



大規模データをより深く・ より速く分析する技術

西川 徹

(株) Preferred Infrastructure

代表取締役

<nishikawa@preferred.jp>

PFIのビジョン：PFIはどのような会社か？

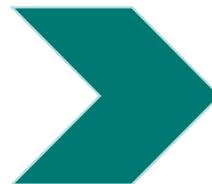
**テクノロジーとサービス・プロダクトとの間のギャップを埋め
テクノロジーが持つポテンシャルを最大限に引き出すこと**

研究ベースの技術が想定しているものと実世界には大きな壁が存在する

- アカデミックの第一線で研究しているリサーチャーと、アカデミックな研究を理解し、実装・実用化できるエンジニアを集合させる
- 世の中に必要とされている技術の中でも、特に難しい課題を選び抜き、それに対する解を提供していくこと

Basic Technologies

Academic Researches



Products

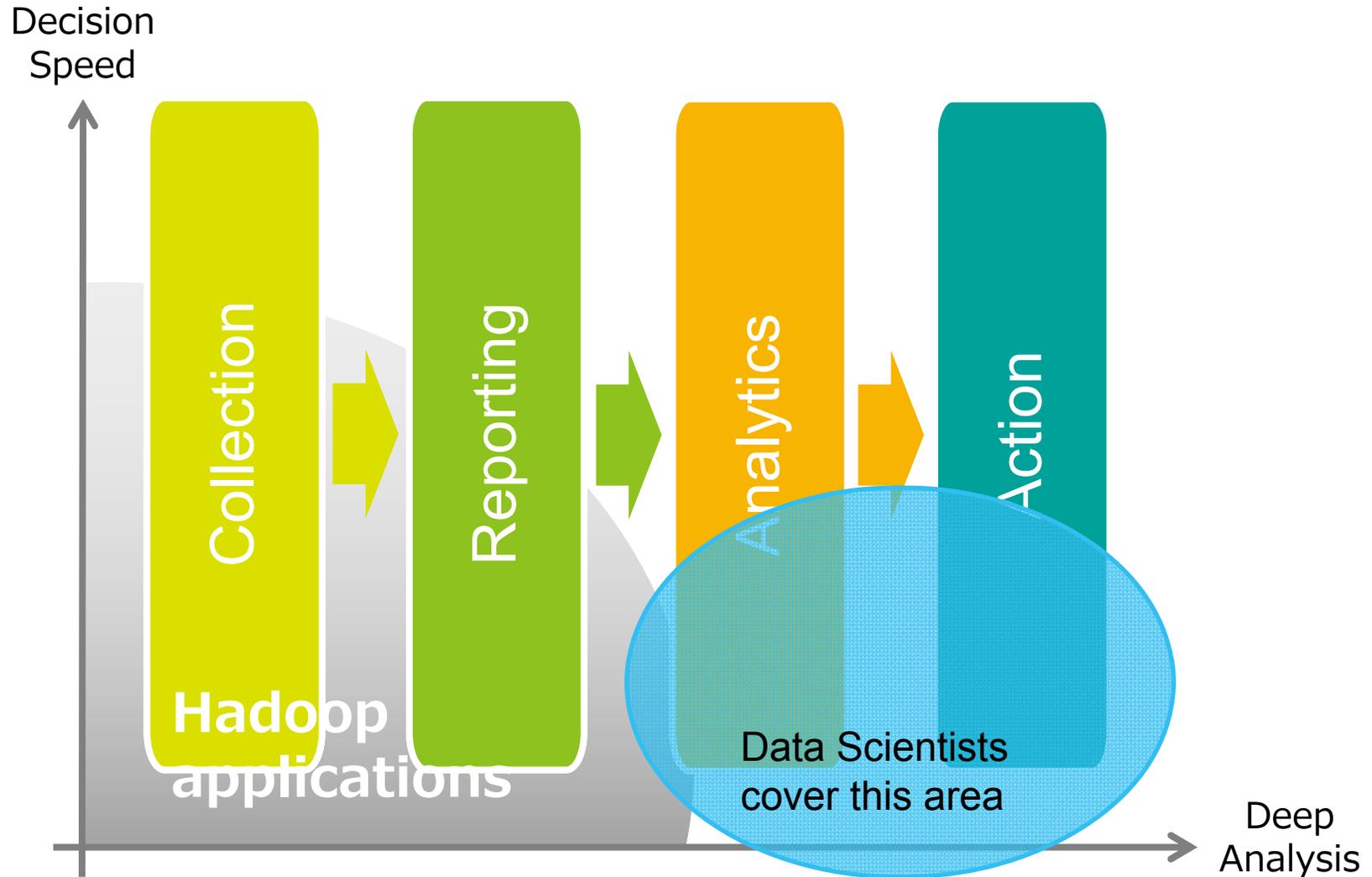
Services

ビジョンを達成するために

- 外部資本を入れない
 - 研究は多くのチャレンジがあつて、よい研究ができる。製品開発も同じ。ベンチャーキャピタルは、そのチャレンジを表面的にしか許容しない（下手したらまったく許容しない）
- フルスクラッチでの受託開発を行わない
 - 製品に繋がるビジネスにこだわる。
- 技術の多様性・人の多様性を重要視する
 - 1人では全分野を把握することは難しい。チーム・組織としての生産性を重要視する。もちろん、1人1人の能力がとても優れている上で。

