



無料電子ブック

学習

bigdata

Free unaffiliated eBook created from
Stack Overflow contributors.

#bigdata

.....	1
1:	2
.....	2
Rich-untamed-Data.....	2
Examples.....	3
.....	3
.....	4
.....	4
2: Big Data / Hadoop Security	6
.....	6
.....	6
Examples.....	6
ACL.....	6
.....	7

You can share this PDF with anyone you feel could benefit from it, downloaded the latest version from: [bigdata](#)

It is an unofficial and free bigdata ebook created for educational purposes. All the content is extracted from [Stack Overflow Documentation](#), which is written by many hardworking individuals at Stack Overflow. It is neither affiliated with Stack Overflow nor official bigdata.

The content is released under Creative Commons BY-SA, and the list of contributors to each chapter are provided in the credits section at the end of this book. Images may be copyright of their respective owners unless otherwise specified. All trademarks and registered trademarks are the property of their respective company owners.

Use the content presented in this book at your own risk; it is not guaranteed to be correct nor accurate, please send your feedback and corrections to info@zzzprojects.com

1: ビッグデータのい

このセクションでは、bigdataのとがなぜそれをいたいのかをします。

ビッグデータは、4Vでけられるデータである。これらは、ボリューム、ベロシティ、バラエティ、およびベラシティです。

1. ボリューム - データがテラバイトやペタバイトのようなボリュームにある。によると、たちは23にの90のデータをしています。
2. ベロシティ - データがシステムをれる。たとえば、ソーシャルネットワーキングサイトにコンテンツをアップロードするものユーザーが、テラバイト/ののデータをします。
3. バラエティ - そのについたさまざまなデータ。いRDBMSのはう、メール、XMLなど、、センサーデータなどがです。
4. Veracity - なデータにのあるをるです。これは、ビジネスののがデータのにするため、データのもなとえることができます。

ビッグデータのとにされるもなプラットフォームは、**Hadoop** フレームワークです。それは2つのものでされています

1. HadoopファイルシステムHDFS - データはHadoop Distributed File SystemHDFSにされます。HDFSは、にサーバーにするなとはなり、ハードウェアのクラスタです。HDFSにあるデータは、とフレームワーク。
2. MapReduceMR - これはHadoopのデフォルトフレームワークです。 [MapReduce](#) Apache Hadoopのです

Hadoopのにより、しいツールがHadoopコミュニティにしました。ものあるツール/フレームワークのない

1. [Apache Spark](#)
2. Apache Storm
3. [Apache Flink](#)

などなど..

なHDFSのストレージメカニズムはほとんどありません。

1. [ハイク](#)
2. [HBase](#)
3. [カサンドラ](#)

などなど..

はビッグデータのにがあり、データをどのようにしているかにきな

いがあることがわかります。パラレルユニバースでは、きなデータを**Rich-untamed-Data**とぶこともできます。このなデータをざんしなければなりません。きなデータでは、のデータのされたをできるかもしれません。

ショッピングウェブサイトでののクリックにおいて、その、クリック、およびそのウェブサイト
にやされたがのがげられ、オンラインにをし、ユーザのについでをするようする。

Examples

ビッグデータとは

ビッグデータは、もなで、データのさまざまなによってされるとしてできます。これらのなるは、

ボリュームのデータ、ベロシティデータフローの、バラエティ、、データとデータに
づいたな

これらのメトリックは、いのリレーショナルデータベースではにくいものでした。しいシステムのがじ、ビッグデータのがされました。くの々がビッグデータのをしています、データのリーダーがするビッグデータのはほとんどありません。

- 「なデータは、にされるハードウェアやソフトウェアツールのをえて、ユーザーのにするえられるとともに、データを、、します。」 Teradata Magazine article、2011
- 「ビッグデータとは、なデータベースソフトウェアツールがキャプチャ、ストア、、するをえたデータセットをします」 The McKinsey Global Institute、2012
- 「ビッグデータは、でなデータセットのであり、のデータベースツールをしてすることがになります」 Wikipedia、2014
- 「ビッグデータは、の、の、プロセスのをにするしいのをとする、、、なです」 Gartner、2012

