

Open Community of Cloud Computing
جامعه آزاد رایانش ابری ایران

کارگروه تاکسونومی و استانداردسازی

نقشه راه توسعه کلان داده در کشور

تدوین مبانی اولیه فلسفه مورد نیاز برای تدوین سندی کلان در حوزه کلان داده

همایش داده های عظیم در تاریخ ۲۴ دی ۱۳۹۳ در محل پژوهشگاه ارتباطات و فناوری اطلاعات (مرکز تحقیقات مخابرات ایران) برگزار گردید که در بر اساس برنامه ریزی های صورت گرفته، به موازات برنامه های همایش، از طرف کارگروه کلان داده جامعه آزاد رایانش ابری یک تیم تخصصی جهت گردآوری و جمع بندی محتوای ارایه شده شروع به کار کردند. ضمن اینکه نتایج آن در این سند منعکس گردیده است، شکاف های موجود شناسایی شده است و برای رفع آنها نقشه راهی پیشنهاد گردید است تا مورد استفاده مدیران و تصمیم گیران قرار بگیرد. با توجه به اینکه یک نسخه از مستند به صورت آزاد از طریق ویکی جامعه در دسترس است، از متخصصین و صاحب نظران دعوت می شود که ضمن بررسی مستندات تهیه شده، نظرات خود را یا بصورت مستقیم در ویکی، و یا از طریق ایمیل به آدرس info@occc.ir اعلام نمایند.

نسخه ۴.۰ / دیماه ۱۳۹۵

به نام خدا



کارگروه تاکسونومی و استانداردسازی

نقشه راه توسعه کلان داده در کشور



به نام خدا



مرتضی جوان

کارگروه تاکسونومی و استانداردسازی

جامعه آزاد رایانش ابری ایران

استانداردها میزان بلوغ یک صنعت را در طول زنجیره ارزش آن از ابعاد مختلف (فنی/تجاری/فرهنگی) نشان می‌دهند و بر اساس بلوغ صنعت به روزرسانی می شوند. نقشه راه یک فناوری اغلب روشی برای شناسایی شکاف بین استانداردهای موجود و استانداردهای مورد نیاز برای رسیدن به وضعیت مطلوب است. کارگروه کلان داده در جامعه آزاد رایانش ابری ایران، متشکل از اعضای مختلفی از صنعت، دولت، دانشگاه و جامعه کاربری، مطالعات گسترده ای را تا کنون در حوزه کلان داده در ایران و جهان انجام داده است و با توجه به تجربیات موفق مشابه در حوزه رایانش ابری، در تیرماه ۱۳۹۴ نسبت به انتشار نسخه اولیه نقشه راه پیشنهادی در موضوع کلان داده با تمرکز بر بعد فرهنگی اقدام نمود و بر اساس آن برنامه ریزی جهت حرکت به سمت وضعیت مطلوب متناسب با ظرفیت های جامعه آزاد رایانش ابری ایران صورت گرفت.

پس از گذشت یک سال و نیم از ارائه سند قبلی، و ضمن رصد آخرین وضعیت کلان داده در سومین رویداد بزرگ سالیانه جامعه آزاد رایانش ابری ایران، از زاویه دید رایانش ابری بعنوان یک فناوری راهبردی و پشتیبان، سند نقشه راه پیشنهادی مورد بازنگری قرارگرفته و نسخه ۴,۰ آن در ادامه تقدیم می گردد. در این نسخه ضمن بررسی استانداردهای جدیدی که در حوزه کلان داده تدوین شده است، اکوسیستم کلان داده و نقش ها و فعالیت‌های اصلی و سطح بالای مطرح در آن بر اساس استاندارد ITU-T Y.3600 بطور خلاصه نشان داده شده است که جزییات بیشتر در گزارش اصلی قابل دسترسی است. همچنین انطباق معماری رایانش ابری بر اساس استاندارد ISO17789 برای پشتیبانی از این خدمات کلان داده در بستر ابر نشان داده شده که در قالب یک مثال کاربردی برخی از ابعاد آن مورد شرح و بررسی قرار گرفته است. بررسی های اخیر می‌تواند به عنوان الگوی مناسبی برای استانداردسازی مفاهیم مورد استفاده در حوزه کلان داده در کشور در نظر گرفته شود تا زمینه بهتری برای توسعه فناوری و خلق ارزش بر اساس یک ادبیات مشترک در بین بازیگران مختلف این صنعت فراهم شود.

فهرست مطالب

۱- مقدمه ۵
۲- وضعیت کلان داده در دنیا ۵
۳- وضعیت کلان داده در ایران ۱۱
۴- چشم انداز کلان داده در دنیا و در کشور ایران ۱۲
۵- چالش های حوزه کلان داده در دنیا و در کشور ایران ۱۵
۶- برنامه ریزی در حوزه کلان داده ۱۸
۷- نقشه راه ۲۰
۸- پروژه های پشتیبان نقشه راه ۲۲
۹- پیوست۱: استاندارد کلان داده مبتنی بر رایانش ابری (ITU-T Y.3600) ۲۳
۱۰- پیوست۲: مثال موردی استفاده از کلان داده در شخصی سازی ۲۵
۱۱- پیوست۳: گزارش آماری ۲۶

۱- مقدمه

ظهور فناوری ها اغلب دارای مبانی فلسفی بسیاری است که روند توسعه و دگرگونی آنها را توجیه می کند. بنابراین شناخت این مبانی فلسفی می‌تواند نقش مهمی در تدوین نقشه راه توسعه هر فناوری داشته باشد. مجموعه هایی که در این موضوع اندیشه کافی کرده باشند، توسعه هر فناوری را بر اساس یک ضرورت و یا یک نیاز واقعی که دارای توجیه کافی از ابعاد مختلف می باشد دنبال می‌کنند و مسیر توسعه آن را بر اساس نیازهای واقعی خود اصلاح می کنند. کشورهای دنباله رو، اغلب بدون شناخت این مبانی فلسفی و نظری، سعی می کنند ضرورت حرکت از مسیر طی شده توسط دیگران را برای رسیدن به یک وضع مطلوب توجیه کنند و یا از الگوهای مشابه پیروی کنند، فارغ از اینکه نیاز واقعی خود را شناسایی کرده باشند و مبانی فلسفی آن را شکل دهند. در این گزارش سعی شده است با این نگاه به تدوین مبانی اولیه فلسفه مورد نیاز برای تدوین سندی کلان در حوزه کلان داده بپردازیم. حال آنکه حرکت در این مسیر فراتر از یک همایش و برگزاری چند نشست تخصصی است و باید با تامل و استمرار بیشتری همراه باشد.

۲- وضعیت کلان داده در دنیا

امروزه داده ها به عنوان یک دارائی ملی شناخته می شوند. همچنین تحلیل و استفاده از داده به عنوان یک عامل کلیدی برای رقابت و رشد برای کلیه شرکت ها تبدیل شده است و موجب مزیت رقابتی در کسب و کارها، محرک نوآوری، افزایش رقابت و اثرات مثبت اجتماعی خواهد شد. بنابراین شرکت ها از نقطه نظر رقابت و توان دریافت ارزش، نیاز به بهره گیری از کلان داده دارند. از جمله مزایای استفاده از کلان داده که بصورت مکرر در ادبیات موضوع ذکر شده است، فراهم آمدن امکان پیشگویی آینده و تصمیم گیری سریع تر، بهتر و دقیق تر در حوزه های مختلف است. مطالعات نشان میدهد بهبود در تصمیم گیری در حوزه های اقتصادی و اجتماعی موجب رشد و توسعه اقتصادی کشورها و بهبود شرایط رقابت برای کسب و کارها است. البته استفاده از داده ها برای پیشگویی آینده خیلی جدید نیست و دولت ها و کسب و کارها همواره از مدل کردن داده برای پیشگویی سناریوها و عدم قطعیت هایی که ممکن است پیامدها را همراهی نمایند استفاده نموده اند. فرق کلان داده این است که انتظار میرود که عموما به دلیل دسترسی به حجم بیشتر اطلاعات، صحت و دقت پیشگویی در آن بالاتر است. از جمله مهمترین فعالیت هایی که در چند سال اخیر در بحث کلان داده در دنیا انجام شده است، برنامه ریزی برای استانداردسازی و ایجاد یک فهم مشترک بین موجودیت های درگیر است که بر اساس آن بتوانند نسبت به تدوین نقشه راه کلان داده به عنوان یک فناوری اقدام کنند و خروجی های متعددی نیز داشته است که مهمترین آنها در ادامه توضیح داده شده است.

در سال ۲۰۱۲ در مجموعه CSA کارگروهی با هدف شناسایی تکنیک های مقیاس پذیر برای مشکلاتی امنیتی و حریم خصوصی داده-محور ایجاد شد که تاکنون گزارش های متعددی از این کارگروه منتشر شده است. انتظار می رود که خروجی های این کارگروه بتواند در شفاف کردن تجارب موفق برای امنیت و حریم خصوصی در کلان داده موثر باشد، همچنین صنعت و دولت ایالات متحده را در پذیرش آن تجارب موفق راهنمایی کند. فعالیت های مرکز استاندارد آمریکا (NIST) نیز از اواسط ۲۰۱۲ با یک برنامه کارگاهی شروع شد و یک سال بعد کارگروهی را بصورت عمومی ایجاد کرد. هدف این کارگروه ایجاد یک توافق عمومی و مشترک در تعاریف، ارتباط بین موجودیت ها، معماری مرجع و نقشه راه فناوری برای تکنیک های تحلیلی و زیرساخت های فناوری می باشد تا بصورت امن و موثر بتواند از پذیرش کلان-داده پشتیبانی کند. کمیته مدیریت داده در ISO/IEC JTC1 نیز

مطالعات خود را بر روی نسل بعدی تکنیک‌های تحلیلی شروع کرده است. موسسه W3C هم چندین کارگروه روی موضوعات مختلف کلان داده تشکیل داده است که منجر به انتشار گزارش هایی در خصوص برخی فناوری های باز مبتنی بر وب قابل استفاده در موضوع کلان داده شده است. یکی از مهمترین گزارش های بدست آمده تا کنون مربوط به ITU-T است که در اواخر سال ۲۰۱۵ پیشنهادات خود را در قالب گزارشی با عنوان ITU-T Y.3600 منتشر کرد که شامل نیازمندی های مبتنی بر رایانش ابری برای کلان داده بوده است؛ خلاصه ای از این گزارش در پیوست مورد بررسی قرار گرفته است. در ادامه به بررسی دقیق تر فعالیت مرکز NIST به عنوان یک نهاد حاکمیتی و اثرگذار در سطح کلان می پردازیم.

۲-۱- فعالیت های انجام شده در NIST

این مجموعه راهبری تدوین نقشه راه فناوری کلان داده را در سطح جهانی بر عهده گرفته است. در این نقشه راه اولویت های کلان در حوزه های مختلف تعامل پذیری، قابلیت حمل، قابلیت استفاده مجدد و قابلیت گسترش برای تکنیک های تحلیلی و نیز فناوری های زیرساختی کلان داده مشخص خواهد شد. به این منظور، کارگروهی برای تدوین استانداردهای حوزه کلان داده مشتمل بر موارد زیر ایجاد شده است:

- تدوین تعاریف مشترک و واحد از کلان داده
- تهیه یک دسته بندی از کلان داده به منظور شناخت مسایل مختلف مطرح در کلان داده و ارتباط بین آنها
- تهیه نیازمندی های فنی و غیرفنی مرتبط با کلان داده
- تهیه نیازمندی های امنیتی و نیازمندی های مرتبط با حریم خصوصی
- تهیه یک معماری مرجع از کلان داده
- تهیه نقشه راه فناوری کلان داده

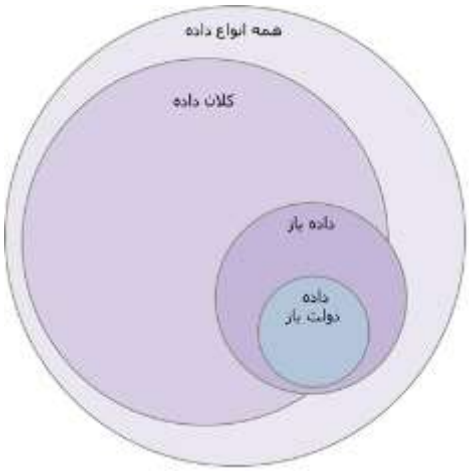
تهیه نقشه راه یک فناوری می تواند نقش موثری در جهت دهی فعالیت ها و برنامه ریزی ها داشته باشد تا بتوان برای مراحل مختلف بلوغ فناوری برنامه ریزی کرد. برای انجام این کار نیاز به ایجاد یک فهم مشترک می باشد. تدوین تعاریف واحد و استاندارد سبب ایجاد یک فهم مشترک بین همه طرفین درگیر می شود و می تواند از بروز بسیاری از ابهامات جلوگیری کند. پس از ایجاد این فهم مشترک است که می توان نسبت به توصیف موجودیت های مختلف و دسته بندی آنها اقدام کرد. این اقدام دید خوبی برای شناخت مسایل مختلف مطرح در کلان داده و ارتباط بین آنها ایجاد میکند و به شناخت نیازمندی های فنی، غیرفنی و امنیتی منجر می شود. با مشخص کردن ارتباط بین اجزا و نیازمندی ها و ارتباط بین آنها می توان به یک معماری مرجع دست پیدا کرد که قادر باشد همه نیازمندی ها را در بر گیرد.