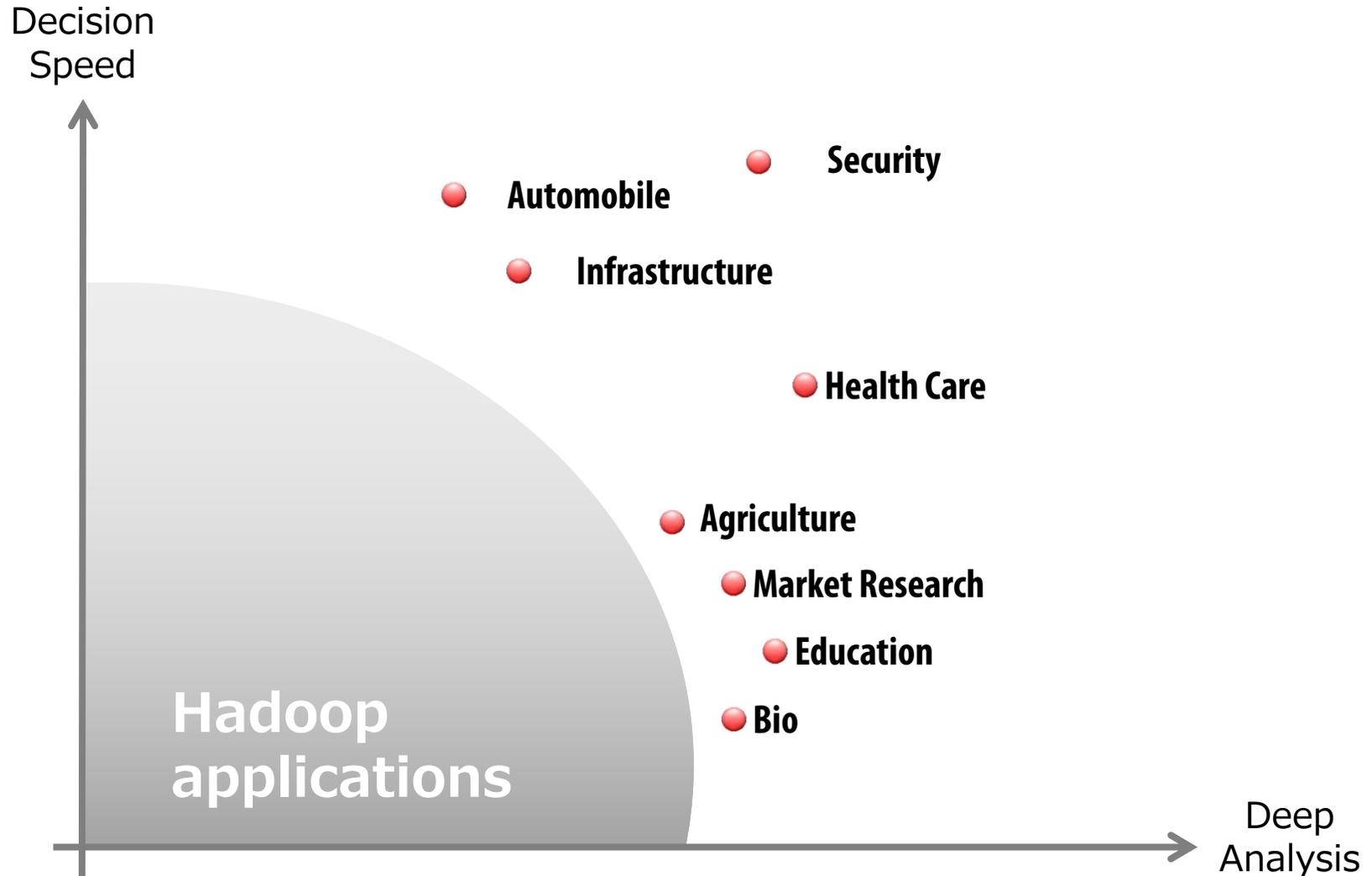
BigData Analysis beyond Hadoop

[Phases of BigData Processing]

