

# مباحث پیشرفته در فناوری اطلاعات جلسه چهارم - کلان داده ها

مرتضی سرگلزایی جوان  
مرکز تحقیقات رایانش



# سرفصل مطالب

- بخش اول: مقدمه ای بر کلان داده ها
- بخش دوم: تعاریف پایه کلان داده ها
- بخش سوم: انواع کلان داده ها
- بخش چهارم: معماری مرجع
- بخش پنجم: وب داده ها
- بخش ششم: امنیت در کلان داده ها
- بخش هفتم: گمنام سازی داده ها



# Data is the new oil





amazon

UBER

Microsoft

Google

f

TESLA

# داده ها در ۶۰ ثانیه (در سال ۲۰۱۷ و ۲۰۱۸)



<http://listen.hatnote.com/>



# تبدیل داده به عمل



# موارد کاربرد

شناسایی افراد مظنون

بهینه‌سازی شبکه

بهبود رضایت مشتری

بهبود کمی و کیفی تولیدات

نگهداری پیشگویانه

رصد اطلاعات

پیشگیری از وقوع جرم

تحلیل شبکه های اجتماعی

تشخیص تقلب

پیش‌بینی رفتار مشتری

شناسایی الگوهای وقوع جرم

مدیریت ریسک

ارزیابی هوشمند

شهر هوشمند

شفاف‌سازی تعاملات

تشخیص هویت افراد

مدیریت ترافیک

پرونده‌های قضایی

• حوزه ارتباطات

• حوزه آموزش

• حوزه سلامت

• حوزه انرژی

• خدمات مالی

• امنیت ملی

• هواشناسی

• راهنمایی و رانندگی

• مسائل انتظامی

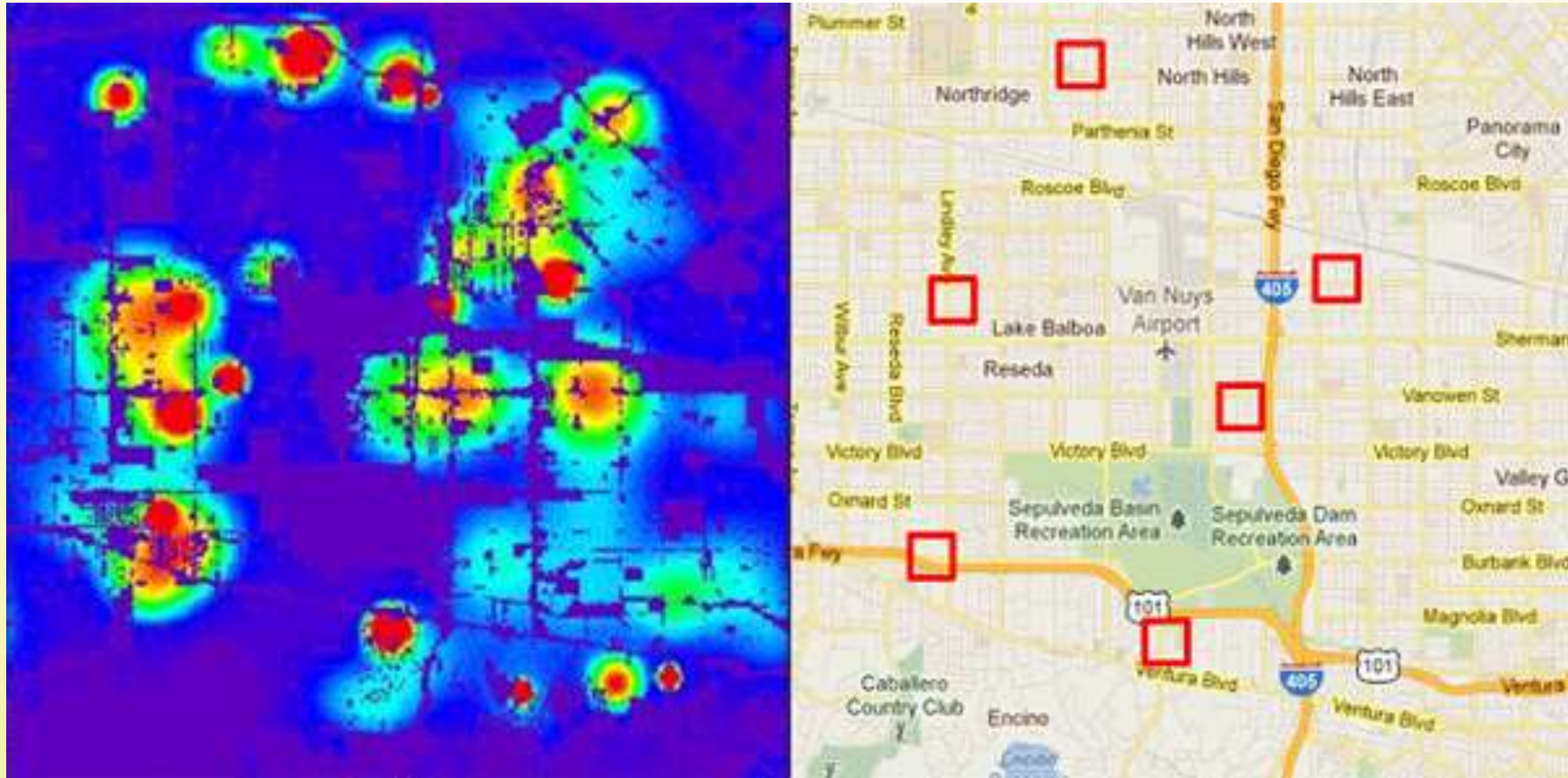
• مسائل زیست‌محیطی

• حوزه کشاورزی

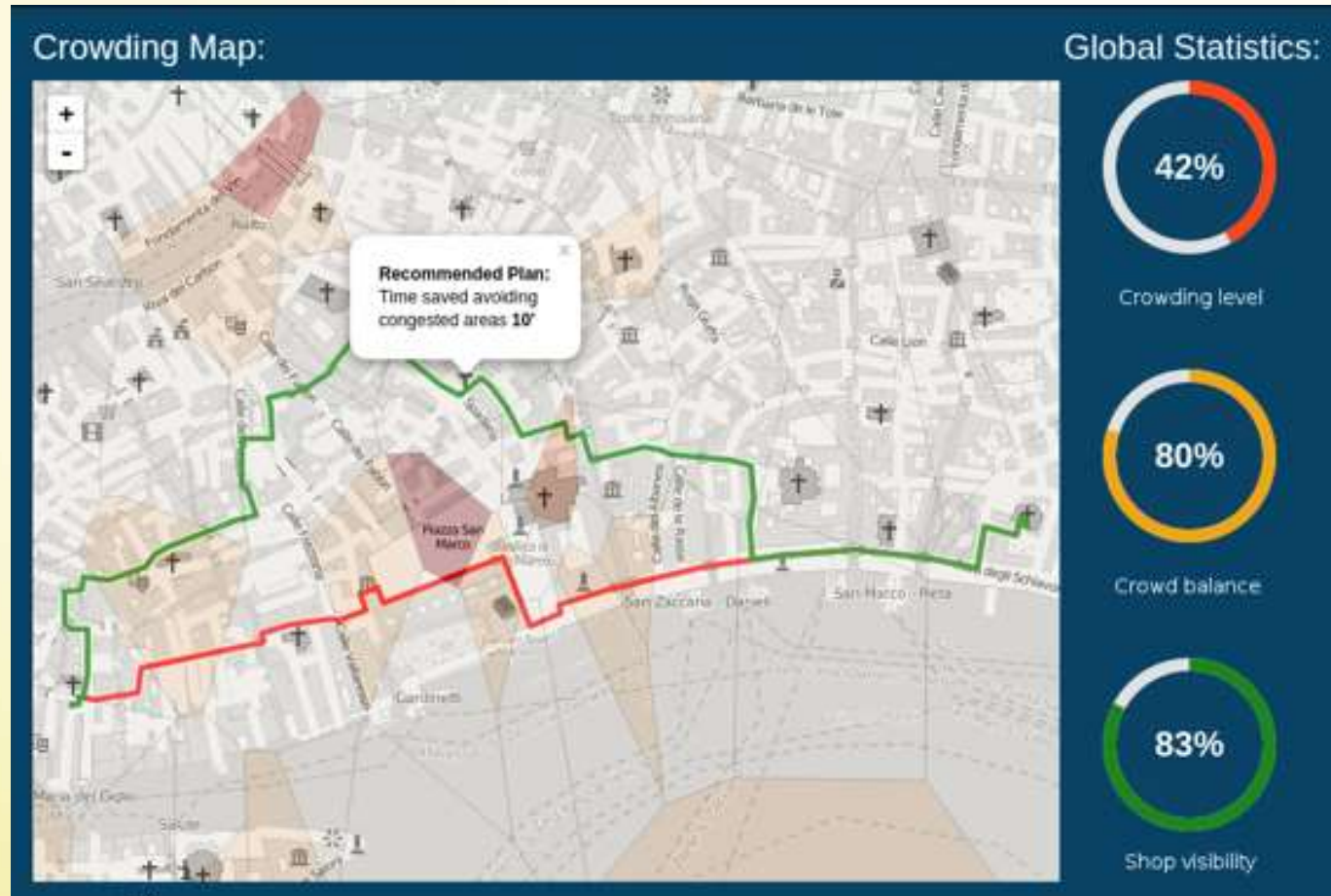




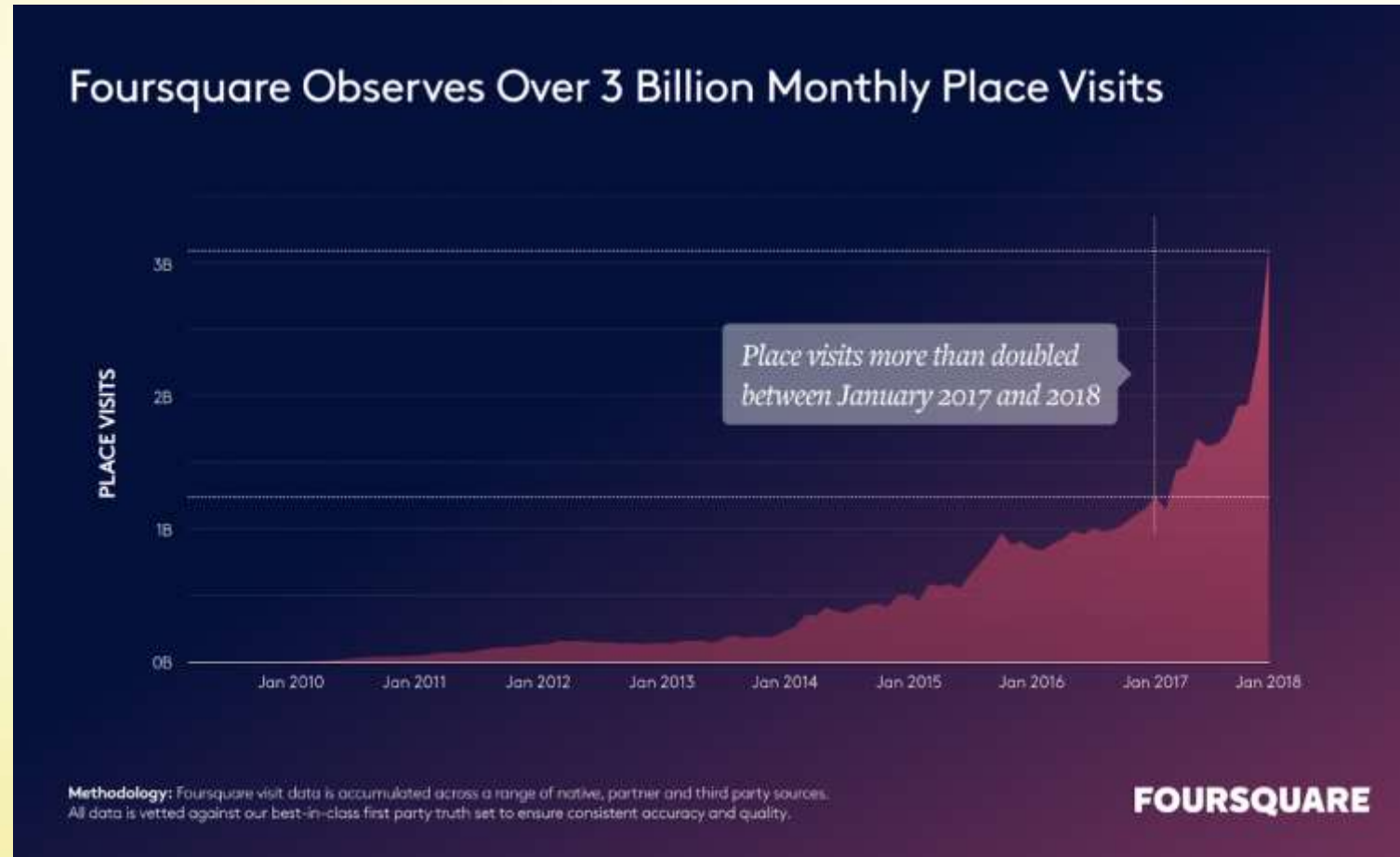
# ارتقای امنیت



# بهبود حمل و نقل



# مدیریت گردشگری



# تحلیل عقاید

## Hillary Clinton ▾

Democratic Party



### Mentions

Total	785
No. of <b>positive</b>	151
No. of <b>negative</b>	155
No. of neutral	479

## Donald Trump ▾

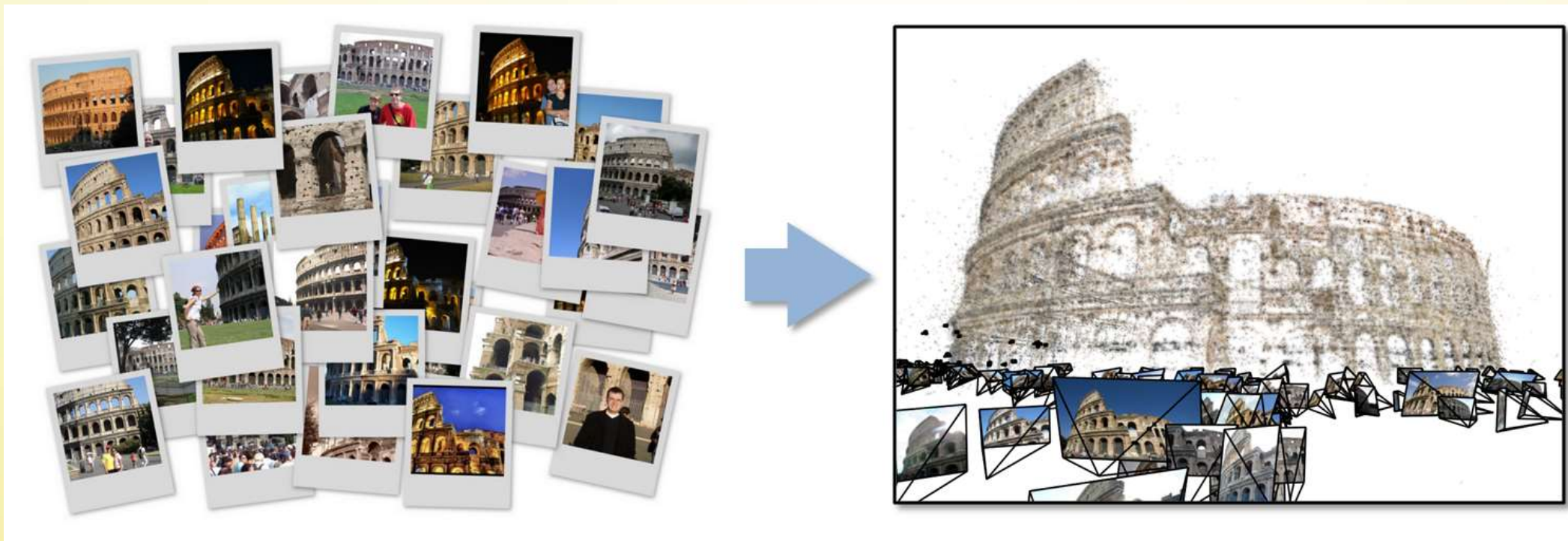
Republican Party



### Mentions

Total	1,412
No. of <b>positive</b>	255
No. of <b>negative</b>	402
No. of neutral	755

