

Big Data Analytics in Healthcare

출처 : <https://www.udacity.com/>

강좌 찾기

[Catalog](#)[Explore](#)[Nanodegree](#)[For Business](#)[Blog](#)[My Classroom](#)

Courses and Nanodegree Programs



CATEGORY

[All](#)[Android](#)[Artificial Intelligence](#)[Data Analytics](#)[Data Science](#)[Deep Learning](#)[Developer Essentials](#)[Digital Marketing](#)[Georgia Tech Masters in CS](#)[iOS](#)

Health Informatics in the Cloud

■ ■ ■ Beginner

This survey course provides a broad, forward-facing overview of contemporary health informatics within the context of the US healthcare delivery system.

IN COLLABORATION WITH: **Georgia Institute of Technology**

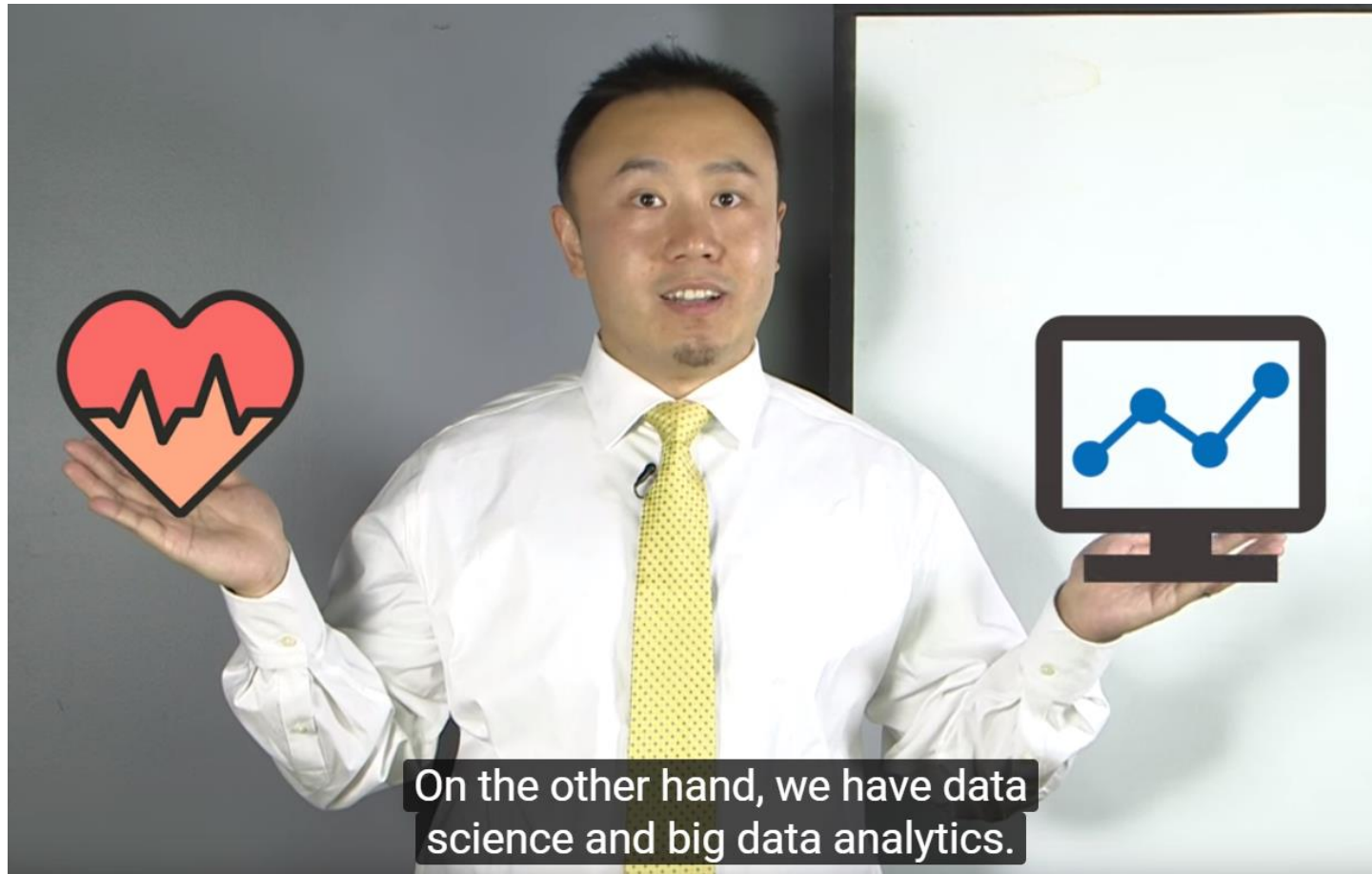


Big Data Analytics in Healthcare

In this course, we introduce the characteristics of medical data and associated data mining challenges on dealing with such data. We cover various algorithms and systems.

IN COLLABORATION WITH: **Georgia Institute of Technology**

Lesson 1: Introduction to Big Data



주로 다룰 주제

1. Healthcare 데이터와 관련 어플리케이션
2. 데이터 과학과 빅데이터 분석
 - 데이터 알고리즘
 - 처리를 위한 시스템
 - 빅데이터 분석 방법
3. 위의 두 주제의 상호관련성과 적용방법

Lesson 1: Introduction to Big Data



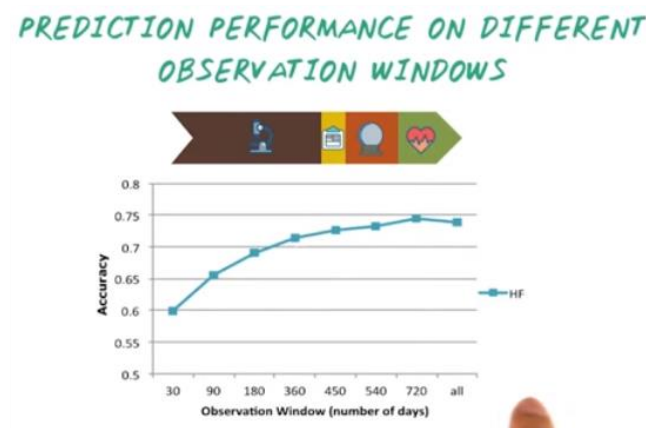
이 강좌의 목표

1. Healthcare 데이터 이해
2. 분석 알고리즘 이해
3. 빅데이터 시스템 이해

Lesson 1: Introduction to Big Data

이 강좌를 마치고 난후의 기대 효과
=> Healthcare 데이터에 기반한 모델을 구축할 수 있다.

예를 들면,
개별 질병 위험 예측 모델
공통된 특징이 있는 환자 군집에 따른
치료 추천



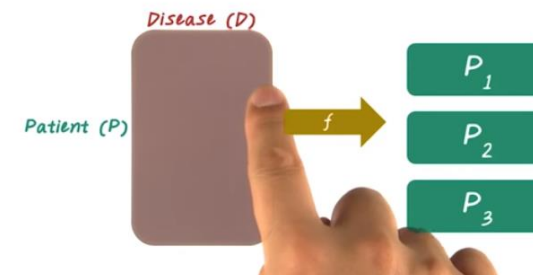
DISEASE SUBTYPES

Phenotype 4 (31.1% of patients)	Phenotype 2 (31.5% of patients)	Phenotype 6 (24.3% of patients)
Hypertension	Hypertension	Hypertension
ACE Inhibitors	Beta Blockers Cardio-Selective	Calcium Channel Blockers
Thiazides and Thiazide-Like Diuretics	Angiotensin II Receptor Antagonists	Antihypertensive Combinations
	Loop Diuretics	Antiadrenergic Antihypertensives
	Potassium	Potassium Sparing Diuretics
	Nitrates	
	Alpha-Beta Blockers	
	Vasodilators	

PATIENT SIMILARITY



WHAT IS CLUSTERING?



