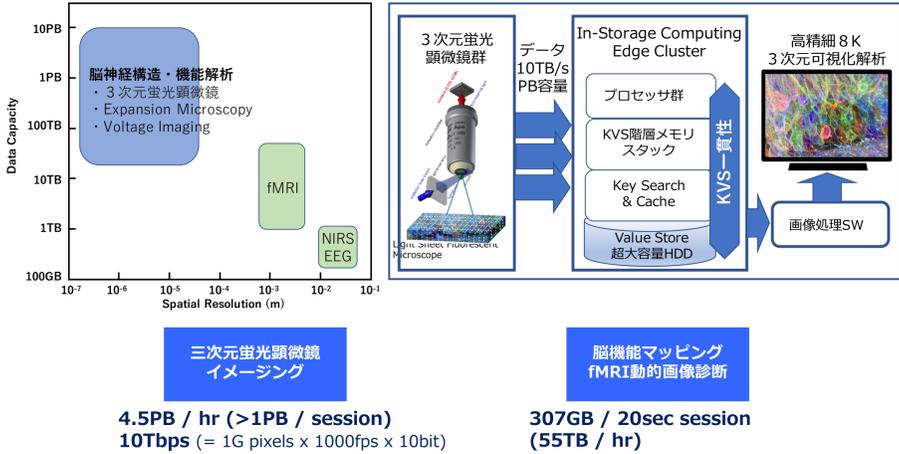


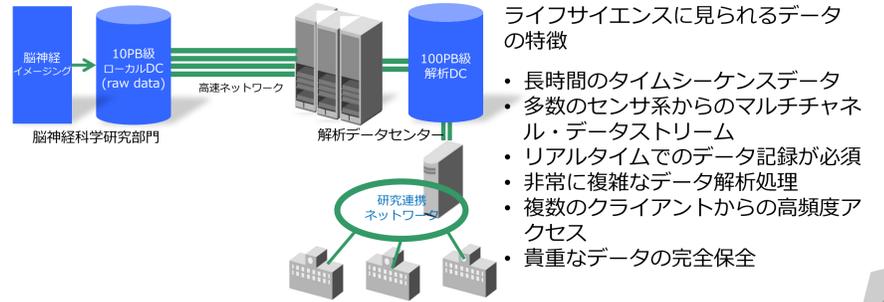
## 課題：脳機能・構造センシングの解析はデータヘビー



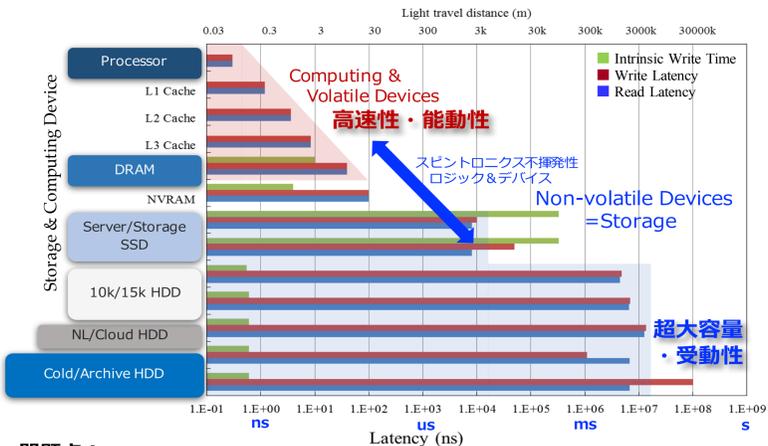
\*Source: Patric Hagmann, "Chairman's introduction: definitions and basic imaging technique", NH7: The human connectome: a comprehensive map of brain connections, ECR 2014 (Courtesy of Dr. Hitoshi Yamagata, Toshiba Medical Co.)  
[https://en.wikipedia.org/wiki/Light\\_sheet\\_fluorescence\\_microscopy](https://en.wikipedia.org/wiki/Light_sheet_fluorescence_microscopy)

## 目的：神経構造解析等の大規模データ解析の課題解決を図るストレージプラットフォームの開発

- ・脳神経構造・機能解析、CTやMRI等のモーダル画像解析研究などのライフサイエンス研究では、膨大なデータ・インテンシブ解析が進められている。
- ・ミクروسケールの高解像度・高時間分解能観察が可能となり、数10TB~PB級の極めて大規模なデータセットの解析が必須。
- ・高解像度・高時間分解能観察では、データ源近傍での大規模データ解析が求められ、データストレージと解析コンピューティングの近接化が重要。



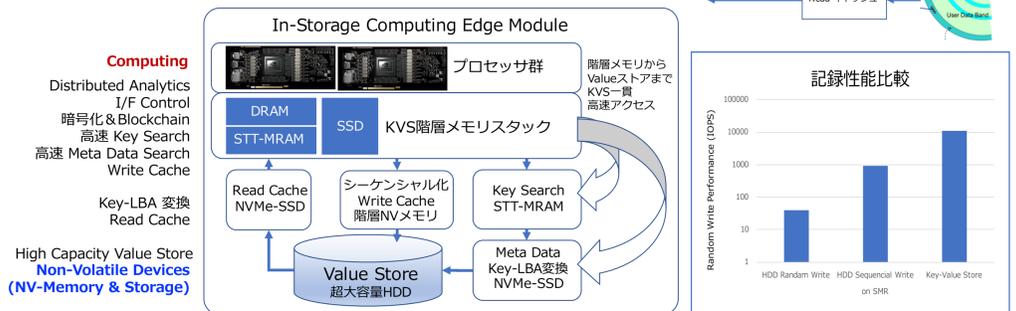
## コンピューティングとデータの距離感 ~latency(遅延)の観点~



Source: Yoichiro Tanaka, Characterizing Advanced Recording Technology Assets with Hyper-Scale Applications, IEEE Trans. Magn., Vol.52, No.2, pp.1-4, 2016  
<http://news.toshiba.com/press-release/business-and-retail-solutions/toshiba-demonstrates-high-performance-object-storage-tec>, also presented at Openstack Summit 2015, Vancouver, May 2015