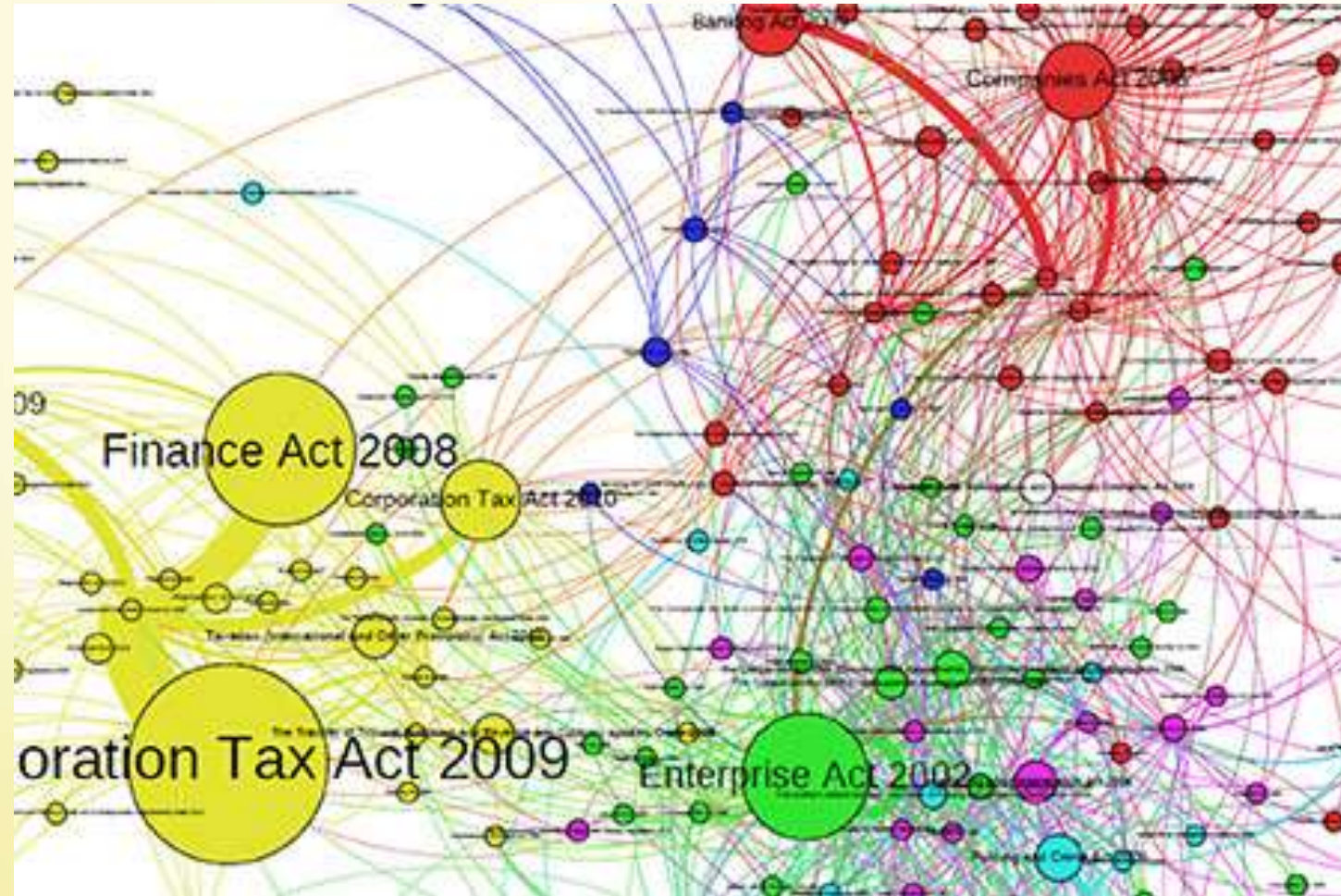
# پردازش تصویر



# دسته بندی مشتریان و تحلیل الگوهای رفتاری



# کلان داده ها در حوزه قانون گذاری

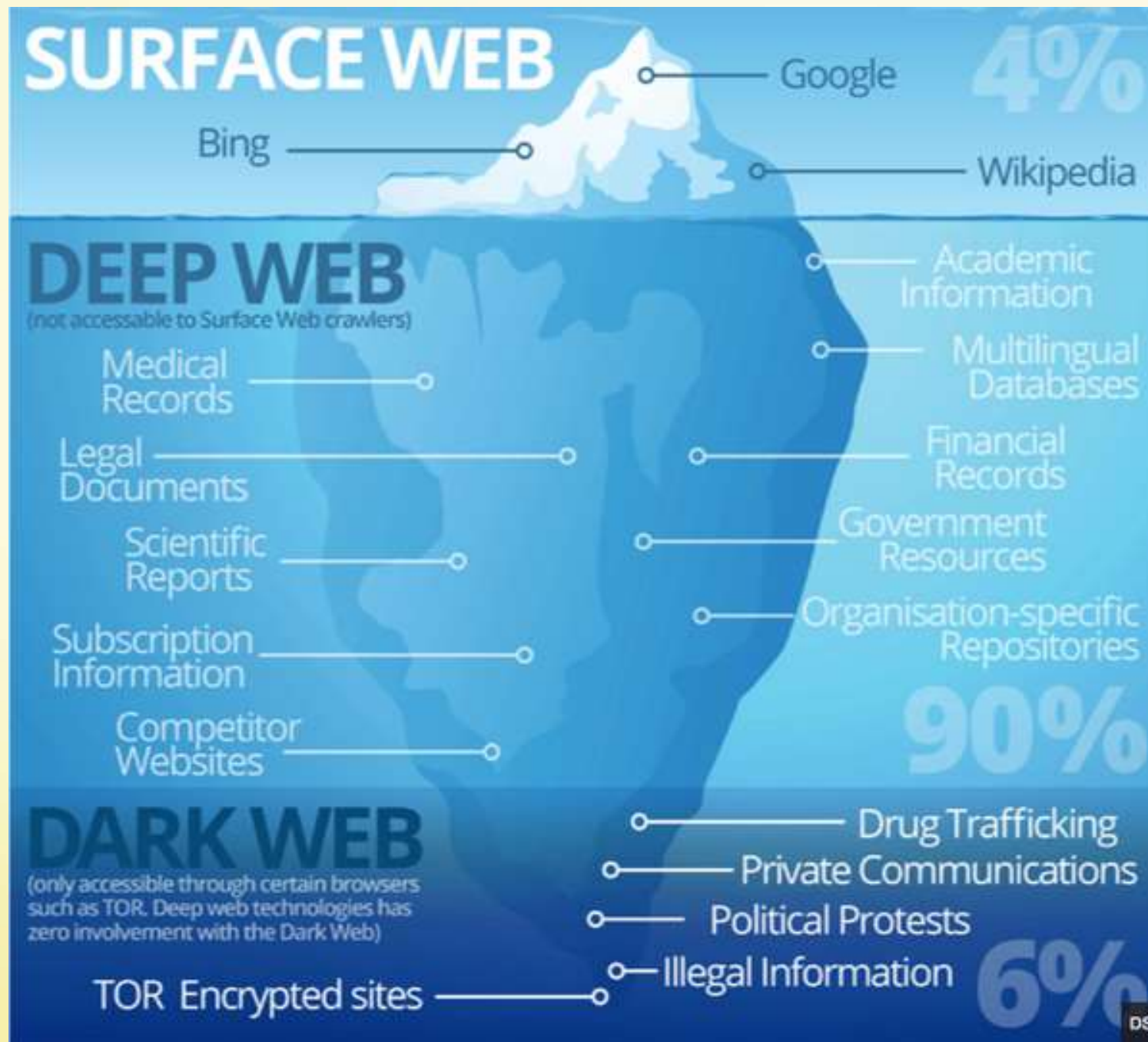


# جستجوی اطلاعات

How can we find something  
in the Web?







# بخش دوم: تعاریف پایه کلان داده ها

مرتضی سرگلزایی جوان  
مرکز تحقیقات رایانش ابری





## داده ها

- داده ها اجزای خام و پردازش نشده ای هستند که از ترکیب قاعده مند آنها اطلاعات حاصل می شود.

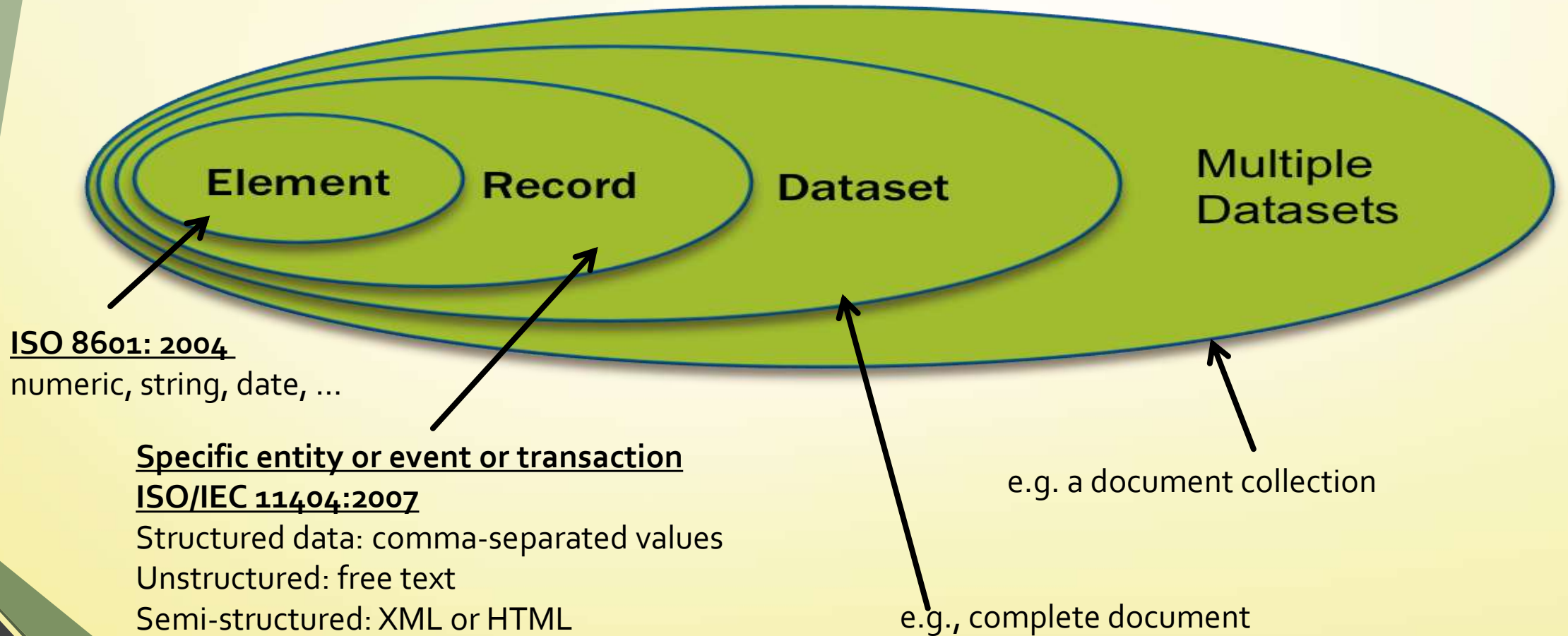
- Turban, Information Technology for Management, Wiley, 2005

- بازنمایی/توصیف اطلاعات به گونه ای که قابلیت انتقال، تفسیر و پردازش داشته باشد.

- ISO/IEC 2382:2015- Information technology Vocabulary



# سلسله مراتب داده ها



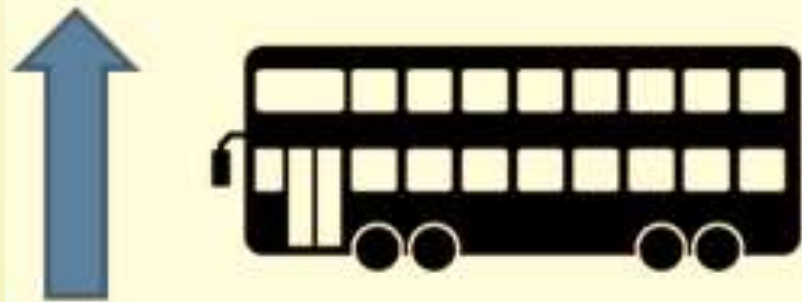
## داده خرد >> داده کلان

- **Small Data**: مجموعه داده هایی هستند که یک تحلیلگر میتواند بطور کامل آنها را بفهمد و ارزیابی کند.

- **Big Data**: مجموعه داده های گسترده از نظر حجم، تنوع، سرعت تولید و یا تغییر پذیری که نیازمند یک معماری مقیاس پذیر برای پردازش باشند.