Lesson 1: Introduction to Big Data

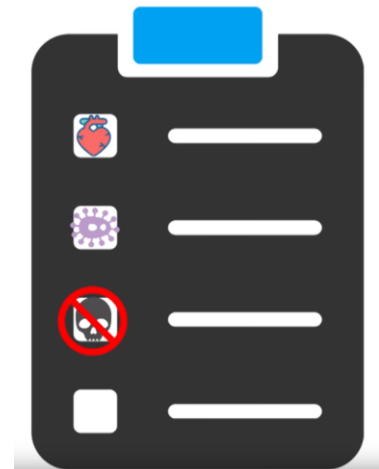
- 현재 Healthcare분야에서의 문제점

1. 미국의 한해 3.8조달러로 건강관리에 소비
 2. 이중에서 7,650억 달러가 낭비
 3. 예방 가능한 사망자가 연간 200 ~ 40000건 발생 (사망원인 3위)
- => 비용은 높고 낭비가 많고 효율이 낮음.

OVERALL SPENDING

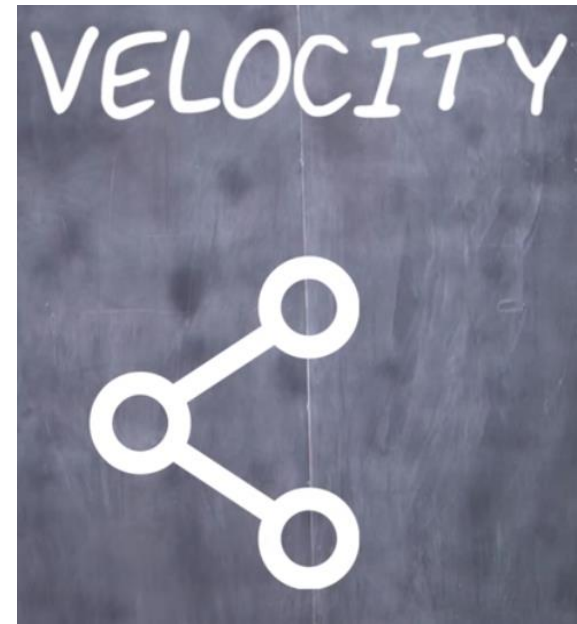
=

3.8 TRILLION



Lesson 1: Introduction to Big Data

- 빅데이터의 4가지 특징 (4 V)



Lesson 1: Introduction to Big Data

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

Who Are These People?

If capitalizing on big data depends on hiring scarce data scientists,

, what data scientists do is make discoveries while swimming in data. I

Data scientists realize that they face technical limitations, but they don't allow that to bog down their search for novel solutions. As they make discoveries, they communicate what they've learned and suggest its implications for new busi



Lesson 1: Introduction to Big Data

Quiz Time

Getting a graduate level degree is necessary for becoming a data scientist.

☐ TRUE

☐ FALSE

First, true or false?

Lesson 1: Introduction to Big Data

Quiz Time

*How much does a data scientist
make on average?*

Lesson 1: Introduction to Big Data

Quiz Time

*What skills do data scientists
need to know?*

A large, empty rectangular box with a white border, intended for the user to write their answer to the quiz question.

Lesson 2: Big Data Course Overview

BIG DATA. BIG PICTURE.



Lesson 2: Big Data Course Overview

HEALTHCARE APPLICATIONS



*Predictive
Modeling*



*Computational
Phenotyping*



*Patient
Similarity*

Lesson 2: Big Data Course Overview

예측 모델 *CHALLENGES*

So much data!

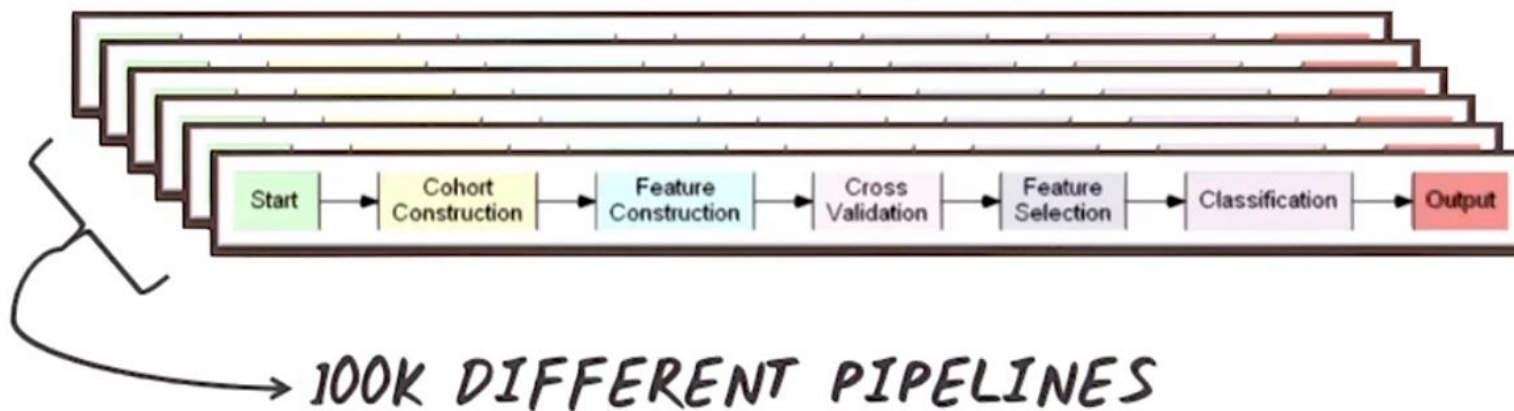


Lesson 2: Big Data Course Overview

예측 모델 *CHALLENGES*

So many models!