۲-۲- بررسی چند پروژه کلان مرتبط با کلان داده

در این بخش چند نمونه از فعالیت هایی که در ارتباط با کلان-داده و مرتبط با موضوعات امنیتی یا استانداردسازی، بطور خاص در سطوح حاکمیتی راهبری می شوند ارائه شده است. مطالعه دقیق چنین فعالیت هایی می تواند دید بهتری برای برنامه ریزی بر روی کلان داده در کشور ایجاد کند.

۲-۲-۱- داده باز (داده آزاد)

عنوان داده باز (Open Data) به داده هایی اطلاق می شود که به صورت رایگان و آزاد در اختیار همه قرار دارد تا بتوانند از آن استفاده کنند و در هر قالب دلخواهی بازنشر دهند. داده باز دارای محدودیت های مربوط به حق کپی، یا موارد محدودکننده مشابه آن نمی باشد و مفاهیم مشابهی با دیگر جنبش های "باز" نظیر نرم افزارهای آزاد/متن‌باز یا محتوای آزاد دارد. همانطور که در شکل زیر نشان داده شده است، داده های باز اغلب از جنس کلان داده هستند، با این حال مجموعه داده های کوچک نیز می توانند بصورت باز منتشر شوند. به همین دلیل این دو تعریف از هم متمایز هستند.



شکل ۱- ارتباط بین داده باز، کلان داده و دیگر انواع داده

داده باز ضمن اینکه بعد جدیدی از عملیات تحلیلی را به کلان داده اضافه کرده است و سبب افزایش نوآوری در آن شده است، باعث ایجاد مفاهیم جدیدی نظیر دولت باز (Open Government) نیز شده است. در حقیقت داده باز یکی از اصول هشتگانه دولت باز می باشد که در صنایع مختلف مالی، سلامت، حمل و نقل، انرژی، آموزش و ... بطور بالقوه می تواند حدود ۳۰۰۰ تا ۵۰۰۰ میلیارد دلار ارزش اقتصادی تولید کند. بسترسازی برای داده باز اصولا باید توسط دولت ها انجام شود تا بتوان ارزش استخراج شده توسط داده ها را بصورت واقعی مورد استفاده قرار داد. از نمونه بسترهای ایجاد شده در این خصوص می توان به پرتال های داده باز اشاره کرد که در دولت های مختلف اجرا شده است و برخی از آنها در ادامه معرفی شده اند.

جدول ۱- نمونه پرتال های داده باز

State / Org.	Website
Belgium	http://data.gov.be/
Ghana	http://data.gov.gh/
India	http://data.gov.in/
Kenya	https://www.opendata.go.ke/
Morocco	http://data.gov.ma/
Russia	http://opengovdata.ru/
UAE	government.ae/web/guest/uae-data
UK	http://data.gov.uk/
US	http://www.data.gov/
EU	http://open-data.europa.eu/
OECD	http://stats.oecd.org/
UN	http://data.un.org/
UNHCR	http://data.unhcr.org/
World Bank	http://data.worldbank.org/

بررسی مدل اجرای داده باز در دولت های مختلف نشان دهنده الگوهای مختلفی در اجرا می باشد و هر کدام دارای تجربیات خاص خود هستند که بطور جداگانه حائز اهمیت است و باید مورد بررسی قرار بگیرد. برای مثال در دولت فدرال آمریکا این برنامه زیر مجموعه برنامه دولت باز مطرح شده است که خود آن نیز در برنامه FASTER پیگیری می شود، و در مجموع سازوکار منظمی برای اجرا و بهره برداری از آن ایجاد شده است.

۲-۲-۲- رایانش ابری و کلان داده

از رایانش ابری در حال حاضر به عنوان صنعت همگانی پنجم یاد می شود و مسیر توسعه آن به سمتی است که کل زیرساخت های فناوری اطلاعات یک شهر یا کشور را برای ارایه فناوری های جدید (نظیر اینترنت اشیاء، رایانش سیار، خدمات سلامت، شبکه های حسگر، کلان داده) شامل می شود. با حرکت از سمت سیستم های سنتی به سمت رایانش ابری، ملاحظات مربوط به قرارگیری اطلاعات حساس در ابرها اهمیت بیشتری پیدا کرده است و روشهای سنتی تامین امنیت برای این مدل کافی نیستند. در DARPA چندین پروژه برای امن سازی ابرها و توسعه شبکه های ابری در مقیاس بالا در حال اجرا است تا بلوغ کافی برای انجام محاسبات و کار با کلان داده در آنها ایجاد شود. برای مثال می توان به پروژه های MRC، CORONET، CRASH و PROCEED اشاره کرد. پروژه کلان داده دیگری با عنوان PRISM توسط NSA از سال ۲۰۰۷ در حال اجرا است که بر اساس آن این آژانس امنیتی می تواند کلیه داده های ذخیره شده در ابرها و داده های مربوط به ارتباطات اینترنتی آنها را جمع آوری و پردازش کند. از جمله دیگر پروژه های این مرکز، ایجاد مرکز داده Utah بعنوان بزرگترین مرکز جاسوسی داده ها و اطلاعات می باشد و حدود ۲ میلیارد دلار تنها برای ایجاد آن هزینه شده است. از جمله جدیدترین فعالیت های انجام شده که همگرایی رایانش ابری و کلان داده را به خوبی نشان می دهد، می توان به پروژه های PLINY و MUSE در DARPA اشاره کرد که عملا مجموعه همه بسترهای توسعه که تا کنون برای توسعه سیستم های متن باز مورد استفاده قرار می گرفتند، حال برای تولید سیستم-های نرم افزاری بصورت سریع و مطمئن در اهداف مختلف از جمله نظامی مورد استفاده قرار می گیرند. همچنین پروژه کلان داده دیگری با عنوان DCGS-A وجود دارد که برای تحلیل سریع اطلاعات در کاربردهای نظامی مورد استفاده قرار می گیرد و توسط DARPA بصورت ابری در سطح دنیا در حال گسترش است و نمونه اولیه آن در سال ۲۰۱۱ در افغانستان مورد استفاده قرار گرفت. فعالیت های بسیار قابل توجه دیگری برای توسعه زیرساخت های ابری در کشورهای مختلف در سطوح کلان حاکمیتی انجام شده است که با توجه به همگرایی رایانش ابری و کلان داده، ضرورت برنامه ریزی مناسب در این خصوص بسیار حیاتی است.

۲-۲-۳- تنظیم مقررات و رگولاتوری

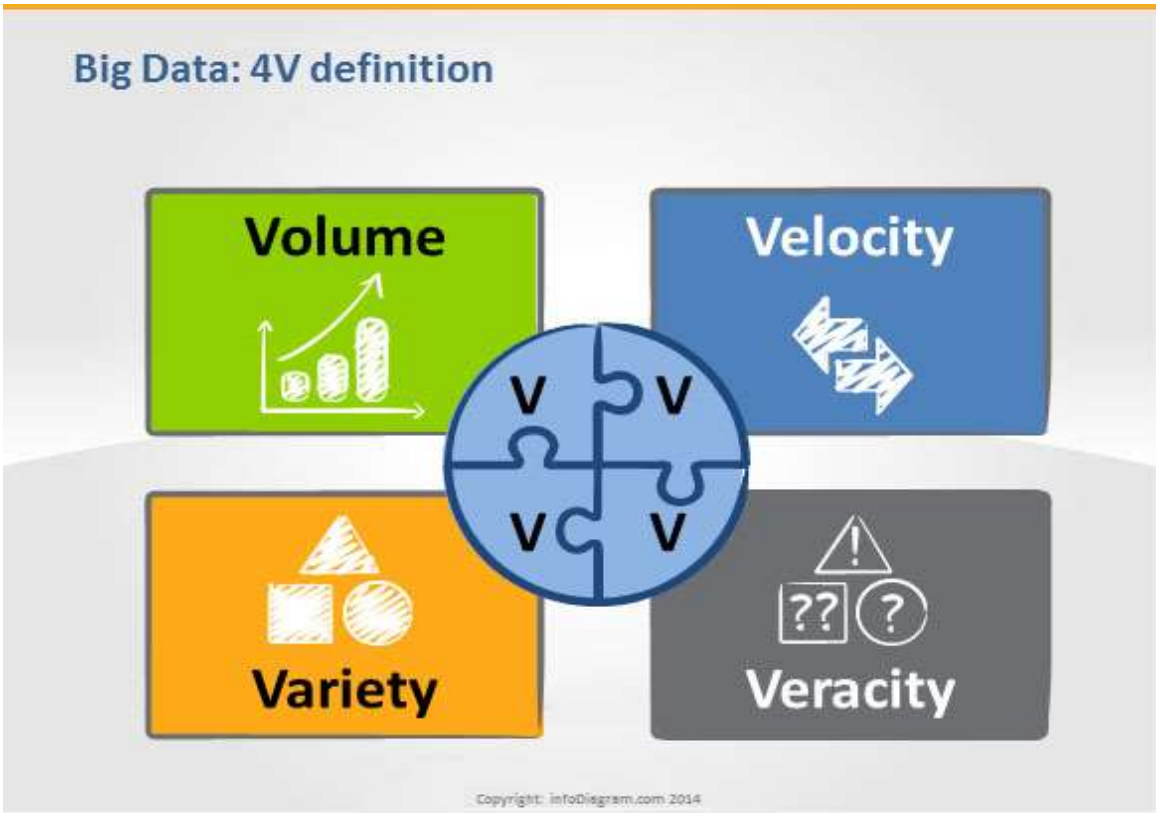
یکی از پروژه هایی که در برنامه داده-دانش-عمل در دولت آمریکا در حال اجرا است، مربوط به برنامه های تنظیم مقررات بیش از ۷۰ آژانس دولت فدرال آمریکا است که با استراتژی باز اجرا می شود، به این شکل که افراد می توانند پیشنهادهای خود را ارسال کنند و دیگران در خصوص آن نظر دهند. در این پروژه تحلیل های قوی بر روی داده انجام می شود تا بینش بیشتری در خصوص آنچه که در متن توضیحات است بدست آید. به این ترتیب که نظر افراد تحلیل می شود و گزارشهای تحلیلی در خصوص نظرهای مشابه، نظرهای تکراری، نظرهای منحصر به فرد، عبارت های رایج مورد استفاده، نحوه توزیع ارایه نظرها در طول زمان و تحلیل های مربوط به روند آنها انجام می شود. این پروژه و دیگر موارد شبیه به آن، نشان دهنده توجه دولت فدرال آمریکا به دریافت بازخورد در بخش های مختلف جهت بهبود مستمر خدمات دولت است که با توجه به ضعف فناوری های پیشین، در صدد بهبود آنها با بکارگیری مفهوم کلان داده است. پس یکی از مفاهیم نهفته در کلان داده، **تلاش برای بهبود وضعیت است** و این

امر زمانی محقق می شود که سازوکار مناسب برای تبدیل دانش بدست آمده به عمل نیز ایجاد شده باشد. به عبارت دیگر در کلان داده زمانیکه داده بیشتری وجود داشته باشد انتظار می رود که بینش بیشتری نیز حاصل شود.

