



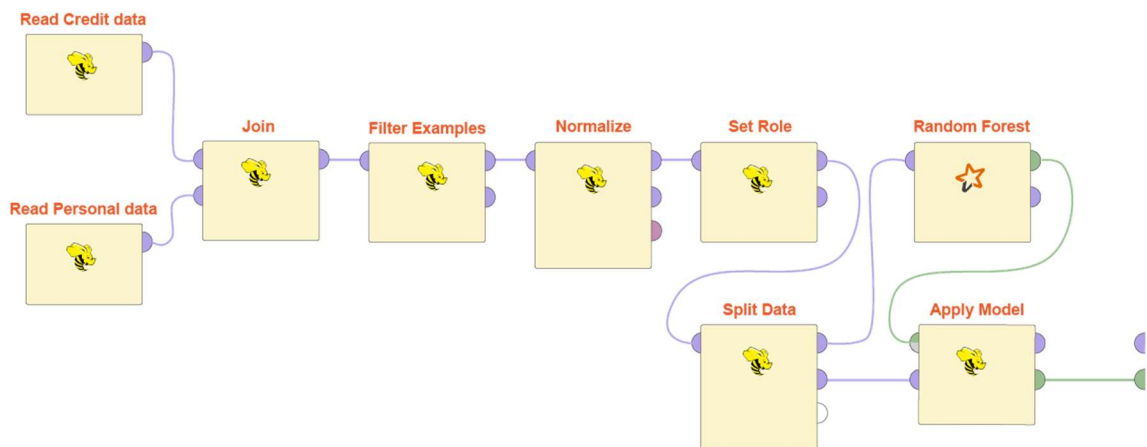
RapidMiner Radoop

ビッグデータ分析が簡単に

データが大きくなったらそのデータを保存して分析するために必要なメモリと処理能力も増えます。パソコンをアップグレードすることはできますが、限界があります。また、アップグレードにはパソコンのダウンタイムが必要です。それには費用がかかり、不便になる可能性があります。1つのパソコンでは十分でないので、分散コンピューティングが役立つことがあります。複数のコンピューターの力を組み合わせると、必要ならメモリと処理能力を簡単にスケールアップまたはスケールダウンすることができます。また、耐障害性と冗長性と計算効率も増えます。しかし、必要なソフトウェアのセットアップと管理は複雑になり、時間がかかります。そこで RapidMiner Radoop が役立ちます！

RapidMiner Radoop とは

Radoop は RapidMiner の製品で Apache Hadoop と Spark と Hive でコードなしの分散型機械学習にできる機能です。



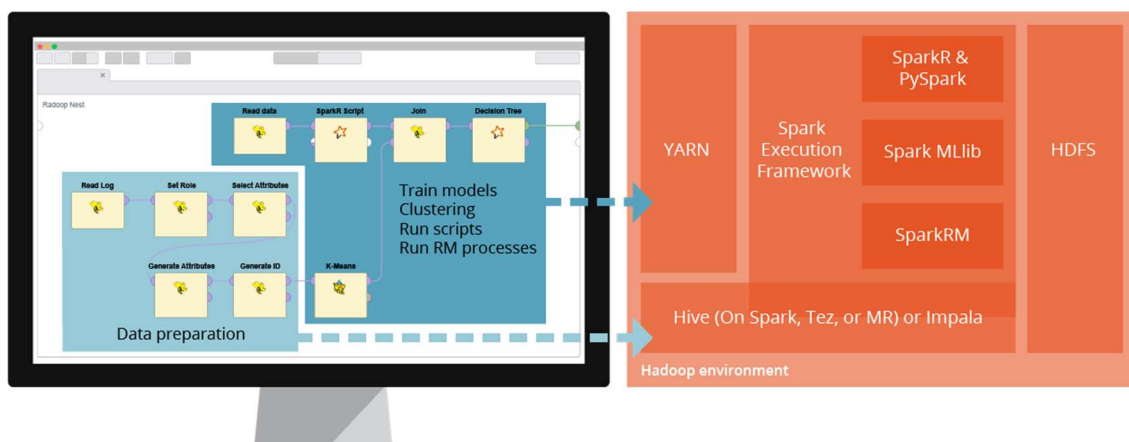
[Apache Hadoop](#) は信頼でき、スケール化できる、分散コンピューティングのために作った、オープンソースのプロジェクトです。Radoop では、分散ファイルシステム (HDFS) とアプリケーションのスケジューリングとクラスタのリソース管理 (YARN) の能力を使うために、Apache Hadoop を使っています。



[Apache Spark](#) は大きいデータ処理の統合分析エンジンです。Radoop では、モデリングのために、Apache Spark の学習ライブラリ (MLlib) を使っています。

[Apache Hive](#) はデータウェアハウスのソフトウェアで、SQL を使って、分散ストレージにある大きなデータセットの読み取りと書き込みと管理のために使います。

Radoop でも、同じ目的に使っています。



利点は何か

- **使いやすさ**：コードなしのプログラミング環境を使って Hadoop と Spark と Hive からデータサイエンスの複雑さを取り除きます。
- **豊富な機能**：70 以上の Hadoop と Spark と Hive のオペレーターがあります。また、Spark MLlib のアルゴリズムも全アクセスがあります。
- **柔軟性**：既存の SparkR や PySpark や Pig や HiveQL のコードの再利用を可能にしたり、Hive Script や Pig Script や Spark Script のオペレーターを使って新しいコードを作れたりもします。

- セキュリティ：コンピュータネットワーク認証プロトコル（Karberos）とデータアクセス認証（Apache Sentry と Apache Ranger）と HDFS 暗号化と Hadoop なりすましのサポートがあります。

さらに、エンタープライズバージョンは [Single Process Pushdown と SparkRM オペレーター](#)で RapidMiner の全 1500 以上のオペレーターを Hadoop で実行する機能が含まれています！

ビッグデータ分析は RapidMiner Radoop で簡単になり、RapidMiner のマーケットプレイスから無料でダウンロードして開始ができます。エンタープライズサポートと機能が必要になれば、KSK アナリティクスがお手伝いします。

オペレーター概要

Data Access	- Hive テーブルからと Hive テーブルに読み取り、保存、追加を行います - CSV ファイルからの読み取りと CSV ファイルへの書き込み（HDFS、Azure Blob または Datalake、Amazon S3、またはローカルファイルシステム） - データベースからの読み取りとデータベースへの書き込み（MySQL、PostgreSQL、
--------------------	---

	Sybase、Oracle、HSQLDB、Ingres、Microsoft SQL Server、または ODBC ブリッジを使用するその他のデータベース)
Blending	- 属性の選択、構築、集計など - 例のサンプリング、フィルタリング、並べ替えなど - ExampleSets をピボットや結合する
Cleansing	- 欠損値を置き換え、重複した例を削除、正規化、PCA など
Modeling	- K-Means クラスタリング、PCA、相関マトリクス、共分散マトリクス、ナイーブベイズ、ロジスティック回帰、決定木など
Scoring	- モデルを適用
Validation	- パフォーマンス評価と分割検証
Utility	- ループ、スクリプティング、サブプロセス、Single Process Pushdown、SparkRM など