Predictive Modeling Pipeline



Lesson 2: Big Data Course Overview

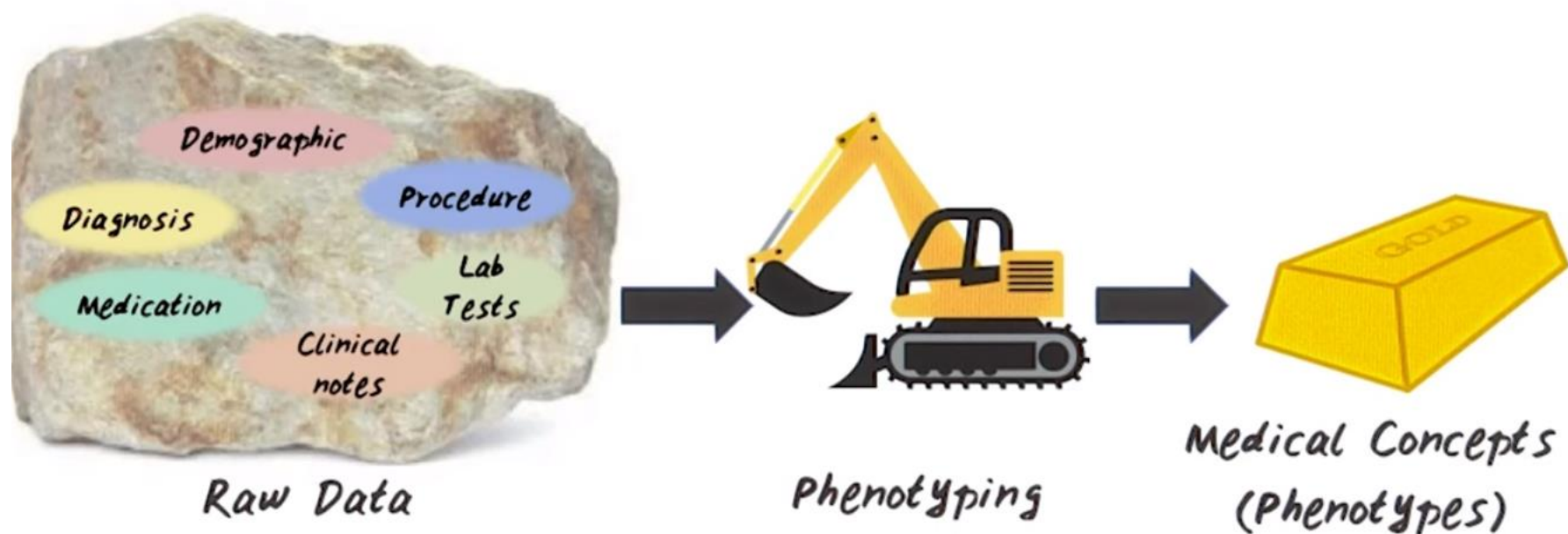
HEALTHCARE APPLICATIONS



*Computational
Phenotyping*

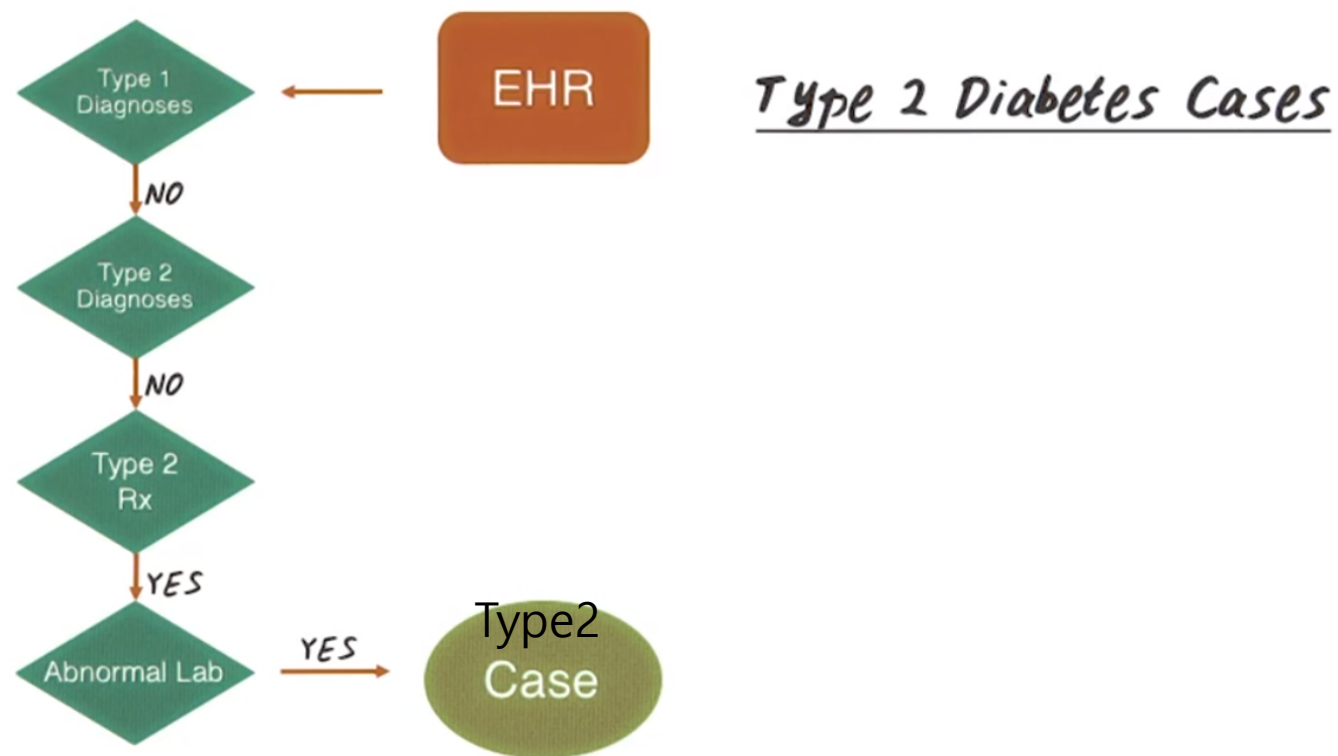
Lesson 2: Big Data Course Overview

COMPUTATIONAL PHENOTYPING

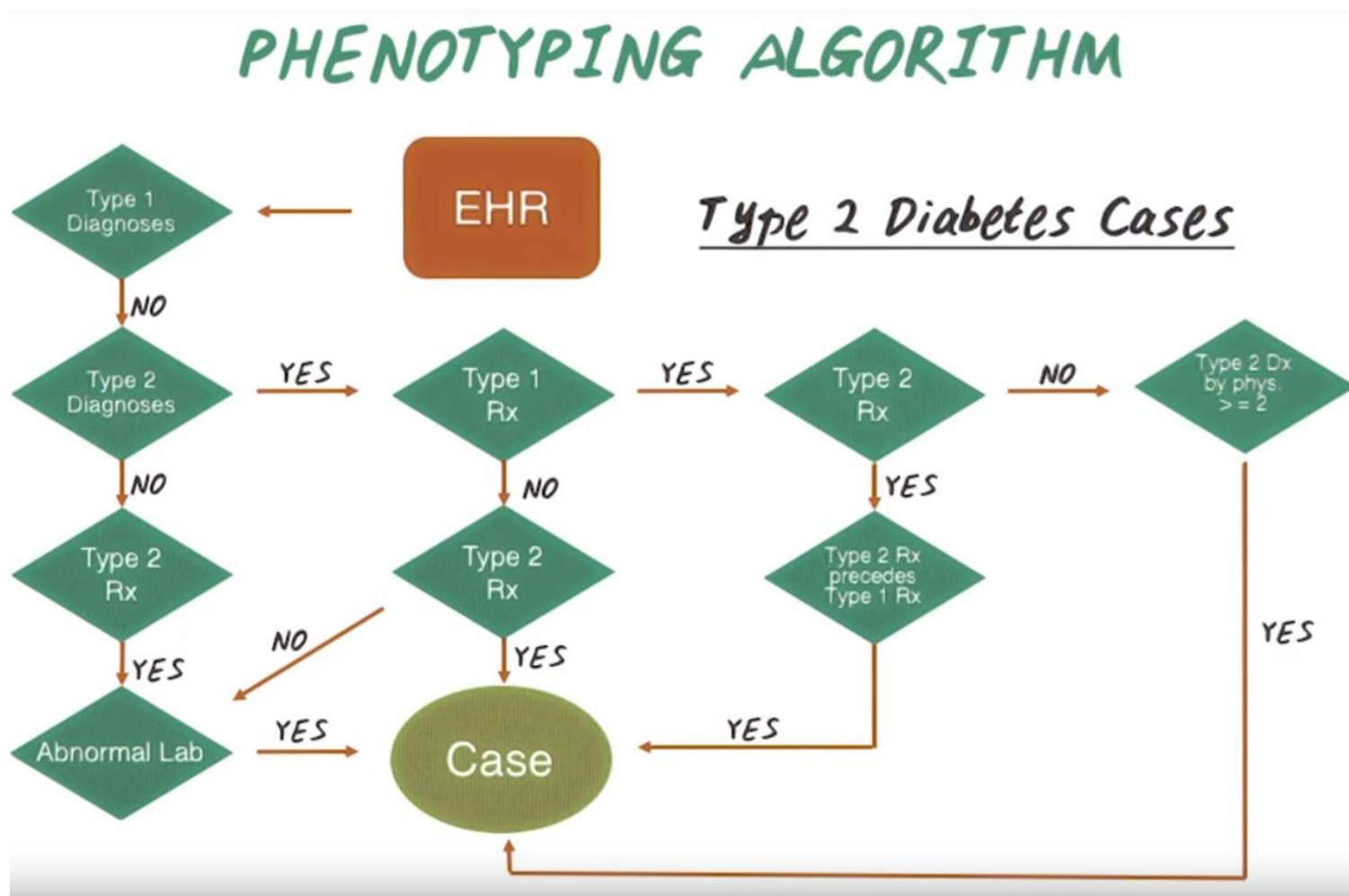


Lesson 2: Big Data Course Overview

PHENOTYPING ALGORITHM



Lesson 2: Big Data Course Overview



이렇게 복잡한 알고리즘이 필요한 이유
⇒ EHR 데이터가 매우 실뢰할 수 없기 때문
⇒ Missing Data, 중복 Data

빅데이터 시스템을 사용해서 이와 같은 알고리즘을 효율적 개발하는 것을 배울 것임.