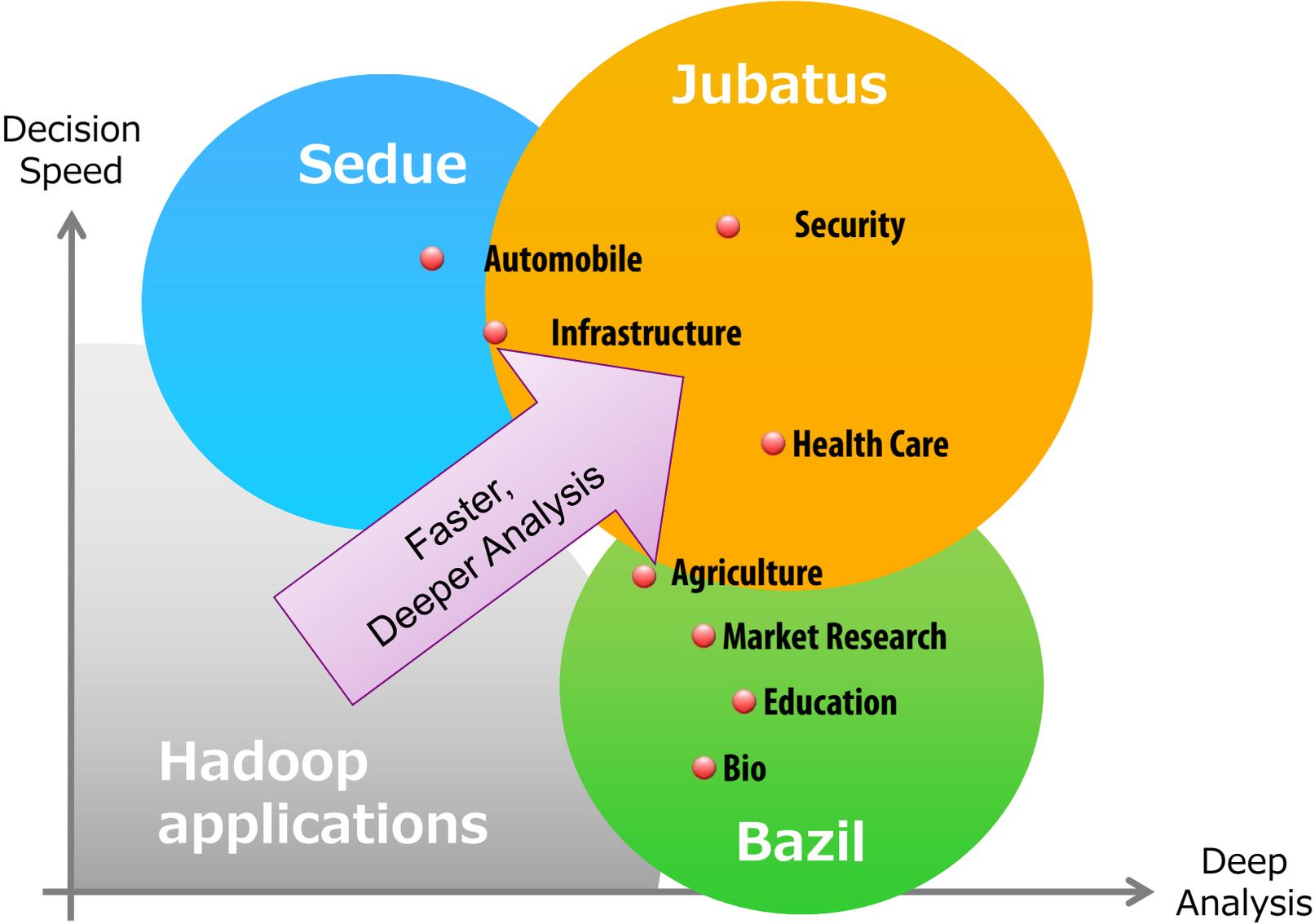


BigData Analysis beyond Hadoop

[Map of the Marketplace]

