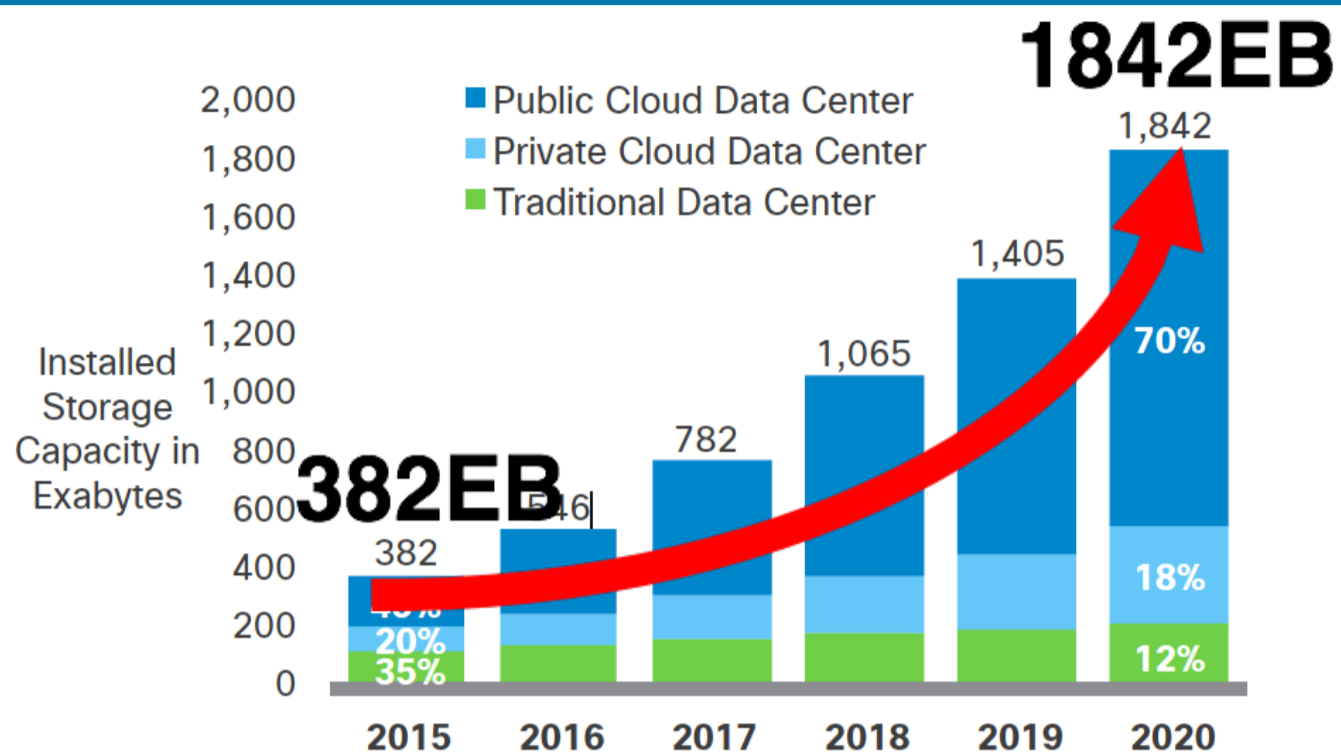


クラウドストレージ用 コールドデータジオレプリケーションの システムアーキテクチャ提案

研究背景



Source: Cisco Global Cloud Index, 2015-2020.

- クラウドストレージはクラウドコンピューティングにおいて、様々なサービスを支えている
 - ストレージの使用容量は指数的に増加
 - その増加の大部分はIoTに普及に伴うものと予想されている
 - IoTのデータはアクセスレートが非常に低い