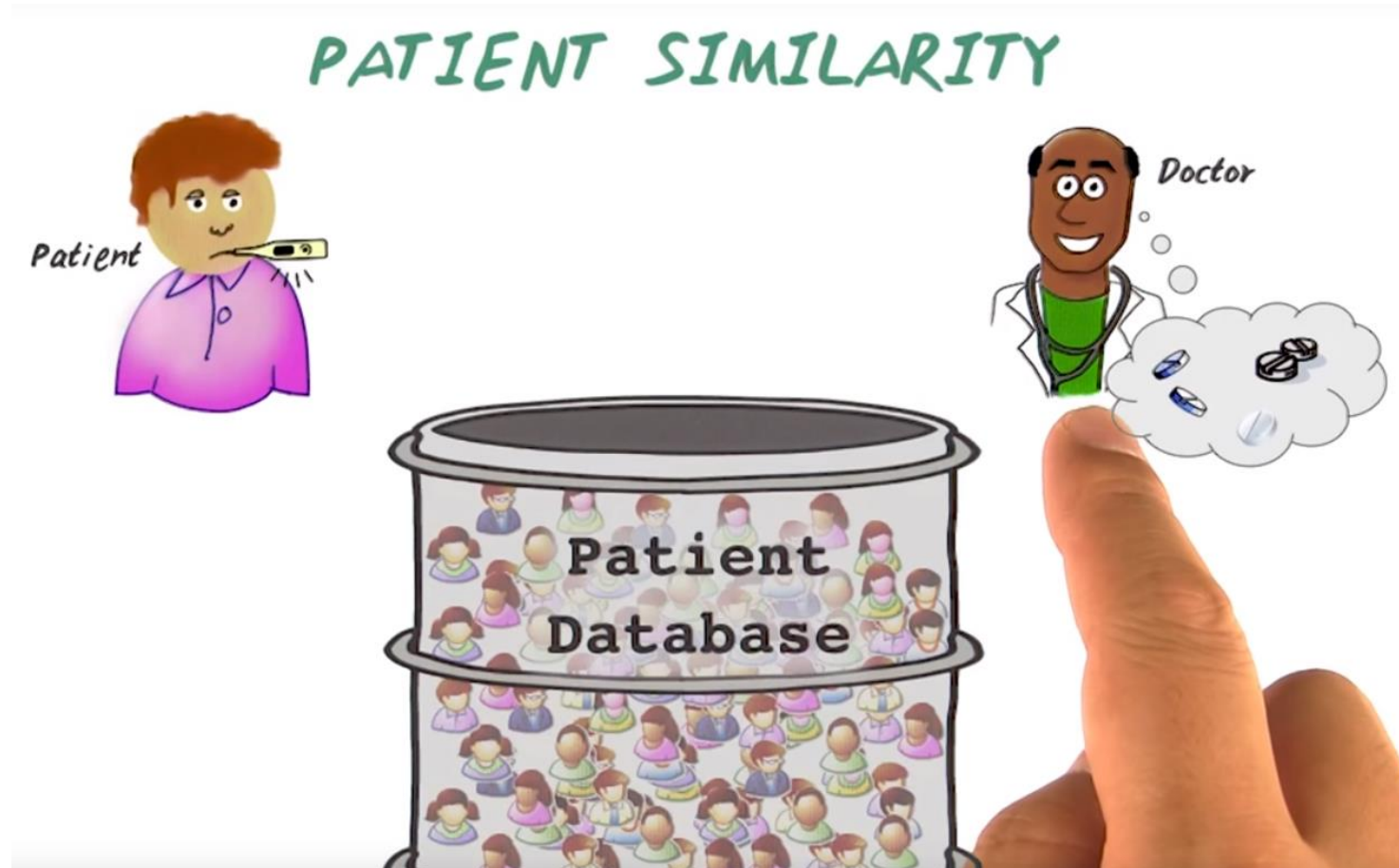
Lesson 2: Big Data Course Overview

PATIENT SIMILARITY QUIZ

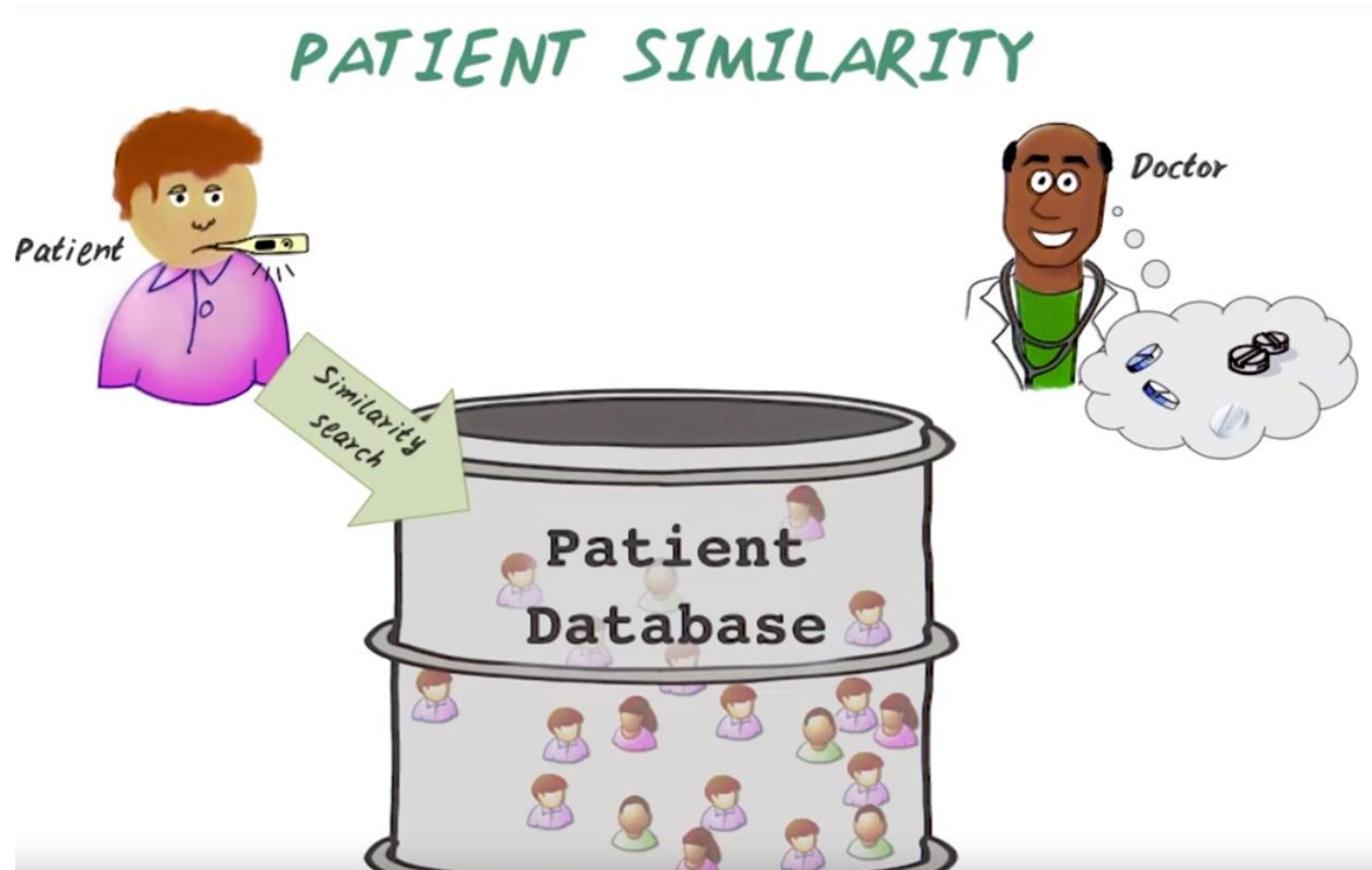
Which of the following types of reasoning do doctors engage most often?

- ☐ *Flowchart reasoning*
- ☐ *Instinct and Intuition*
- ☐ *Comparison to past individual patients*