۲-۳- جمع بندی وضعیت کلان داده در دنیا

جامعه ای را تصور کنید که در آن جمعیت بطور نمایی در حال افزایش است، اما خدمات و زیرساخت های عمومی آن نتواند پاسخگوی رشد جمعیت باشد و از عهده مدیریت آن برآید. چنین شرایطی در حوزه داده در حال وقوع است. بنابراین نیازمند توسعه زیرساخت های فنی برای مدیریت داده و رشد آن در بخش هایی نظیر جمع آوری، ذخیره سازی، جستجو، به اشتراک گذاری و تحلیل می باشیم. چنین مسائلی دارای مشخصه هایی هستند که می بایست روی آنها تمرکز کنیم. شکل ۲ ابعاد عمومی مسائل کلان داده در دنیا را نشان می دهد البته ابعاد دیگری نیز وجود دارد که بصورت موردی در برخی مسائل مطرح میشود:

- **حجم داده (Volume):** حجم داده ها به صورت نمایی در حال رشد می باشد. منابع مختلفی نظیر شبکه های اجتماعی، لاگ سرورهای وب، جریان های ترافیک، تصاویر ماهواره ای، جریان های صوتی، تراکنش های بانکی، محتوای صفحات وب، اسناد دولتی و ... وجود دارد که حجم داده بسیار زیادی تولید می کنند.
- **نرخ تولید (Velocity):** داده ها از طریق برنامه های کاربردی و سنسورهای بسیار زیادی که در محیط وجود دارند با سرعت بسیار زیاد و به صورت بلادرنگ تولید می شوند. بسیاری از کاربردها نیاز دارند به محض ورود داده به درخواست کاربر پاسخ دهند. ممکن است در برخی موارد نتوانیم به اندازه کافی صبر کنیم تا مثلا یک گزارش در سیستم برای مدت طولانی پردازش شود.



شکل ۲- ابعاد عمومی مسائل کلان داده

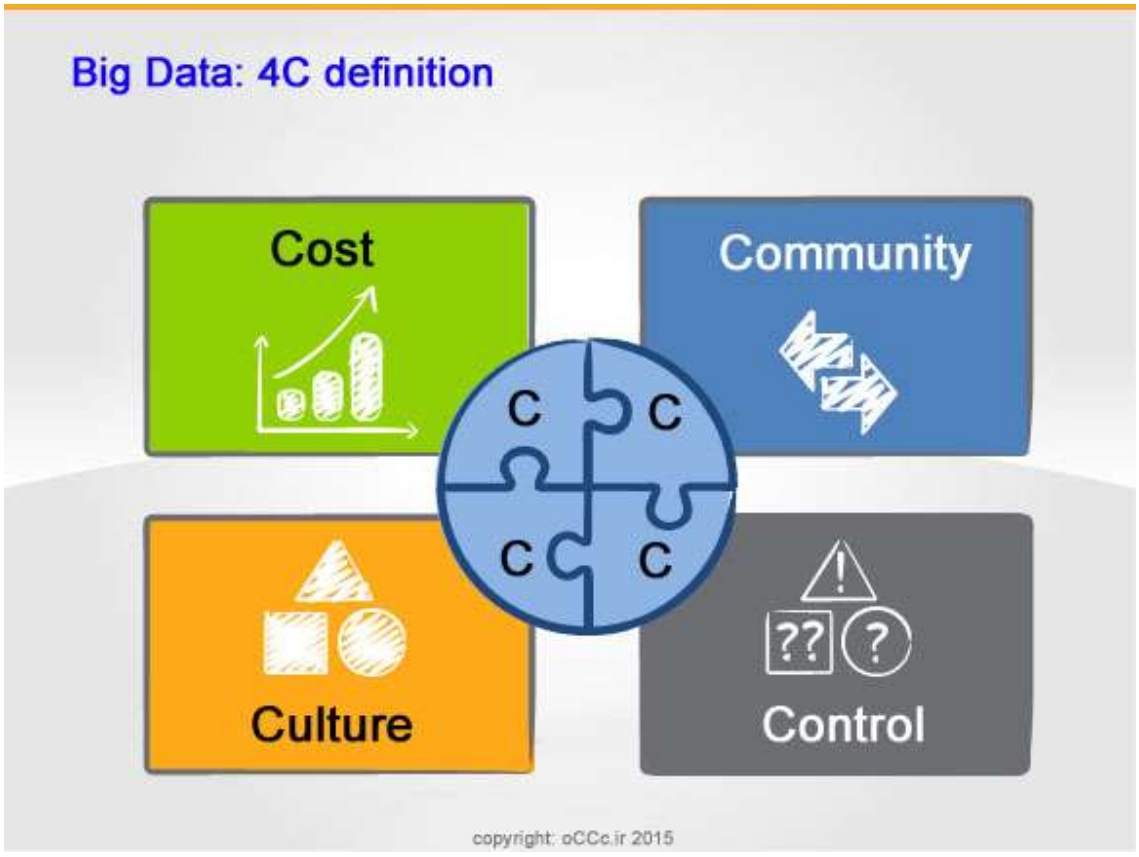
- **تنوع (Variety)**: انواع منابع داده و تنوع در نوع داده بسیار زیاد می باشد که در نتیجه ساختارهای داده ای بسیار زیادی وجود دارد. مثلا در وب، افراد از نرم افزارها و مرورگرهای مختلفی برای ارسال اطلاعات استفاده می کنند. بسیاری از اطلاعات مستقیما از انسان دریافت میشود و بنابراین وجود خطا اجتناب ناپذیر است. این تنوع سبب میشود جامعیت داده تحت تاثیر قرار بگیرد. زیرا هرچه تنوع بیشتری وجود داشته باشد، احتمال بروز خطای بیشتری نیز وجود خواهد داشت.
- **صحت (Veracity)**: با توجه به اینکه داده ها از منابع مختلف دریافت میشوند، ممکن است نتوان به همه آنها اعتماد کرد. مثلا در یک شبکه اجتماعی، ممکن است نظرهای زیادی در خصوص یک موضوع خاص ارائه شود. اما اینکه آیا همه آنها صحیح و قابل اطمینان هستند، موضوعی است که نمیتوان به سادگی از کنار آن در حجم بسیار زیادی از اطلاعات گذشت. البته بعضی از تحقیقات صحت را به معنای حفظ همه مشخصه های داده اصلی بیان کرده اند که باید حفظ شود تا بتوان کیفیت و صحت داده را تضمین کرد. البته تعریف دوم در مولدهای کلان داده صدق میکند تا بتوان داده ای تولید کرد که نشان دهنده ویژگی های داده اصلی باشد.
- **اعتبار (Validity)**: با فرض اینکه دیتا صحیح باشد، ممکن است برای برخی کاربردها مناسب نباشد یا به عبارت دیگر از اعتبار کافی برای استفاده در برخی از کاربردها برخوردار نباشد.
- **نوسان (Volatility)**: سرعت تغییر ارزش داده های مختلف در طول زمان میتواند متفاوت باشد. در یک سیستم معمولی تجارت الکترونیک، سرعت نوسان داده ها زیاد نیست و ممکن است داده های موجود مثلا برای یک سال ارزش خود را حفظ کنند، اما در کاربردهایی نظیر تحلیل ارز و بورس، داده با نوسان زیادی مواجه هستند و داده ها به سرعت ارزش خود را از دست میدهند و مقادیر جدیدی به خود می گیرند. اگرچه نگهداری اطلاعات در زمان طولانی به منظور تحلیل تغییرات و نوسان داده ها حائز اهمیت است. افزایش دوره نگهداری اطلاعات، مسلما هزینه های پیاده سازی زیادی را دربر خواهد داشت که باید در نظر گرفته شود.
- **نمایش (Visualization)**: یکی از کارهای مشکل در حوزه کلان داده، نمایش اطلاعات است. اینکه بخواهیم کاری کنیم که حجم عظیم اطلاعات با ارتباطات پیچیده، به خوبی قابل فهم و قابل مطالعه باشد از طریق روش های تحلیلی و بصری سازی مناسب اطلاعات امکان پذیری است.
- **ارزش (Value)**: این موضوع دلالت بر این دارد که از نظر اطلاعاتی برای تصمیم گیری چقدر داده حائز ارزش است. بعبارت دیگر آیا هزینه ای که برای نگهداری داده و پردازش آنها میشود، ارزش آن را از نظر تصمیم گیری دارد یا نه. معمولا داده ها میتوانند در لایه های مختلف جابجا شوند. لایه های بالاتر به معنای ارزش بیشتر داده می باشند. بنابراین برخی از سازمانها میتوانند هزینه بالای نگهداری مربوط به لایه های بالاتر را قبول کنند.

این مدل به این معناست که مثلا "هنگامی که با اطلاعات بیشتری در حوزه سلامت مواجه باشیم، با بازدهی بیشتری سلامت را ارتقا دهیم"، "در شرایطی که خطرات امنیتی افزایش پیدا میکند، سطح امنیت بیشتری را فراهم کنیم"، "وقتی که با رویدادهای بیشتری از نظر آب و هوایی مواجه باشیم، توان پیش بینی دقیقتر و بهتری بدست آوریم"، "در دنیایی با خودروهای بیشتر، آمار تصادفات و حوادث را کاهش دهیم"، "تعداد تراکنش های بانکی، بیمه و مالی افزایش پیدا کند، ولی تقلب کمتری را شاهد باشیم"، "با منابع طبیعی کمتر، به انرژی بیشتر و ارزانتری دسترسی داشته باشیم" و بسیاری موارد دیگر از این قبیل که اهمیت پنهان کلان داده را نشان می دهد.

۳- وضعیت کلان داده در ایران

از نظر فلسفی در حال حاضر در کشور با ضعف زنجیره فناوری اطلاعات مواجه هستیم. بطور خاص در بحث اینترنت، با سرعت و کیفیت پایین، نبود SLA، ضعف رگولاتوری، فقدان محتوای قابل عرضه و گردش ترافیک اطلاعاتی و داده ها به سمت بیرون مرزها میتوان اشاره کرد. بحث کلان داده در حالی در کشور مطرح شده است که هنوز روش های سنتی تحلیل داده در سلسله مراتب تصمیم گیری مورد استفاده قرار نگرفته است و حتی در بسیاری از بخش ها با نبود سیستم های اطلاعاتی، نبود شفافیت، و بطور خلاصه با ناتوانی در مدیریت اطلاعات مواجه هستیم. این مشکلات در مدلی با عنوان 4C نشان داده شده است:

- **Cost (هزینه):** نگاه به داده به عنوان یک منبع هزینه بجای یک منبع ارزش
- **Culture (فرهنگ):** نگاه ابزاری به مساله و استفاده از راه کارهای کلان داده برای حل مجدد همان مسائل سنتی
- **Control (کنترل):** ناتوانی در بکارگیری سیستم های اطلاعاتی و تحلیل داده در سازمان و وجود رویکردهای سنتی فناوری اطلاعات
- **Community (جامعه):** ضعف زنجیره فناوری اطلاعات برای مدیریت فناوری و چرخه عمر آن



شکل ۳- ابعاد عمومی مسائل کلان داده در ایران

در حقیقت این مدل بیان میکند که بجای اینکه ما با **مسائل "کلان داده"** مواجه باشیم، با **"مسائل کلان" داده** مواجه هستیم. این موضوع سبب شده است که بحث کلان داده به موضوعی برای طرح همان مسائل سنتی با بکارگیری فناوری ها و

ابزارهای مدرن تبدیل شود. به این ترتیب مشاهده میشود که حتی مسائلی ساده نظیر اتوماسیون یک فرآیند سازمانی یا بکارگیری ابزارهای هوش تجاری جهت انجام تحلیل در سازمان نیز بعنوان مسائلی کلان جلوه میکنند و تلاش برای حل آنها بعنوان نمونه هایی از مسائل کلان داده مطرح میشود.