# مقیاس پذیری



Scale Up- Vertical Scaling

مقیاس پذیری عمودی

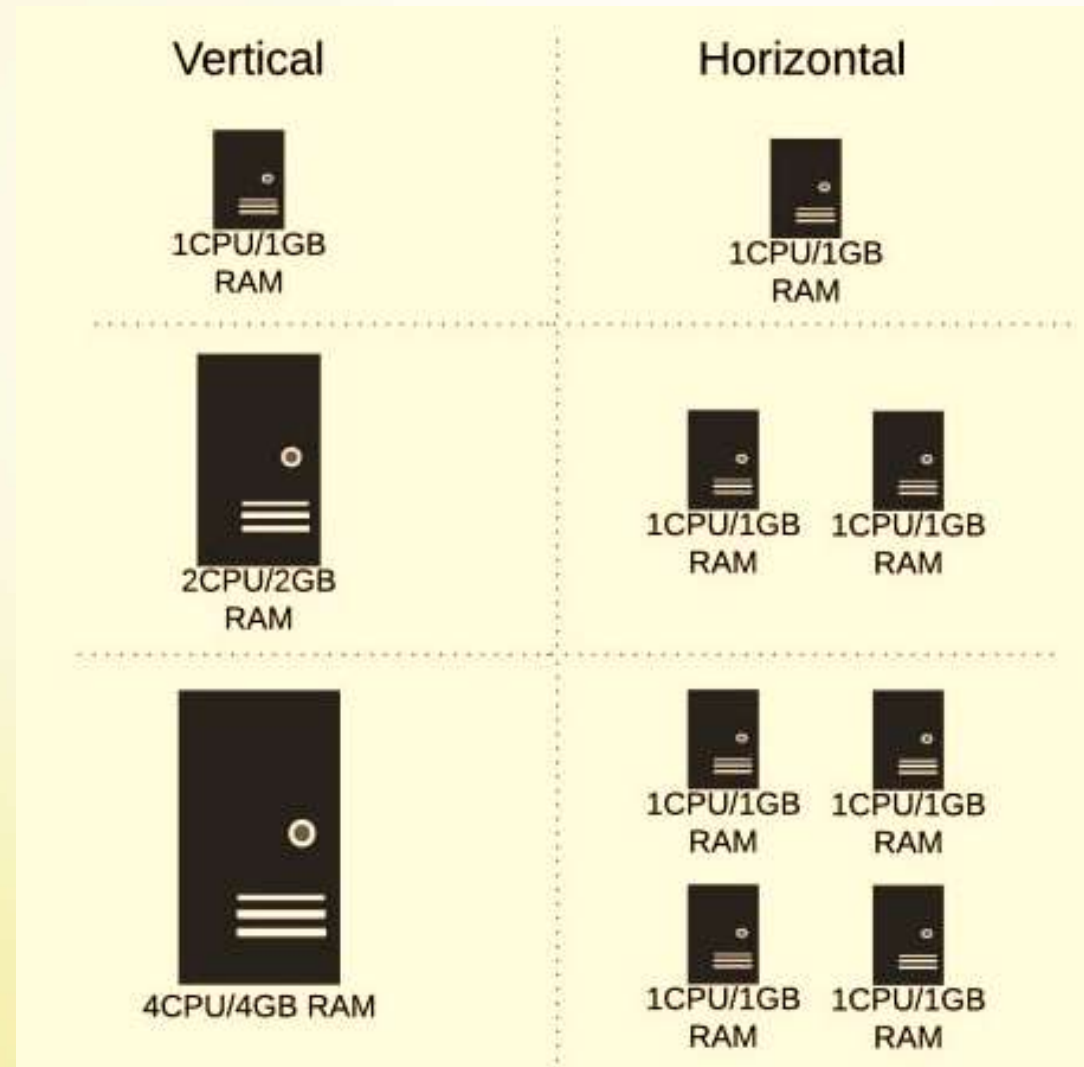


Scale Out- Horizontal Scaling

مقیاس پذیری افقی



# مقیاس پذیری



# مقیاس پذیری افقی







Country: Iran  
Year: 2011  
Cores: 4600  
AMD Opteron 2.3 GHz  
9 TB RAM  
42 TFlop/s



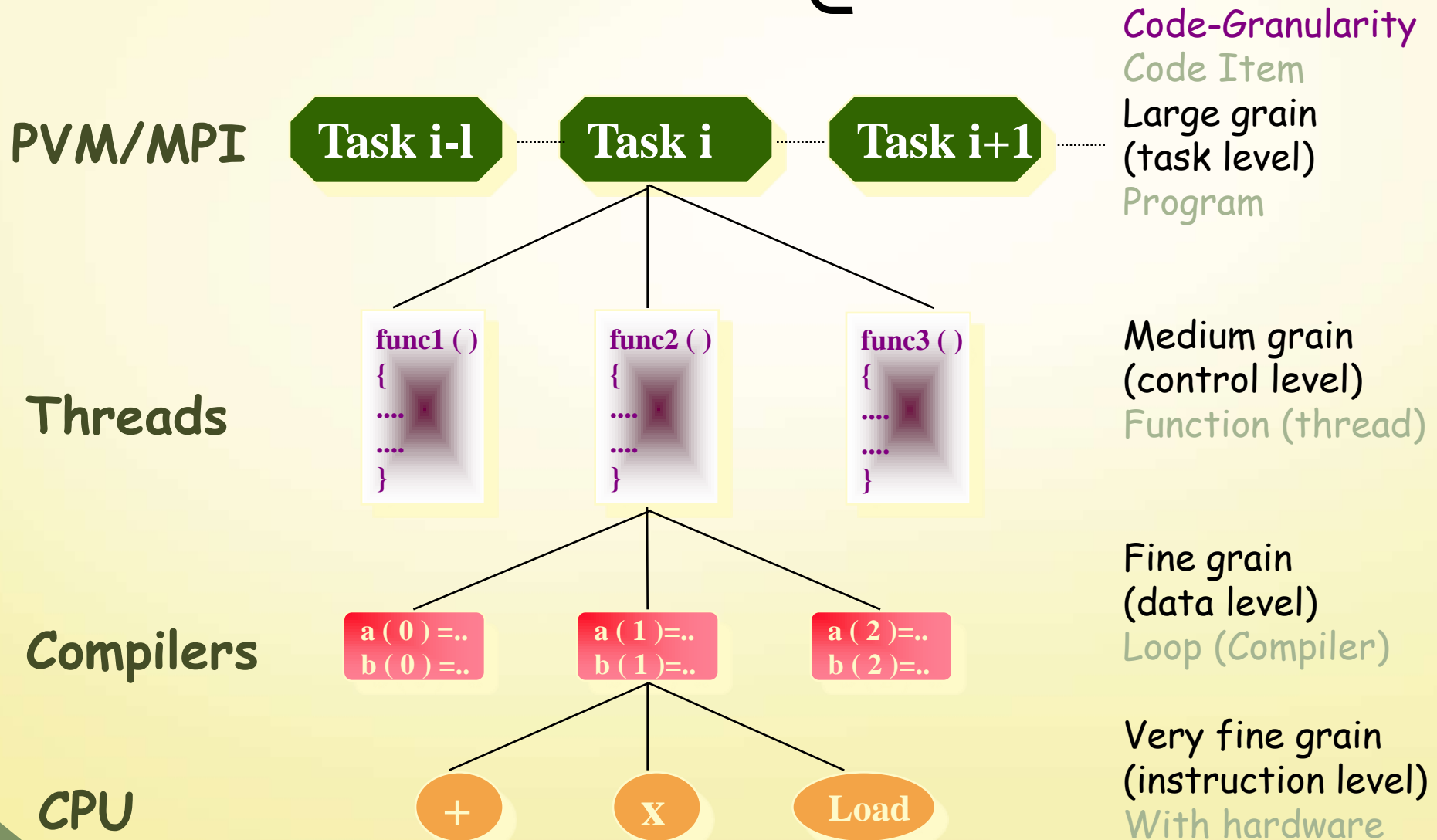
Country: USA  
Year: 2001  
Nodes: 708  
Processors: 1416  
Pentium III  
364 GB RAM  
1416 GFlop/s



Country: China  
Year: 2015  
Cores: 3,120,000  
Xeon E5-2692v2  
1.0 PB RAM  
34 PFlop/s



# سطوح موازی سازی



# انواع مدل های دسترسی به حافظه

- Shared-memory model

The communication between processes in a **multiprocessor environment** is achieved through shared (or global) memory.

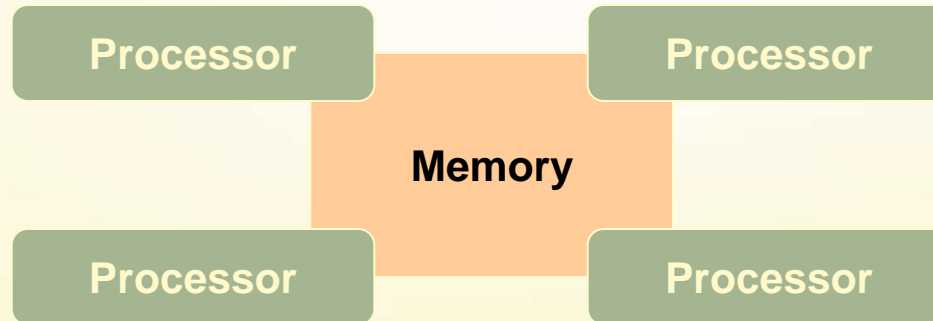


- Message-passing model

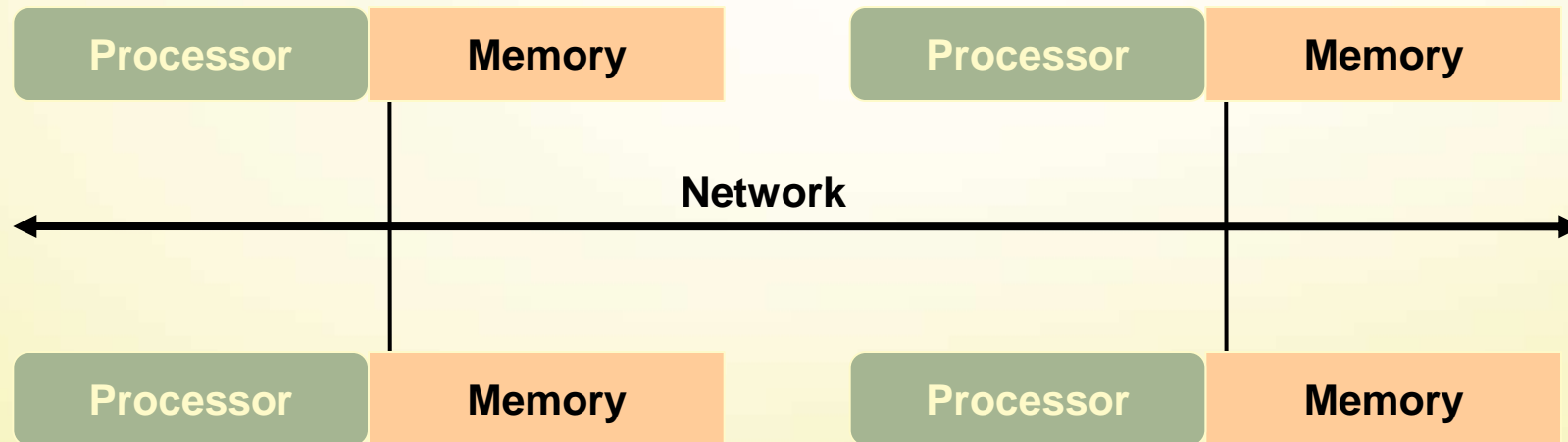
Communication between processes in a **multicomputer environment** is achieved through some kind of message-switching mechanism.



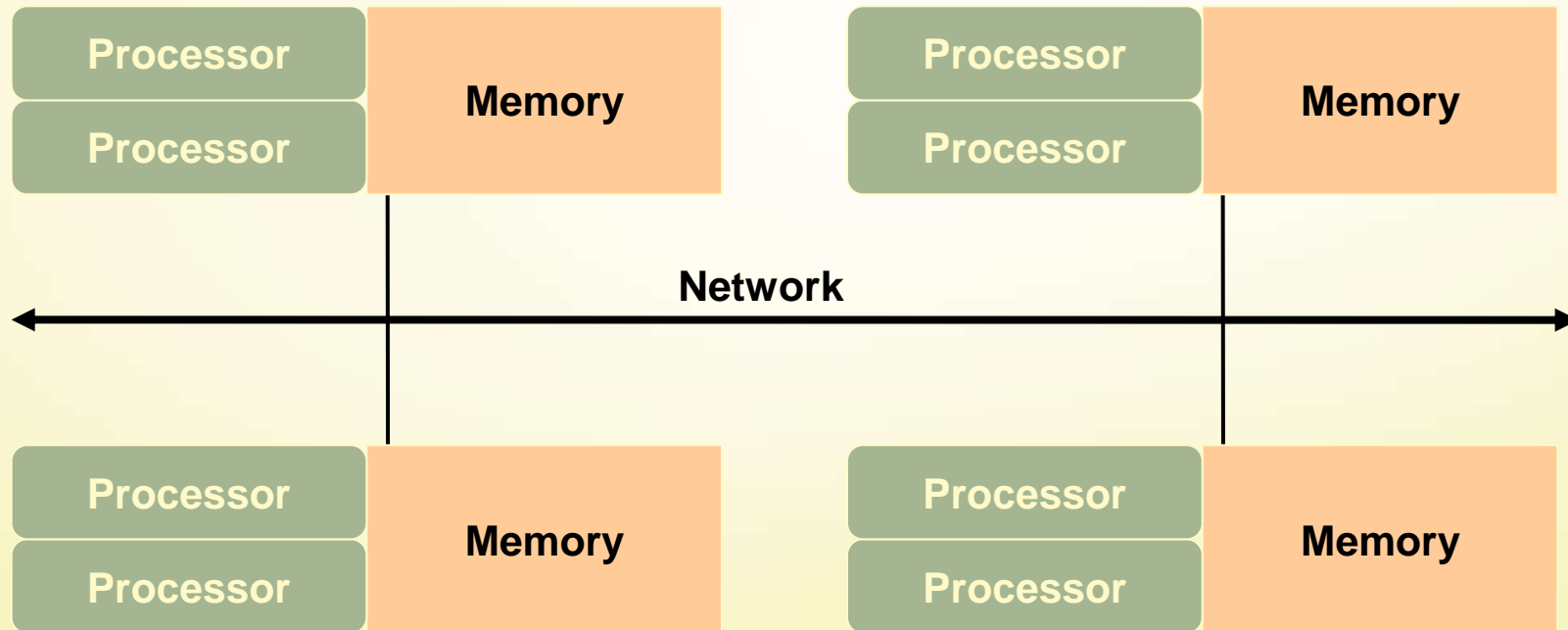
# Memory Typology: Shared



# Memory Typology: Distributed



# Memory Typology: Hybrid





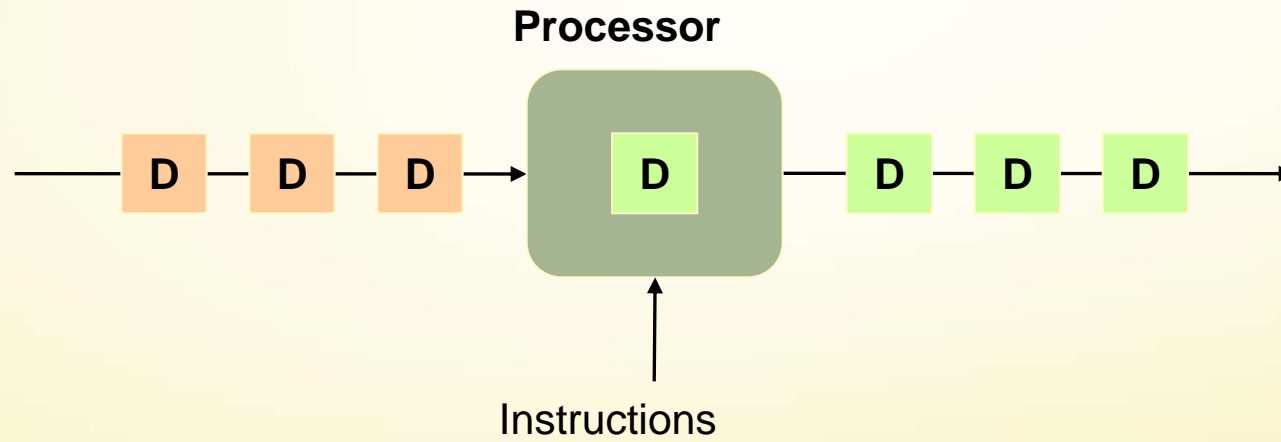
# معماری های مختلف سامانه های پردازشی

		Instructions	
		Single (SI)	Multiple (MI)
Data	Single (SD)	<b>SISD</b> Single-threaded process	<b>MISD</b> Pipeline architecture
	Multiple (MD)	<b>SIMD</b> Vector Processing and GPUs	<b>MIMD</b> Multi-threaded Programming and <b>Multi-Core Processors</b>

