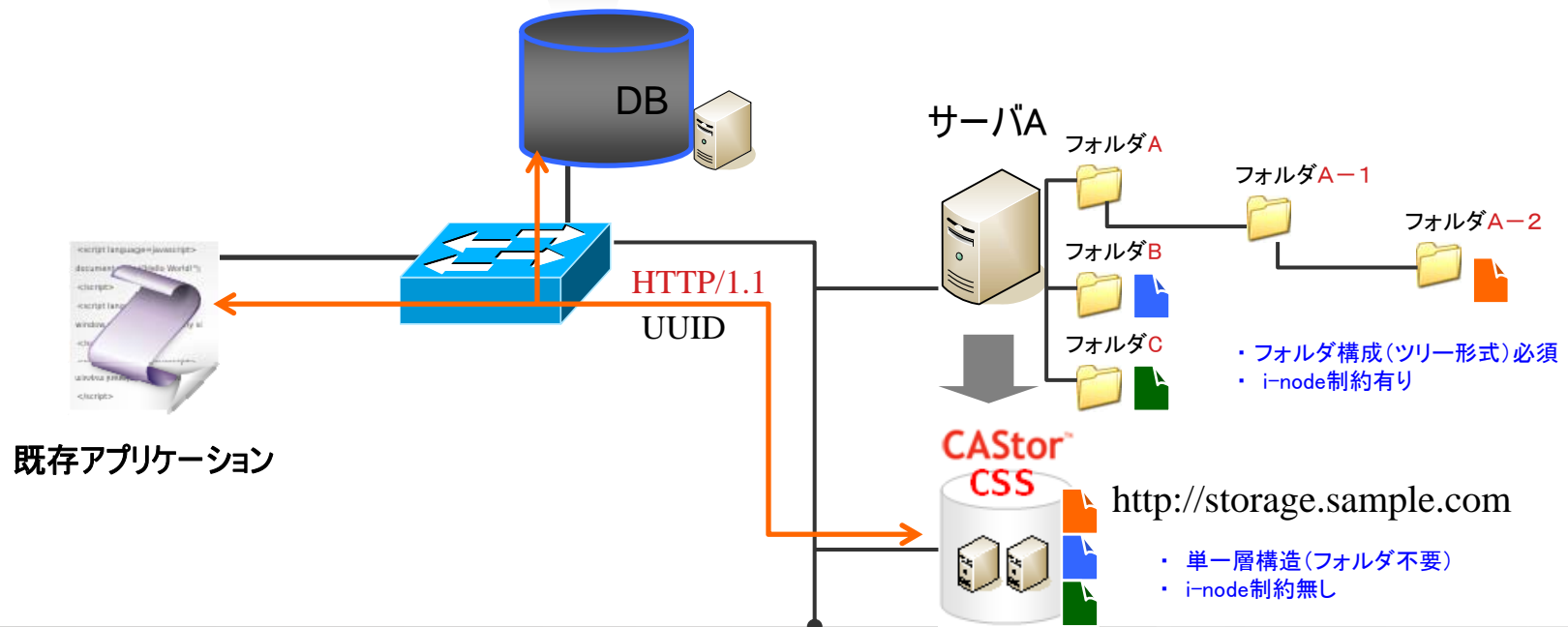
CASstor™ CFSを経由する場合は、既存App対応は基本的に必要ありません。

## 【既存管理情報(DB)】

SeqNo	Date	UserID	UserName	FileName	FilePath	Cs
N000001	20090801	yamada	山田	証明写真.JPG	※サーバA*フォルダA	C
N000002	20090801	tanaka	田中	証明写真.JPG	※サーバA*フォルダB	C
N000003	20090801	nakajima	中島	証明写真.JPG	※サーバA*フォルダC	C

## 【対応後管理情報(DB)】

SeqNo	Date	UserID	UserName	FileName	UUID	FilePath	Cs
N000001	20090801	yamada	山田	証明写真.JPG	B00123	storage.sample.com	C
N000002	20090801	tanaka	田中	証明写真.JPG	AB0014	storage.sample.com	C
N000003	20090801	nakajima	中島	証明写真.JPG	12345A	storage.sample.com	C



# HW推奨スペック(例)およびCAStor™ シリーズソフトウェア価格

## アーカイブ専用ソフトウェア



CPU	Intel Xeon、AMD Opteronまたは同等な64-bit CPU
Disk	SATA、SASまたはSCSIディスク
RAM	1GB(／1TB容量または200万オブジェクト)
Network Interface	Gigabit Ethernet x 2
価格(ソフトウェアのみ)	¥680,000(単位:ストレージ容量1TB、税別)

## 遠隔自動レプリケーションソフトウェア



CPU	Intel Xeon、AMD Opteronまたは同等な64-bit CPU
Disk	SATA、SASまたはSCSIディスク
RAM	1GB以上
Network Interface	Gigabit Ethernet
OS	Ubuntu Server 64-bit version 8.10 以上(別途ご用意ください)
価格(ソフトウェアのみ)	¥240,000(単位:1サイト、税別)

## CAStorアクセス専用ゲートウェイ



CPU	Intel Xeon、AMD Opteronまたは同等な64-bit CPU
Disk	SATA、SASまたはSCSIディスク + CD-ROM
RAM	2GB以上
Network Interface	Gigabit Ethernet
OS	Ubuntu Server 64-bit version 8.10 以上(別途ご用意ください)
価格(ソフトウェアのみ)	¥2,480,000(税別)