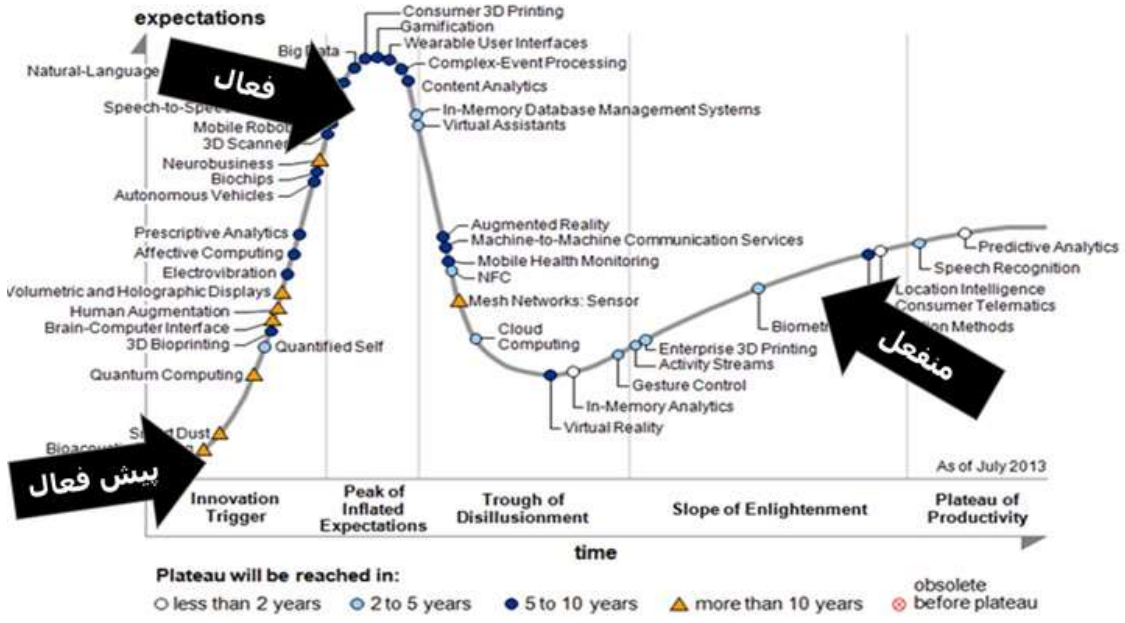
در ادامه نگاهی می اندازیم به وضعیت کلان داده در کشور در بخش های مختلف:

- **وضعیت در بخش کاربردها و راهکارهای ذخیره و مدیریت کلان داده**
  - ضعف دانش بومی در حوزه توسعه فناوری و وابستگی عمده به ابزارها
  - ضعف صنعت فناوری اطلاعات کشور در حوزه تولید فناوری های نرم افزاری جدید
  - عدم وجود تجربه موفق در حوزه بومی سازی در کشور
  - بلوغ پایین سیستم های اطلاعاتی در سازمان ها
  - نبود بسترهای اولیه برای دسترسی به داده های موجود (نظیر پرتال های داده باز)
  - محدودیت در دسترسی به داده ها
  - ضعیف بودن جوامع کاربری فعال در حوزه کلان داده
  - ضعف در حوزه تولید محتوا
- **وضعیت در بخش سرمایه گذاری، راهبرد و الگوهای مهاجرت به کلان داده**
  - نبود نظام مدیریت هماهنگ برای توسعه فناوری
  - نبود فرهنگ شفافیت اطلاعاتی
  - نبود فرهنگ به اشتراک گذاری داده ها
  - نبود الگوهای موفق در حوزه مدیریت چرخه فناوری در کشور
  - ریسک بالای سرمایه گذاری
- **وضعیت در بخش زیرساخت ها و بسترهای کلان داده**
  - ضعف زنجیره فناوری اطلاعات
  - ضعف در حوزه ارتباطات شبکه
  - ضعف در حوزه مدیریت انرژی در حوزه فناوری اطلاعات
  - فقدان سرویس دهنده های ابری با بلوغ مناسب
- وضعیت در بخش امنیت، مسائل حقوقی و استانداردهای کلان داده
  - ضعف رگولاتوری
  - نبود متولی مشخص برای مدیریت فناوری
  - عدم بلوغ جامعه کاربری برای تدوین مسائل حقوقی و استانداردها

۴- چشم انداز کلان داده در دنیا و در کشور ایران

با پیش بینی از روند برنامه ریزی یک مجموعه اغلب میتوان چشم اندازی را متصور شد که در آینده ای نزدیک محقق شود. آنچه در شکل زیر نشان داده شده است، نموداری است از موسسه تحقیقات گارتنر که بصورت مستمر وضعیت بلوغ فناوری های

مختلف را گزارش می کند. در محور افقی وضعیت بلوغ فناوری-ها نشان داده می شود که پس از ظهور و گذر از سطح انتظارات مختلف، اگر مورد پذیرش واقع شوند، به مرحله تجاری سازی رسیده و مورد استفاده قرار می گیرند. در محور عمودی سطح انتظارات نشان داده می شود که در ابتدای معرفی فناوری بصورت کاذب اوج می گیرد و پس از آن حباب شکسته می شود و سطح انتظارات به شدت سقوط می کند، و در مرحله ای دیگر بصورت واقع بینانه مورد استفاده قرار می گیرد.



شکل ۴- مدل برنامه ریزی مجموعه های توسعه یافته و در حال توسعه بر روی منحنی بلوغ فناوری ها

موضوعی که در اینجا برای ما حائز اهمیت است، نگاشت مدل برنامه ریزی یک مجموعه (شامل سازمان های دولتی و یا شرکتهای تجاری) نسبت به توسعه فناوری است. مجموعه های توسعه یافته معمولا نقشی پیش فعال دارند و با ارائه دستاوردهای جدید خود، فناوری های را وارد این چرخه می کنند که به عنوان فناوری های نوظهور مطرح می شود. اما مجموعه های در حال توسعه، بدلیل فقدان نوآوری اغلب در دو نقطه دیگر از این نمودار قرار دارند. آن هایی که فعال تر هستند، اغلب در زمان اوج گرفتن انتظارات سعی می کنند با فناوری هایی که بیشتر مطرح هستند آشنا شوند و با صرف بودجه های تحقیقاتی نسبت به شناخت فناوری اقدام می کنند. اما مجموعه های منفعل تنها مصرف کننده هستند و وقتی فناوری با بلوغ خوبی به مرحله تجاری سازی رسید تنها از آن استفاده می-کنند.

این مدل به خوبی نشان می دهد که مدل سنتی تحقیقات در کشورهای در حال توسعه، اصولا نمی تواند پیشرفت آنها را در عرصه نوآوری و تولید فناوری تضمین کند و به مرور توان رقابتی آنها کاهش می یابد. زیرا پیوسته در حال شناخت موضوعات جدیدی هستند که با شتاب زیادی مطرح می شوند و فرصتی برای نوآوری و تولید بدست نمی آورند و اگر هم چنین فرصتی دست دهد، فاقد بلوغ کافی و یا صرفه اقتصادی خواهد بود.

۴-۱- چشم اندازه کلان داده در دنیا

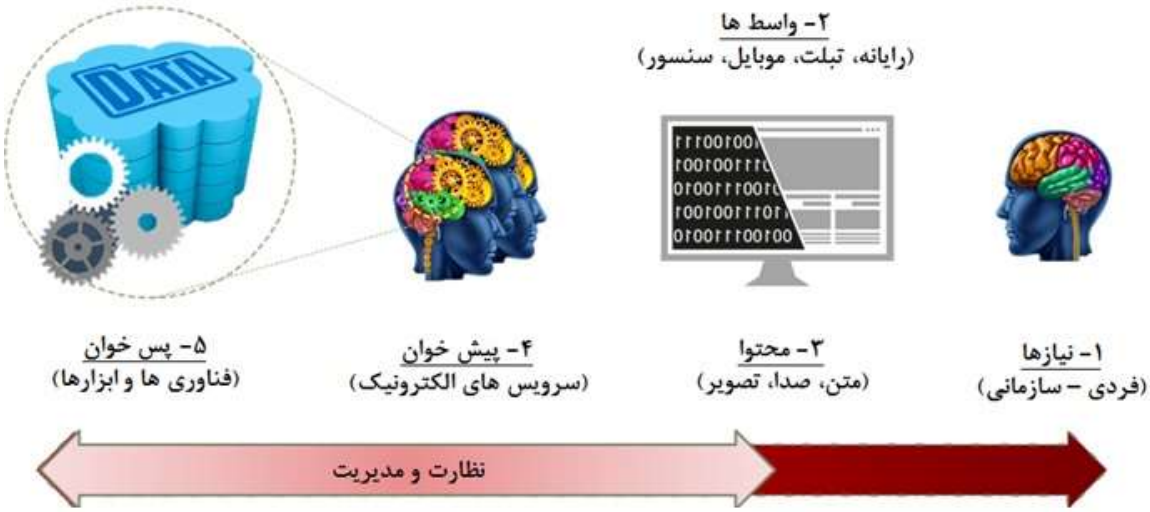
روند توسعه پیشرفت های حوزه فناوری اطلاعات نشان می دهد که دنیای ما در حال تبدیل به یک محیط فراگیر از نظر ارتباطات و محاسبات است. این حرکت باعث گسترش کاربرد فناوری اطلاعات در صنایع مختلف و به وجود آمدن فرصتها و مزایایی

شده که زندگی روزمره ما را نسبت به گذشته بهبود خواهد بخشید. ولی از طرفی دیگر مسائل جدیدی مطرح خواهد شد. کلان داده ماهیت ذاتی این محیط رایانشی فراگیر خواهد بود. از سویی دیگر، کلان داده موجب استخراج دانش نهفته در داده های موجود می شود که به واسطه آن رشد و شتاب بیشتری در نوآوری و توسعه علم و فناوری در حوزه های مختلف انتظار می رود. با توجه به ماهیت موضوع کلان داده، نوآوری و پیشرفت را در دو حوزه می توان انتظار داشت:

- **توسعه فناوری های مرتبط با کلان داده:** در این بخش شاهد ارایه تکنیک ها، الگوریتم ها و ابزارهای بیشتری برای پوشش دادن طیف بیشتری از مسائل خواهیم بود.
- **توسعه کاربردهای مرتبط با کلان داده:** در این بخش شاهد نوآوری و پیشرفت در حوزه های مختلف زندگی بشر (سلامت، حمل و نقل، کشاورزی، بانک، بیمه، نظامی، آموزش و ...) خواهیم بود.

پس علاوه بر ظهور استارتاپ های جدید در حوزه کلان داده، میتوان کلان داده را موتور محرکت بسیاری از فناوری ها در نظر گرفت. برای مثال پذیرش "رایانش ابری" در دنیا با شروع مباحث کلان داده، شتاب بیشتری به خود گرفته است و با ظهور "اینترنت اشیاء"، شاهد گسترش هر چه بیشتر مفهوم کلان داده خواهیم بود و باید انتظار ارایه پدیدهای جدیدتری را نیز در ادامه داشته باشیم. حال اگر کمی فلسفی به روند پیشرفت فعلی نگاه کنیم، ابعاد دیگری از این چشم انداز نیز نمایان خواهد شد. توسعه فناوری اطلاعات نشان دهنده وجود پنج بعد اصلی است که در شکل ۵ ارتباط بین آنها نشان داده شده است:

- نیازها (زندگی فیزیکی)
- واسط ها
- محتوا
- کاربردها و خدمات (پیش خوان فناوری)
- بستر ها و زیرساخت ها (پس خوان فناوری)



شکل ۵- ابعاد برنامه ریزی در حوزه فناوری اطلاعات

همانطور که در تصویر نشان داده شده است، هرچه توان بیشتری در تولید فناوری های مربوط به این حوزه (خصوص بخش های مربوط به پس خوان فناوری و نیز واسط های ارتباطی) وجود داشته باشد، قدرت بیشتری در دخل و تصرف وجود دارد که

اغلب این کار با تلاش بر نفوذ بیشتر در بازار و تدوین استانداردها همراه است. بنابراین مشاهده می شود که در دولت های توسعه یافته نظیر آمریکا (بطور خاص در بحث فناوری اطلاعات) برنامه ریزی های کلانی در این خصوص در حال انجام است. برای مثال میتوان به کلان برنامه NITRD و یا پروژه FASTER و یا پروژه های دیگری نظیر Pliny و MUSE که به منظور افزایش شتاب در توسعه فناوری از ابعاد مختلف در حال انجام است اشاره کرد. همچنین مشاهده می شود که مجموعه هایی نظیر NIST هماهنگ با سایر مجموعه ها نظیر ISO و ITU-T در تدوین استانداردهای مرتبط با فناوری های نوظهور پیشگام هستند. این تلاش سبب افزایش نظارت و مدیریت دولت های توسعه یافته بر حوزه های مختلف فناوری می شود و با ترویج مباحث جدید نظیر رایانش ابری، اینترنت اشیا ، رایانش سیار ، رایانش فراگیر و مباحثی از این قبیل، قادر خواهند بود بیشترین سطح کنترل را در آن بدست آورند. کامل بدیهی است که با توجه به همگرایی علوم و فناوری ها از یک طرف و از طرف دیگر نفوذ فناوری اطلاعات در ابعاد مختلف زندگی بصورت واقعی (نظیر تاثیرگذاری رسانه ها و شبکه های اجتماعی در خانواده و جامعه، نظریه بازی ها در اقتصاد و سیاست، هوش تجاری در کسب و کار، رایانش ابری در ارائه خدمات فناوری اطلاعات سازمان ها و دولت ها، نفوذ سنسورهای مختلف در زندگی روزمره، حتی حوزه سلامت و ...) این نظارت و مدیریت قابل گسترش به زندگی حقیقی نیز می باشد. پس امکان کنترل دنیای فیزیکی به واسطه افزایش توان مدیریت فضای مجازی، انگیزه بالایی برای راهبری فناوری اطلاعات و استانداردسازی آن توسط دولت های توسعه یافته ایجاد می کند.

