

کلان داده Big data

- ۱_ کلان داده یا Big Data از موضوع های جدید فناوری اطلاعات است.
- ۲_ فعالیتهای مردم در فضای مجازی، اثر و رد دیجیتالی به جا می گذارد که امکان تصمیم گیری و راه حل های کلان را به ما می دهد. لذا رویکرد ما به سمت کلان داده اجتناب ناپذیر است.
- ۳_ در حوزه کلان داده، مرزی بین داده حساس و غیر حساس نداریم و تمایزی برای کیفیت داده ها نیست و این یک چالش مهم در این حوزه است.
- ۴_ حجم اطلاعات شخصی در حال انبوه شدن است و بخشی از داده هایی که از گوشی های موبایل ارسال و دریافت می شود ناخواسته و حاصل نرم افزار های نصب شده است و این داده ها در بخش خصوصی جمع آوری می شود.
- ۵_ اطلاعاتی که google از "من" دارد خیلی بیشتر از اطلاعاتی است که مملکت خودم از "من" دارد.
- ۶_ یکی از چالش های کلان داده این است که حجم بالای داده با کیفیت پایین ممکن است منجر به تصمیم گیری غلط شود.
- ۷_ چالش دیگر اینکه افراد با شرایط یکسان ولی دارای دسترسی های متفاوت هستند که این نابرابری است.
- ۲_ به قول آقای Tom Peters، هر سازمانی که اهمیت کلان داده را درک نکند، در اقتصاد جاری محکوم به فناست.
- ۳_ استفاده بهینه از داده ها، مدیریت جامع دسترسی پذیری و امنیت داده ها برای منافع سازمان مفید است. لذا باید ساختار حکمرانی داشته باشیم. آیا داده ها در جایگاه اصلی خود هستند؟
- ۴_ تا سال ۲۰۲۵، کل داده تولید شده ۱۷۵ ZB خواهد بود $ZB=10^{21}$ Byte و این به طور نمایی در حال افزایش است.
- ۵_ تقاضا برای داده های قابل اعتماد در حال افزایش است .
- ۶_ ازدیاد سیستم های متمرکز بر روی داده ها مثل ERP, CRM, ECM داریم.
- ۷_ بدون تحلیل داده ها، حکمرانی آنها قابل انجام نیست.
- ۸_ اشتراک گذاری داده مستلزم سازماندهی داده است.
- ۹_ کیفیت داده (حس اعتماد، ثبات، ارائه به موقع داده روزانه) در تصمیم سازی سطوح مختلف کارساز است.
- ۱۰_ استقرار سیستم یکپارچه که به واسطه ساماندهی داده ها، اطلاعات قابل قبول برای مدیریت را فراهم می نماید.
- ۱۱_ چارچوب حکمرانی داده ها (تشکیلات، سیاست ها، فرآیندها، استانداردها)
- ۱۲_ موارد مهم در حکمرانی کلان داده شامل حفظ اطلاعات شخصی، حریم خصوصی، ظهور تک قطبی، انحصار، تضمین کیفیت داده و عدم دسترسی به منبع داده است و الان مایکروسافت و گوگل نمونه هایی از حکمرانی داده هستند.

۱۳_ ویژگی های داده های کلان داده : timeliness , Trustfulness , meaningfulness , sufficiency

۱۴_ باید بتوان نتایج معنی دار را ارائه نمود .

۱۵_ مدیریت چرخه عمر داده انجام شود.(نمایه سازی داده ها)

۱۶_ در سطح ملی، آزاد سازی کلان داده های ارزشمند دولتی و توسعه صنعت مبتنی بر داده مطرح است.

۱۷_ در سطح سازمانی، توسعه فرهنگ داده محور در بخش و دستگاههای دولتی، درآمدزایی از داده های سازمانی و ارزشمند سازی داده ها مطرح می شود.

۱۸_ سند IGIF به عنوان چارچوب تعامل پذیری دولت الکترونیک

۱۹_ تدوین نقشه راه کلان داده ها در پژوهشگاه ارتباطات انجام می شود.

۲۰_ رگولاتوری کلان داده در سازمان تنظیم مقررات انجام می شود.

۲۱_ لایحه حکمرانی الکترونیکی داده در دست اقدام است.

۲۲_ همایش حکمرانی داده در سال ۱۳۹۸ خواهد بود شامل توافقات و مقررات تاثیر گذار، بررسی اسناد بالا دستی، اسناد پشتیبانی فنی، الزامات قانونی و حقوقی

۲۳_ سایت bigdata.itrc.ac.ir در دسترسی است.