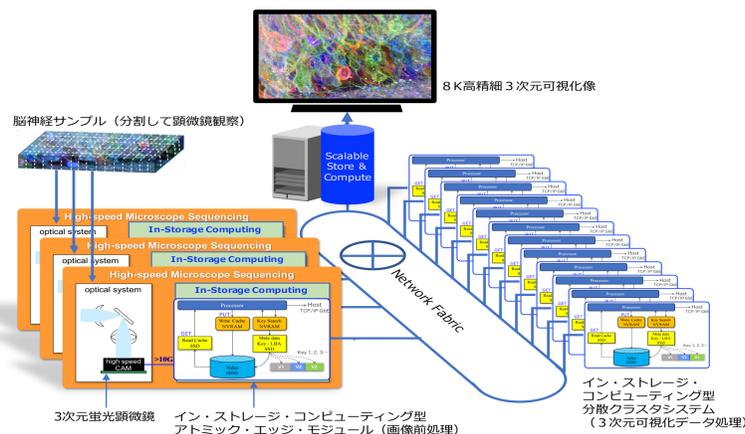
## "In-Storage Computing" Edge Moduleによるコンピューティングとストレージの近接化

- ・データストレージ上にコンピューティング能力を配置、PB級大規模データを高性能解析
- ・非構造化データに対応するKey Value型分散オブジェクト「In-Storage Computing」
- ・データ移動を最小にするアーキテクチャによるリアルタイム解析
- ・システム拡張性
- ・ストレージで閉じたセキュリティ機能 (カプセル化)

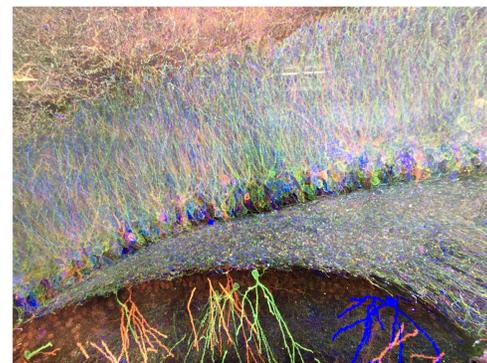


## "In-Storage Computing" クラスタ解析システムの構成

- ・NVRAM/SSD/HDDストレージの役割分担と計算能力の一体化
- ・Key-Value Storeによるアクセス性能とフラットポロジ化
- ・Edge Deviceとしてデータ近傍にコンピューティング能力を配置



## Mouse Brain Hippocampus 3D画像解析の一例

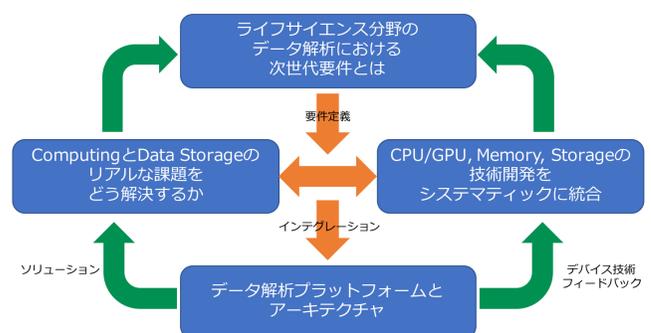


Macroscopic & microscopic 8K images (85" display) of mouse hippocampus neuron structures, scanned by Light Sheet Microscope  
Data set size; 6TB ~ 60TB  
Specimen: 1.5mm x 0.8mm x 0.2mm<sup>3</sup>  
2000 scans per 0.2mm<sup>3</sup> (100nm<sup>3</sup> each)  
Scan section 80um x 8.3um

Source: Brain Data Center Project; MIT, NHK, Toshiba Memory, Yoichiro Tanaka, July 2017

## まとめ

1. ストレージは、「受動」から「コンピューティング一体化の能動」へ、新たな変革のフェーズ。
2. ライフサイエンスにおけるデータ解析では、大規模、準リアルタイムな高性能解析、多様な非構造化データとメタデータ、セキュリティの確保が大きな課題である。データ解析プラットフォームとして、データ源に近い場所で、より強力なエッジコンピューティングが望まれる。
3. 大量に生成される多様なデータから効率良く価値情報を抽出するためには、コンピューティングをよりデータに近接させる概念 "In-Storage Computing" が効果的と考えられる。
4. ビッグデータのアプリケーション側から、ストレージとコンピューティングのシステム機能を定義するアプローチにより、ライフサイエンス次世代データ解析プラットフォームとアーキテクチャを目指す。



## Acknowledgements

本研究は、東芝メモリ株式会社 (現キオクシア株式会社)、MIT Media Lab.、NHK技研との連携協力、及び東北大学ヨットインフォマティクス研究センター2019年度研究助成により実施されたものです。