Lesson 2: Big Data Course Overview

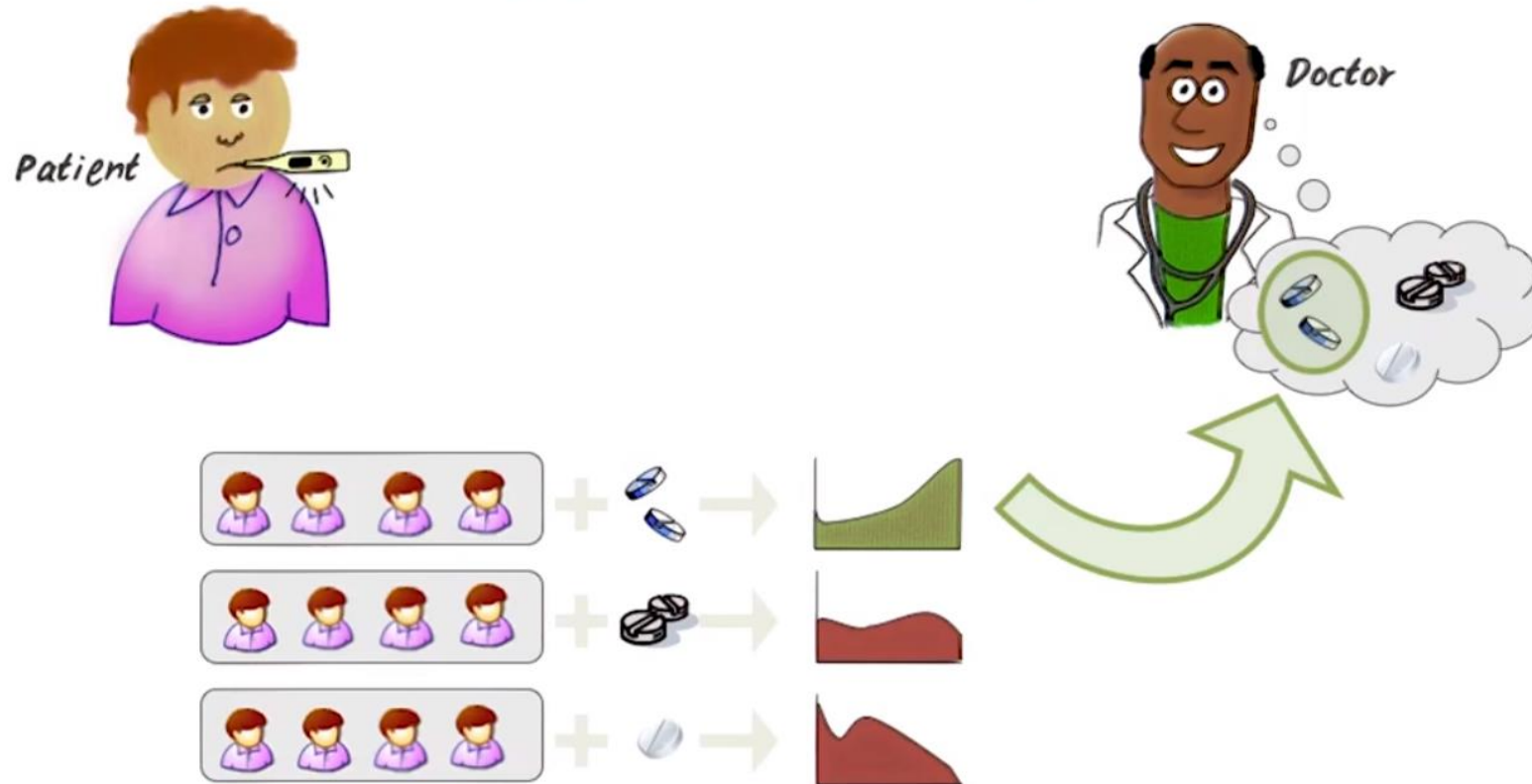


Lesson 2: Big Data Course Overview



Lesson 2: Big Data Course Overview

PATIENT SIMILARITY



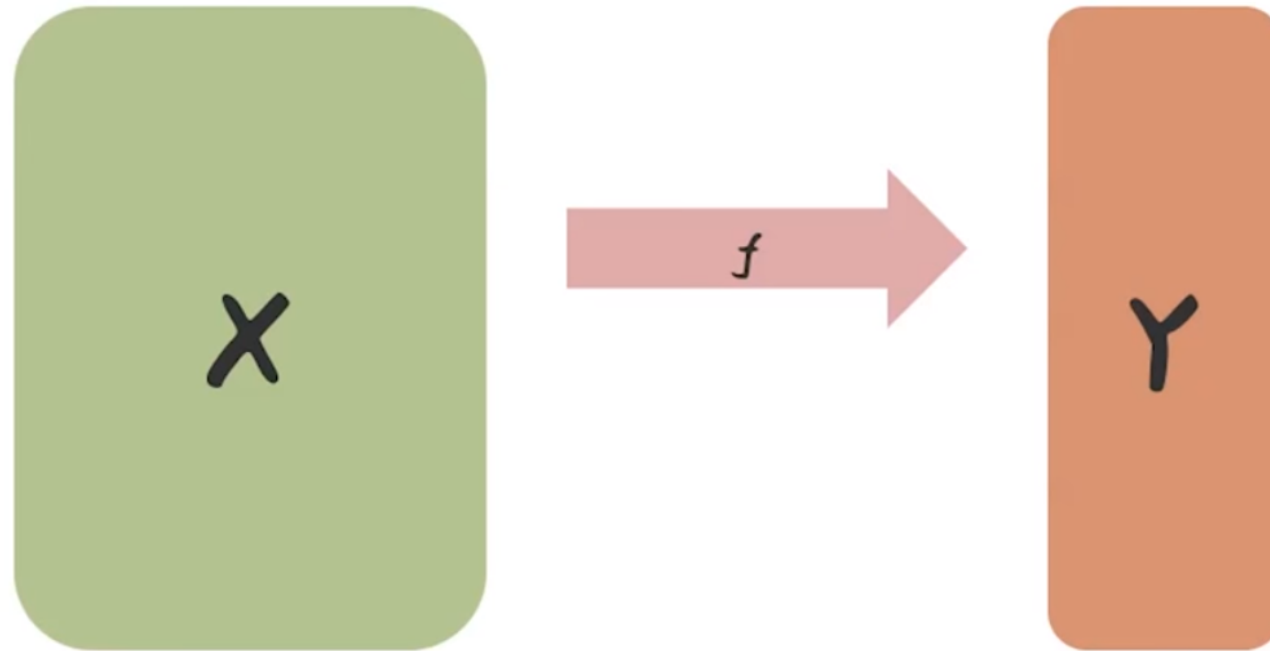
Lesson 2: Big Data Course Overview

BIG DATA ALGORITHMS

- *Classification*
- *Clustering*
- *Dimensionality reduction*
- *Graph analysis*