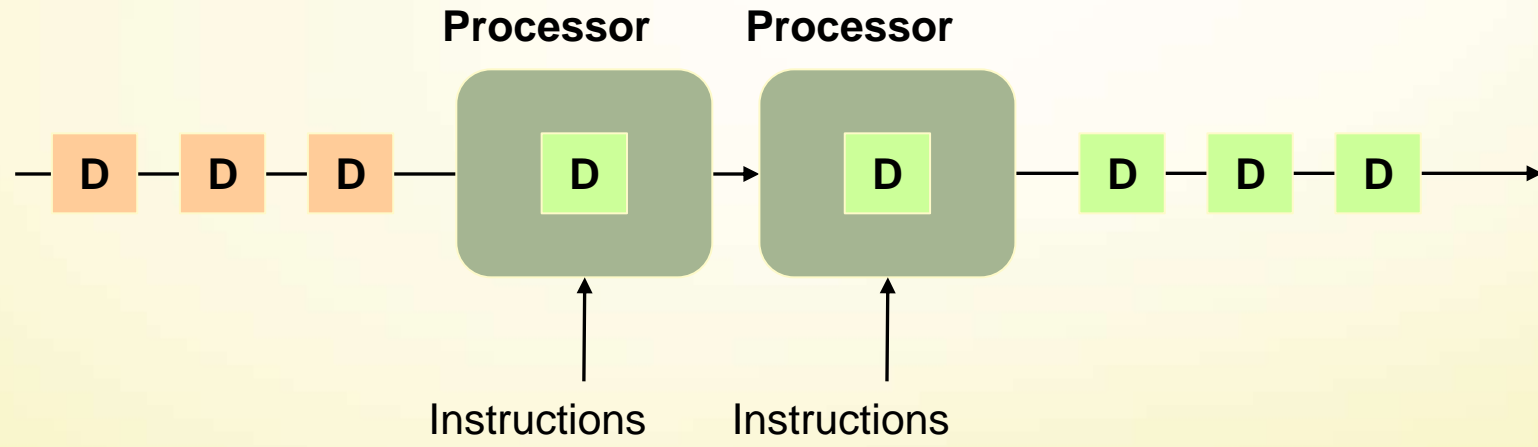




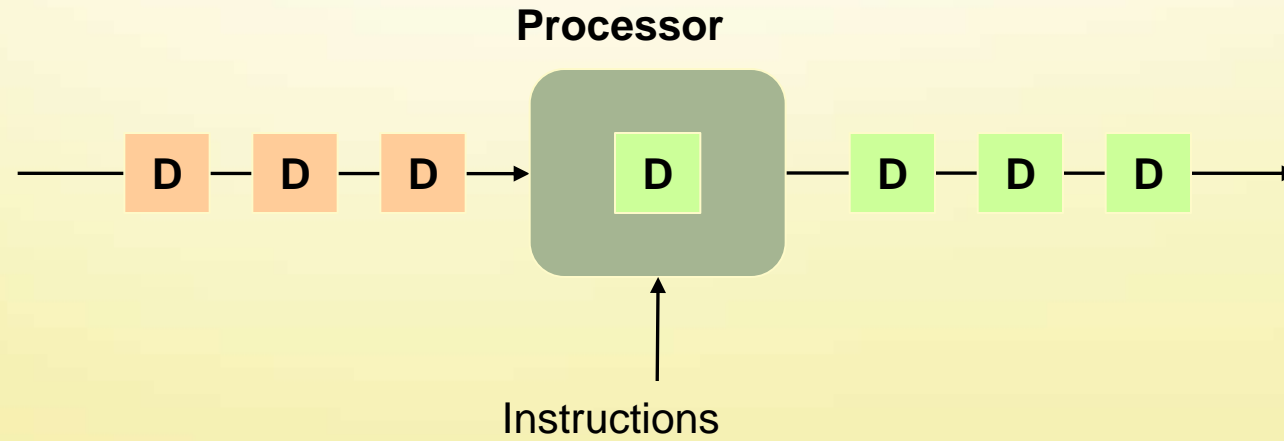
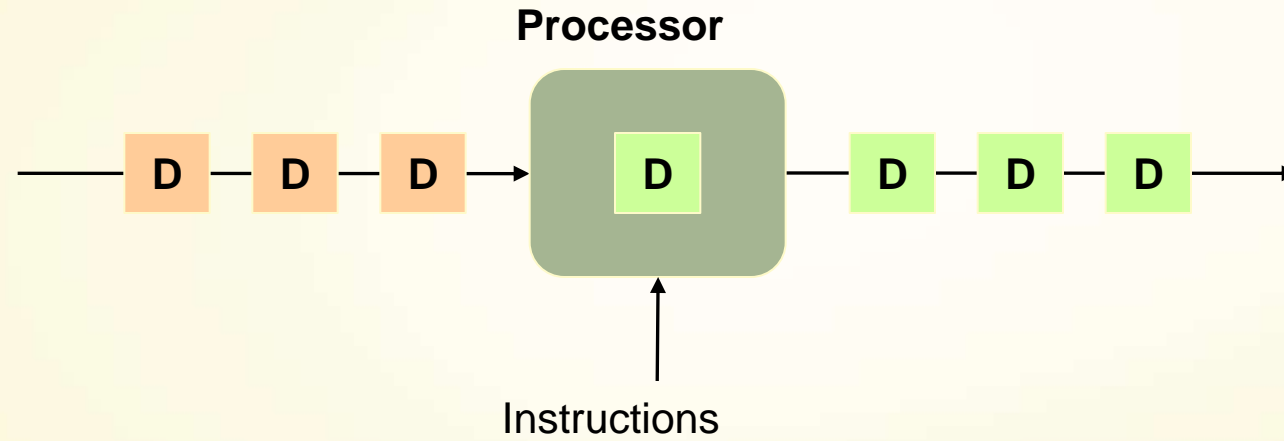
# SISD



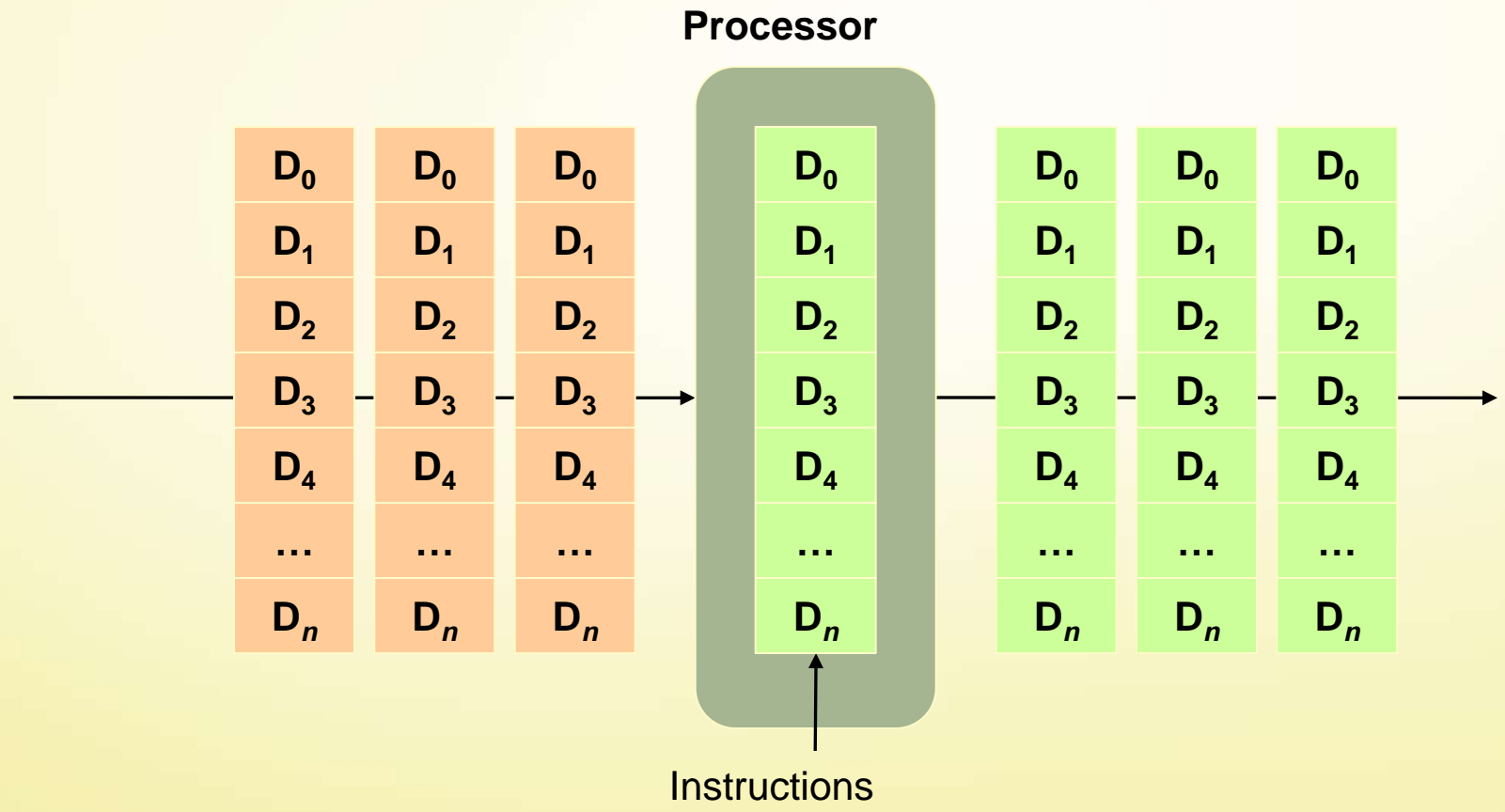
# MISD



# MIMD



# SIMD

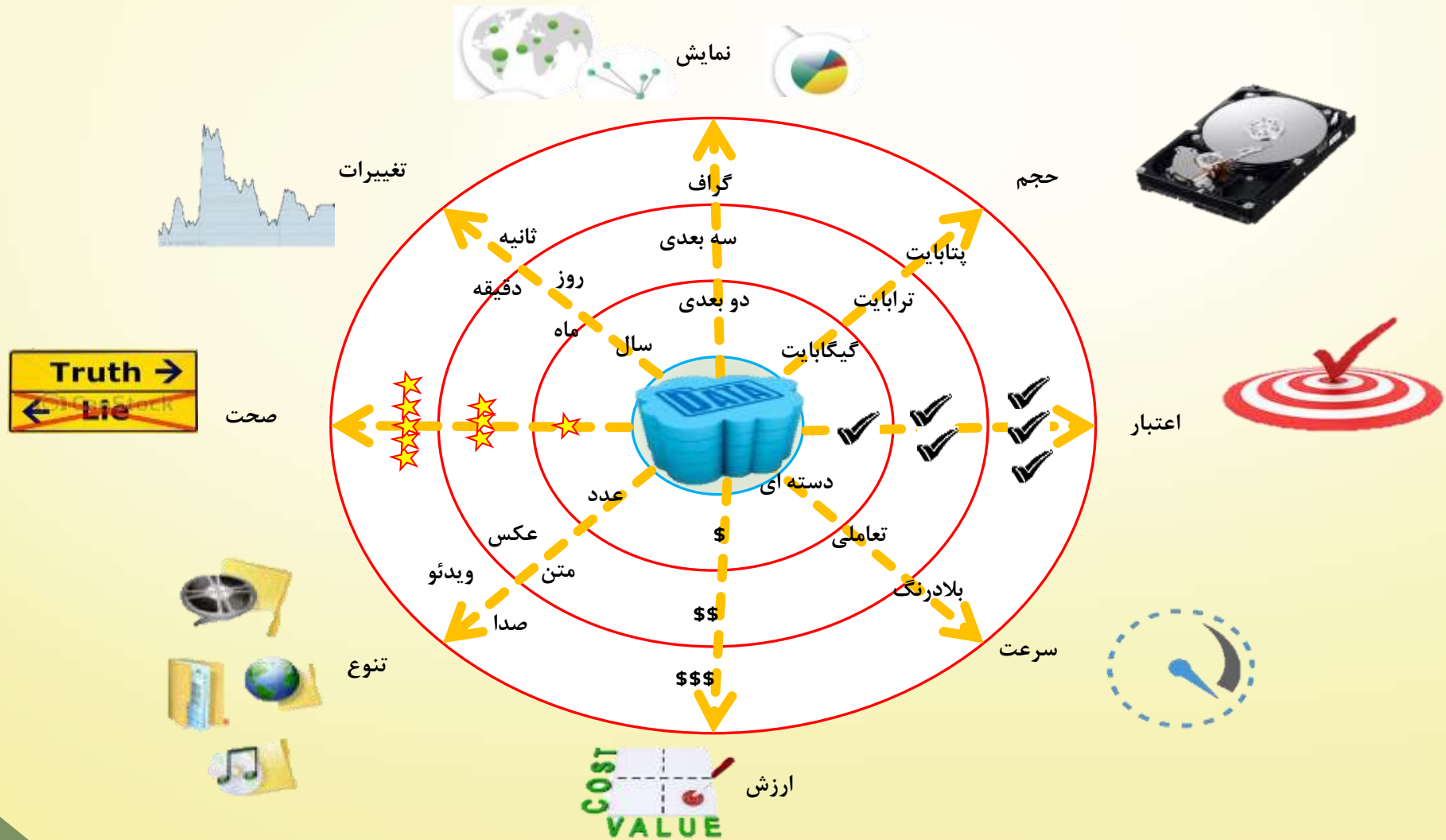


# بخش سوم: مشخصات کلان داده ها

مرتضی سرگلزایی جوان  
مرکز تحقیقات رایانش ابری



# مروری بر مشخصه ها



# کلان داده ها در صنایع مختلف

صنعت	حجم داده ها	تنوع داده ها	نرخ تولید داده ها	ارزش داده ها	حجم داده های تاریک
صنایع مالی و بانکی	۴	۲	۵	۵	۲
صنایع تولیدی	۳	۳	۴	۵	۵
خرده فروشی / عمده فروشی	۵	۴	۵	۵	۴
ارتباطات / فناوری اطلاعات / رسانه	۵	۴	۵	۵	۳
خدمات همگانی (آب و برق و گاز)	۵	۴	۵	۴	۳
حمل و نقل	۳	۲	۳	۲	۲
دولت	۳	۳	۳	۳	۵
آموزش	۱	۲	۲	۲	۵
سلامت	۳	۳	۳	۲	۳
بیمه	۳	۳	۳	۳	۱



