AI융합 연계전공(010982)-인공지능의 이해 3강
Data Engineering for Artificial Intelligence

Jaewook Byun

Ph.D., Assistant Professor, Department of Software, Sejong University
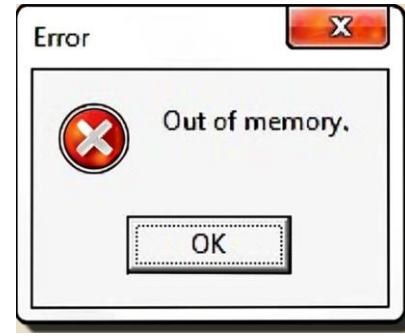
Table of Contents

1. Introduction to Data Engineering for Artificial Intelligence (Big Data Processing)
2. Data Collection & Processing

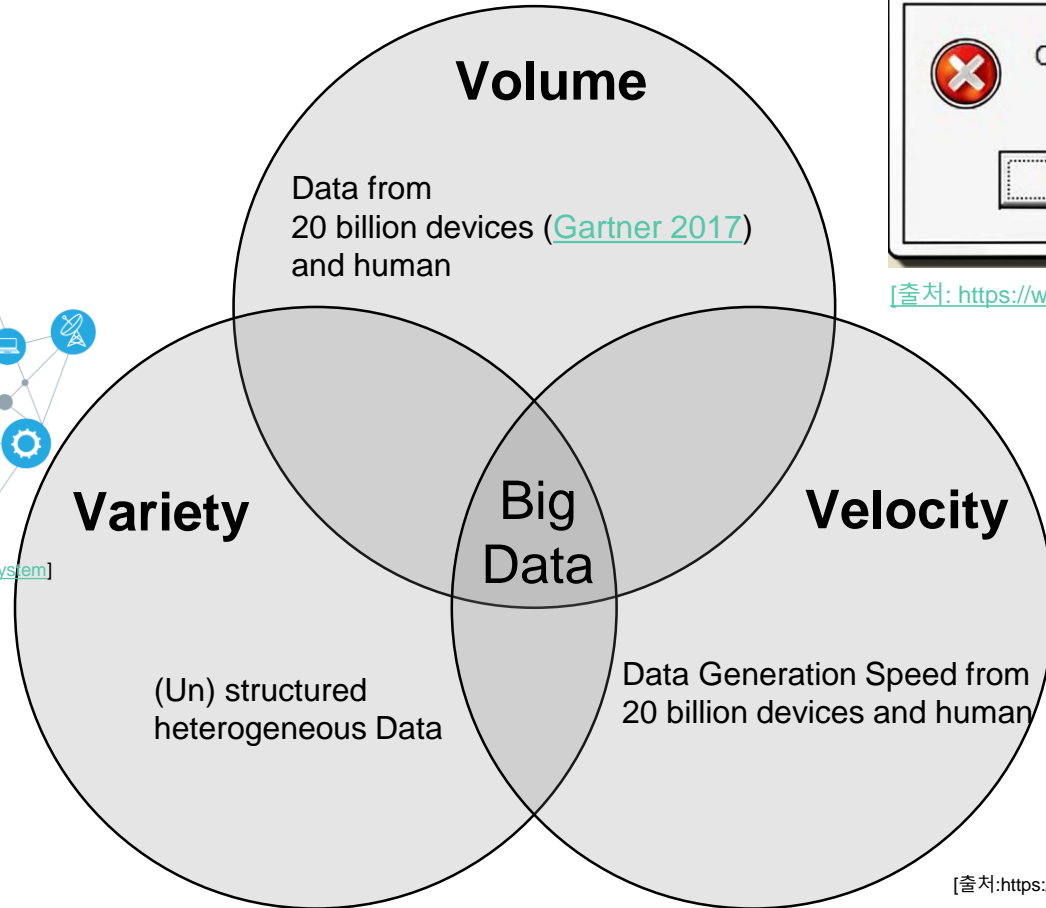# Introduction to Data Engineering for Artificial Intelligence

Big Data
&
Big Data
Techniques

→

Artificial
Intelligence

[1][2][3]

Introduction to Big Data

Volume

Data from
20 billion devices (Gartner 2017)
and human

Variety

Big Data

Velocity

[출처: https://www.sensorsexpo.com/iot-ecosystem]

