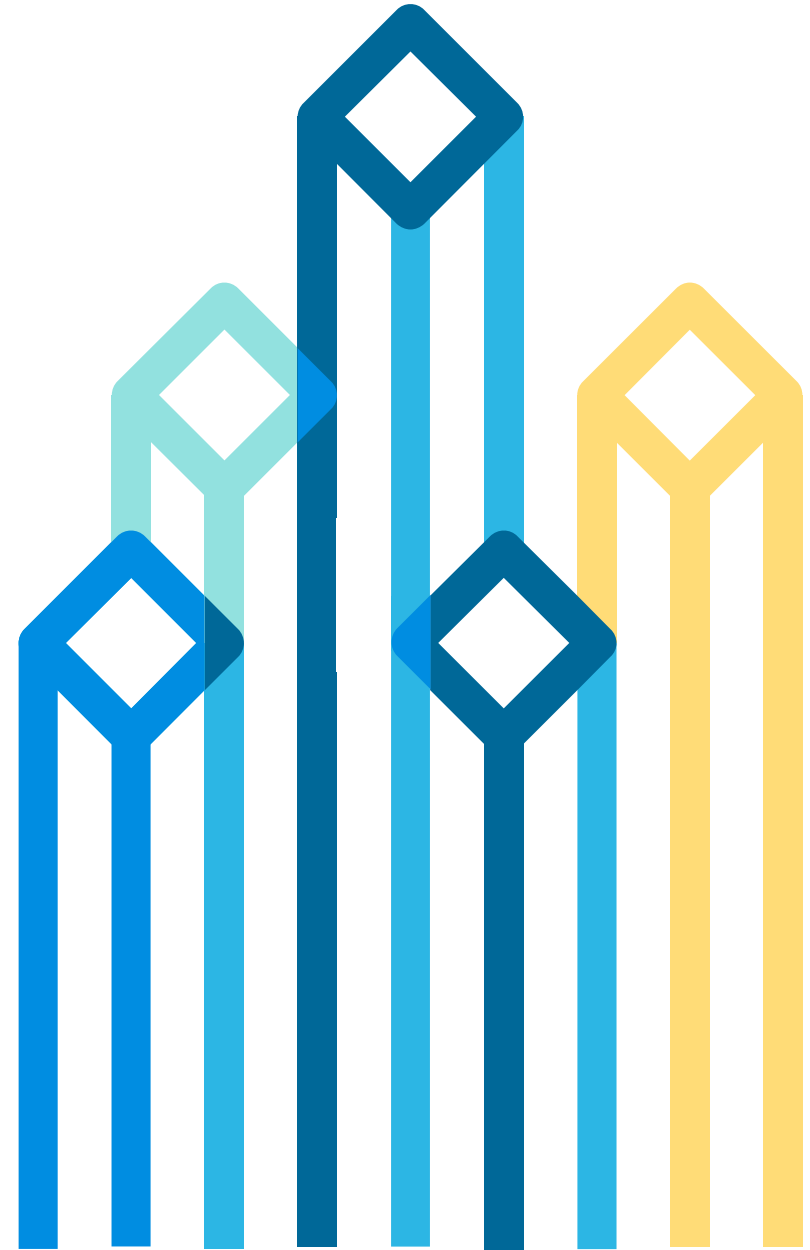




10年目を迎えた Apache Hadoop

Doug Cutting | チーフアーキテクト
@cutting



今年は〇〇10周年！

Hadoop

クラウド

そして、デジタル・トランスフォーメーション

かつては ...

- ハードウェア： 特殊製品
- ソフトウェア： RDBMS、プロプライエタリ
- データ： スキーマありき
- 対象領域： 帳簿管理
- デザイン： ウォーターフォール型

... RDBMS に入っていないものはデータではない

データ革命

ムーアの法則などとともに、データは指数的に増加

デジタル・トランスフォーメーションによって、データのビジネス価値も拡大

波及効果：

- 今やほとんどのデータが、トランザクションではなくマシンによって生成されている
- ビジネスのあらゆる面に影響
- 拡張性に富む、（低コストで）汎用的なデータシステムが必要

Lucene の教訓

- 1998年に最初に書かれた
- 2000年にオープンソース化
- 2001年にApacheプロジェクト入り
- 今は検索の標準

教訓： コミュニティによるオープンソース開発が、ソフトウェアの導入を加速し、その品質をいっそう高めることになる

Hadoop の歴史

Lucene に触発され ...

- 2002年： Nutch プロジェクトを開始
- 2003-4年： Google が GFS および MapReduce に関する論文を発表
- 2005年： Nutch にこのアイデアを取り込む
- 2006年： Nutch が Yahoo 入り、 独立した Hadoopプロジェクトを開始
- 2007年： 信頼性と拡張性に富む、オープンソースデータプラットフォームの完成！
- 2008年： Cloudera 創立
- 2009年： Cloudera に参画
- 2009年 ~ 現在： Hadoop エコシステムが、新たなデータ標準に

今では ...

- ハードウェア：コモディティ
- ソフトウェア：柔軟、オープンソース
- データ：共有、スキーマ・オンデマンド
- 対象領域：ビジネス全体（営業、製造、マーケティングなど）
- デザイン：アジャイル

... あらゆるデータから恩恵を受けることが可能に！

新しいデータソフトウェアエコシステム

- オープンソースプロジェクト間の緩い同盟関係
- 集中管理や調整なし
- 適者生存
- ユーザーニーズに対応
- 以前のやり方よりもスピードが改善
- より強力な長期的アーキテクチャー

汎用的なエンジンとして、Spark が MapReduce をリプレースしようとしている
Kudu と オブジェクトストアが HDFS を補完

将来

... クラウド

... コンテナ

... ハードウェア

... ?

ありがとうございました！



@cutting