



INTHEFOREST会社紹介

～超大規模データ・並列分散処理に強みのあるデータエンジニアリングカンパニー～

株式会社INTHEFOREST

Mail: sales@intheforest.co.jp / Tel:03-5848-2424
〒176-0023 東京都練馬区中村北1-13-13 OHD練馬ビル501

会社紹介

会社名 株式会社INTHEFOREST
(インザフォレスト)

設立 平成23年1月

所在地 東京都練馬区中村北1-13-13
OHD練馬ビル 5F

TEL 03-5848-2424

URL <http://www.intheforest.co.jp/>

代表取締役 富田 和孝

主要お取引先様

国立研究開発法人 理化学研究所様、コネクト株式会社様、株式会社Two Five様、伊藤忠テクノソリューションズ株式会社様、ヤフー株式会社様、株式会社ワークスアプリケーションズ様、楽天株式会社様、カルチュア・コンビニエンス・クラブ様、KSKアナリティクス様



事業内容

- Webサービス・データベースなど
分散処理OSSを中心に超大規模データ処理基盤の
構築・運用・保守コンサルティング
- 分散DB Apache Cassandra商用サポート

経営チーム紹介



富田 和孝(TOMITA Kazutaka)

INTHEFOREST代表取締役

日本Cassandraコミュニティーメンバー

DBエンジニア・システムアーキテクト

経歴：ぐるなび、シンプレックス（FX）、ISPなど、
DB中心としたシステム構築・運用・保守を担当。
大規模なデータ処理基盤作りに強み。
日本のCassandra第一人者。



高木 宏明(TAKAGI Hiroaki)

INTHEFOREST 取締役副社長

慶應ビジネススクール 非常勤講師「データサイエンス」担当

経歴：国立 高松高専専攻科修了、慶應ビジネススクール(MBA)
日立システムズ 主任、KSKアナリティクス（機械学習コンサル）
ビジネス推進部部長を経て現職。
IT・データ分析分野のビジネス開発・セールス・マーケティング
活動に長年従事。2019年4月よりINTHEFORESTに参画。

エンジニアチーム紹介



サハ・ベナルファ

ミドルウェアエンジニア・サーバサイドエンジニア

元カルタゴ大学助教授、コンピューターサイエンスの専門家

(IFでは会計システム、外国為替取引システム設計・開発・構築に従事)



シャルバニ・コエンデ

プロジェクトマネージャ・サーバサイドエンジニア

ムンバイ大学卒、インド・日本にて15年以上の金融などのシステム開発経験あり。

(IFでは国為替取引システム設計・開発・構築に従事)



有本 絵麻

フロントサイドエンジニア

企業内研究機関にて研究員向けデータ分析環境の開発及び分析業務支援に従事。

(IFでは日本語解析システム、行動履歴取得システム、など多数システムの構築に従事)



村岡 志保

基盤エンジニア・データベースエンジニア

分散DB Cassandraスペシャリスト、OSSを組み合わせたデータ処理環境構築に強み。

(IFではレコメンデーション基盤など、ビッグデータ保有の企業向けコンサルティングに多数従事)