۴-۲- چشم انداز کلان داده در ایران

کشور ایران در زمره کشورهای در حال توسعه ای است که سرمایه های داده ای میتواند به عنوان یک فرصت برای بخش های مختلف در آن مورد استفاده قرار بگیرد. بر اساس برنامه ها و پروژه های ملی در حوزه فناوری اطلاعات و ارتباطات نظیر شبکه ملی اطلاعات و توسعه دولت الکترونیک، گذر به نظام آدرس دهی IPv6، توسعه دسترسی پهن باند و مهاجرت اپراتورهای همراه به نسل های جدید و همچنین عرضه روز افزون تجهیزات هوشمند در بازار ایران، تحقق نسبی اینترنت اشیاء در کشور و ضریب نفوذ ۱۰۰ درصد پهنای باند برای کاربران انسانی تا دو سال آینده دور از ذهن نیست. با تحقق نسبی اینترنت اشیاء و اجرای پروژه های ملی در حوزه های اقتصادی صنعتی و خدمات عمومی مبتنی بر فناوری اطلاعات و ارتباطات و ظهور رسانه ها و شبکه های اجتماعی بومی تا پایان برنامه ششم توسعه (سال ۲۰۲۰ میلادی)، کشور در آستانه تولید داخلی داده های ناشی از بهره برداری از برنامه ها و پروژه های فوق تا ۵ اگزابایت قرار خواهد داد. این در حالی است که مسیری که بطور عام در ایران در حوزه توسعه علم و فناوری طی میشود، با در نظر گرفتن شکاف فناوری (از نظر نوآوری و تولید فناوری در حوزه فناوری اطلاعات) چشم انداز مطلوبی را ترسیم نمیکند. زیرا استراتژی قالب بر بخش های مختلف مبتنی بر یکی از دو حالت فعال یا پیش فعال می باشد که پیشرفت کشور را تضمین نمیکند. پس باید سعی شود پس از طرح تصویری کلان از شرایط مطلوب، نسبت به برنامه ریزی مناسب برای اصلاح مسیر فعلی گام برداشت.

۵- چالش های حوزه کلان داده در دنیا و در کشور ایران

۵-۱- چالش های فنی و زیرساختی

- پایگاه داده های قدیمی و رابطه ای پاسخگوی نیازهای داده های عظیم نیستند
- کند بودن دسترسی به داده ها

- سربارهای پایگاه داده های رابطه ای
- جمع آوری داده های عظیم
- پردازش داده های عظیم
- به واسطه ظهور داده های عظیم الگوی ترافیک شبکه ها در حال تغییر است:
- الگوهایی مانند ترافیک های انفجاری در حال افزایش ااست
- نیاز به تجهیزات مقتضی برای زیرساخت ارتباطی
- نیاز به مدیریت پویا تر شبکه (مانند ساختار های SDN)
- بهینه سازی ترافیک داده
- کاستی ابزاری در داده های عظیم:
- ابزارهای موجود مانند هادوپ، در عین اینکه استفاده زیادی دارند، دارای مشکلات فراوانی نیز هستند، برای مثال دربرخی موارد مانند سرعت پردازش کارآمدی لازم را ندارند
- ابزارهای موجود برای کارایی لازم را برای تمامی امور ندارند
- برقراری امنیت با توجه به حجم زیاد داده ها و تراکنش ها
- حفظ محرمانگی
- تشخیص و جلوگیری از نفوذ
- کنترل دسترسی
- پراکنده‌سازی در توزیع داده‌های بزرگ
- ایجاد گمنامی در پایگاه داده‌ها
- وجود چالش ها در زمینه الگوریتم های پردازش و تحلیل اطلاعات
- وجود چالش در سخت افزار پردازش و ذخیره سازی
- توزیع جغرافیایی
- همزمان سازی داده ها

۵-۲- چالش های تحقیقاتی، مدیریتی و حقوقی

- تحقیقات و توسعه در این حوزه با توجه به چند وجهی بودن و پیچیدگی دانش مرتبط
- مشخص نبودن میزان نسبی آمادگی کشور در مواجهه با پدیده داده های عظیم
- برنامه ریزی و اجرای آن برای مواجهه با داده ای عظیم
- نحوه پیاده سازی صحیح و هدفمند داده های عظیم در کمترین زمان ممکن در کشورهای در حال توسعه
- استفاده از شبکه های اجتماعی و داده های عظیم تولید شده در آن برای تجارت
- سرمایه گذاری کلان در بخش داده های عظیم و جذب سرمایه
- در فضایی که داده های عظیم مطرح می شود، شرکت ها چگونه باید با آن برخورد کنند
- تولید حجم زیاد داده در کشورها در حالی که بسیاری هنوز آمادگی و دانش مدیریت آن را ندارند
- کاربردهای داده های عظیم هنوز به درستی شناخته و استفاده نشده است

- داده ها در حال افزایش اند و به صورت داده های کلان تبدیل شده اند . استفاده از این داده ها می تواند مزایای فراوانی را در بر داشته باشد، اما:
- کمبود امکانات زیرساختی وجود دارد و ذخیره این داده ها به خودی خود یک مسئله قابل توج است
- هزینه بالای تحقیقات و تأمین زیرساخت های لازم برای داده های عظیم
- ارزش واقعی داده های عظیم درک نشده و لذا برنامه ای برای بهره برداری و استفاده از آن وجود ندارد
- الگوی مشخصی از تولید و استفاده کلان داده وجود ندارد
- عدم بلوغ مدلهای داده های عظیم متناسب با مدلهای کسب و کار موجود
- امکان عدم بازگشت سرمایه
- تأمین هزینه:
- هزینه بالای نیروی انسانی متخصص و تجهیزات
- مشکل جذب سرمایه گذار
- حفظ حریم شخصی افراد
- نبود و یا نقص قوانین
- قوانین مالکیت معنوی و حقوق مالکیت فکری
- قوانین مربوط به تجارت الکترونیک
- قوانین جرائم و تخلفات سایبری
- در حوزه مصادیق
- در حوزه برخورد
- لایسنس کردن محصولات توسعه یافته در کشور
- توسعه محصولات داخلی بر پایه محصولاتی با لایسنس متن باز (مانند GPL، BSDF و انواع مشابه) بدون داشتن دانش و شناخت کافی از این لایسنس ها

۵-۳- سایر چالش ها

- عدم وجود تجربه موفق در حوزه بومی سازی در کشور
- عدم بلوغ کافی در این حوزه در اکوسیستم جهانی (اکوسیستم در حال شکل گیری است)
- عدم وجود انگیزه و آگاهی لازم در شرکت ها و سازمانها برای استفاده از داده های عظیم
- ضعیف بودن رزومه کاری فعالان در حوزه داده های عظیم
- وابستگی به فناوری های وارداتی و عدم بلوغ آنها (تحمیل هزینه و پیچیدگی زیاد)
- سرمایه گذاری های کلان، مدیریت راهبردی دقیق و منظم و با ریسک بالا
- عدم راه اندازی و احتمال شکست پروژه
- در دسترس نبودن داده (NoData Problem)

۶- برنامه ریزی در حوزه کلان داده

۶-۱- سیاست های کلی و راهبردهای کشور ایران در حوزه کلان داده

- در نظر گرفتن داده های عظیم به عنوان سرمایه
- سرمایه گذاری در این حوزه
- هدایت ترافیک داده ها به سمت داخل کشور
- مدیریت و هدایت داده های تولیدی در کشور
- سرمایه گذاری و بهره برداری از داده های عظیم
- طراحی و ساخت زیرساخت های لازم (حداقل در حوزه نرم افزار) به صورت کاملاً بومی
- حمایت از نخبگان جهت ایجاد و حفظ متخصصان داخلی
- برنامه ریزی برای تحقیقات در حوزه داده های عظیم با رویکرد کاربرد گرایانه

۶-۲- چارچوب برنامه اجرایی پیشنهادی

- نیاز سنجی و شناخت کاربردها ی داد های عظیم در بخش های مختلف کشور
- تحقیقات در حوزه داده های عظیم در راستای نیازسنجی های انجام شده
- ایجاد بستر ارتباطی و استفاده از بسترهای موجودی برای تعامل اجزای اکوسیستم داده های عظیم در ایران
- جذب سرمایه گذاری و ایجاد بستر کاری مشترک میان دولت و بخش خصوصی در این زمینه
- ایجاد کارگروه هایی تخصصی تر جهت برگزاری جلسات منظم و انجام ارایه های تخصصی توسط عوامل درگیر جهت تبادل نظر و رسیدن به یک زبان مشترک: ارایه های تخصصی برای اینکه اثربخش باشند نباید محدود به یک همایش باشد و بجای آن باید بصورت مستمر در جلسات هم اندیشی وجود داشته باشد. در جلسات هم اندیشی هم نباید صرفا مباحثات پراکنده انجام شود و باید فرآیندی جهت ثبت و پردازش مباحثات بصورت مستمر وجود داشته باشد.
- شناخت مدل راهبری فناوری های حوزه کلان داده و ارتباط آن با دیگر فناوری ها در کشورهای دیگر (در سطوح آکادمیک / صنعتی و حاکمیتی)
- شناسایی شرکت های فعال در حوزه کلان داده و شناخت مسائل و نیازمندی ها
- اولویت دهی محورها بر اساس کاربردهای مطرح در کشور
- لزوم توجیه مسئولین و مدیران ارشد و میانی شرکت های بزرگ و سازمانها
- در دسترس قرار گرفتن محتوای مربوط به این حوزه برای محققین و دانشجویان
- لزوم برگزاری جلسات هدف مند و مستندسازی نتایج

۶-۳- چشم انداز تکالیف و مأموریت های دولت در حوزه کلان داده

- ایجاد اکوسیستم بومی کلان داده (استانداردسازی شامل ایجاد فهم مشترک، شناخت نیازمندیها و ایجاد مدل مرجع و تکسونومی کلان داده و ایجاد نقشه راه فناوری در کشور در حوزه کلان داده)

- رویکرد اثرگذاری به جای اثرپذیری با شروع از ایجاد اکوسیستم
- ایجاد ارتباط بین عوامل مؤثر در چرخه (صنعت، دانشگاه و کاربران پایانی و...) بر اساس مدل دانش آزاد
- شناخت نیازهای واقعی، ایجاد زمینه های پژوهشی و برنامه ریزی برای تولید محصولات مورد نیاز
- حمایت از پروژه ها و پایان نامه های این حوزه
- دولت خود مصرف کننده داده های عظیم است، پس نیاز به سرمایه گذاری در این بخش دارد
- دولت در ارائه داده های عظیم به عنوان یک سرویس و خدمت نقش دارد
- با توجه به بحث های مختلف مطرح شده در همایش، دولت می بایست بخش خصوصی را هم شریک کند و از قابلیت های آن در این حوزه استفاده نماید
- بررسی ارتباط موضوع با حوزه شرکت های دانش بنیان و چگونگی ورود به این حوزه
- بررسی ارتباط موضوع با نهادهای مختلف نظیر شورای عتف و نحوه اثرگذاری
- لزوم بررسی مسایل مرتبط با امنیت (هم از بعد چالشی خود کلان داده نظیر حریم خصوصی و هم از بعد کاربرد کلان داده نظیر کشف تقلب و کشف جرم و ...) از منظر قانون گذاری و تعریف فرآیندهای مرتبط در حوزه قضائی
- برنامه ریزی به سمت دانش آزاد: ایجاد فرآیندهای مورد نیاز برای تعامل صنعت و دانشگاه برای ایجاد کارگروهای مختلف داده های عظیم بر اساس مدلهای open data و جامعه آزاد (برای جلوگیری از انحصار). مثال: http://data.gov (پلتفرم دسترسی به مجموعه داده های صنایع مختلف)
- بهره گیری از فرصتهای موجود:
  - مثال اجرایی: همگرایی با برنامه قطب فناوری‌های مرتبط با رایانش ابری در مرکز تحقیقات مخابرات جهت حمایت از توسعه راهکارهای بومی
  - ارتقا و تأمین زیرساختهای مورد نیاز (بسترهای ارتباطی، ذخیره سازی و پردازشی و ...)
  - تجهیز آزمایشگاهها و فراهم کردن بستر سخت افزاری مناسب در این زمینه
  - حمایت از محققین داخلی

۶-۴- چشم انداز مأموریت دانشگاه ها و مراکز تحقیقاتی در حوزه کلان داده

- تلاش برای ارائه مدل های جدید برای حل مسائل داخلی
- آسیب شناسی توسعه فناوری در کشور از طریق شناخت تجربیات و الگوهای موفق، شناخت تجربیات و الگوهای ناموفق به منظور بهره برداری در حوزه داده های عظیم
- ایجاد سازوکاری جهت اندازه گیری میزان ارزش استخراج شده از داده های موجود در کشور با روش های سنتی فعلی و روش های مدرن پیشنهادی
- لزوم بررسی مسایل مرتبط با امنیت (هم از بعد چالشی خود داده های عظیم نظیر حریم خصوصی و هم از بعد کاربرد داده های عظیم نظیر کشف تقلب و کشف جرم و ...) از نظر فنی
- آینده پژوهی (مثال اجرایی: برنامه رصد خانه فناوری اطلاعات)