(Un) structured
heterogeneous Data

Data Generation Speed from
20 billion devices and human
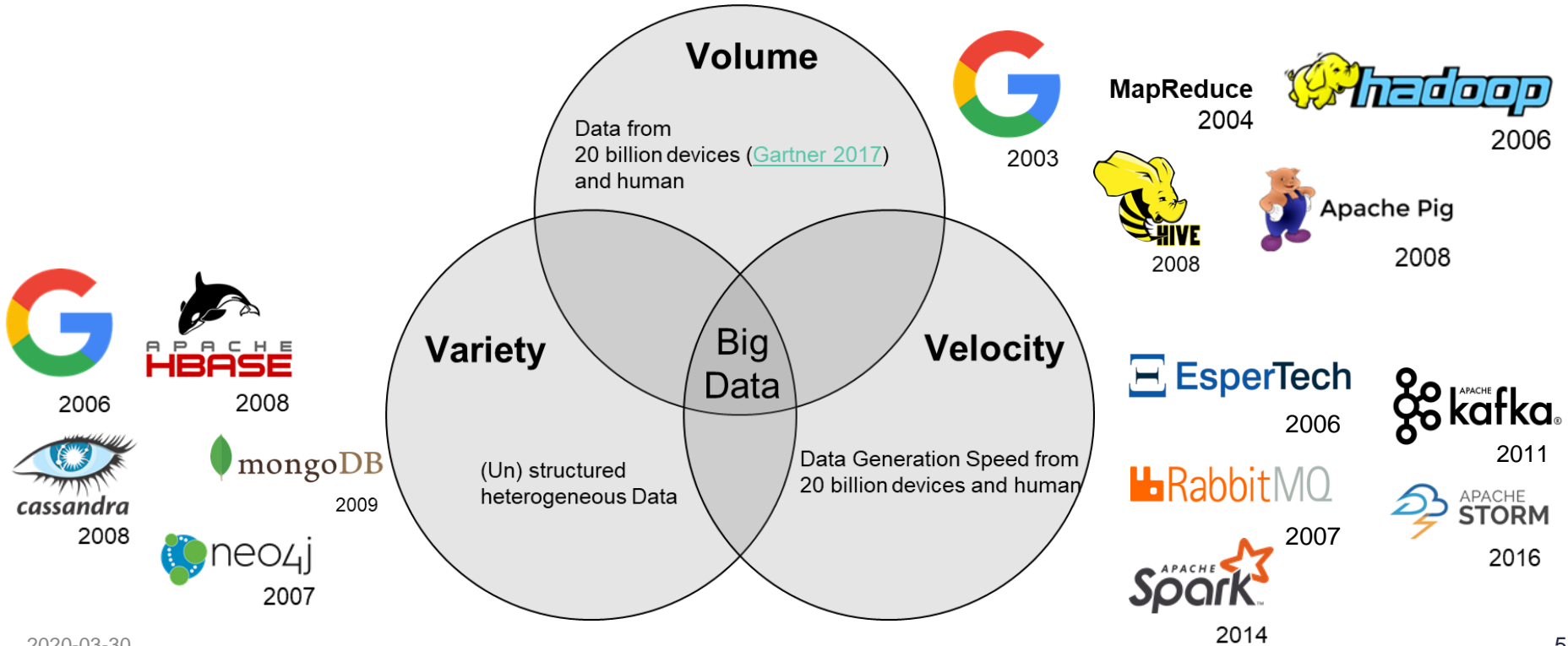
[출처: https://www.youtube.com/watch?v=qE133EHwoI0]

Every 60 seconds

98,000+ tweets

695,000 status updates

11 million instant messages

698,445 Google searches

168 million+ emails sent

1,820TB of data created

217 new mobile web users

[출처:https://pbs.twimg.com/media/CJpQUc2WUAQHztb.png]