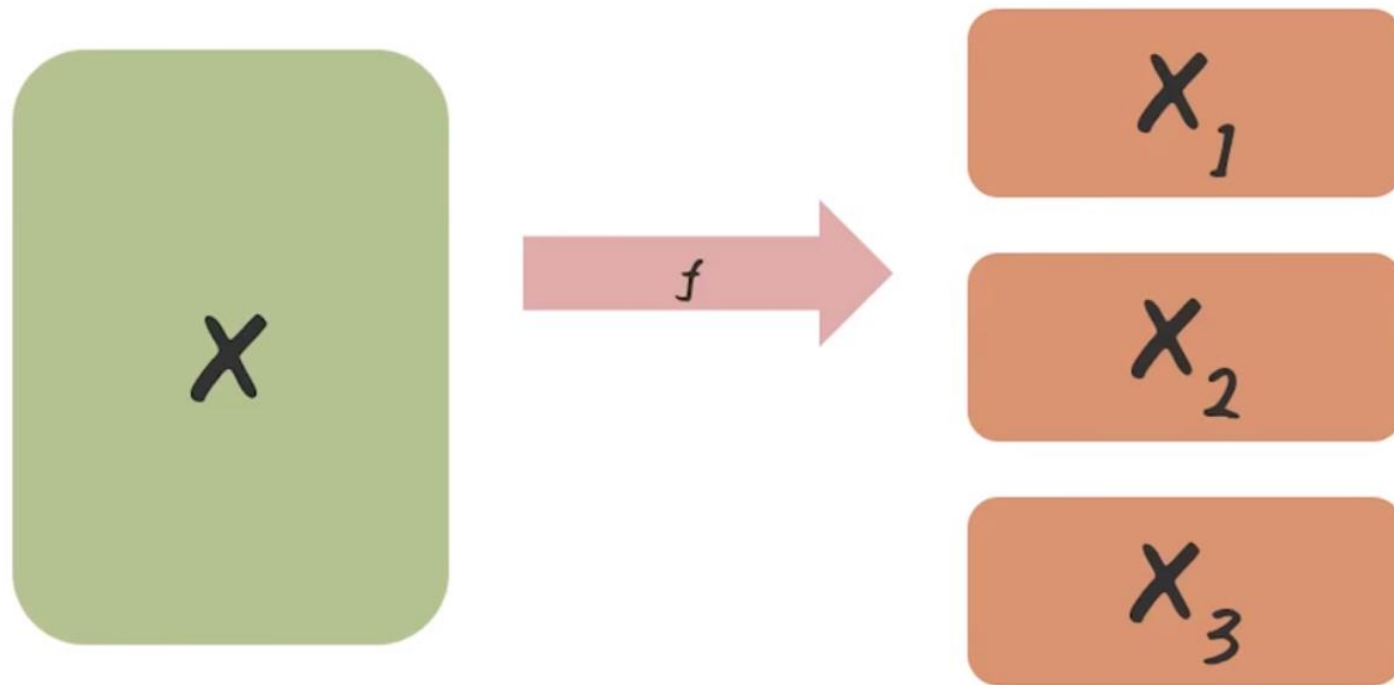
Lesson 2: Big Data Course Overview

CLASSIFICATION



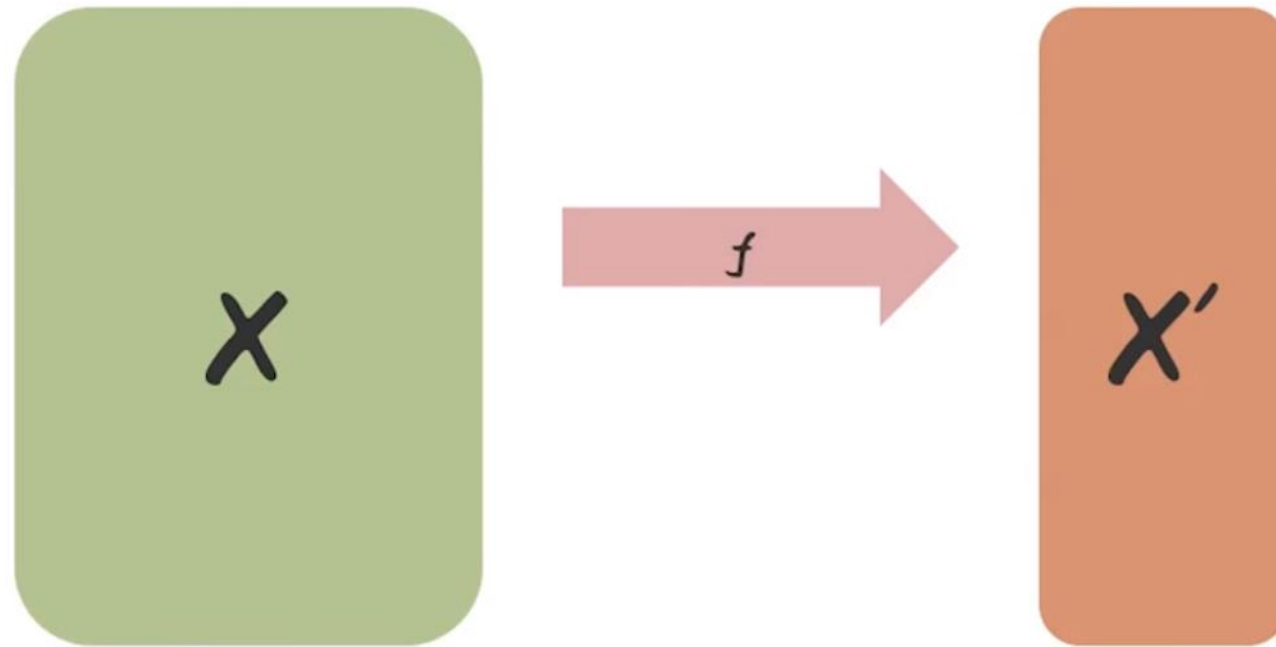
Lesson 2: Big Data Course Overview

CLUSTERING



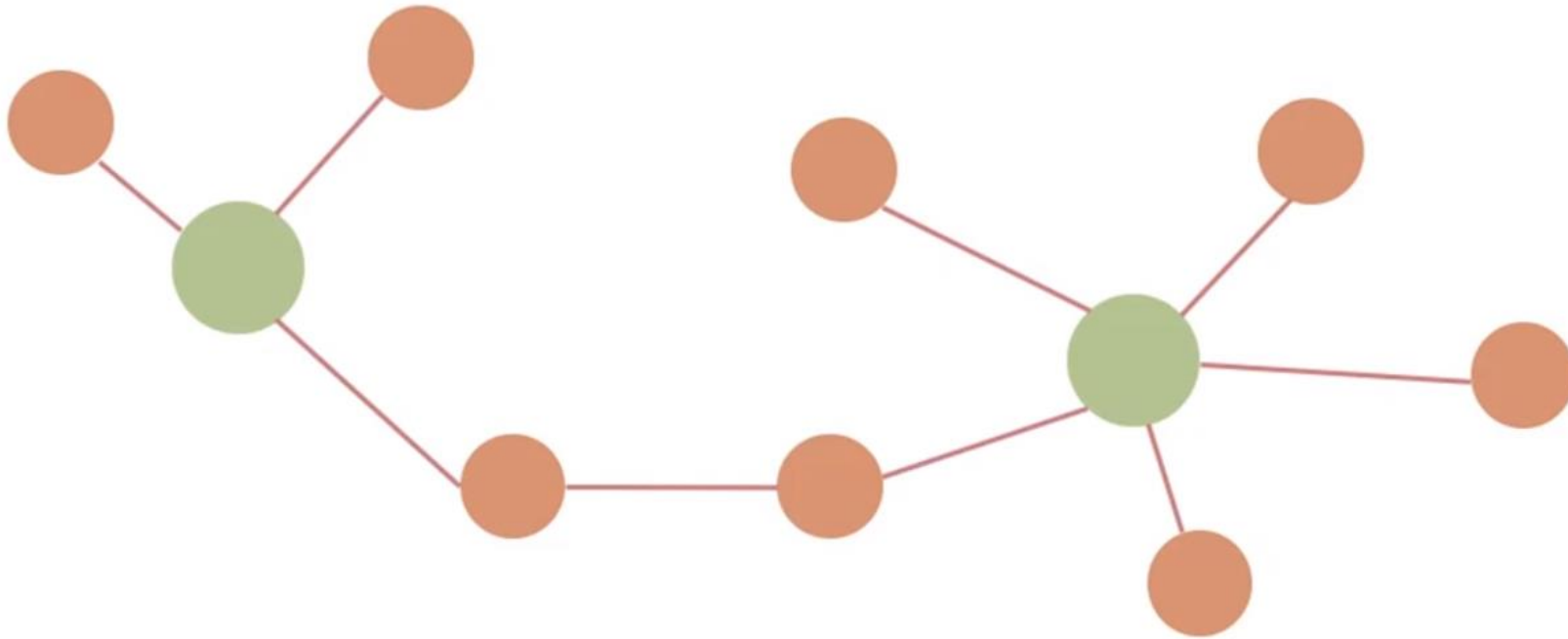
Lesson 2: Big Data Course Overview

DIMENSIONALITY REDUCTION



Lesson 2: Big Data Course Overview

GRAPH ANALYSIS



Lesson 2: Big Data Course Overview

BIG DATA SYSTEMS

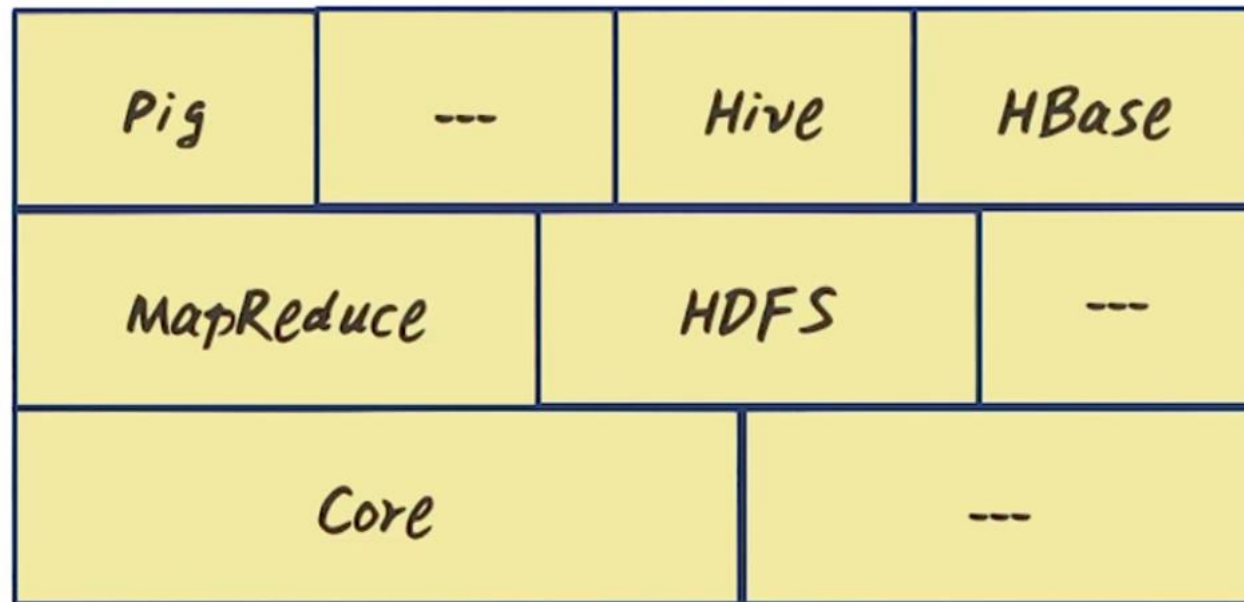


*Distributed disk-based
big data system*

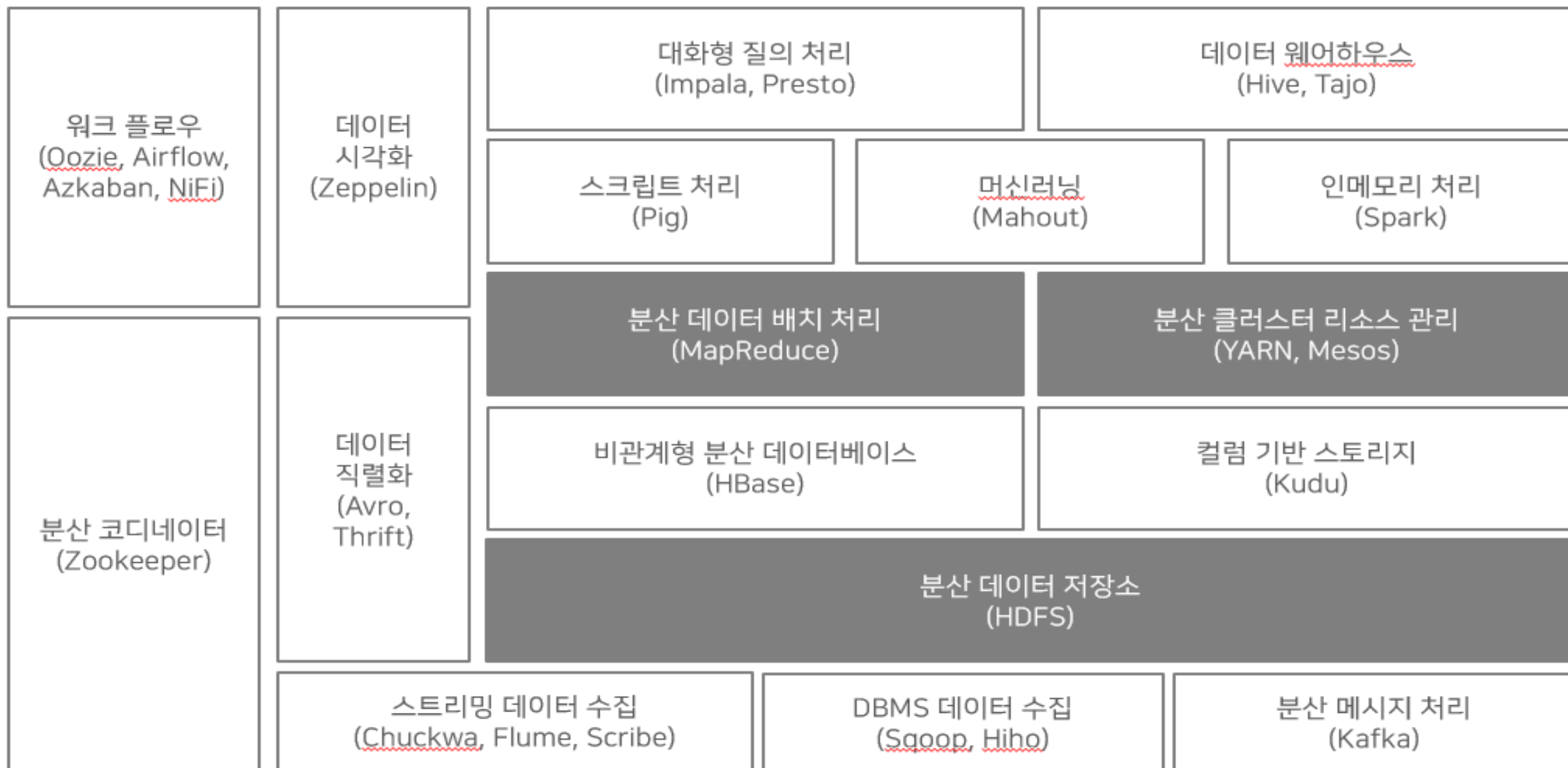


*Distributed in-memory
big data system*

Lesson 2: Big Data Course Overview



Hadoop 2.0과 Ecosystem



Hadoop Ecosystem

데이터 시각화

SAS Visual Analytics

Tableau

Qlik

SAP Lumina

R

D3.JS

iCHARTs

TimeLine JS

Apache Zeppelin

시스템 배포

Apache ambari

Cloudera Manager

Apache Mesos

Marathon

Apache Bigtop

Apache Eagle

Buildoop

Brooklyn

Apache Helix

데이터 수집

Apache Flume

Apache Sqoop

Facebook Scribe

Apache Kafa

Netflix Suro

Apache Samza

Cloudera Morphline

Apache NiFi

HIHO

서비스프로그래밍

Apache Thrift

Apache Zookeeper

Apache Avro

Apache Karaf

Elephant Bird

Norbert

스케줄링/복제

Apache Oozie

Linkedin Azkaban

Apache Falcon

보안

Apache Sentry

Apache Ranger

프레임워크

SpingXD

Jumbune

메타데이터

