Big Data For Dummies

Part 1 : Getting Started with Big Data

In this part

- Trace the evolution of data management
- Define big data and its technology components
- Understand the different types of big data
- Integrate structured and unstructured data.
- Understand the difference between real-time and non-real-time data
- Scale your big data operation with distributed computing

Chapter 1 : Grasping the Fundamentals of Big Data

In this Chapter

- Looking at a history of data management
- Understanding why big data matters to business
- Applying big data to business effectiveness
- Defining the foundational elements of big data
- Examining big data's role int the future

Chapter 2 : Examining Big Data Types

In this Chapter

- Identifying structured and unstructured data
- Recognizing real-time and non-real-time requirements for data types
- Integrating data types into a big data environment

Although data management has been around for a long time, two factors are new in the big data world:

- Some sources of big data are actually new like the data generated from sensors, smartphone, and tablets
- Previously produced data hadn't been captured or stored and analyzed in a usable way. the main reason for this is that the technology wasn't there to do so. In other words, we didn't have a cost-effective way to deal with all that data.

Defining Structured Data

Exploring sources of big structured data

- **Computer- or machine-generated:** Machine-generated data generally refers to data that is

created by a machine without human intervention.
- **Human-generated:** This is data that humans, in interaction with computers, supply.

Understanding the role of relational databases in big data

Defining Unstructured Data

Exploring sources of unstructured data

Here are some examples of machine-generated unstructured data:

- **Satellite images:** This includes weather data or the data that the government captures in its satellite surveillance imagery.
- **Scientific data:** This includes seismic imagery, atmospheric data, and high energy physics.

The following list shows a few examples of human-generated unstructured data:

- **Text internal to your company:** Think of all the text within documents, logs, survey results, and e-mails. Enterprise information actually represents a large percent of the text information in the world today.
- **Social media data:** This data is generated from the social media platforms such as YouTube, Facebook, Twitter, LinkedIn, and Flickr.
- **Mobile data:** This includes data such as text messages and location information.
- **Website content:** This comes from any site delivering unstructured content, like YouTube, flickr, or Instagram.

Understanding the role of a CMS in big data management

- CMS[1)]
- AIIM[2)] : www.aiim.org
- ECM[3)]

Looking at Real-Time and Non-Real-Time Requirements

- Monitoring for an exception with a new piece of information, like fraud/intelligence
- Monitoring news feeds and social media to determine events that may impact financial markets, such as a customer reaction to a new product announcement
- Changing your ad placement during a big sporting event based on real-time Twitter streams
- Providing a coupon to a customer based on what he bought at the point of sale

---

관련 문서

- Machine Learning

1)

Content management systems
2)

Association for information and Image Management
3)

Enterprise Content Management