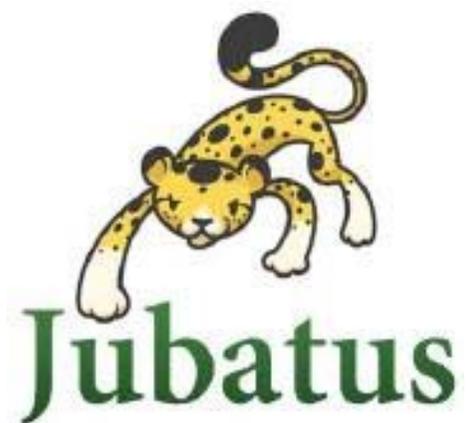


PFI Products beyond Hadoop



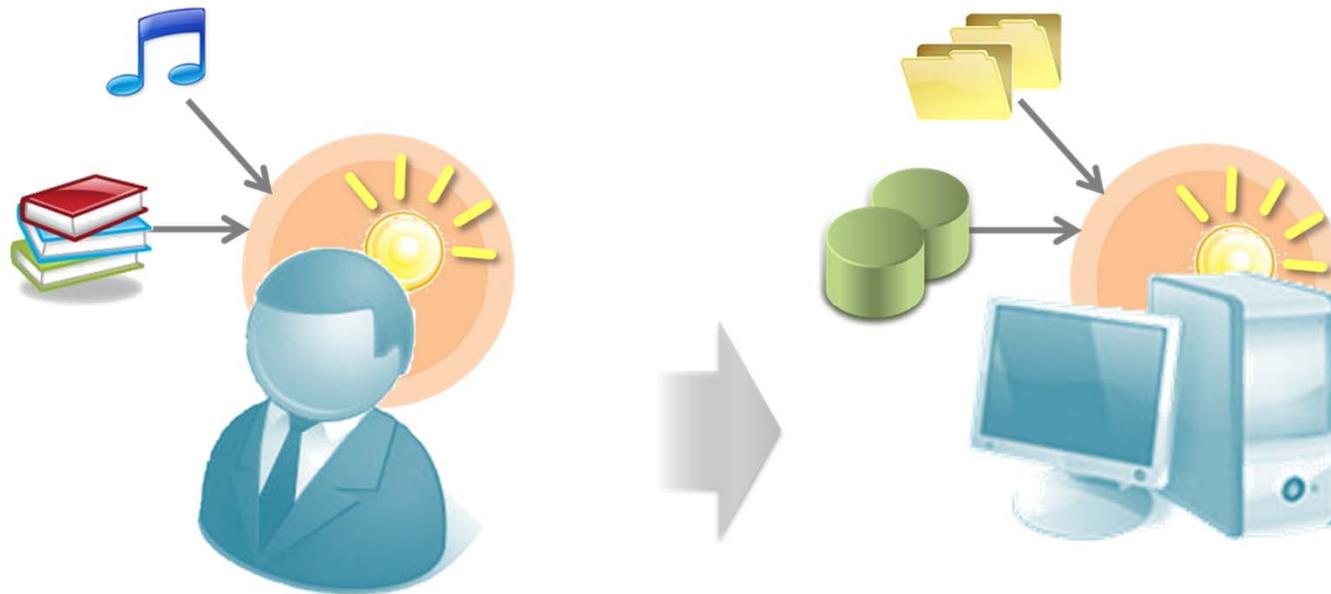
大規模なデータに立ち向かう

- マシンを横にたくさん並べる ← あたりまえ
- ハードウェア・デバイスの進化を生かす
- アルゴリズム的なブレークスルー
- アーキテクチャの進化



大規模・リアルタイムな機械学習を実現する

機械学習技術を活用することにより、
データに対する深い分析の実現を目指す

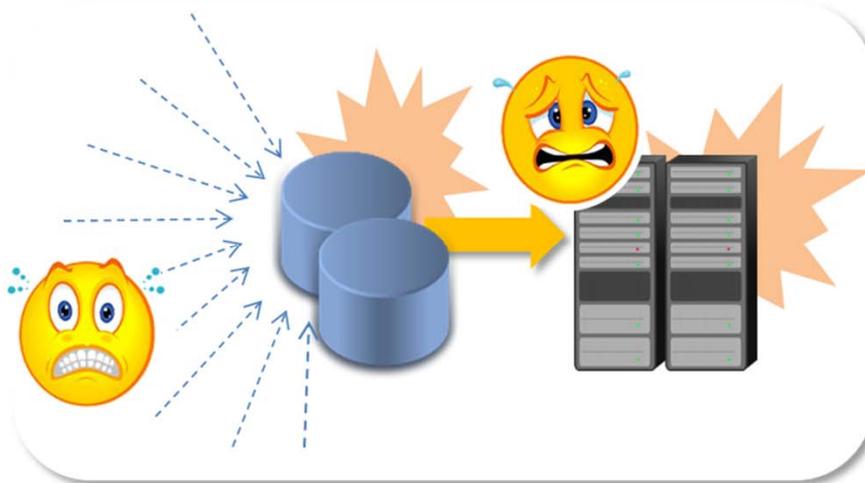