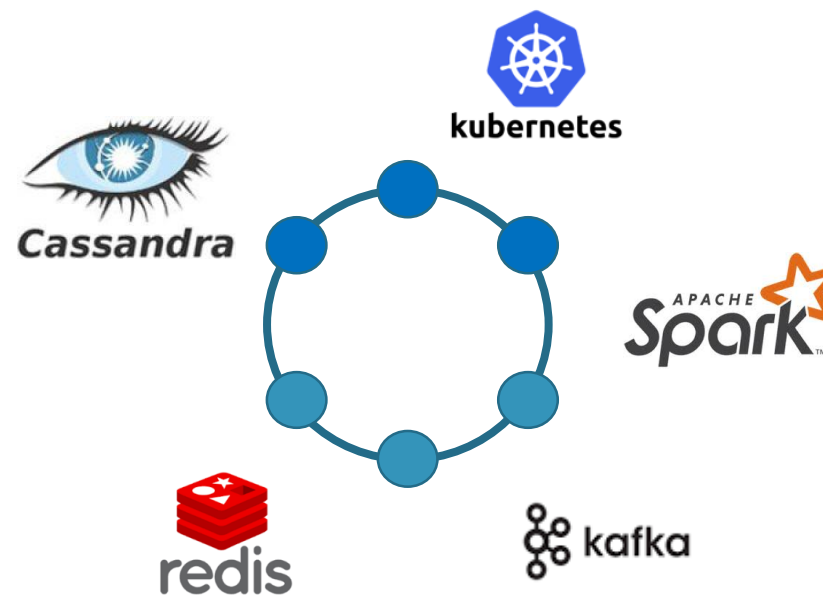
ミハイ・シュテウ

アナリティクスアドバイザー

インペリアルカレッジ・ロンドン博士課程在籍※
研究領域：Machine Learning/Deep Learning on Time Series

※ The Times Higher Education World University Rankings 世界総合8位 2018年

スキル・バックグラウンド



- ▶ 超大規模Web・基幹システムの構築経験
- ▶ 上場企業・金融監査に耐えるシステム内部統制対応

- ▶ Cassandraはじめとした分散処理・リソース管理OSSの世界レベルのエンジニアリング能力
- ▶ ITアーキテクトとしてインフラ～ミドルウェア～フロントエンドまで一気通貫した総合的知見

弊社データマネージドサービス

- ▶ 分散処理・リソース管理OSSを中心に弊社がシステム構築・運用・保守を行う
- ▶ お客様は”一時SI費”+“運用保守代行費”をサービス利用料としてお支払い



クラウド環境・ネットワーク環境構築含む
(AWS/Azure/オンプレ環境どちらにも対応)

Apache Zeppelin / Jupyter Notebook
Apache Mesos / Apache Hadoop(Yarn)
Presto/drill / impala Spark/Hadoop MR MongoDB/Redis/Cassandra Oracle/MySQL/PostgreSQL
Kubernetes/Docker Swarm
Docker
OpenStack / DC/OS
Xen/KVM/VMWare/VirtualBox
Debian (Linux) /CentOS(Linux)/FreeBSD
Hard Ware

他社とのコスト比較

- ▶ 他社：各分野のエンジニアを集めて大人数で設計・構築（大人数＋マネジメントオーバーヘッド発生）
- ▶ 弊社IF：OSSを用いてスペシャリストが少人数・短期間で設計・構築

他社:エンタープライズ構築
(Enterprise + SI)



他社: OSS構築
(OSS + SI)



INTHEFOERST: OSS構築



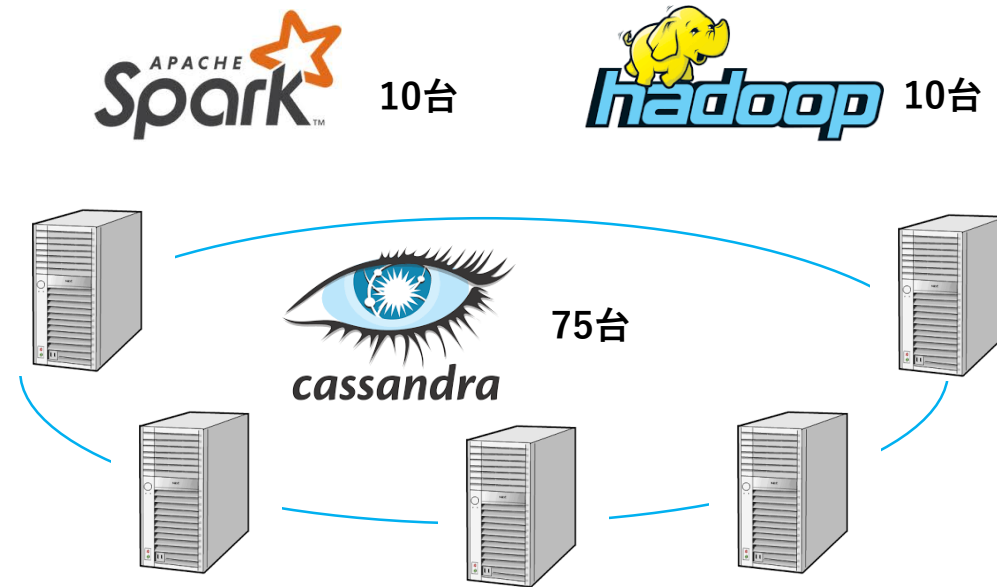
※Enterprise製品を用いた場合、OSSを用いた場合の2パターンと比較

ケース (大手ポイントメーカー様)

- ▶ 競合大手SI：費用2億円以上、約1年以上の設計構築期間
- ▶ 弊社IF：**総額4000万円、期間3ヶ月、人員3名**（データ基盤2名+予測モデル1名）
非常に厳しい**個人情報保護・セキュリティポリシー**にも対応