۶-۵- چشم انداز مأموریت صنایع، شرکت های دانش بنیان و واحدهای خصوصی در حوزه کلان داده

- تلاش برای افزایش توان داخل جهت حل مسائل مرتبط با داده
- تولید و صادرات فناوری های مرتبط با کلان داده در سطح منطقه
- برقراری ارتباط موثر با مراکز تحقیقاتی و دانشگاهی برای حفظ پویایی
- ارایه تجربیات موفق جهت بهبود اکوسیستم

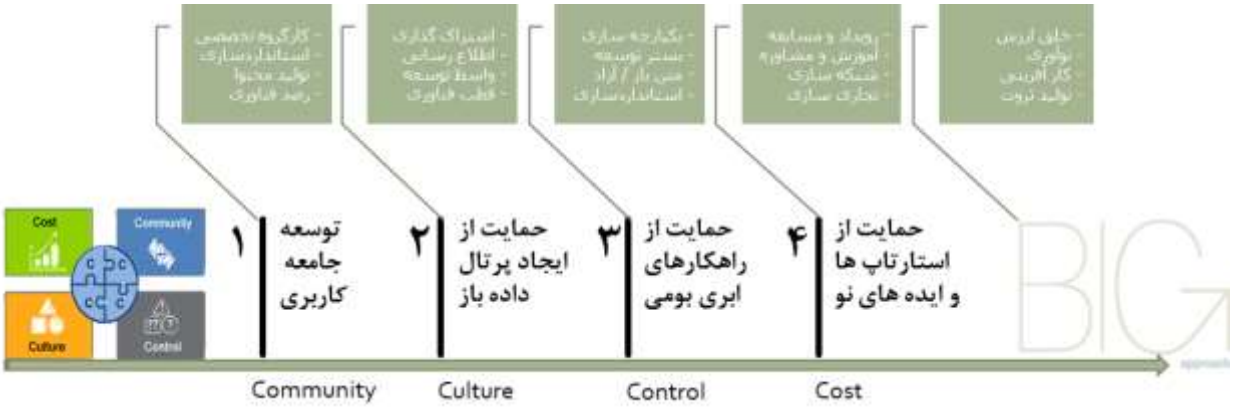
۷- نقشه راه

وضعیت فعلی کلان داده در کشور که بصورت مدلی با عنوان 4C گردید، توصیفی بر موانع اصلی توسعه فناوری های مرتبط با کلان داده در کشور می‌باشد. با توجه به اینکه این موانع عمدتا از جنس فرهنگی می‌باشند، می‌بایست در برنامه نقشه راه، راهبردی اتخاذ شود که سبب رفع تدریجی این موانع فرهنگی شده و به تدریج شرایط برای نوآوری، خلق ارزش و تولید ثروت فراهم شود. شمایی از نظام هماهنگ مورد انتظار در حوزه کلان داده در شکل زیر نشان داده شده است که بر اساس آن، شرایطی فراهم می‌شود که توسعه و توانمندسازی بر اساس نیاز واقعی در خود صنعت صورت پذیرد.



شکل ۶- شمایی از وضعیت مورد انتظار

همانطور که نشان داده شده است، در این نقشه بر روی مباحثی از قبیل کسب توانمندی در تولید ابزارها و فناوری های مرتبط با کلان داده، مدیریت ترافیک شبکه و پهنای باند و حرکت به سمت مدیریت جریان داده و اطلاعات در داخل کشور، تولید محتوا، نوآوری و کارآفرینی و مسائل فرهنگی مرتبط تمرکز شده است. در ادامه شمایی از نقشه راه توسعه کلان داده در کشور ارایه شده است که نقطه شروع آن وضعیت 4C (به عنوان مجموعه موانع فرهنگی) می‌باشد و چهار گام اصلی بر اساس راهبرد **"توسعه علمی مبتنی بر رفع موانع فرهنگی"** در نظر گرفته شده است که سبب رسیدن به وضعیت مطلوب مورد انتظار خواهد شد. جزئیات هر گام از نقشه راه به همراه زیر پروژه ها و فعالیت های مرتبط در ادامه توضیح داده شده است..



شکل ۷- نقشه راه برنامه ریزی و سیاست گذاری در حوزه کلان داده در کشور

۷-۱- توسعه جامعه کاربری:

- ایجاد کارگروه استانداردسازی در حوزه کلان داده (متشکل از واحدهای فناور)
- حمایت از گروه ها و انجمن های تخصصی در حوزه کلان داده
- حمایت از ایجاد پروژه های آزاد در حوزه کلان داده و تلاش برای تقویت جامعه کاربری آنها
- شناسایی و گزارش تجربیات موفق
- حمایت از اجرای برنامه ها و رویدادهای آموزشی و فرهنگ سازی
- حمایت از ایجاد برنامه رصد در حوزه کلان داده

۷-۲- حمایت از ایجاد پرتال داده باز:

- سیاست گذاری و حمایت از ایجاد پرتال داده باز
- بررسی و رفع موانع حقوقی و امنیتی برای آزاد سازی داده ها
- برگزاری نشست های تخصصی برای بحث پیرامون مسائل حقوقی
- حمایت از تولید و به اشتراک گذاری محتوا

۷-۳- حمایت از راهکارهای ابری بومی:

- حمایت از ایجاد قطب فناوری های مرتبط با رایانش ابری
- حمایت از توسعه ابزارهای کلان داده مبتنی بر رایانش ابری
- حمایت از توسعه سرویس های ابری مبتنی بر کلان داده
- حمایت از ایجاد پروژه های آزاد در حوزه رایانش ابری و تلاش برای تقویت جوامع کاربری

۷-۴- حمایت از استارتاپ ها و ایده های نو:

- حمایت از سرمایه گذاری و مشارکت در پروژه های متن باز و آزاد و تولید محتوای آزاد
- حمایت از شکل گیری استارتاپ ها در حوزه کلان داده
- حمایت از شکل گیری استارتاپ های کلان داده با تمرکز بر زبان فارسی

- حمایت از اجرای برنامه ها و رویدادهای آموزشی و فرهنگ سازی

۸- پروژه های پشتیبان نقشه راه

در اجرای این نقشه راه، برنامه ها و پروژه هایی وجود دارد که می‌بایست برای حمایت از اجرای سیاست های در نظر گرفته شده در برنامه اصلی انجام شوند. این پروژه ها شامل موارد زیر می شوند:

۸-۱- برنامه ها و پروژه های پشتیبان فاز اول:

- اجرای برنامه رصد فناوری در حوزه کلان داده (تجربه مشابه در حوزه رایانش ابری وجود دارد)
- ایجاد دبیرخانه جامعه آزاد کلان داده و راه اندازی ابزارهای مورد نیاز (وبسایت، گروه پستی و رسانه های مرتبط)
- توسعه API ها و ابزارهای نظارت بر جامعه و گزارش‌گیری برخط

۸-۲- برنامه ها و پروژه های پشتیبان فاز دوم:

- ایجاد هسته اولیه پرتال داده باز بصورت متن باز در قطب فناوری های مرتبط با رایانش ابری
- توسعه API های مورد نیاز برای دسترسی به پرتال داده باز و توسعه ابزارها و واسط های گزارش‌گیری
- امن سازی پرتال داده باز

۸-۳- برنامه ها و پروژه های پشتیبان فاز سوم:

- ایجاد قطب فناوری های مرتبط با کلان داده (همگرایی با برنامه قطب رایانش ابری)
- یکپارچه سازی پرتال داده باز با پلت فرم ابری توسعه داده شده در قطب رایانش ابری
- ایجاد سامانه نظارت بر پروژه های آزاد مرتبط با کلان داده
- ایجاد آزمایشگاه بررسی و ارزیابی پروژه های کلان داده
- ایجاد محک های مرتبط با زبان فارسی

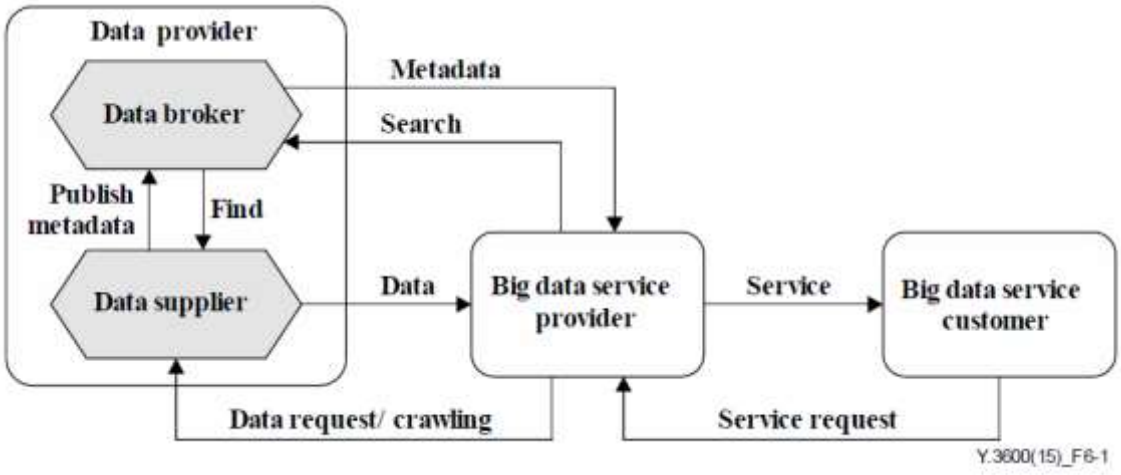
۸-۴- برنامه ها و پروژه های پشتیبان فاز چهارم:

- ایجاد دبیرخانه رویدادهای استارتاپ و رقابت های مرتبط با کلان داده
- پروژه مهاجرت دولت به کلان داده (با عنوان دولت باز یا Open Government)

۹- پیوست۱: استاندارد کلان داده مبتنی بر رایانش ابری (ITU-T Y.3600)

بر اساس استاندارد ITU-T Y.3500، رایانش ابری بعنوان یک پارادایم برای دسترسی به مجموعه ای از منابع اشتراکی بصورت مقیاس پذیر، خودکار، بر حسب تقاضا و از طریق شبکه مطرح شده است. همچنین بر اساس استاندارد ITU-T Y.3600، کلان داده به عنوان یک پارادایم برای جمع آوری، ذخیره سازی، مدیریت، تحلیل و مصورسازی اطلاعات بصورت بلادرنگ و بر روی مجموعه وسیعی از منابع داده ای ناهمگن معرفی شده است. همچنین اشاره شده است که چنانچه یک سرویس ابری بتواند قابلیت هایی نظیر جمع آوری، ذخیره سازی، تحلیل، مصورسازی و مدیریت داده بر روی کلان داده را برای یک مشتری انجام دهد، به آن BDaaS[1] گفته می شود. در اکوسیستم کلان داده همانطور که در شکل ۸ نشان داده شده است، نقش های زیر تعریف شده است:

- **فراهم کننده داده**[2]: که شامل دو زیر نقش تامین کننده داده[3] و کارگزار داده[4] می باشد. تامین کننده داده وظیفه تامین داده از منابع مختلف را برای کارگزار دارد، ضمن اینکه فراداده مورد نیاز را هم ممکن است ایجاد نماید. کارگزار داده ارتباط بین سرویس دهنده کلان داده را با فراهم کننده داده برقرار می‌کند تا سرویس دهنده بتواند داده های مورد نیاز خود را جستجو کند و بدست آورد.
- **سرویس دهنده کلان داده**[5]: سرویس دهنده قابلیت های مختلف نظیر تحلیلی یا زیرساختی را فراهم می‌کند. در حقیقت سرویس دهنده مشابه یک پلت فرم است که می‌تواند به پلت فرم های های تحلیل داده موجود اضافه شود تا امکان پشتیبانی از کلان داده را فراهم کند.
- **مشتری سرویس کلان داده**[6]: کاربر نهایی سیستم کلان داده است که از خدمات و یا نتایج سرویس دهنده بهره برداری میکند. البته مشتری ممکن است دانش یا سرویس جدیدی را بر اساس فعالیت خود خارج از اکوسیستم کلان داده نیز تولید کند.