Introduction to Big Data Technologies

- Big Data Technologies (Definition of IDC 2011)
  - A new generation of technologies and architectures extract value from Big Data
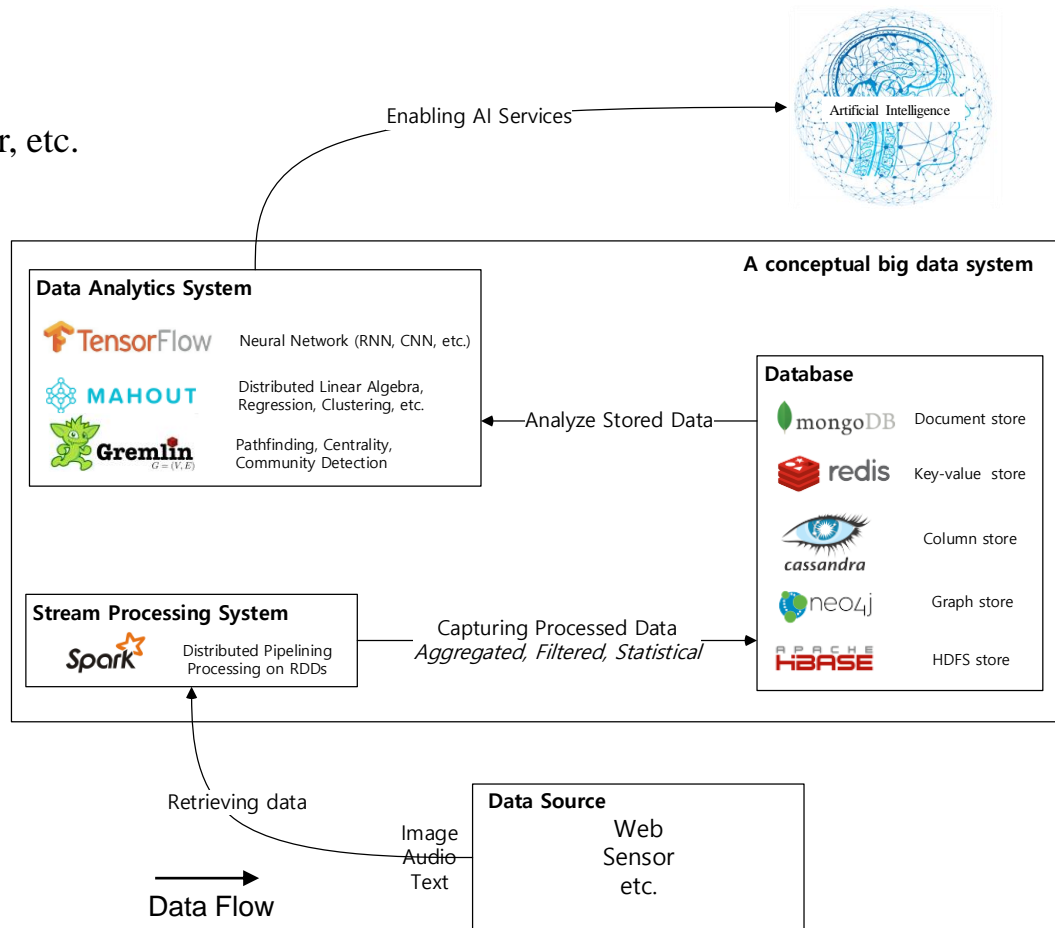
[4]

6

Conceptual Architecture

- Data Source
  - Image, Audio, Text, etc. from Web, Sensor, etc.
- Stream Processing System
  - e.g., Distributed pipelining processing on RDDs
- Database
  - Document store
  - Key-value store
  - Column store
  - Graph store
  - HDFS store
- Data Processing System
  - Image/audio Processing
  - Text Processing
  - Graph Processing
- Applications
  - AI Services

Enabling AI Services

Artificial Intelligence

A conceptual big data system

Data Analytics System

TensorFlow — Neural Network (RNN, CNN, etc.)

MAHOUT — Distributed Linear Algebra, Regression, Clustering, etc.

Gremlin $G = (V, E)$ — Pathfinding, Centrality, Community Detection

Database

mongoDB — Document store

redis — Key-value store

cassandra — Column store

neo4j — Graph store

HBASE — HDFS store

Analyze Stored Data

Stream Processing System

Spark — Distributed Pipelining Processing on RDDs

Capturing Processed Data
Aggregated, Filtered, Statistical

Retrieving data

Data Source
Image
Audio
Text
Web
Sensor
etc.

Data Flow

Reference

- [1] -https://www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data/#6c2f83b44f6d

- [2] - http://economychosun.com/client/news/view.php?boardName=C00&t_num=9318

- [3] - https://www.samsungsds.com/global/ko/support/insights/091517_CX_CVP3.html

- [4] – https://mattturck.com/data2019/

- Super Mario World - https://www.youtube.com/watch?v=qv6UVOQ0F44
- Snake - https://www.youtube.com/watch?v=zIkBYwdkuTk

Thank you for listening