目的：商品レコメンデーション基盤構築

サーバ100台規模の並列分散処理環境
(インフラAWS、全てOSSで構築)



ケース（公益財団法人 高輝度光科学研究センター様）

- ▶ 大型放射光施設SPring-8において、加速器とビームライン制御システム MADOCA II のログ保管DB **Cassandra**の設計支援・サポートを弊社が担当

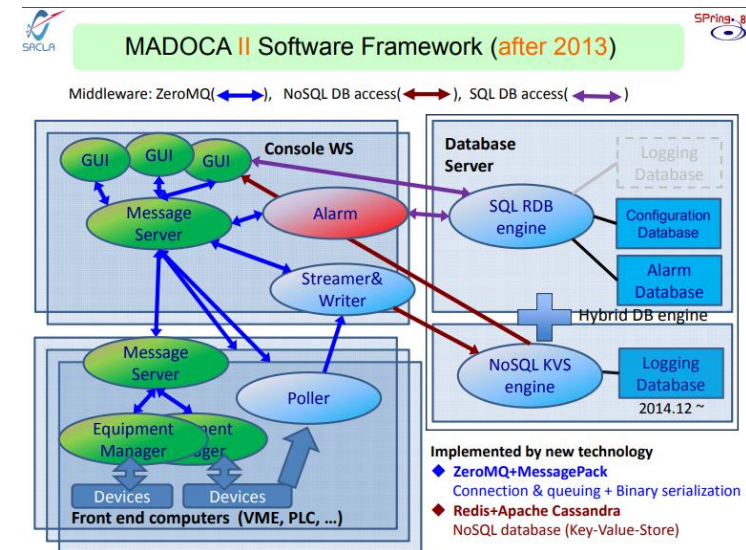
目的：放射光（センサーデータ）の保管



24時間切れ目なく50,000クエリ/秒を要求する
高負荷処理

MADUCA II Software Framework

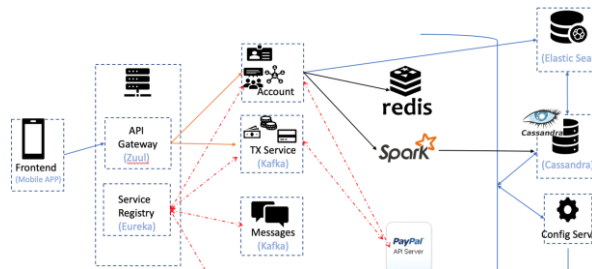
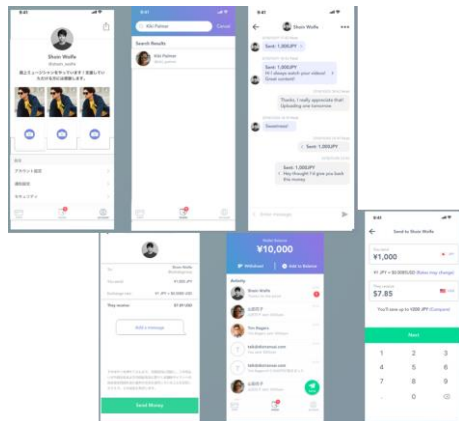
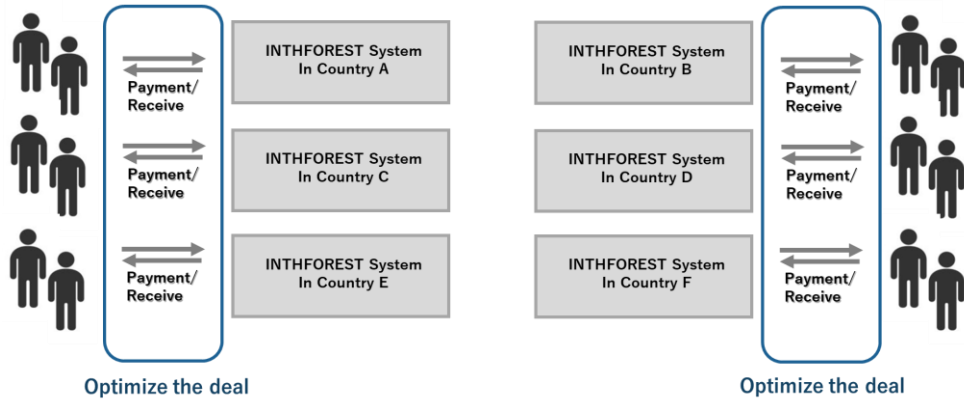
(MADUCA : Message And Database Oriented Control Architecture)