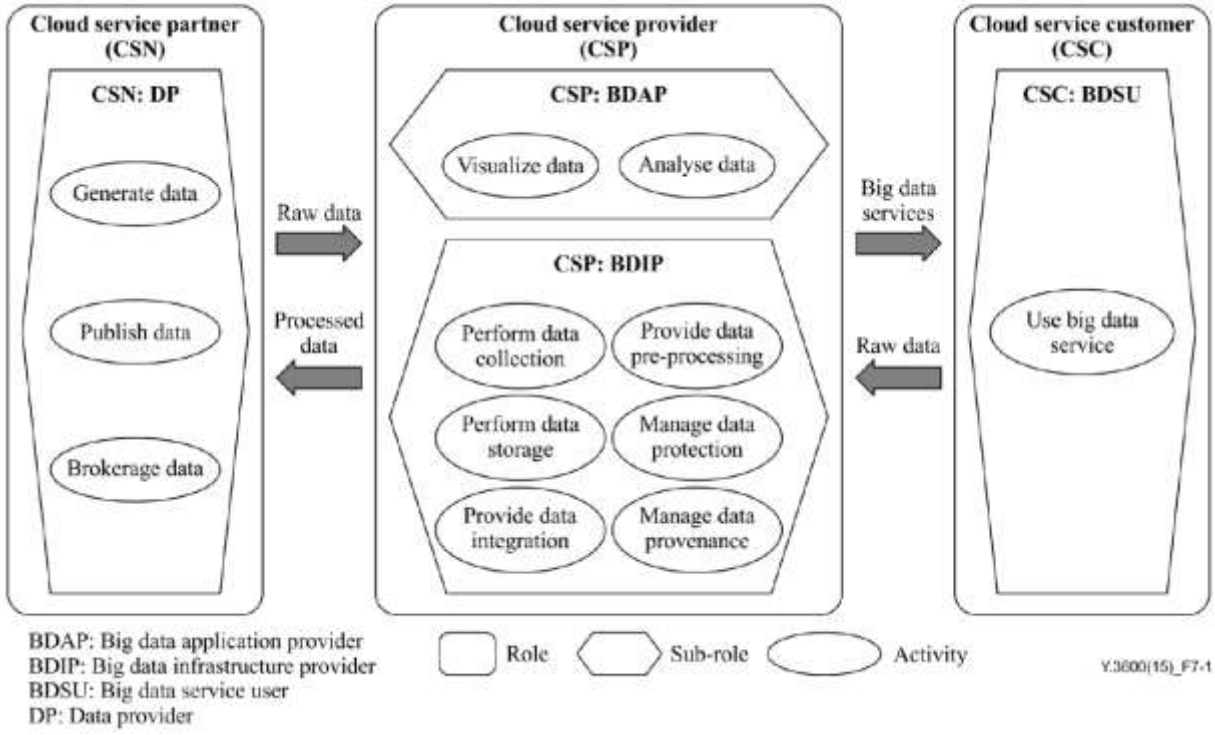


شکل ۸-اکوسیستم کلان داده

[1] Big Data as a Service
[2] Data Provider (DP)
[3] Data Supplier
[4] Data Broker
[5] Big Data Service Provider
[6] Big Data Service Customer

کلان داده نیاز به کارآیی بالا برای پردازش داده و ذخیره سازی توزیع شده بر حسب تقاضا دارد. همچنین مجموعه متنوعی از ابزارها برای پشتیبانی از فعالیت‌های مختلف در اکوسیستم مورد نیاز آن خواهد بود. رایانش ابری بر اساس مشخصه های ذاتی خود می‌تواند پاسخ مناسبی برای حل این چالش ها ارائه دهد. ماهیت انفجاری کلان داده، بسیار متناسب است با خاصیت مقیاس پذیری و بر حسب تقاضا بودن رایانش ابری؛ ترکیب اکوسیستم کلان داده و معماری مرجع رایانش ابری (ISO17789:2014) در شکل ۹ نشان داده شده است. معماری بدست آمده شامل نقش های زیر می باشد:

- **زیر نقش CSN:DP**: زیر نقش فراهم کننده داده[1]، به عنوان یکی از نقش‌های شریک سرویس در معماری مرجع رایانش ابری قابل نگاشت است
- **زیر نقش CSP:BDAP**: زیر نقش سرویس دهنده کاربردی کلان داده[2] به عنوان یکی از نقش های سرویس دهنده، وظیفه کارهای تحلیلی و مصورسازی را بر عهده دارد.
- **زیر نقش CSP:BDIP**: زیر نقش سرویس دهنده زیرساخت کلان داده[3] به عنوان یکی از نقش های سرویس دهنده، وظیفه تامین منابع ابری جهت پردازش کلان داده با ملاحظاتی نظیر امنیت، جامعیت و مدت نگهداشت داده بر عهده دارد.
- **زیر نقش CSC:BDSU**: زیر نقش کاربر سرویس کلان داده[4]، کاربر نهایی سیستم است که از خدمات CSP:BDAP و CSP:BDIP استفاده می کند.

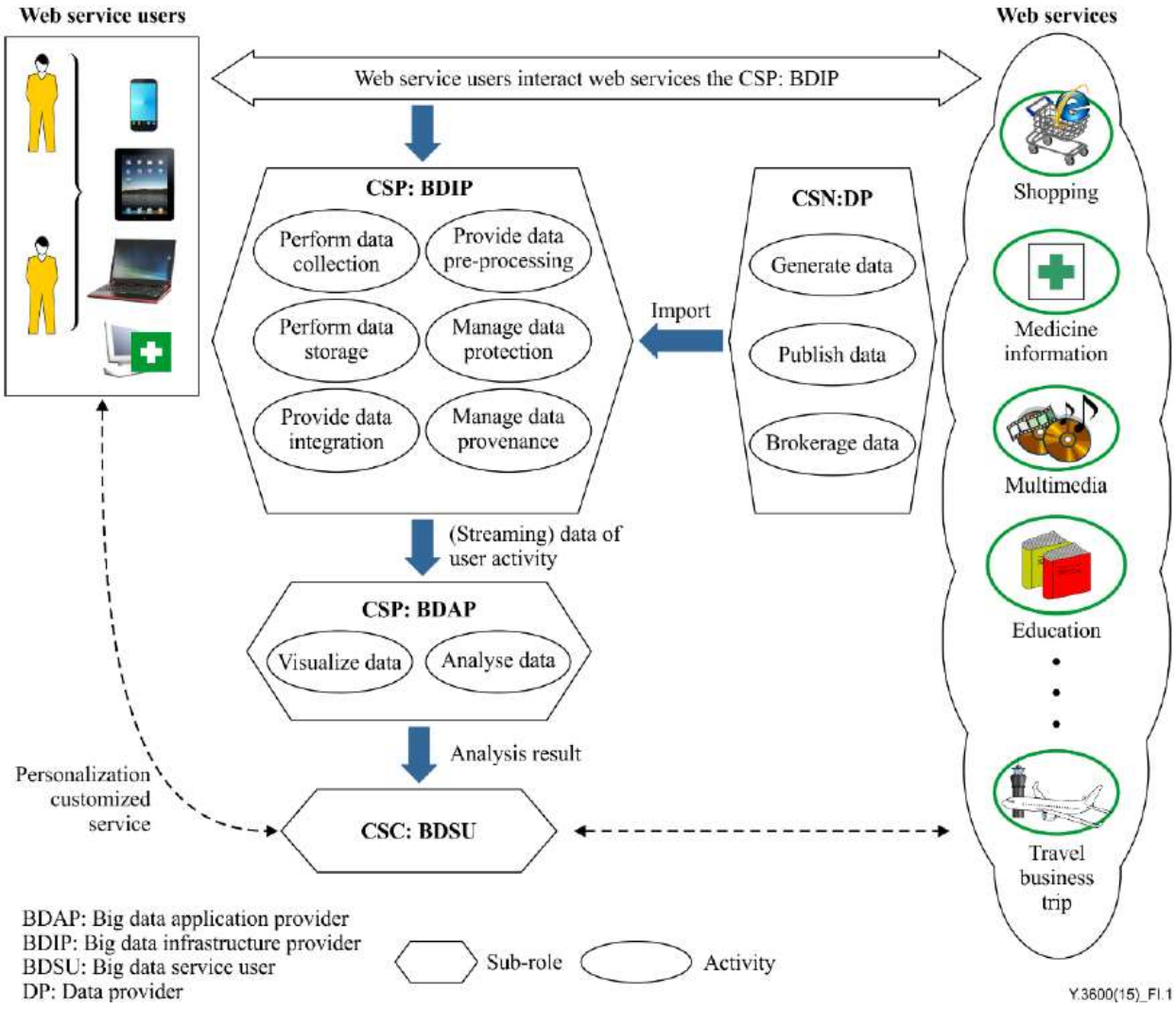


شکل ۹- رایانش ابری مبتنی بر یک سیستم کلان داده

[1] CSN:Data Provider
[2] CSP:Big Data Application Provider
[3] CSP:Big Data Infrastructure Provider
[4] CSC:Big Data Service User

۱۰- پیوست۲: مثال موردی استفاده از کلان داده در شخصی سازی

تجهیزات زیادی نظیر موبایل، تبلت، لپ تاپ و ... به اینترنت متصل هستند و از سرویس های برخط استفاده میکنند. استفاده کاربران از این سرویس ها لاگهای زیادی تولید می‌کند که حجم و تنوع بسیار زیادی دارند و با تحلیل آنها میتوان به الگوی رفتاری کاربران جهت ارائه خدمات شخصی سازی شده (نظیر تبلیغات) دست پیدا کرد. معماری زیر نشان می دهد که در صورت وجود سازوکاری جهت مشاهده لاگها و یا مبدا و مقصد بسته ها توسط CSP:BDIP، و با توجه به اینکه خدمات و منابع داده‌ای در اینترنت قابل بازیابی مجدد است، تحلیل های مختلف میتواند توسط CSP:BDAP بر روی این حجم عظیم داده ها در ابر انجام شود و از نتیجه آن برای ارائه خدمات شخصی سازی شده به کاربران استفاده کرد.



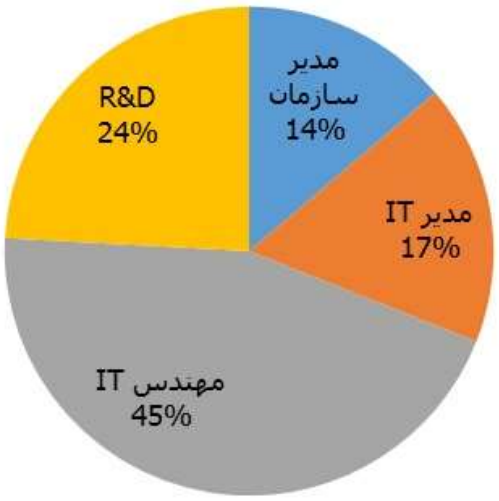
از جمله تحلیل هایی که در اینجا توسط CSP:BDAP قابل انجام است میتوان به تشخیص الگوهای علاقه مندی هر کاربر بر اساس فعالیت هایی که انجام میدهد اشاره کرد. نتیجه این تحلیل ها در اختیار CSC:BDSU قرار می گیرد تا بتواند برای معرفی سرویس یا پشتیبانی بهتر کاربران در استفاده از خدمات وب مورد استفاده قرار گیرد. در اینجا CSC:BDSU مثلا می تواند یک سایت تجارت الکترونیک در نظر گرفته شود.

۱۱- پیوست۳: گزارش آماری

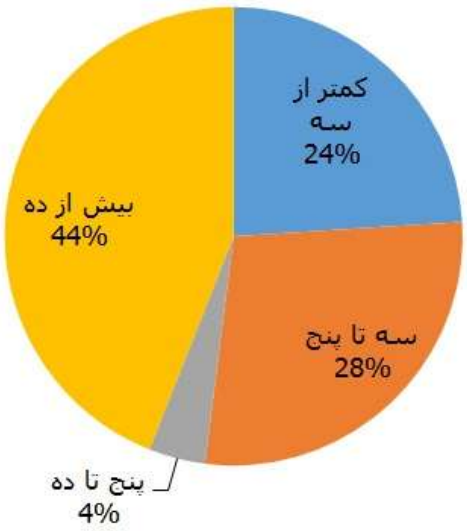
نتایج نظرسنجی های انجام شده در همایش داده های عظیم (دیماه ۱۳۹۳) در این قسمت ارایه شده است.