Metascope

Apache Tika

머신러닝

Apache Mahout

WEKA

Spark MLLIB

Deeplearni ng4j

MADLib

H2O

Sparkling Water

Apache SystemML

분산 프로그래밍

Apache Pig

Apache MapReduce

Apache Spark

Apache Storm

Apache Flink

Apache Apex

Facebook Corona

Apache REEF

Apache Twill

Apache Hama

Apache Tez

Apache DataFu

Kangaroo

Apache Beam

JQAL

TinkerPop

SQL on Hadoop

Apache Hive

Cloudera Impala

Apache HCatalog

Apache HAWQ

Facebook Presto

Kylin

Apache Tajo

Apache Phoenix

NoSQL Databases

Wide Column

Apache HBase

Apache Cassandra

Apache Parquet

Apache Kudu

Document

MongoDB

CouchDB

DunamoDB

Gemfire

Key-Value

Redis

OpenTSDB

RocksDB

Linkedin Voldemort

Graph

Giraph

Neo4j

TitanDB

OrientDB

NewSQL Databases

TokuDB

SAP HANA

VoltDB

InfluxDB

Haeinsa

Drizzle

분산 파일 시스템

Apache HDFS

Red Hat GlusterFS

Lustre File System

GridGain

XtreemFS

Quantcast File System

Ceph File System

Brooklyn

Cloudera Enterprise

데이터 시각화	SAS Visual Analytics	Tableau	Qlik	SAP Lumina	R	D3.JS	iCHARTs	TimeLine JS	Apache Zeppelin
시스템 배포	Apache ambari	Cloudera Manager	Apache Mesos	Marathon	Apache Biotop	Apache Eagle	Buildoop	Brooklyn	Apache Helix
데이터 수집	Apache Flume	Apache Sqoop	Facebook Scribe	Apache Kafa	Netflix Suro	Apache Samza	Cloudera Morphline	Apache NiFi	HIHO