Ryotaro Tanaka "Conceptual Design of the Control System for SPring-8-II" 10th International Workshop on Personal Computers and Particle Accelerator Controls 14th - 17th October, 2014, Karlsruhe, Germany

ケース (CtoC国際送金システム)

目的：個人間の国際送金プラットフォーム



- ▶ **各国間リアルタイムのデータ処理連携**
- ▶ **ユーザ数万～数千万～数億人規模に対応した分散処理システム**
(シームレスなスケール)
- ▶ **為替価格ロジックを組み込んだシステム設計**
(対抗可能な企業が日本で数社)
- ▶ **圧倒的なコストパフォーマンス**
(構築コスト 50億円～100億円 が1/10以下へ)
- ▶ **処理規模に合わせてシステム運用コストが可変**
(ノード数増減で調整)
- ▶ **金融庁監査に耐えるシステム設計・監査ログ取得**
(法令・監査対応)

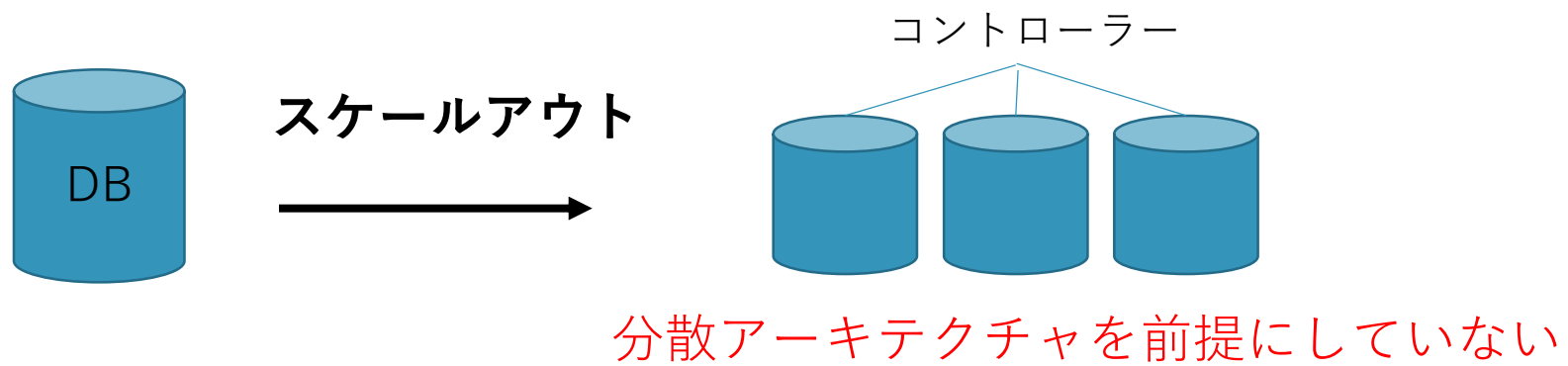
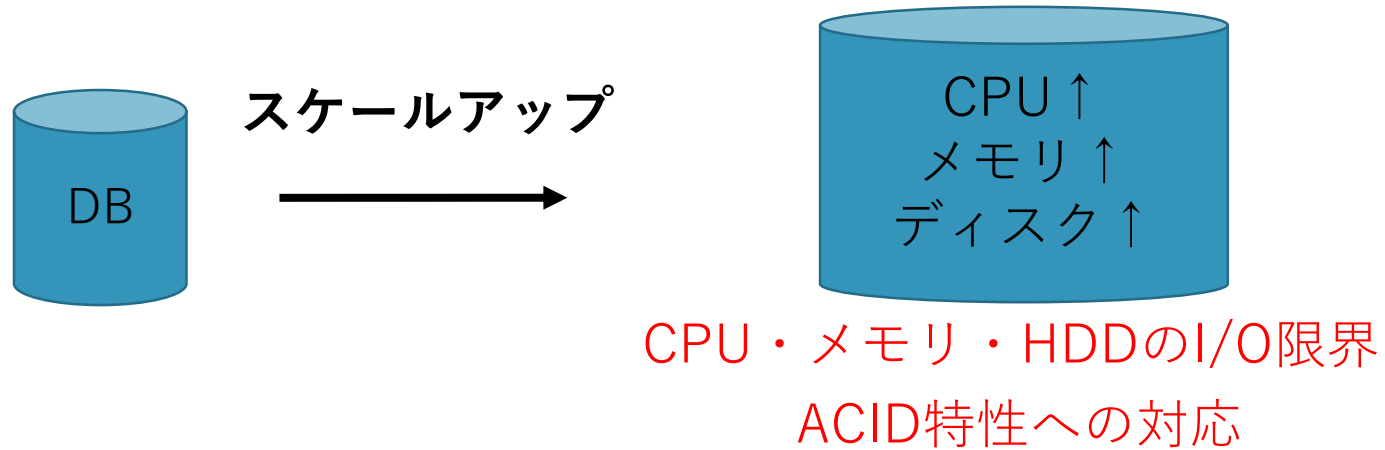