データが **"Big"** になったら



IOPS: Input/Output Operations Per Second

ビッグデータの

ビッグデータとは、のデータアプリケーションではえないほどでなデータセットのです。、キャプチャ、データのキュレーション、、、ストレージ、ビジュアライゼーション、クエリ、、、のプライバシーなどのがあります。

きなデータのな

ソーシャルネットワーキングサイトのFacebookにされたデータ。FacebookはテラバイトTBのデータをしています。されたデータは、、、などであってもよい。データは、されていないものからされていないものまで々である。たちがそのをはっきりとっているように、ている、かちっている、あるいはしているかもしれません、またははににわないデータです。これらのデータはすべてにBigDataをします。

ビッグデータにはがありますか

きなデータには、さまざまなデバイスやアプリケーションによってされたデータがまれます。ピ

ビッグデータののにあるフィールドのいくつかをにします。

- ブラックボックスデータヘリコプター、ジェットなどのコンポーネントで、の、マイクとイヤフォンの、のをキャプチャします。
- ソーシャルメディアのデータFacebookやTwitterなどのソーシャルメディアは、のもの々がしたとをしています。
- データデータには、がしたなるのについてわれた「い」および「」にするがされています。
- グリッドデータグリッドデータは、にしてのノードによってされるをする。
- データデータには、のモデル、、、およびがまれます。
- エンジンデータエンジンは、さまざまなデータベースからくのデータをします。
- センサデータセンサでするさまざまなデバイスからのデータ、およびデータ、データ、、などデータ。



したがって、ビッグデータはな、、およびななデータをむ。そののデータは3あります。

1. Structured data : Mostly data from Relational Databases.
2. Semi Structured data : XML data, email data.
3. Unstructured data : Word, PDF, Text, Media Logs.

オンラインでビッグデータのいをむ <https://riptutorial.com/ja/bigdata/topic/6890>ビッグデータのい

2: Big Data / Hadoop Securityをいめる

き

さまざまなHadoopでデータをすることができます。それぞれのにはのがあります。よりいをするためにのをみわせることもできます。このトピックでは、メソッドのについてします

1. Kerberosはネットワークプロトコルです。

a. エントリレベルでユーザーをします。

b. Kerberosは、のないユーザーによるへのアクセスをします。ログイン、テーブル、カラム、フォルダ、ファイルレベルなどのレベルのはされません

2. Apache Sentryは、のロールベースをするためのシステムです

a. Hive、Impala、Solrなどのアプリケーションレベルの。のユーザー/グループのDB、テーブル、レベルのアクセスをします。

b. Hive、ImpalaなどのアプリケーションのにあるHDFSフォルダをすることはできません。
/user/hive/warehouse/prod.db/table1にされているHiveテーブルprod.table1。 Hueでのセントリのは、Hueのテーブル/アクセスのみをしますが、ユーザーはHDFSのフォルダにアクセスすることができます

c. Hive、ImpalaなどとのないHDFSフォルダはされません

3. アクセスリスト ACLは、アクセスエントリ**ACE**のリストです。 **ACL**の**ACE**はトラスティをし、そのトラスティにして、、またはされるアクセスをします

a. フォルダレベルのアクセスは、

4. HDFSは、HDFSからみきされるデータのなエンドツーエンドをします

a. データをすることにより、さらにレベルのセキュリティがされます。に、データのは、のなる、、によってされています

Examples

ACLの

```
hadoop fs -setfacl
```

オンラインでBig Data / Hadoop Securityをいめるをむ

<https://riptutorial.com/ja/bigdata/topic/9869/big-data---hadoop-security>をいめる

クレジット

S. No		Contributors
1	ビッグデータのい	Ani Menon , Community , Mr. P , NeoWelkin , Sayali Sonawane
2	Big Data / Hadoop Securityをいめる	saranvisa