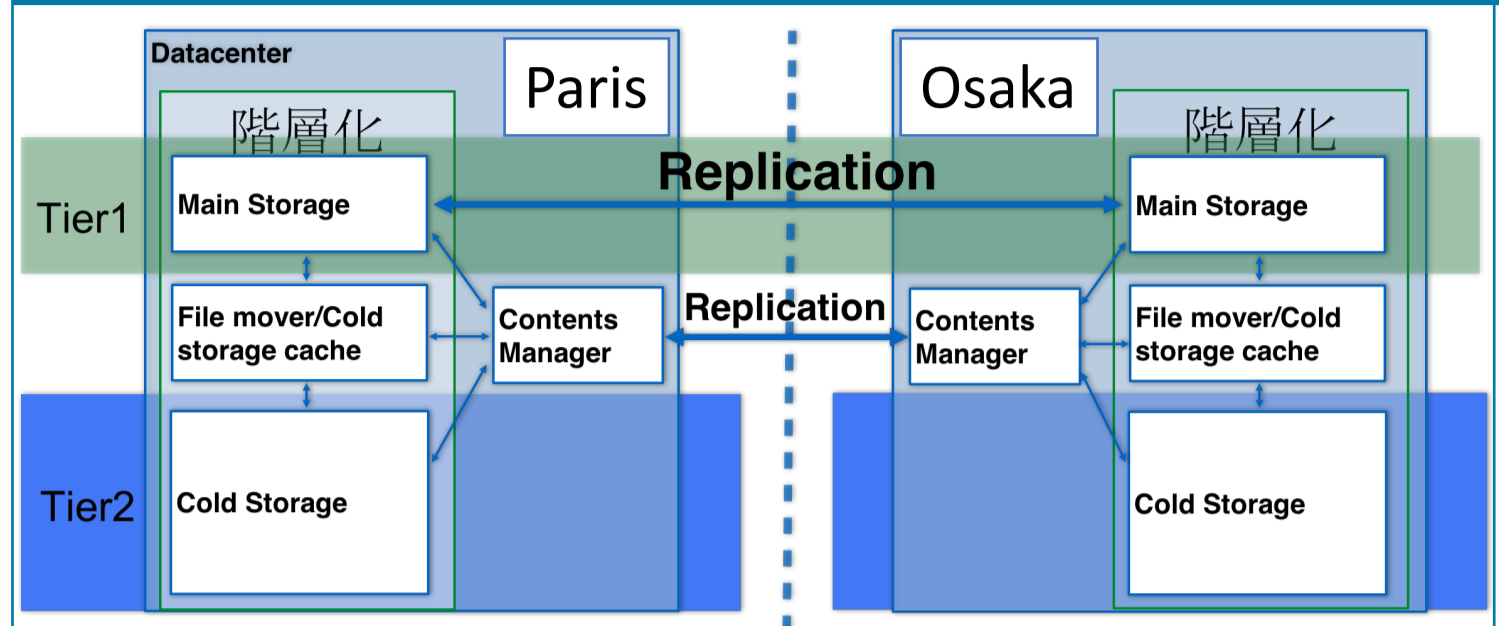
効率的なファイルストレージが必要

- 効率的なストレージ技術として以下
 - 階層化ストレージ
 - データをコストとパフォーマンスの異なる媒体、SSD、HDD、Blu-rayなどに分けて格納
 - IoTのデータにおいて、アクセスレートが非常に低いデータはBlu-rayなどの低コスト、低パフォーマンスの媒体に
 - ジオレプリケーション
 - データ損失防止のため、地理的に離れたストレージ筐体間での複製
 - 国や大陸を跨いで複製する事で、ストレージとユーザの距離を小さくし、空間局所性 (Spatial Locality) を実現可能

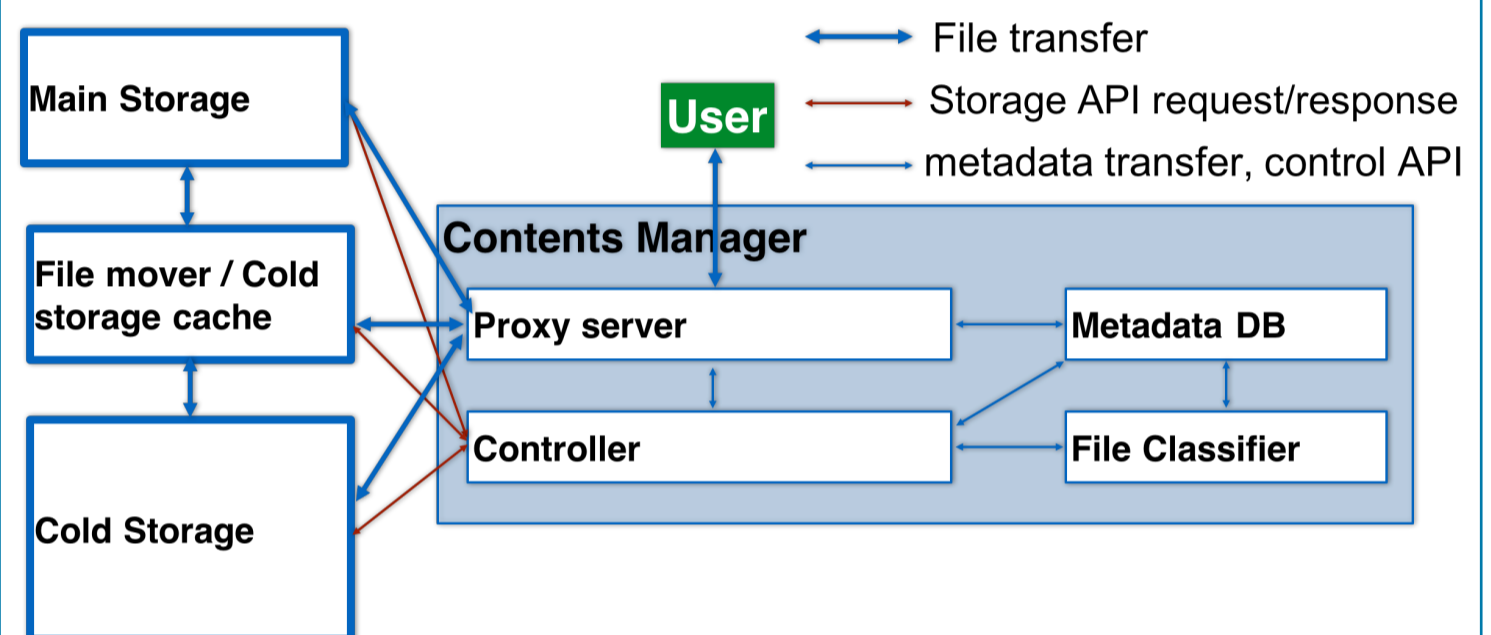
研究目的

本研究では、コールドデータのためのクラウドストレージのシステムアーキテクチャを提案し、実証システムを構築

システムアーキテクチャ



- 階層化ストレージとContents Managerで構成
- 実証システムでは、RINGストレージとBlu-rayアーカイブに階層化されている



- Contents Managerがデータの管理を行う → データ毎に用いるストレージの自動最適化
- ファイルを各ストレージ階層に自動でかつ正しく分類することが重要



機械学習を使ったファイルの自動分類手法

ファイル自動分類手法の評価

- 機械学習を用いて、ファイルを自動分類
- ファイル名に基づいて、1回だけアクセスされたファイルと2回以上アクセスされたファイルに分類
- DNNを用いることで、新しいファイルに対して、78%の精度でファイルのアクセス数に関する分類が可能

機械学習を用いて、ファイルの適切な保存先ストレージを決定することが可能

今後の課題

- 実証システムの大規模化
- IoTのデータやBig dataを分類するためのデータ収集と学習方法を確立