Appendix

現行データベースの限界(RDBMSの限界)

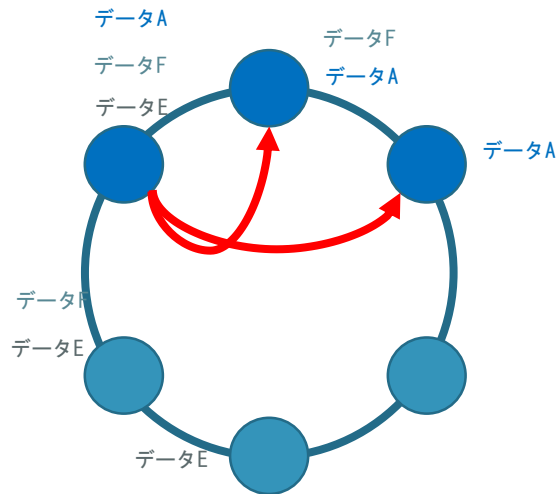
- ▶ **単一システムが前提**であり、複数環境（マルチクラウドなど）との整合性を保った状態でデータ連携を行うことをそもそものアーキテクチャとして想定されていない
- ▶ データサイズ **10～100Tbyte程度、秒間1000～5,000トランザクション**が処理限界
- ▶ 結果、ユーザ数十万人程度でアーキテクチャとして対応不可となる可能性
（大規模トランザクションに不向き）
- ▶ **今後、既存DBのリプレイス・データ基盤の再整備が必須**

現行データベースの限界(RDBMSの限界)



分散DB Apache Cassandra

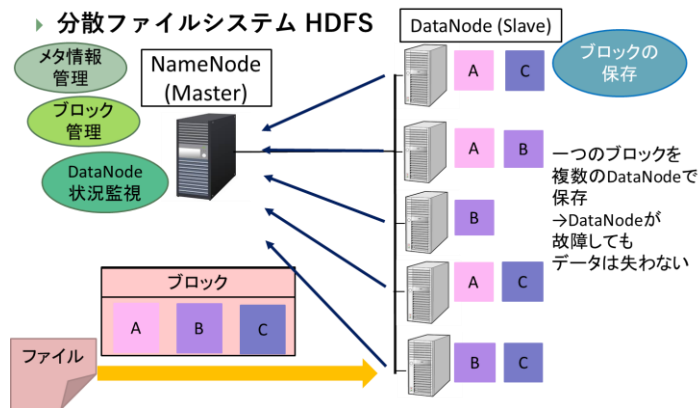
既存のデータベースに代わる次世代のPtoP型分散DB



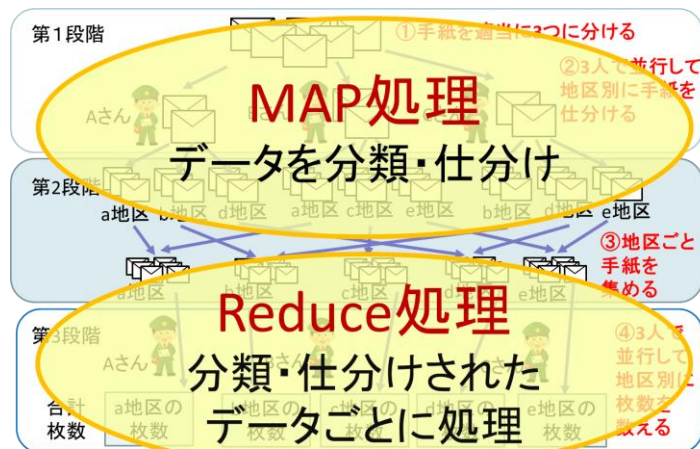
- ▶ iTunes・Instagram・NETFLIXなど億人ユーザ、ペタバイト・エクサバイト級システムで広く活用
※1ペタバイト=1024テラバイト、1エクサバイト=1024ペタバイト
- ▶ 超大規模・高可用性・高パフォーマンス
(1000万クエリ/秒以上も対応可能)
- ▶ データセンター間連携が可能
(マルチクラウド：AWS-Azure-オンプレ間のデータ連携も可能)
- ▶ 非常に高い可用性
(複数ノードが落ちても問題発生せず、障害に強い)
- ▶ リニアなスケールパフォーマンス
(トランザクション・データ量増大に対してもCassandraノード数を増やすことで対応可能)