「学習 = 人間が知識や知能を獲得する過程」を
コンピュータ上で実現しようとする

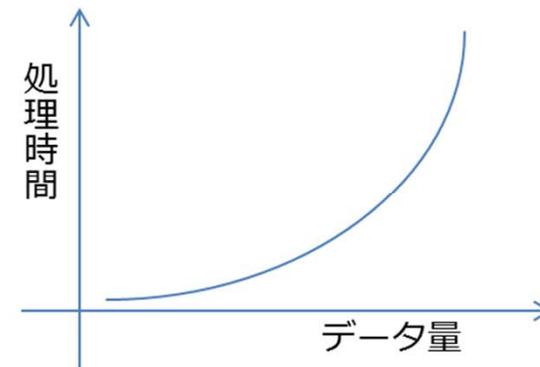
機械学習を大規模化する上での2つの難しい課題



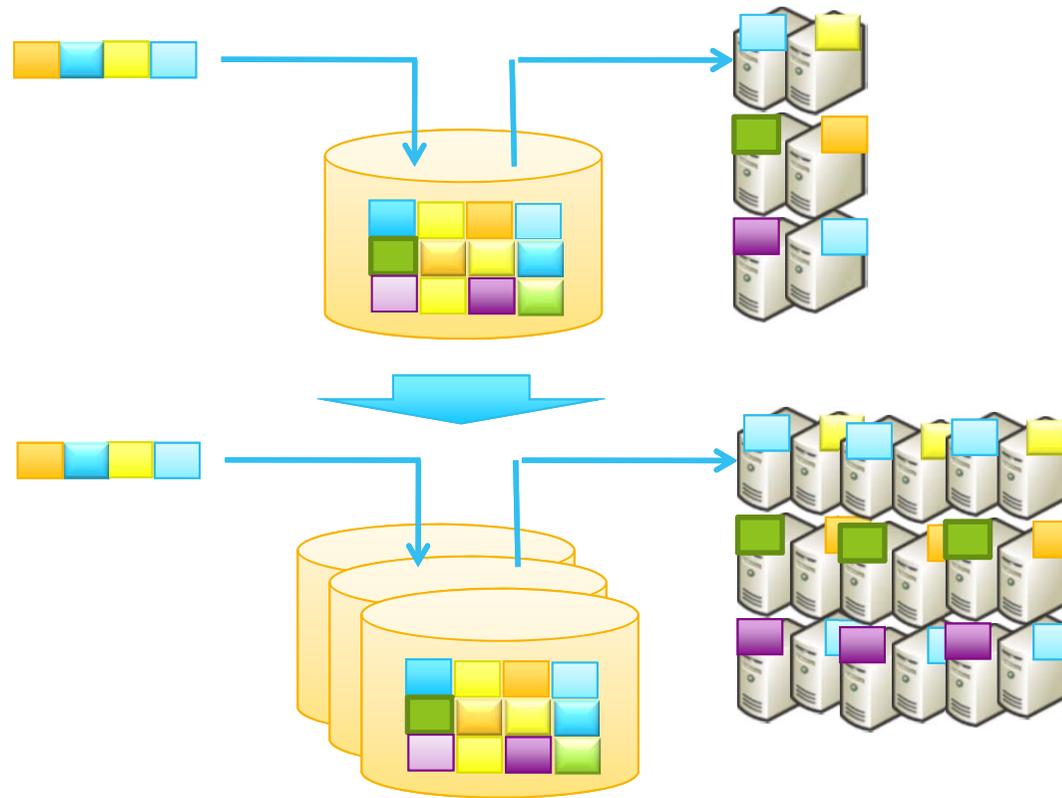
入出力がボトルネックとなる



計算量がボトルネックとなる

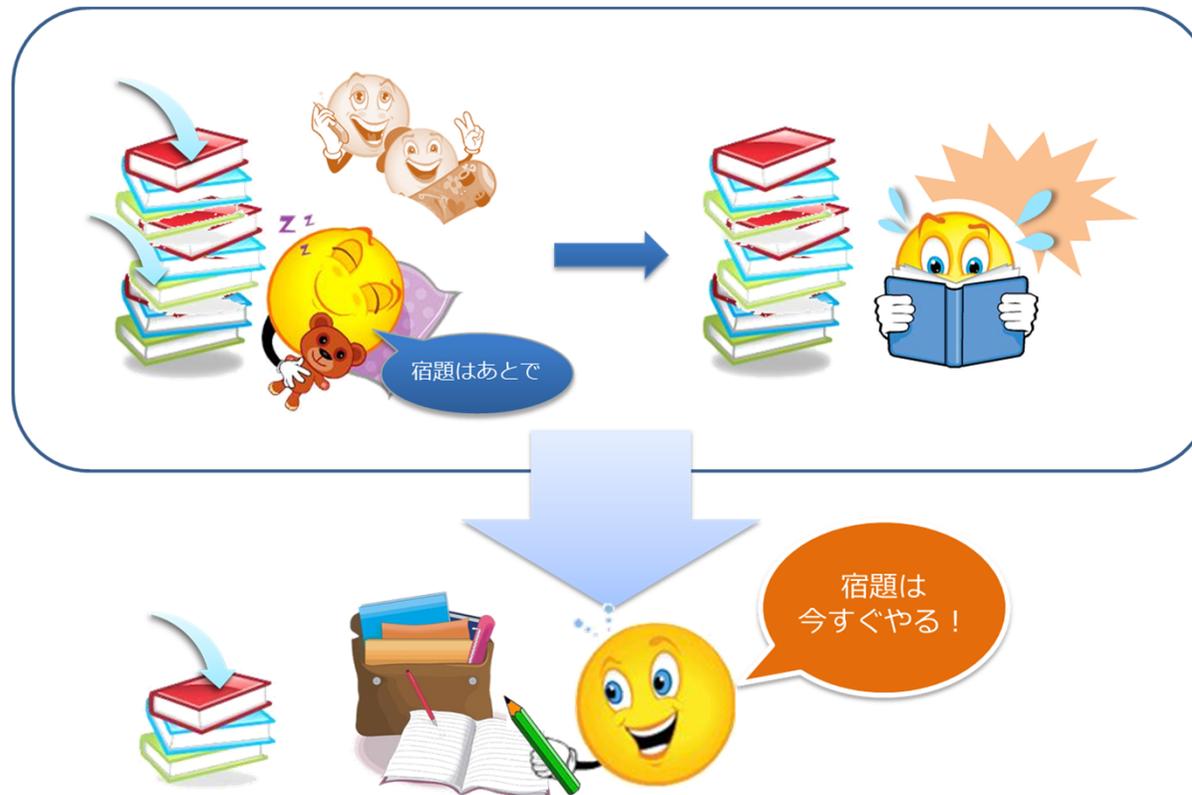


一つの解決策は、大規模分散システム化



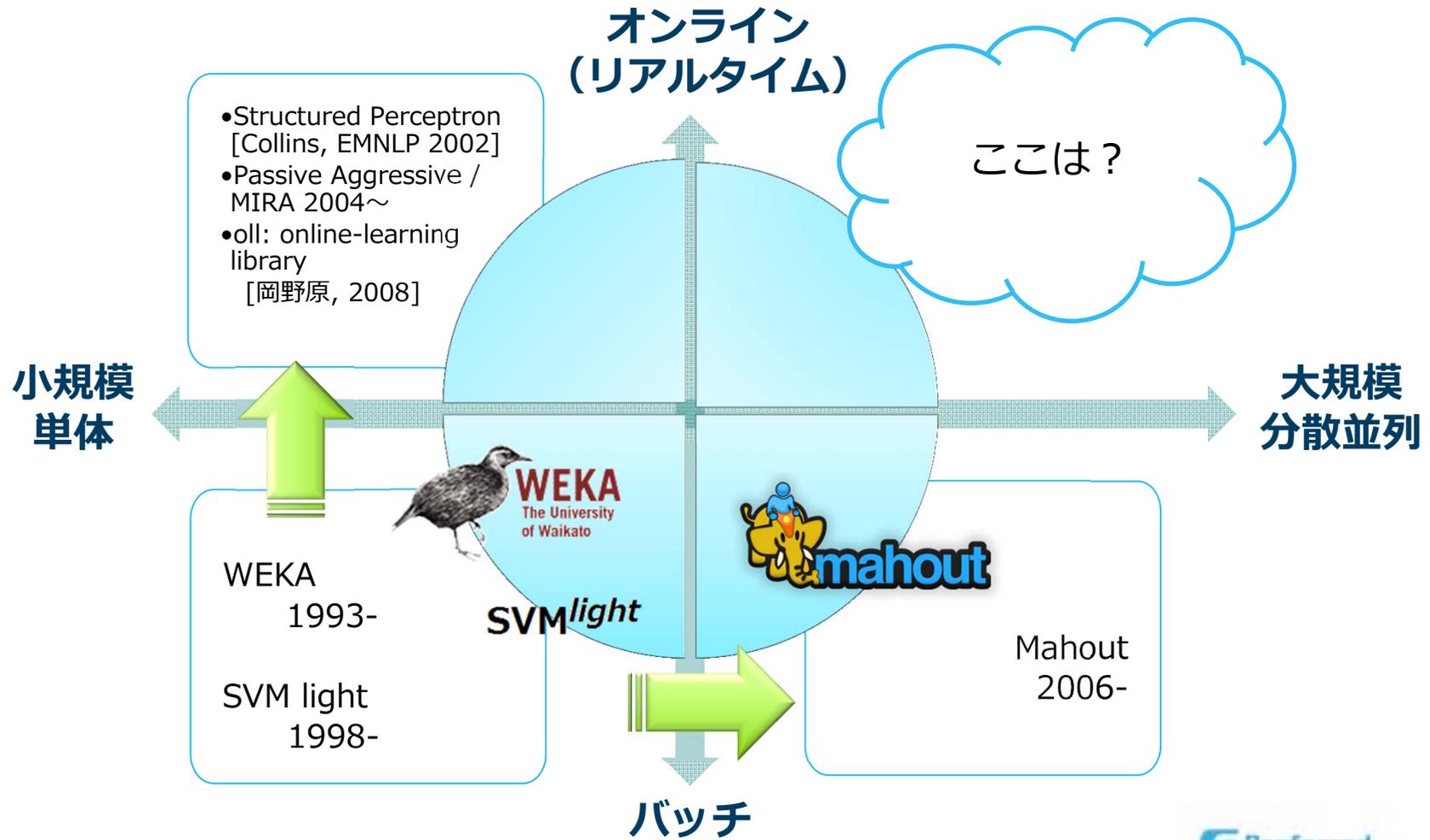
しかし、
アルゴリズムの計算量の問題は依然として残る、、