۲۴_ کلان داده ، فراتر از گنجایش دیتابیس های معمولی است و نرم افزارهای آن برای آنالیز هزاران گیگابایت طراحی شده اند.

۲۵_ میزان کلان داده ذخیره شده در جهان به Peta Byte :

حجم < ۳۵۰۰ (آمریکای شمالی) _ حجم < ۲۰۰۰ (اروپا) _ حجم < ۲۵۰ (چین)

حجم < ۵۰ (آمریکای لاتین) _ حجم < ۲۰۰ (خاورمیانه) _ حجم < ۵۰ (هند)

۲۶_ صنایع کلان داده عبارتند از: Banking , Discrete manufacturing , Process manufacturing

۲۷- قوانین کلان داده (Big Data Rule) در حوزه های بانک داری، تحصیل، حاکمیت، تجاری، ارتباطات ، سلامتی، کارخانه، حمل و نقل، رسانه های اجتماعی و IOT مطرح می شود.

Big Data service provider

Data protection law

Who control data protection?

Austria : mix of federal and state legislation

EURO union: general data protection regulation

SINGAPOR: personal data protection act 2012

۲۸-چالش های کلان داده :

Structured V.S unstructured

Data protection

Data Privacy

Data ownership

"Data is The new oil"

Data is being hailed as "the new oil" .

Big data and its application promise to transform the way we live and work

۲۹- شرکت هایی هستند که منتظر باز شدن داده هستند تا کسب و کار داده راه بیاندازند و شهرداری مشهد این کار را کرده است. open Data

۳۰- چارچوب رگولاتوری کلان داده: مطالعه حقوقی داخلی، مطالعه حقوقی تطبیقی و نتیجه آن شناسایی حوزه های قانونی کلان داده ها ، شناسایی خلاء ها و چالش های قانونی، شناسایی راهکارهای اتخاذ شده توسط سایر کشورها و بومی سازی.

۳۱- ارتباط کلان داده ها با داده ها از منظر حقوقی: -کلان داده، یک فضای داده محور است که اهمیت دارد. -اصول حمایت از داده، اصول کلیدی و پایه مورد نیاز برای حمایت در حوزه کلان داده ها است.

۳۲- گزارش ITU در سال ۲۰۱۷ چالش های حقوقی درباره کلان داده شامل -موضوعات پایه، موضوعات تکمیلی، داده های دولت باز ، تبادل داده میان نهادهای داخلی، حمایت از داده شخصی و جریان اطلاعات بین کشورها است

۳۳-چالش ها:

- امنیت و حریم خصوصی و حمایت داده ها
- داده دولت باز OPEN DATA
- مالکیت داده (آیا هر استفاده ای از داده ها امکان پذیر است؟)

۳۴- بیش از ۵۰ مورد کاربرد در زمینه کلان داده هست: که شامل دولت: سرشماری ، تجاری : موتور جستجو ، نظامی : تشخیص و ردگیری اشیاء ، سلامت : محاسبات ژنتیک ، شبکه اجتماعی : نظیر تویتر ، تحقیقاتی : جستجوی منابع ، فیزیک : شبیه سازی ، محیط زیست : هواشناسی ، انرژی : شبکه هوشمند . حجم داده های در پردازش تصاویر پزشکی ، ده ها ترابایت در روز و در اسناد تحقیقاتی، یک ترابایت در ماه و در سرویس ویدیویی، دو پتا بایت و در تشخیص تقلب در تراکنش های بانکی، چندین گیگابایت در کسری از ثانیه است