分散処理フレームワーク Apache Hadoop

分散ファイルシステム：Hadoop Distributed File System、分散処理管理：MapReduce Framework

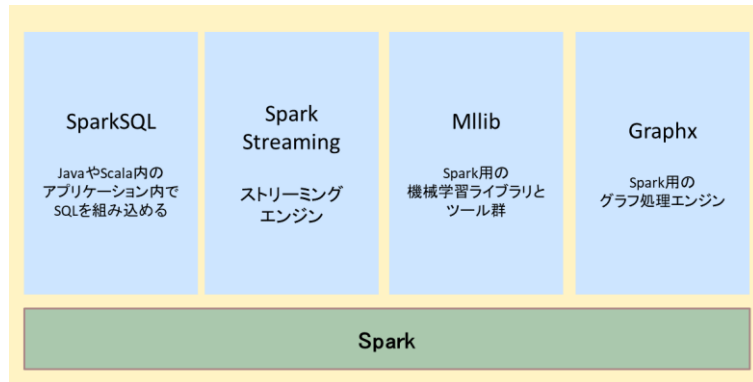


- ▶ 複数のIAサーバを束ねて、一つの大きな処理システムとして利用 (特に大量データの格納・処理に最適化)
- ▶ Map処理、Reduce処理のみを指定すればあとはフレームワークが並列分散処理を実現
- ▶ ノード数を増やせば、基本スケールする
- ▶ サーバが故障しても、ジョブは実行される



分散処理エンジン Apache Spark

MapReduceに限らず、DAG(有向非循環グラフ)型で柔軟・高速に実行できる並列分散処理エンジン

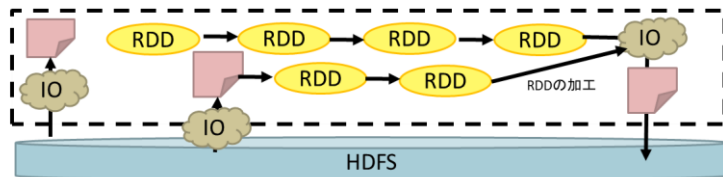


▶ 様々な分散処理のライブラリ群

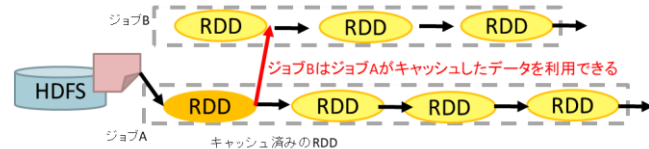
▶ Hadoopが普及にするにつれて、MapReduceフレームワークの処理効率が課題となっていた

▶ データ管理には向かないが、高速なデータ処理に向く (Hadoopの逆)

①ジョブが多段に構成される場合、複雑な処理を少ないジョブ数で実現できる



②複数のジョブで何度も同じデータを利用する場合
何度も利用するRDDは複数のサーバのメモリに分割してキャッシュできる



データエンジニア

- ▶ 海外のITエンジニアは**コンピューターサイエンス**を修め、
（医者・弁護士と同様に）**非常に高度な専門家**として地位が確立
- ▶ 特にビッグデータテクノロジーを取り扱う**データエンジニア**はデータサイエンティストより**貴重な存在**※

Data Engineer vs Data Scientist



※ 最近の米国求人情報検索サイトでの調査でもデータエンジニアの方がデータサイエンティストよりも
年収が高く平均13.5万ドル/年（約1500万円）、求人数も約5倍



**データ基盤・分散システム
構築・運用・サポートなど
お気軽にお問合せください**

**株式会社INTHEFOREST 担当 富田・高木
sales@intheforest.co.jp**