서비스프로그래밍	Apache Thrift	Apache Karaf	Apache Zookeeper	Elephant Bird	Apache Avro	Norbert			
스케줄링/복제	Apache Oozie	Linkedin Azkaban	Apache Falcon						
보안	Apache Sentry	Apache Ranger							
프레임워크	SpringXD	Jumbune							
메타데이터	Metascope	Apache Tika							
머신러닝	Apache Mahout	WEKA	Spark MLLIB	Deeplearni ng4j	MADLib	H2O	Sparkling Water	Apache SystemML	
분산 프로그래밍									
Apache Pig	Apache Flink	Apache Twill	Kangaroo	Apache MapReduce	Apache Apex	Apache Hama	Apache Beam	Apache Spark	Facebook Corona
Apache Storm	Apache REEF	Apache DataFu	TinkerPop						
SQL on Hadoop									
Apache Hive	Facebook Presto	Cloudera Impala	Kylin	Apache HCatalog	Apache Tajo	Apache HAWQ	Apache Phoenix		
NoSQL Databases									
Wide Column	Document	Key-Value	Graph						
Apache HBase	MongoDB	Redis	Giraph						
Apache Cassandra	CouchDB	OpenTSDB	Neo4j						
Apache Parquet	DunamoDB	RocksDB	TitanDB						
Apache Kudu	Gemfire	Linkedin Voldemort	OrientDB						
NewSQL Databases									
TokuDB	InfluxDB	SAP HANA	Haeinsa	VoltDB	Drizzle				
분산 파일 시스템									
Apache HDFS	Red Hat GlusterFS	Lustre File System	GridGain	XtreemFS	Quantcast File System	Ceph File System	Brooklyn		

Lesson 2: Big Data Course Overview