- ۳۵_ تعیین تکلیف در حوزه کلان داده و تبیین خط مشی کلان داده در کشور مهم است.
- ۳۶_ مباحث جدیدی در حوزه ICT در حال ظهور است که مهمترین آن اقتصاد دیجیتالی و مبتنی بر ICT و یکی از مولفه های مهم اقتصاد دیجیتالی **big data** است.
- ۳۷_ در سال ۲۰۱۷ هر دقیقه ۲۵۰۰۰۰ تویت تولید می شود و یا دقیقه ای ۱۵۰۰۰۰ ایمیل و همچنین در هر ثانیه تعداد بسیاری کلید واژه در گوگل جستجو می شود.
- ۳۸_ چهار شرکت گوگل، فیس بوک، آمازون و مایکروسافت کل داده های دنیا را جمع آوری می کنند و این حکمرانی شرکت ها است.
- ۳۹_ استفاده از **big data** به عنوان اصل پذیرفته شده است و ما باید به سمت استفاده از آن برویم.
- ۴۰_ مهمترین چالش، شفاف نبودن فرایند پردازش این داده ها است و چالش دوم ، حجم بسیار بالای این داده ها است و اینکه نمی توان تشخیص داد که حساس هستند یا حساس نیستند.
- ۴۱_ شرکت هایی که از کلان داده استفاده می کنند معمولاً بی طرف نیستند و متناسب با سلیقه خود نتیجه را ارائه می دهند. بحث بی طرفی را رعایت نمی کنند.
- ۴۲_ اصل عدم تبعیض هم رعایت نمی شود (کنار اصل بی طرفی)
- ۴۳_ حفاظت داده های شخصی افراد ، مهمترین چالش است.
- ۴۴_ چالش مختص کشور ما، عدم مدیریت و تصمیم گیری صحیح و نداشتن نهاد تصمیم گیر مشخص در حوزه **big data** است. رگولاتوری می تواند عهده دار این مسئولیت باشد.
- ۴۵_ در اروپا: کمیسیون تنظیم مقررات اروپا این موضوع را منتشر کرده است.
- متولی حوزه کلان داده در کشور مشخص شود و رگولاتوری می تواند اینکار را انجام دهد ولی تصمیم گیر اصلی نیست و در حد اختیارات قانونی می تواند باشد و نیاز به تصمیم گیر کلان هست.
- ۴۶_ پروژه ای در وزارت ارتباطات وجود دارد که سیاست گذاری **big data** را انجام می دهد و بحث دسترسی داده های دولتی در دستور کار آن قرار گرفته است.
- ۴۷_ مقررات استفاده از کلان داده در حال بررسی در حوزه رگولاتوری است که با پژوهشگاه در حال پیشروی است.
- مرز بین کلان داده و غیر کلان داده کجاست؟
- ۴۸- **Small data**: داده هایی که یک تحلیلگر می تواند به طور کامل آنها را بفهمد و ارزیابی کند. داده خرد است.
- ۴۹- **Big data**: مجموعه داده هایی که نیازمند یک معماری مقیاس پذیر برای پردازش می باشند داده کلان است

۵۰- سلسله مراتب کلان داده:

	data lake		
Data set	data set	data set	
	Record		
Element	element	element	

۵۱- مدل مرجع ISO/IEC 20547

Data is the new oil?

همچنان که از دکل نفت اکتشاف و پردازش صورت می گیرد از Data هم استخراج اطلاعات مفید و پردازش صورت می گیرد.

فراهم کننده کلان داده - جمع آوری - تحلیل - مصور سازی - دسترسی - مصرف کننده کلان داده

۵۲- نمای کارکردی: ISO/IEC 20547

فراهم کننده داده مصرف کننده داده

لایه کاربردی

لایه پردازش

لایه سکو

لایه زیرساخت

۵۳- انتخاب شرکت ها و کشورها برای مطالعات تطبیقی:

Hp, ibm, fujitsu, oracle, emc , Microsoft, uber , google , amazon , facebook

۵۴-فراهم کننده کاربرد داده های کلان : ayasdi , bloomreach

۵۵-الگوی بررسی خدمات کلان داده : (نوع سرویس) : dashboard , data base

۵۶-معماری مرجع از نمای کارکردی: لایه کاربردی، لایه نمایش، لایه سکو و لایه زیرساخت

۵۷-کیفیت سرویس دهنده در حوزه کلان داده : از فراهم کننده شروع می شود و به مصرف کننده داده ها خاتمه می یابد.

۵۸-موقعیت جغرافیایی پردازش داده ها: الگوی محاسبه سطح دسترسی، هر سرویس یک SLA دارد.

۵۹-در کلان داده Confidentiality , Integrity , Availability سه ضلع مثلث امنیت آن هستند.

۶۰- گمنام سازی: داده را به شخص خاصی اختصاص نمی دهیم. و آن فیلد مربوط به حریم خصوصی را حذف می کنیم. مثلا در اطلاعات مشخصات پزشکی، همه مشخصات هست به غیر از نام بیمار.

K: anonymity

L:diversity

T:closeness

جمع بندی:

در خصوص کلان داده با توجه به اینکه روندهای رشد نمایی در IOT داریم و تعداد device ها آن دائم در حال افزایش است و لذا data رشد می کند. توان پردازشی نیز رشد نمایی دارد. رشد جمعیت باعث بالا رفتن connectivity بین انسانها می شود و بنابراین نیاز به پردازش داریم و همه اینها نشان دهنده ارتباطات نزدیک این موضوعات با همدیگر است.

Data is power

ثاقب ۹۷/۱۰/۱۵