# حجم

 شرح	عنوان کاربرد
۳۸۰ ترابایت در ۷۵ سال	سرشماری
دهها ترابایت در روز	پردازش تصاویر پزشکی
۱ ترابایت در ماه	سازماندهی اسناد تحقیقاتی
۲ پتابایت تا سال ۲۰۱۳	سرویس ویدئوی Netflix





# سرعت

 شرح	عنوان کاربرد
چندین گیگابایت در کسری از ثانیه	تشخیص تقلب در تراکنش‌های بانکی
حدود ۱ ترابایت ویدئو و ده‌ها گیگابایت تصویر در ساعت	جستجوی وب
۳۰ فریم با حجم تقریبی ۲۰ مگابایت در ثانیه	رهگیری اجسام در ویدئو
چند صد گیگابایت در روز	تجمیع اطلاعات تویتر
نظر دهی افراد در خصوص محصولات و کالاها ۱۰۰۰ نظر در روز	دیجیکالا



# تنوع

 شرح	عنوان کاربرد
طیف وسیعی از داده‌ها از جمله فیلم، عکس، متن ایمیل و ...	تحقیقات بازاریابی
متن، تصویر، اسناد اسکن شده و ...	ذخیره، بازیابی و جستجوی اطلاعات دولتی
داده‌های بدون ساختار (عکسهای رادیولوژی، گزارش های پزشکی و ...)	تحلیل الگوی سلامت
داده‌های ساختار یافته و بدون ساختار (اطلاعات مکانی، متن، فیلم، صوت و ...)	تجمیع اطلاعات تویتر



# تغییرات / نوسان

شرح	عنوان کاربرد
 <p>در وبسایت‌هایی که دائما محتوای خود را به‌روز می‌کنند داده‌ها نوسان بالایی دارند</p>	جستجوی وب
از آنجایی که در این کاربرد معمولا شیء مورد نظر در حال تحرک است لذا داده‌ها نوسان بالایی دارند.	رهگیری اجسام در ویدئو
در صورتی که کاربر دائما در حال تغییر موقعیت باشد داده‌ها نوسان بالایی دارند.	سرویس‌های مبتنی بر مکان
داده‌ها دائما در حال نوسان هستند و می‌توان گفت داده‌های قدیمی اعتبار کمی دارند	تجمیع اطلاعات توییت‌ر



# صحت

 شرح	عنوان کاربرد
امکان آلوده شدن داده‌ها به نویز وجود دارد. همچنین ابزارهای سنجش داده عدم قطعیت دارند. بنابراین می‌توان گفت درجه صحت داده‌ها پایین است	پردازش اطلاعات رادار به منظور کاربردهای GIS
عدم قطعیت در دریافت اطلاعات دولتی و همینطور امکان آلوده شدن فایلها به ویروس صحت داده‌ها را تحت تاثیر قرار می‌دهد	ذخیره، بازیابی و جستجوی اطلاعات دولتی
امکان وجود نویز و همینطور عدم قطعیت در دستگاه های عکسبرداری و سنجش پزشکی موجب افت صحت داده‌ها می‌شود	تحلیل الگوی سلامت
نظر افراد در خصوص کالا یا محصول	دیجی کالا



# اعتبار

 شرح	عنوان کاربرد
نشانه گذاری داده عامل موثری در افزایش اعتبار داده‌هاست.	جستجوی وب
لازم به پاکسازی داده‌های تصویری برای استخراج داده‌های معتبر (مکان شیء رهگیری شده) است.	رهگیری اجسام در ویدئو
داده‌های دریافت شده از تجهیزات سیار اعتبار بالایی دارند	سرویس های مبتنی بر مکان
بسیاری از داده‌ها بی اعتبار هستند.	اطلاعات توییت



# نمایش

 شرح	عنوان کاربرد
حجم بالای داده‌ها موجب چالش در مصورسازی می‌شود.	پردازش اطلاعات رادار به منظور کاربردهای GIS
تنوع و عدم وجود ساختار در داده‌های این نوع کاربرد از چالش‌های اساسی در مصورسازی این نوع داده‌هاست.	تحلیل الگوی سلامت
از آنجایی که این داده‌ها دائما در حال تغییر است و از طرفی حجم داده‌ها بسیار بالاست مصورسازی با چالش روبروست.	اطلاعات تویتر



# ارزش

<b>COST VALUE</b>	شرح	عنوان کاربرد
	داده‌های پزشکی منابع مشخصی دارند. لذا داده‌ها در این کاربرد عموماً ارزش بالایی دارند.	کنترل روند بیماری دیابت
	منبع داده کاملاً مشخص و تایید شده است لذا داده‌های تولید شده در این کاربرد ارزش بالایی دارند.	پردازش اطلاعات رادار به منظور کاربردهای GIS
	ممکن است بسیاری از داده‌ها به کلی فاقد ارزش یا دارای اجزایی فاقد ارزش باشند.	تجمیع اطلاعات توپوگرافیک



# بخش چهارم: معماری مرجع کلان داده ها

مرتضی سرگلزایی جوان  
مرکز تحقیقات رایانش ابری

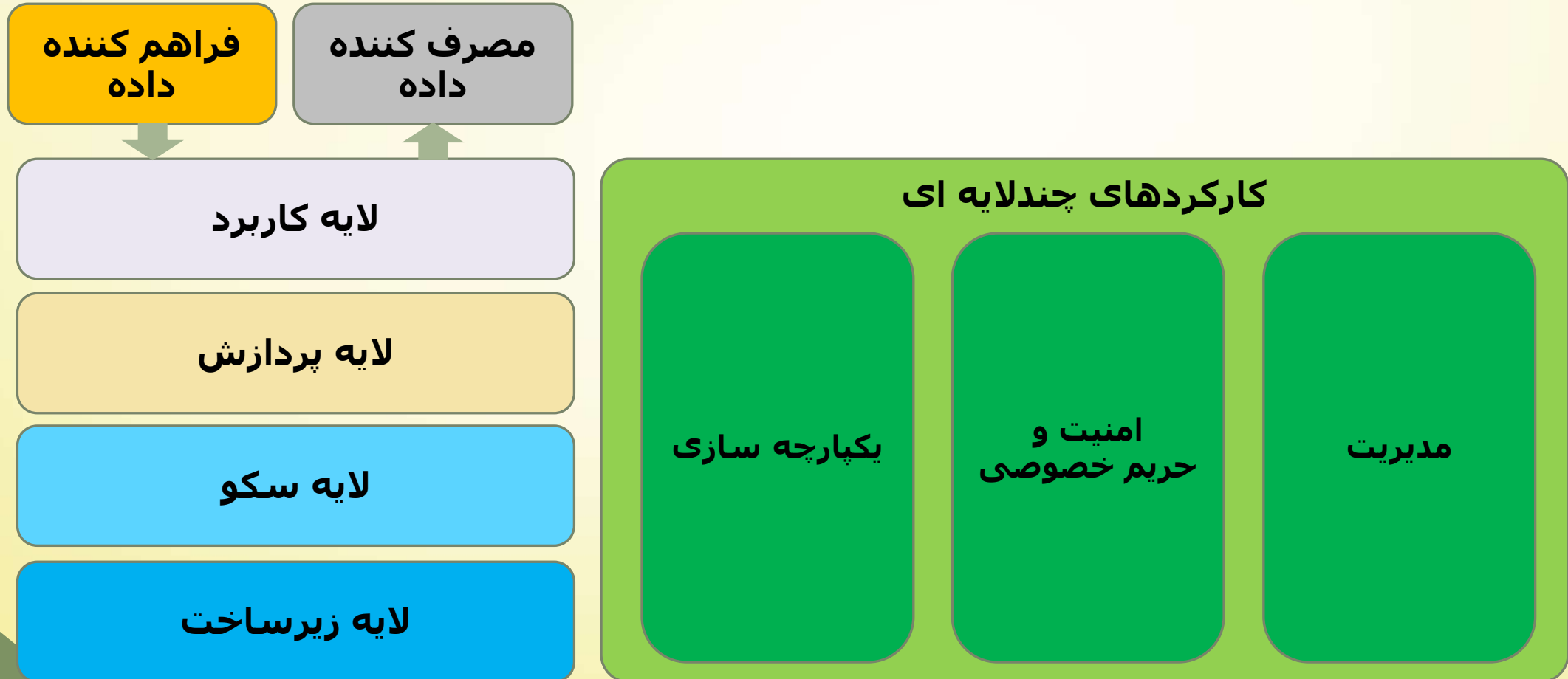


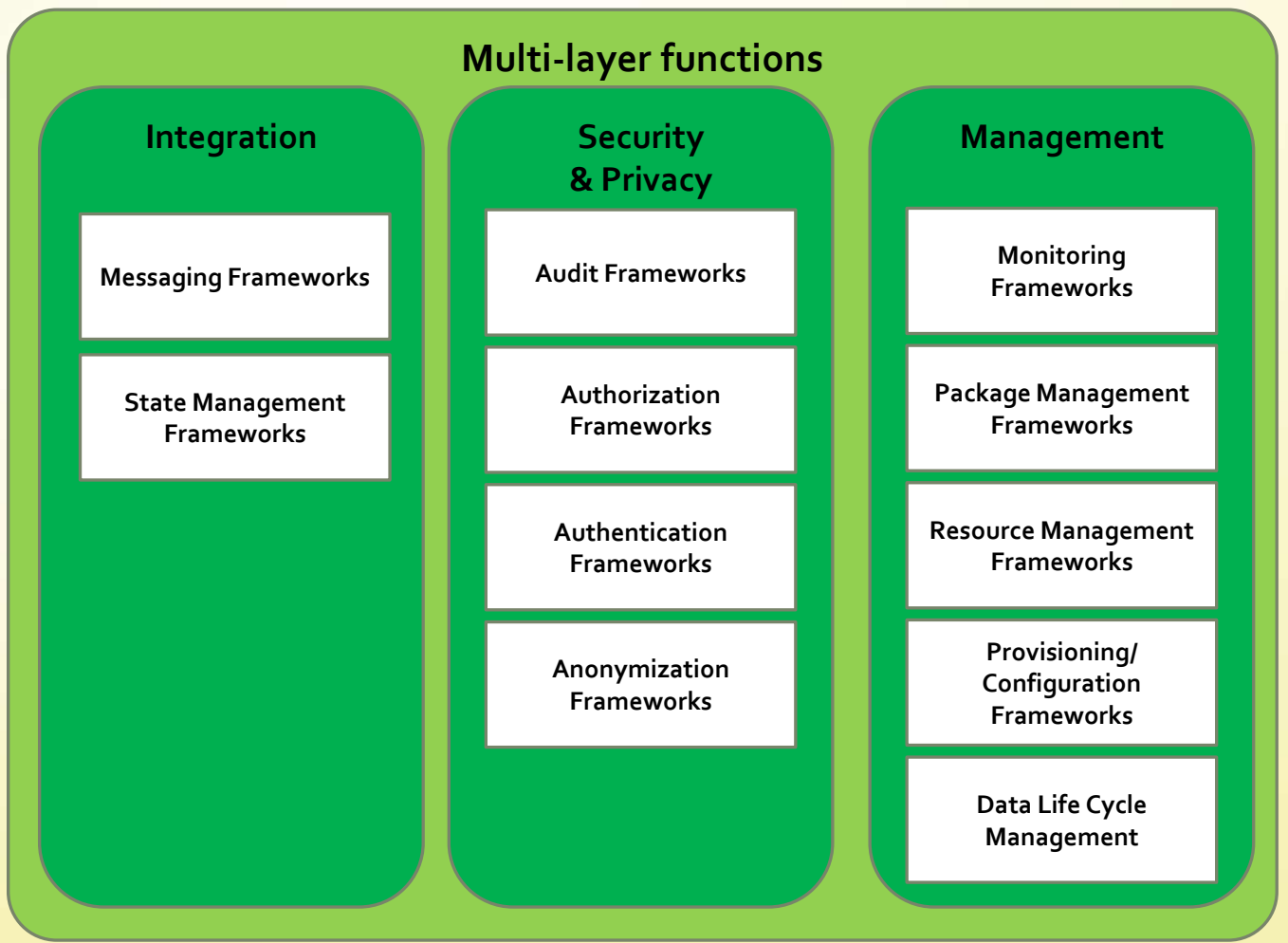
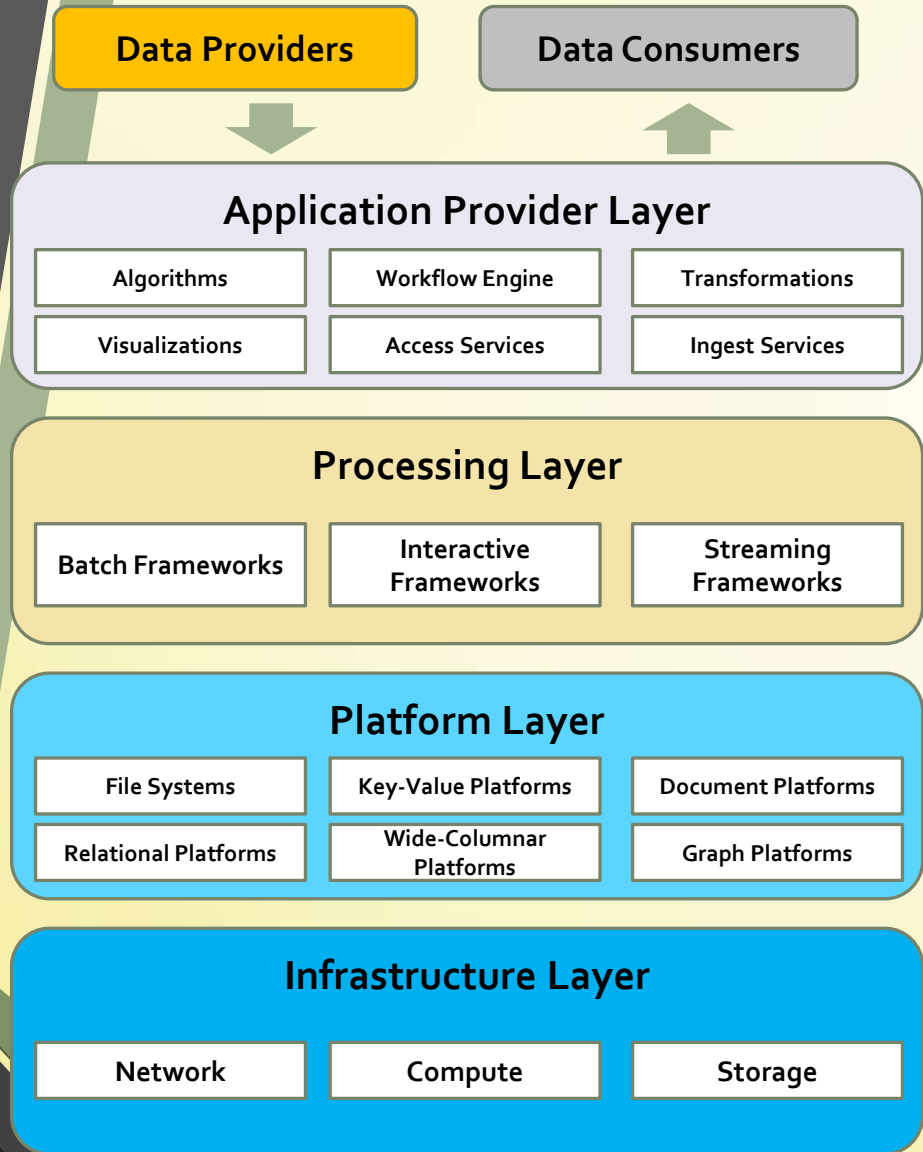