۱۱-۱-وضعیت آماری شرکت کنندگان:

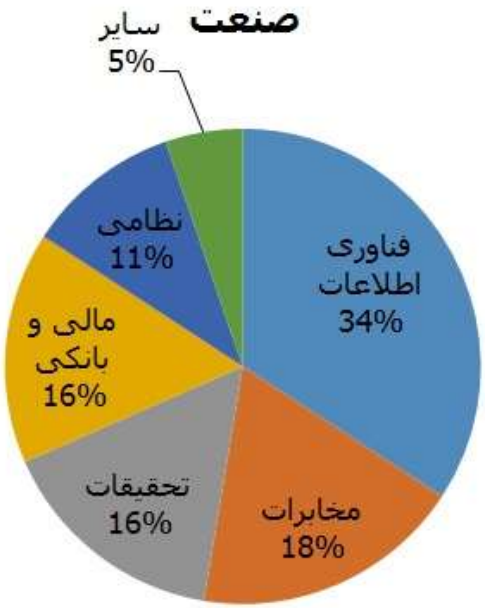
نقش سازمانی



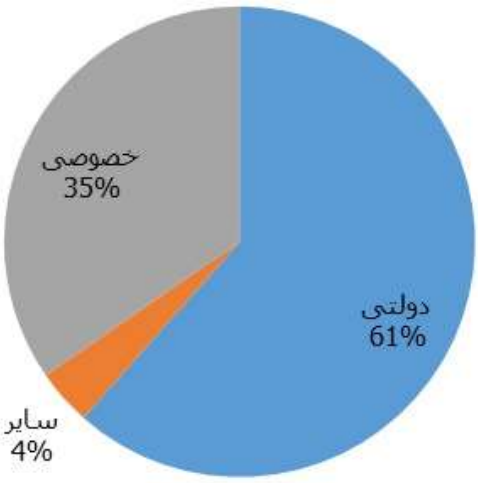
سابقه کار



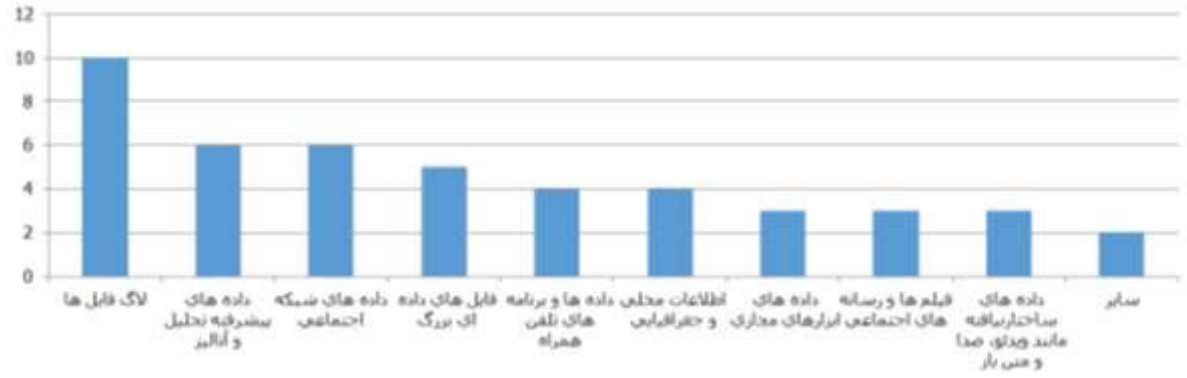
صنعت



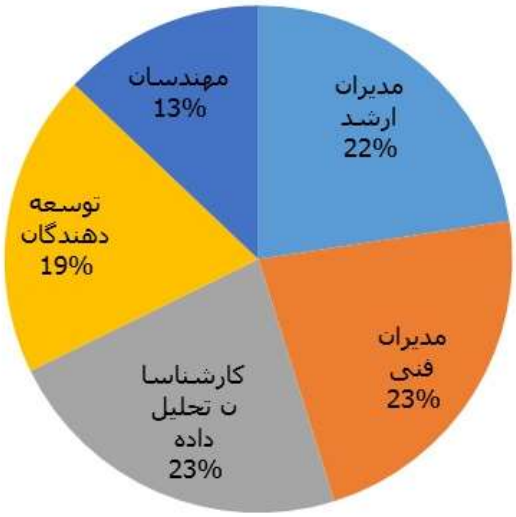
نوع سازمان



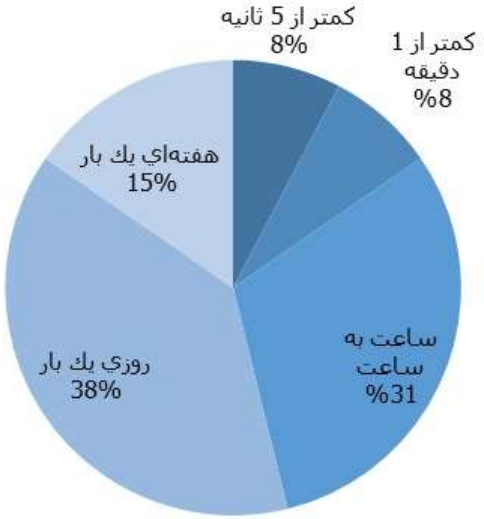
۱۱-۲-ویژگی های کلان داده در سازمانها:



شکل ۱۰- انواع کلان داده در سازمان شما



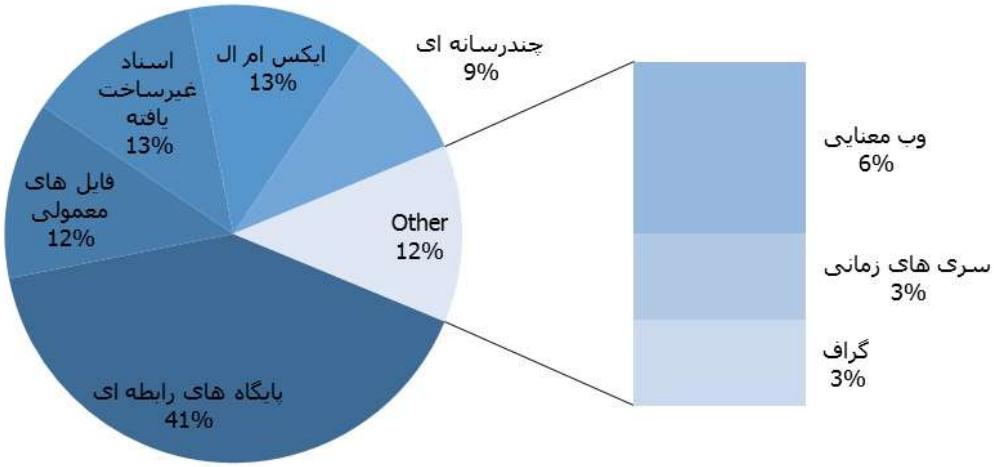
شکل ۱۱- متولی کلان داده در سازمان



شکل ۱۲- بازه های زمانی پردازش داده



شکل ۱۳- ویژگی های مطرح در بکارگیری کلان داده

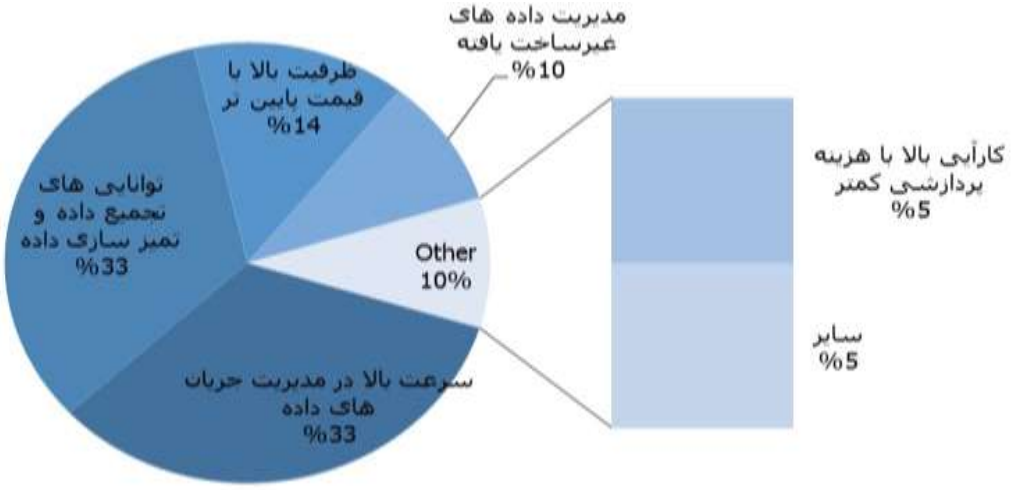


شکل ۱۴- قالب های داده موجود در سازمان

۱۱-۳-فناوری ها و زیرساخت های مورد نیاز یا مورد استفاده



شکل ۱۵- از چه تکنولوژی هایی در حوزه کلان داده استفاده می کنید



شکل ۱۶- چه بسترهایی برای کلان داده مورد نیاز شماست

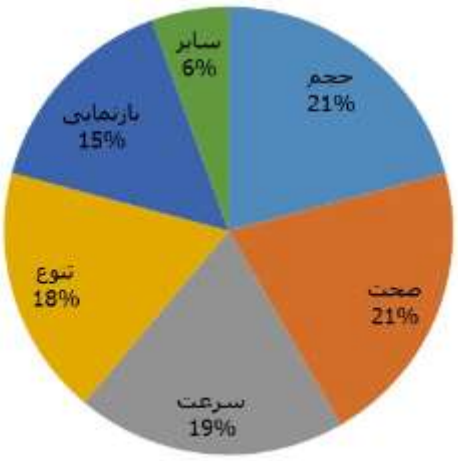
۱۱-۴-چالش ها و فرصت ها



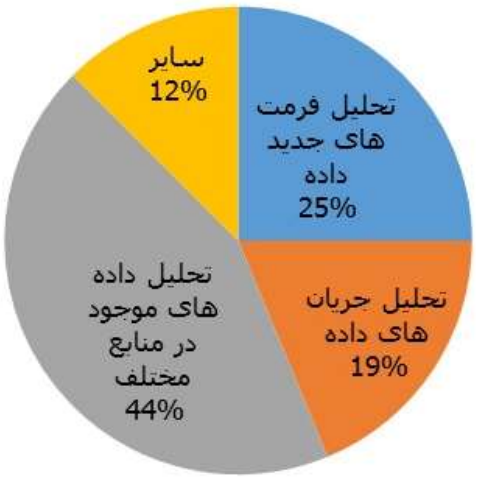
شکل ۱۷- مزیت های استفاده از کلان داده



شکل ۱۸- چالش های اصلی پیاده سازی کلان داده در سازمان



شکل ۱۹- مشخص های کلان داده



شکل ۲۰- چالش های سه سال آینده

شناسنامه سند

نسخه	تاریخ	شرح
نسخه ۱,۰	دیماه ۱۳۹۳	نسخه اولیه پیش از برگزاری همایش داده های عظیم
نسخه ۲,۰	بهمن ۱۳۹۳	تکمیل مستند بر اساس نتایج بدست آمده در همایش داده های عظیم
نسخه ۳,۰	تیرماه ۱۳۹۴	افزودن برنامه های راهبردی و عملیاتی به نقشه راه
نسخه ۴,۰	دیماه ۱۳۹۵	افزودن استانداردهای کلان داده در رایانش ابری و مطالعات موردی

ایجـاد هماهنگـی در بیـن اجـزای اکوسیسـتم یـک صنعـت (متشـکل از دولـت، دانشـگاه و صنعـت) همـواره یـک چالـش جـدی در کشـور بـوده اسـت. نقـش سـازمانهای حمایتـی (نظیـر جامعـه آزاد رایانـش ابـری) کمـک بـه ایجـاد ایـن هماهنگـی از طریـق فرهنـگ سـازی، آمـوزش و ایجـاد بسـترهای اجرایـی لازم اسـت کـه عمـلا توسـط خـود اعضـای جامعـه ولـی در قالـب یـک عنـوان و هـدف واحـد صـورت میگیـرد. برنامـه هـای جامعـه بطـور عمـده در قالـب جلسـات منظـم، ارائـه هـای فنـی، بحـث هـای آزاد، تولیـد محتـوا و تشـکیل کارگـروه هـای تخصصـی اجـرا میشـود و تصمیـم گیـری هـا توسـط اعضـای فعـال در جامعـه (از دولـت، صنعت و دانشگاه) صورت می پذیرد.

بـرای مشـاهده دیگـر اسـناد تهیـه شـده در جامعـه آزاد رایانـش ابـری ایران میتوانید به آدرس http://press.occc.ir مراجعه نمایید. این گزارشها شامل موارد زیر می باشد:

- گزارشی از وضعیت خدمات ابری در ایران
- گزارشی از وضعیت پذیرش رایانش ابری در ایران
- ویژه نامه های مربوط به رصد فناوری اطلاعات در حوزه رایانش ابری

کارگروه تاکسونومی و استانداردسازی

نسخه ۴.۰

دیماه ۱۳۹۵