分析処理のオンライン化・ストリーム化

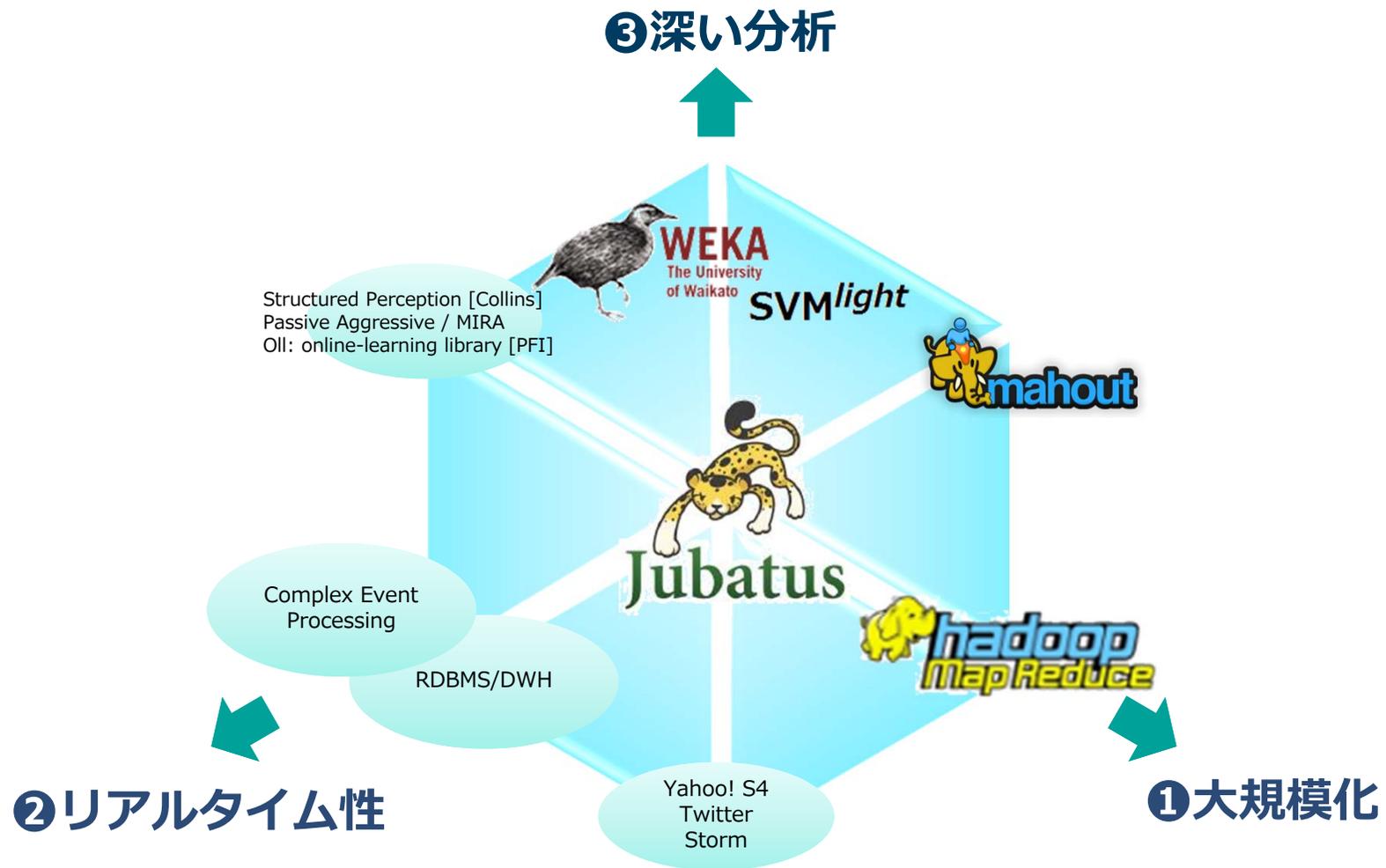


省スペース・効率的なアルゴリズムを使う

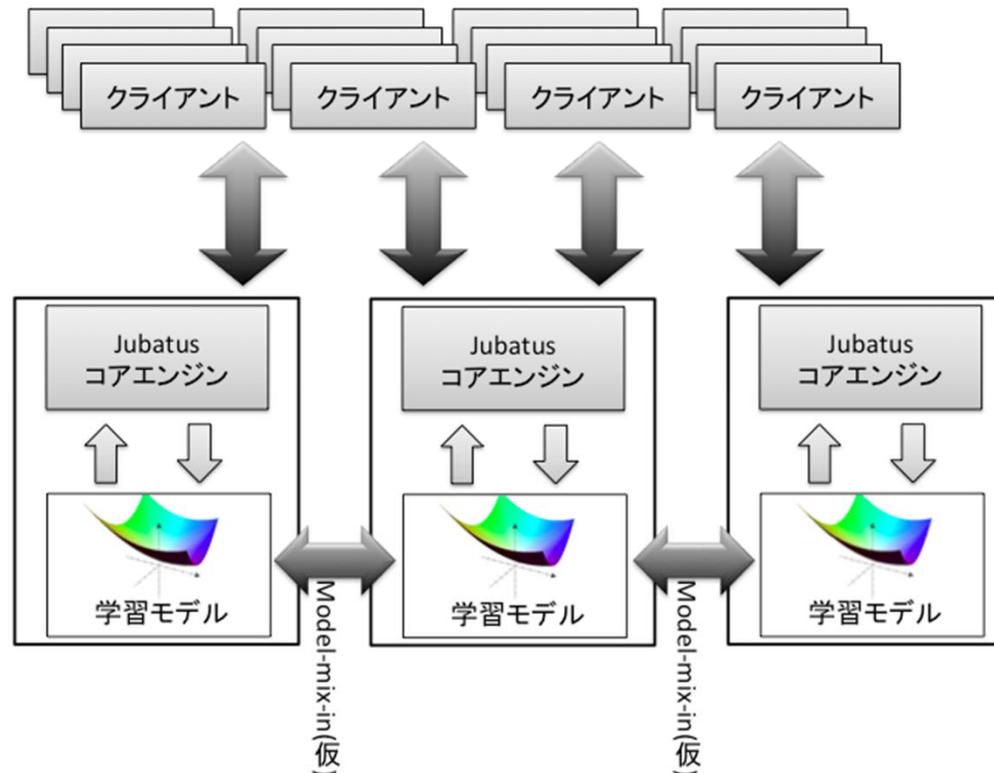
大規模データに機械学習を適用する、2つのアプローチ



Jubatus



機械学習に特化したデータ交換モデル



生データに対して、学習モデルは小さい
小さい部分だけ効率的に、かつ、非同期的に交換する



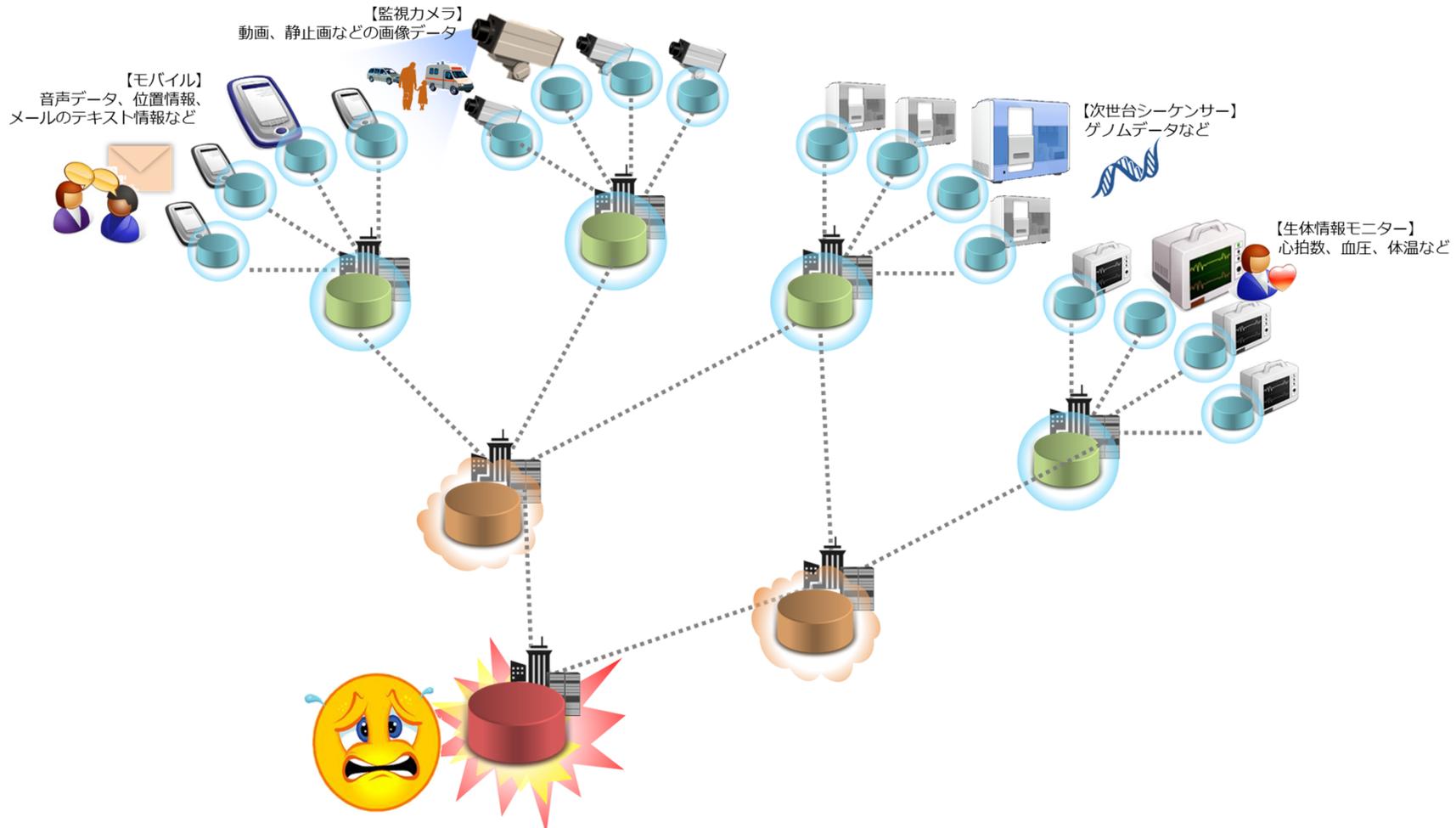
新しいアーキテクチャの創出に向けて

データはもっと増える！

- 監視カメラ市場：750万台（2011年）
 - 100GB/カメラとすると、年間750PB
- 国内スマホ市場 2010万台
 - 10GB/1台、年間2000万台売れているので、年間200PB
- 国内で蓄積される総データ量は年間400PBと推定されており、多くのデータは既にエッジ（縁）に存在している



データを集約するアプローチは、 近い将来うまくいかなくなる



丸山宏, 岡野原大輔 Edge-Heavy Data: CPS・ビッグデータ・クラウド・スマホがもたらす次世代アーキテクチャ
GICTF総会 特別講演 2012, <http://www.gictf.jp/doc/20120709GICTF.pdf>