# مدل مفهومی کلان داده ها



# معماری کارکردی بر اساس ISO 20547-3





# جزئیات کارکردی لایه پردازش



داده ها

ورود

تحلیل

انتشار



نتایج / اطلاعات



پردازش دسته ای

پردازش تعاملی

پردازش جریانی



# معماری کارکردی بر اساس ISO 20547-3

فراهم کننده داده

مصرف کننده داده

## لایه کاربرد

الگوریتم ها

موتور جریان کاری

تبدیلات

مصورسازی

خدمات دسترسی

خدمات ورود داده

لایه پردازش

لایه سکو

لایه زیرساخت

## کارکردهای چندلایه ای

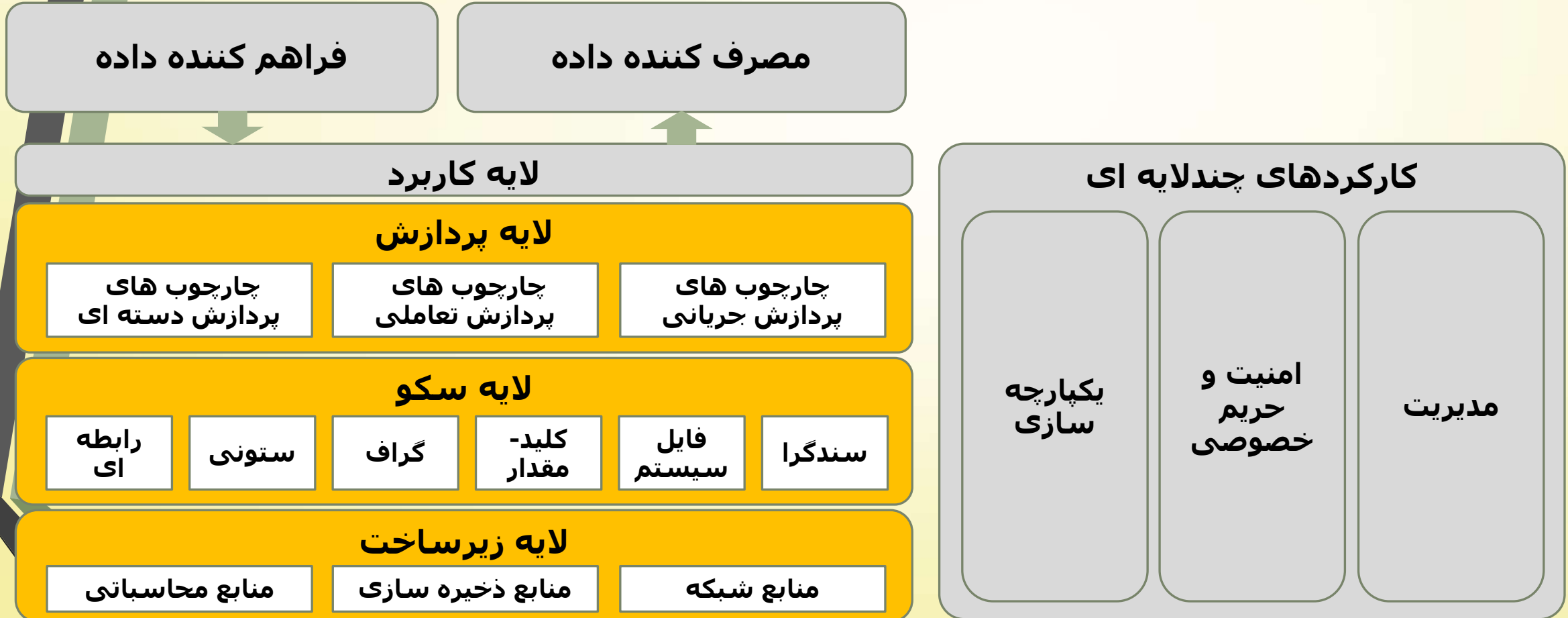
یکپارچه  
سازی

امنیت و  
حریم  
خصوصی

مدیریت



# معماری کارکردی بر اساس ISO 20547-3



# معماری کارکردی بر اساس ISO 20547-3

فراهم کننده  
داده

مصرف کننده  
داده

لایه کاربرد

لایه پردازش

لایه سکو

لایه زیرساخت

## کارکردهای چندلایه ای

امنیت و  
حریم خصوصی

چارچوب های  
اعتبارسنجی

چارچوب های  
بازرسی

چارچوب های  
گمنام سازی

چارچوب های  
احراز هویت

یکپارچه  
سازی

مدیریت



# معماری کارکردی بر اساس ISO 20547-3

فراهم کننده  
داده

مصرف کننده  
داده

لایه کاربرد

لایه پردازش

لایه سکو

لایه زیرساخت

## کارکردهای چندلایه ای

یکپارچه  
سازی

امنیت و  
حریم  
خصوصی

### مدیریت

چارچوب های  
تامین و پیکربندی  
منابع

چارچوب های نظارت

چارچوب های  
مدیریت منابع

چارچوب های  
مدیریت  
چرخه حیات داده

چارچوب های  
مدیریت بسته ها

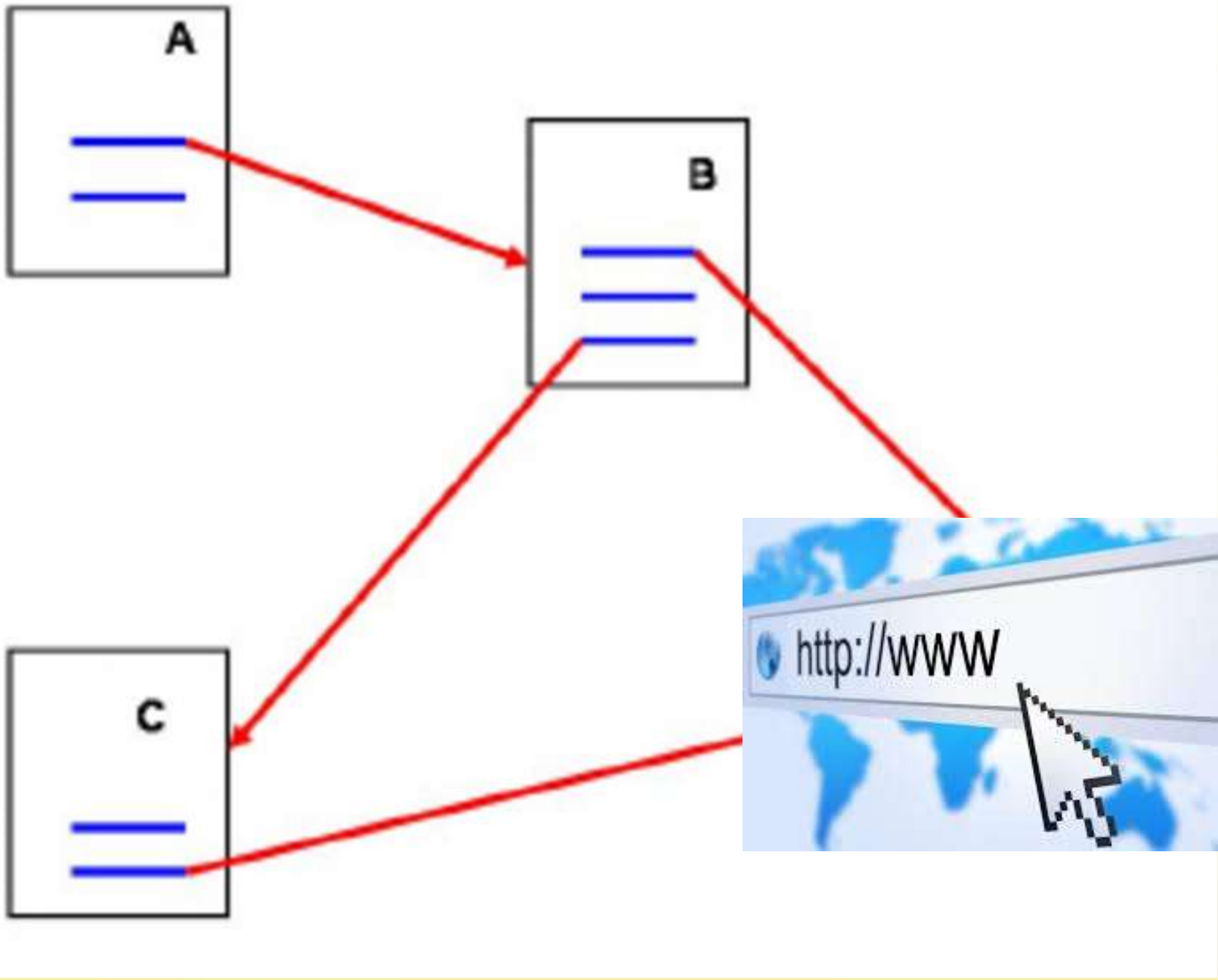
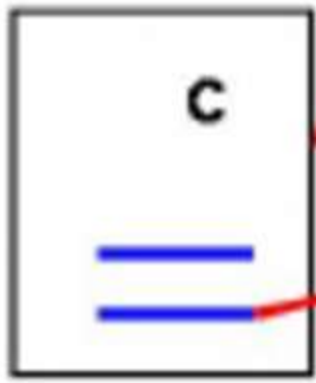
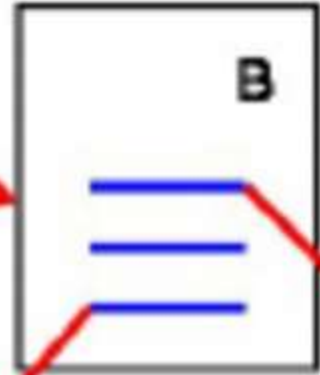




# بخش پنجم: وب معنایی (وب داده)

مرتضی سرگلزایی جوان  
مرکز تحقیقات رایانش ابری





# URI, Resource and Representation

10

URI

`http://www.tour-eiffel.fr/index.html`

identifies

representation

represents (*stands for*)