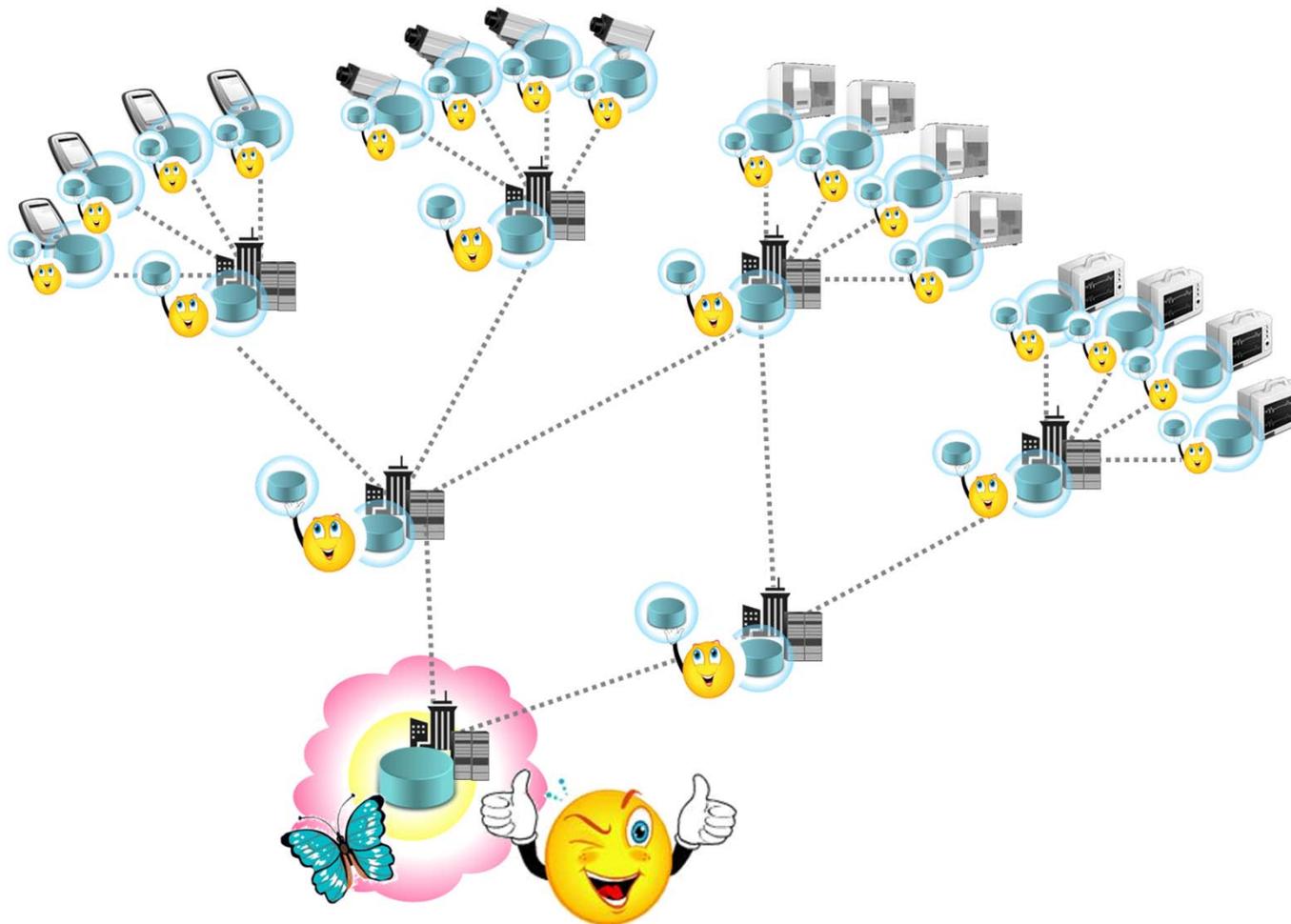
分野によっては、法律的な問題もある

- データを集めようとする、法律や政治の問題が避けられない
- それを解決しようとかんばっている内に、諸外国に技術的に遅れをとってしまう

Edge-Heavy Data

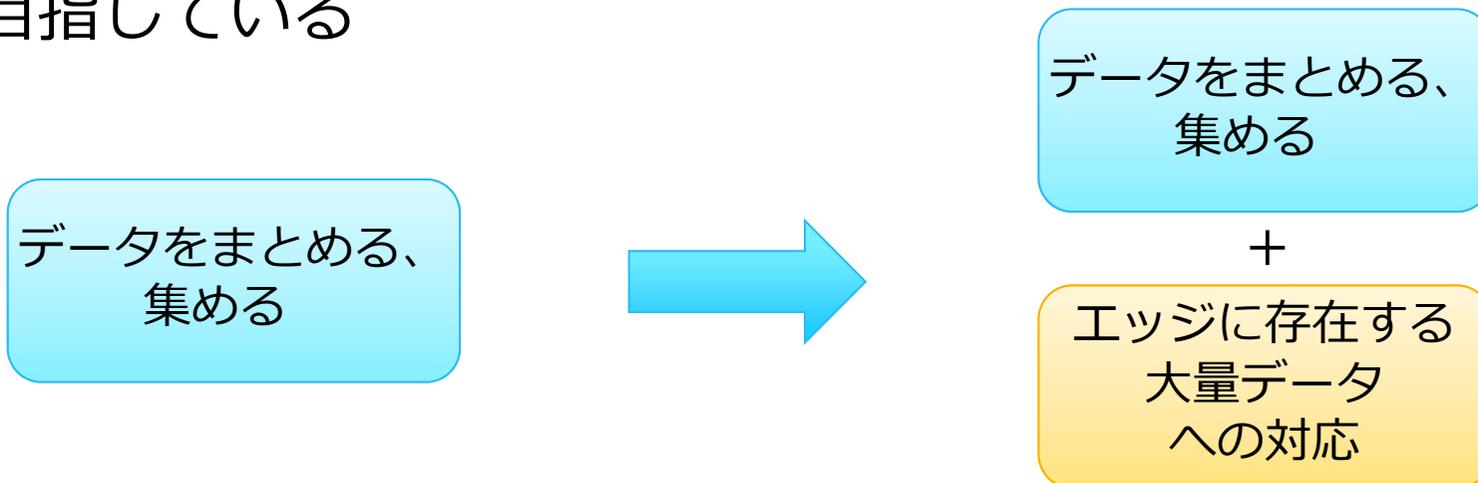
データを「貯めない」、「一カ所に集めない」。
その前提のもとで、
深い分析を実現するアーキテクチャを創る。

エッジ側が、もっとIntelligentになった、 階層型のアーキテクチャ



既存のアーキテクチャでは間違いなく 処理しきれなくなる

- 今の手法の延長線上でなんとか耐えぬくか？
- 新しい波に乗るか？
- Jubatusは、アーキテクチャの転換を引き起こすことを目指している





Copyright © 2006-2011

[Preferred Infrastructure All Right Reserved.](#)