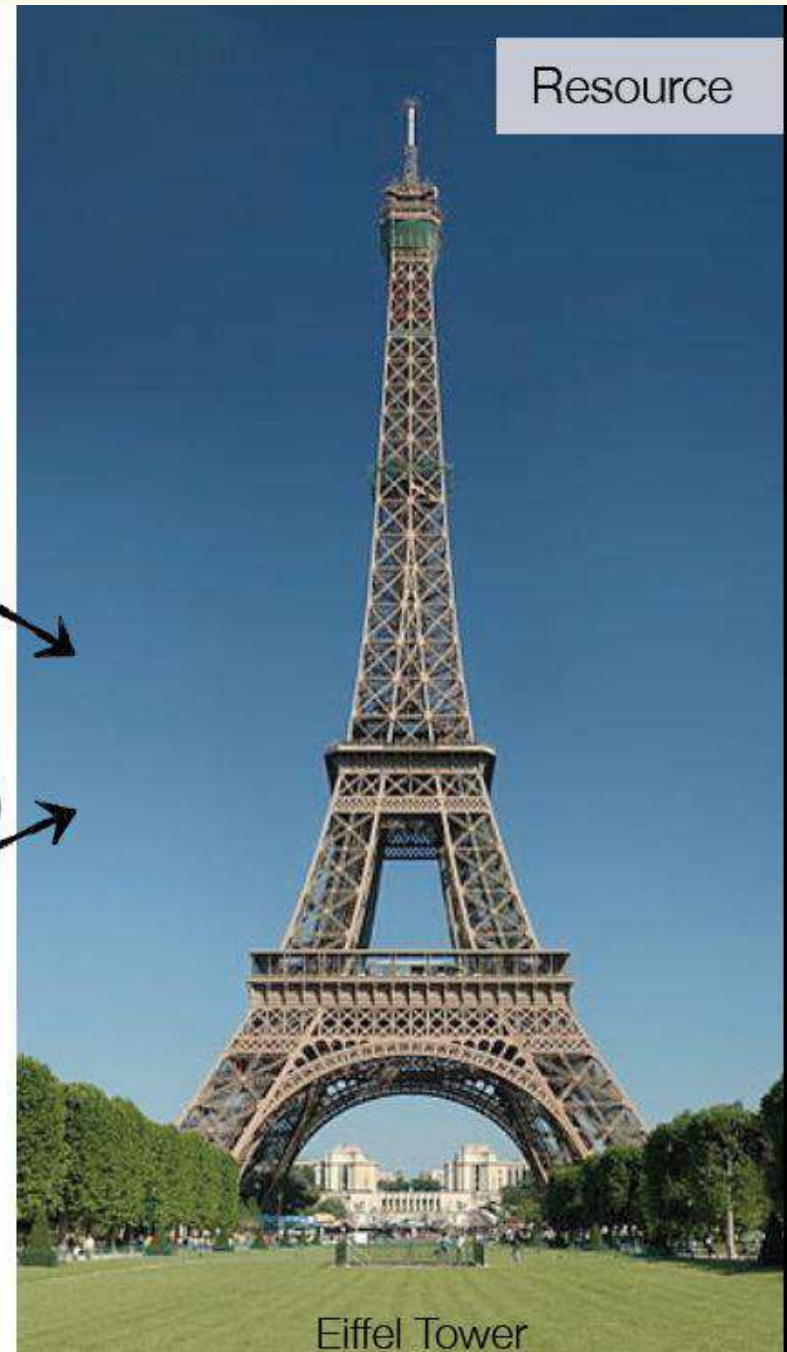
**Metadata:**

Content-type: text/html

**Data:**

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01  
Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">  
<html>  
  <head>  
    <title>Le site officiel de la Tour Eiffel</title>  
  ...  
</html>
```

Resource



Eiffel Tower



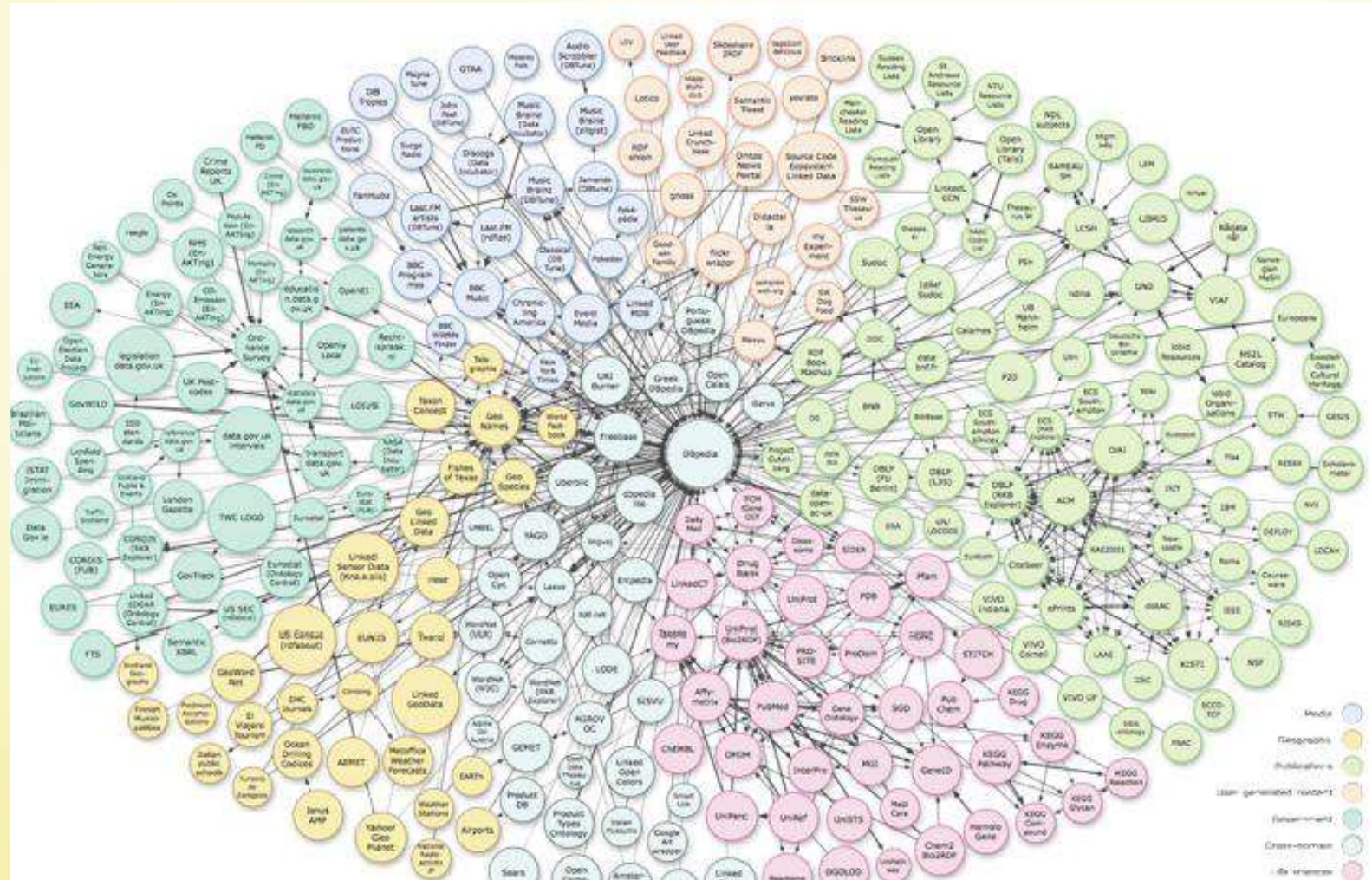
# W3C.org

In addition to the classic “Web of documents” W3C is helping to build a technology stack to support a “Web of data,” the sort of data you find in databases.

The term “Semantic Web” refers to W3C’s vision of the Web of linked data.



# Web of Data



# Real World Example

Search Engine  
Use Case

Google

Iran



All

Images

Maps

News

Videos

M

About 601,000,000 results (0.63 seconds)

Silo's of Documents  
(Traditional Web)

Collection of Facts  
(Semantic Web)

**Iran**  
Country in the Middle East

Iran is an Islamic republic on the Persian (Arabic) Gulf with historical sites dating to the Persian Empire. Extensive marble ruins mark Persopolis, the empire's capital founded by Darius I in the 6th century B.C. The modern capital, Tehran, is home to opulent Golestan Palace, seat of the Qajar Dynasty (1794–1925), plus modern landmarks such as the 435m-high Milad Tower.

**Capital:** Tehran  
**Supreme leader:** Ali Khamenei  
**Population:** 77.45 million (2013) (most recent)  
**Currency:** Iranian rial  
**Official language:** Persian  
**Neighboring countries:** Turkey, Iraq, Pakistan, Afghanistan, Turkmenistan, Armenia, Azerbaijan

**Destinations** [View 15+ more](#)

Mashhad Tehran Isfahan Shiraz Qom, Iran





# Resource Description Framework

- **Resources:**

Objects that can be addressed via **URI**

- **Properties:**

Attributes for the description of resources

- **Statements (RDF-Triple):**

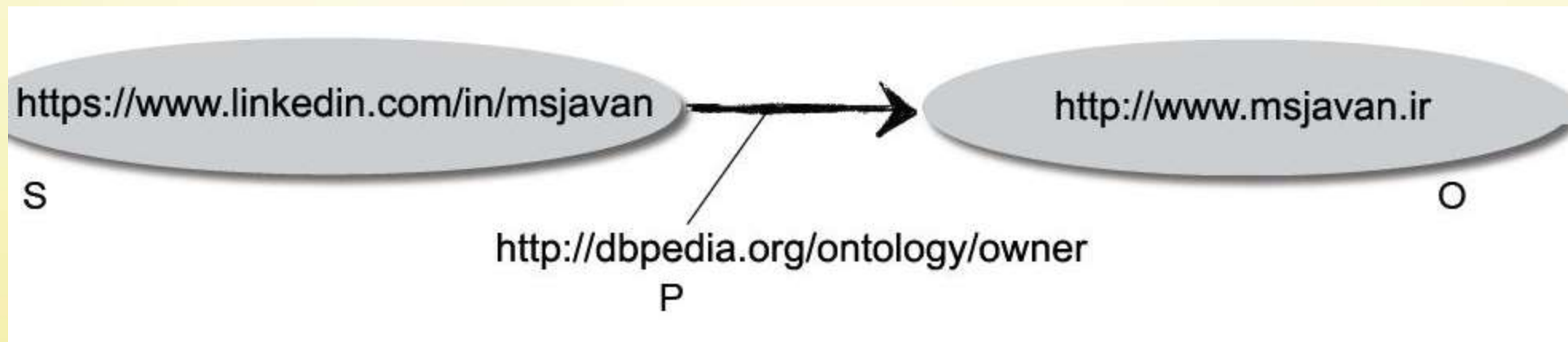
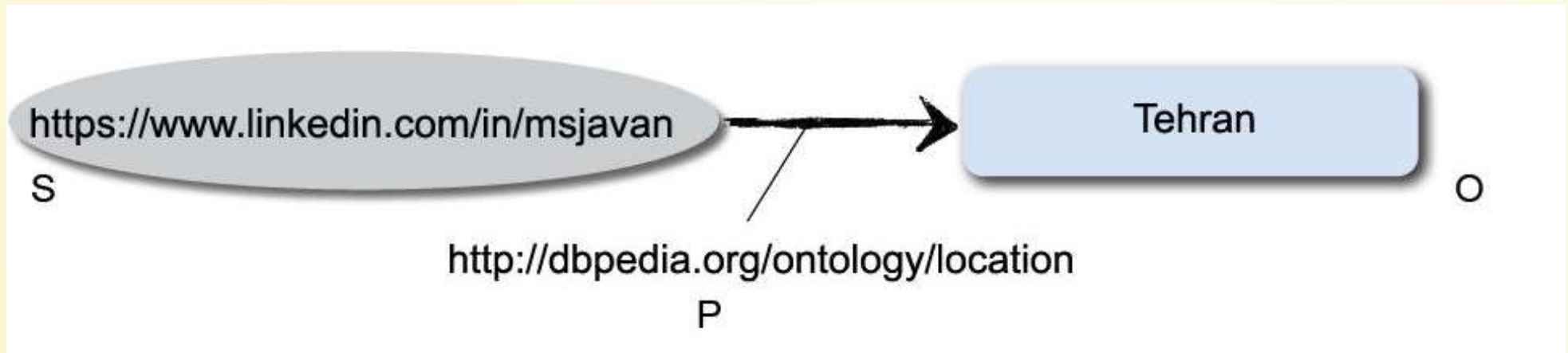
Resource + Property + Object / Value

**URI**

**URI**

**URI / Literal**







dbpedia.org



# https://dbpedia.org/sparql



The screenshot shows the Virtuoso SPARQL Query Editor interface. At the top, the browser address bar displays "Secure | https://dbpedia.org/sparql". Below this, the page title is "Virtuoso SPARQL Query Editor". There is a text input field for the "Default Data Set Name (Graph IRI)" containing "http://dbpedia.org". The main area is a text editor for the "Query Text" containing the SPARQL query: `select distinct ?Concept where {[] a ?Concept} LIMIT 100`. Below the query editor, there are settings for "Results Format" (set to "HTML"), "Execution timeout" (set to "30000" milliseconds), and "Options" (including checkboxes for "Strict checking of void variables", "Log debug info at the end of output", and "Generate SPARQL compilation report"). At the bottom, there are "Run Query" and "Reset" buttons. A note at the bottom states: "(The result can only be sent back to browser, not saved on the server, see [details](#))".

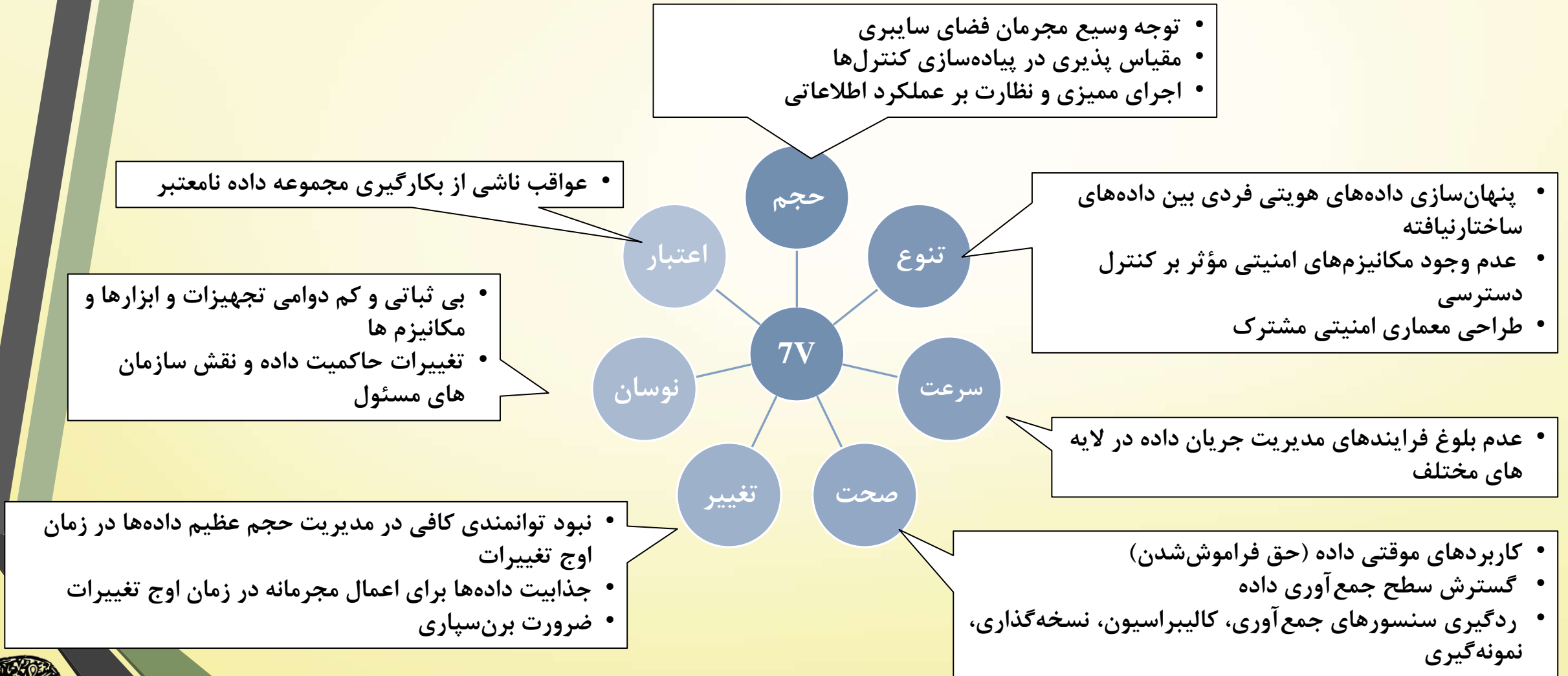


# بخش ششم: امنیت در کلان داده ها

مرتضی سرگلزایی جوان  
مرکز تحقیقات رایانش ابری



# تأثیر ویژگیهای کلان داده ها بر امنیت اطلاعات





# General Data Protection Regulation

- Data (Personal / Sensitive / Children)
- Scope of application (Controller / Processor / Joint / ...)
- Lawful grounds for processing
- Transparency requirements (Providing Notice and Collecting Consent)
- Data protection principles and accountability (Minimization / ...)
- Data subject rights (Access / Portability / ...)
- Data security (Standards)
- Data breaches (Response / Procedures / ...)
- International data transfers (Data flow mapping/ ...)
- Other controller obligations (Privacy by Design / DPO / PIA / ...)
- Other processor obligations (DPO / Sub-processors / ...)



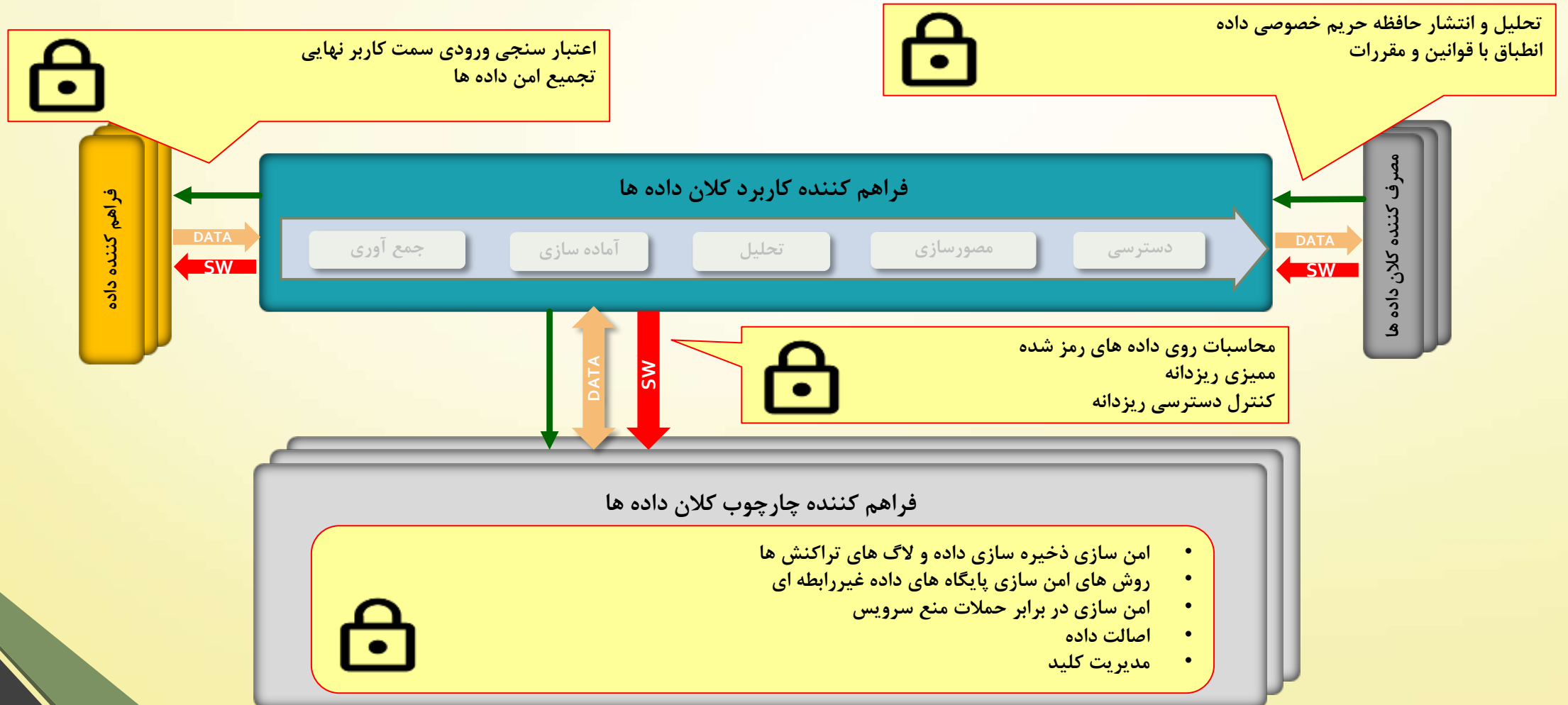
# برخی روش های مطابقت با GDPR



- گمنام سازی
- اعلانهای حریم خصوصی
- ارزیابی تاثیر حریم خصوصی (PIA)
- طراحی با ملاحظات حریم خصوصی
- صدور نشانها و گواهینامه های حریم خصوصی
- قابلیت حمل داده شخصی
- شفافیت الگوریتمها



# نیازمندی های امنیتی در مدل مرجع





# بخش هفتم: گمنام سازی داده ها

مرتضی سرگلزایی جوان  
مرکز تحقیقات رایانش ابری

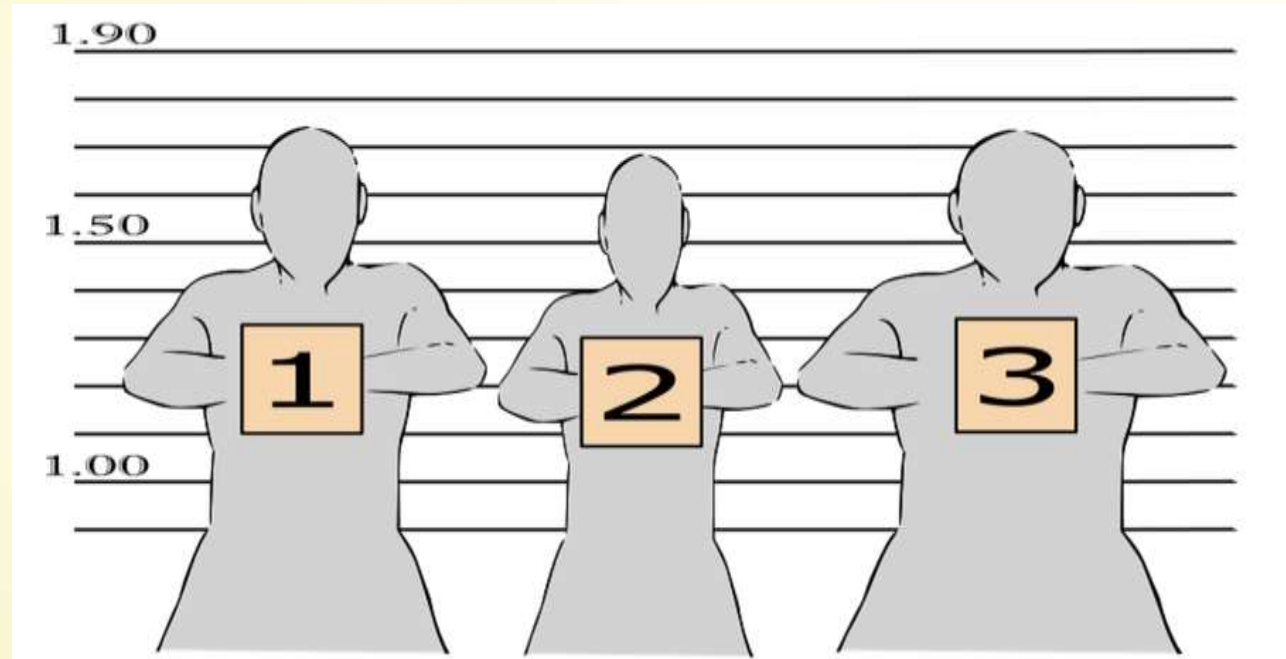


# گمنام سازی

شناسه کلیدی	شبه شناسه			شناسه حساس
Name	DOB	Gender	Zipcode	Disease
Andre	1/21/76	Male	53715	Heart Disease
Beth	4/13/86	Female	53715	Hepatitis
Carol	2/28/76	Male	53703	Brochitis
Dan	1/21/76	Male	53703	Broken Arm
Ellen	4/13/86	Female	53706	Flu
Eric	2/28/76	Female	53706	Hang Nail



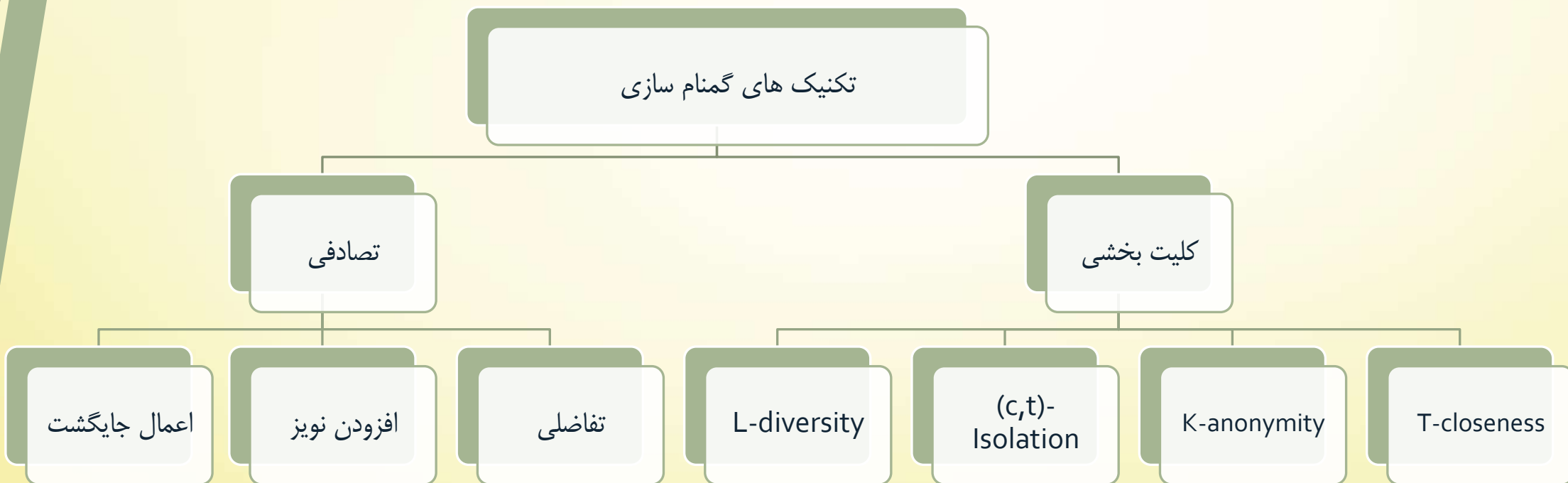
# گمنام سازی



نام	مستعارسازی	داده‌های گمنام سازی شده	شناسه حساس
Clyde	qOerd	xxxxxx	1.60
Marco	Loqfh	xxxxxx	1.72
Les	Mcv	xxxxxx	1.70



# تکنیک های گمنام سازی



# k-Anonymity

Name	Age	Zipcode	Disease
Brian	21	12000	dyspepsia
Alan	22	14000	bronchitis
Albert	24	18000	flu
Dan	23	25000	gastritis
Graeme	41	20000	flu
Harry	36	27000	gastritis
Jean	37	33000	dyspepsia
Kurt	40	35000	flu
Lily	43	26000	gastritis
Peter	52	33000	dyspepsia
Sasha	56	34000	gastritis

Age	Zipcode	Disease
[21,22]	[12000,14000]	dyspepsia
[21,22]	[12000,14000]	bronchitis
[23,24]	[18000,25000]	flu
[23,24]	[18000,25000]	gastritis
[36,41]	[20000,27000]	flu
[36,41]	[20000,27000]	gastritis
[37,40]	[33000,35000]	dyspepsia
[37,40]	[33000,35000]	flu
[43,52]	[26000,33000]	gastritis
[43,52]	[26000,33000]	dyspepsia

2-Anonymous



# k-Anonymity چالش

## Homogeneity Attack

Bob	
<b>Zipcode</b>	<b>Age</b>
47678	27

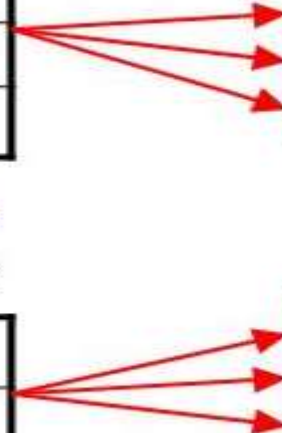
## Background Knowledge Attack

Carl does not have heart disease

Carl	
<b>Zipcode</b>	<b>Age</b>
47673	36

## A 3-anonymous patient table

Zipcode	Age	Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
4790*	≥40	Flu
4790*	≥40	Heart Disease
4790*	≥40	Cancer
476**	3*	Heart Disease
476**	3*	Cancer
476**	3*	Cancer



# L-Diversity

	ZIP Code	Age	Salary	Disease
1	47677	29	3K	gastric ulcer
2	47602	22	4K	gastritis
3	47678	27	5K	stomach cancer
4	47905	43	6K	gastritis
5	47909	52	11K	flu
6	47906	47	8K	bronchitis
7	47605	30	7K	bronchitis
8	47673	36	9K	pneumonia
9	47607	32	10K	stomach cancer

	ZIP Code	Age	Salary	Disease
1	476**	2*	3K	gastric ulcer
2	476**	2*	4K	gastritis
3	476**	2*	5K	stomach cancer
4	4790*	$\geq 40$	6K	gastritis
5	4790*	$\geq 40$	11K	flu
6	4790*	$\geq 40$	8K	bronchitis
7	476**	3*	7K	bronchitis
8	476**	3*	9K	pneumonia
9	476**	3*	10K	stomach cancer

3-Diverse




# L-Diversity چالش

## Similarity Attack

Bob	
<i>Zip</i>	<i>Age</i>
47678	27

## A 3-diverse patient table

Zipcode	Age	Salary	Disease
476**	2*	20K	Gastric Ulcer
476**	2*	30K	Gastritis
476**	2*	40K	Stomach Cancer
4790*	≥40	50K	Gastritis
4790*	≥40	100K	Flu
4790*	≥40	70K	Bronchitis
476**	3*	60K	Bronchitis
476**	3*	80K	Pneumonia
476**	3*	90K	Stomach Cancer





# جمع بندی

- آشنایی با کلان داده ها، تعاریف پایه و ویژگی ها
- آشنایی با سیستم های توزیع شده
- بررسی معماری مرجع و اجزای کارکردی کلان داده ها
- آشنایی با وب معنایی و داده های پیوندی
- بررسی ملاحظات امنیتی و حریم خصوصی در کلان داده ها
- آشنایی با برخی تکنیک های گمنام سازی داده ها



# منابع تکمیلی

- <https://bigdataawg.nist.gov/usecases.php>

**NIST Big Data Program**  
Welcome to NIST Big Data Public Working Group (NBD-PWG)!

<a href="#">Home</a> <a href="#">NBD-WG/Subgroups</a> <a href="#">Charter</a> <a href="#">Co-Chairs</a> <a href="#">Guidelines</a> <a href="#">All WG Meeting</a>	<b>Use Cases and Requirements</b>  <b>Use Case V2.0 Submission</b> (click <a href="#">M0621</a> Use Case V2.0 Template for submission)  <b>Use Case V1.0 Collection</b> [Requirements: <a href="#">UseCase</a>   <a href="#">Summary</a>   <a href="#">General</a>   <a href="#">Gen+Ref</a>   <a href="#">Gen+Ref+Gaps</a>   <a href="#">Gen+Detail</a> ] ( <a href="#">M0237</a> Use Case V1.0 Template is no longer use, <a href="#">M0180</a> to download all use cases, and <a href="#">M0203</a> for full high-level use case descriptions)
<b>Documents</b> <a href="#">Version 2 Final Docs</a> <a href="#">Version 1 Final Docs</a> <a href="#">Docs Repository</a> <a href="#">Use Cases Listing</a> <a href="#">Upload Document</a>	<b>■ Government Operation</b> <ol style="list-style-type: none"><li>1. <a href="#">M0147</a>: <i>Census 2010 and 2000 -- Title 13 Big Data</i>; Vivek Navale and Quyen Nguyen, NARA</li><li>2. <a href="#">M0148</a>: <i>National Archives and Records Administration Accession NARA, Search, Retrieve, Preservation</i>; Vivek Navale &amp; Quyen Nguyen, NARA</li><li>3. <a href="#">M0219</a>: <i>Statistical Survey Response Improvement (Adaptive Design)</i>; Cavan Capps, U.S. Census Bureau</li><li>4. <a href="#">M0222</a>: <i>Non-Traditional Data in Statistical Survey Response Improvement (Adaptive Design)</i>; Cavan Capps, U.S. Census Bureau</li></ol>
<b>Registration</b> <a href="#">New User</a> <a href="#">Update Profile</a>	<b>■ Commerical</b> <ol style="list-style-type: none"><li>5. <a href="#">M0175</a>: <i>Cloud Eco-System, for Financial Industries (Banking, Securities &amp; Investments, Insurance) transacting business within the United States</i>; Pw Carey, Compliance Partners, LLC</li><li>6. <a href="#">M0161</a>: <i>Mendeley -- An International Network of Research</i>; William Gunn, Mendeley</li><li>7. <a href="#">M0164</a>: <i>Netflix Movie Service</i>; Geoffrey Fox, Indiana University</li><li>8. <a href="#">M0165</a>: <i>Web Search</i>; Geoffrey Fox, Indiana University</li><li>9. <a href="#">M0137</a>: <i>IaaS (Infrastructure as a Service) Big Data Business Continuity &amp; Disaster Recovery (BC/DR) Within A Cloud Eco-System</i>; Pw Carey, Compliance Partners, LLC</li><li>10. <a href="#">M0103</a>: <i>Cargo Shipping</i>; William Miller, MaCT USA</li><li>11. <a href="#">M0162</a>: <i>Materials Data for Manufacturing</i>; John Rumble, R&amp;R Data Services</li><li>12. <a href="#">M0176</a>: <i>Simulation driven Materials Genomics</i>; David Skinner, LBNL</li></ol>
<b>Points of Contact</b> <b>Wo Chang</b> NIST / ITL Digital Data Advisor  <b>James St Pierre</b> NIST / ITL Deputy Director	

