فصلنامهٔ امنیت پژوهی (علمی-پژوهشی)
سال بیستم: شماره ۷۳- بهار ۱٤۰۰
صص: ۵۷ - ۷۹

*Amniat Pajouhi: Security Research Quarterly*
*Vol.20, No.73, Spring 2021*
P-P: 57 - 79

تاریخ دریافت: ۱۳۹۹/۱۰/۲۰ تاریخ پذیرش: ۱۳۹۹/۱۱/۱۱

# ارائه چارچوب معماری برای کلان‌داده و بررسی تأثیر آن بر شاخص‌های امنیتی این حوزه

**خداداد هلیلی**[۱]
**محمود دی‌پیر**[۲]

## چکیده

در فنّاوری نوظهور کلان‌داده، جمع‌آوری، پردازش، تحلیل و استخراج دانش از داده‌های متنوع و بسیار حجیم، توسط سامانه‌های نرم‌افزاری خاصی مانند هادوپ، صورت می‌گیرد. استفاده بهینه از فرصت‌های ایجادشده و مدیریت تهدیدات و چالش‌های امنیتی این حوزه، نیازمند لحاظ نمودن همه جنبه‌ها و فرایندهای آن در قالب یک چارچوب معماری است. هدف از انجام این پژوهش، ارائه یک چارچوب معماری برای کلان‌داده و بررسی تأثیر آن بر شاخص‌های امنیتی این حوزه است. بدین منظور ابتدا اجزاء و ویژگی‌های چارچوب اولیه، پس از بررسی و مقایسه چند چارچوب مهم احصاء شد؛ سپس چارچوب پیشنهادی، بر اساس مؤلفه‌های مطرح‌شده درزمینهٔ امنیت کلان‌داده، با نظرسنجی از خبرگان این حوزه، به‌صورت کمّی مورد ارزیابی قرار گرفت. برای این کار، از روش مدل‌سازی معادلات ساختاری و نرم‌افزار اسمارت پی.آل. اس. استفاده شده است. این پژوهش ازنظر هدف، کاربردی و ازنظر روش، توصیفی- پیمایشی است. جامعه آماری آن نیز شامل خبرگان معماری سازمانی و امنیت فضای سایبر، به تعداد ۵۵ نفر است که از میان آن‌ها ۳۵ نفر به‌عنوان نمونه آماری انتخاب گردید. یافته‌های پژوهش برازش مناسب چارچوب معماری پیشنهادی و تأثیر معنادار مؤلفه‌های آن را بر امنیت کلان‌داده تأیید می‌کنند.

**کلید واژگان:**
کلان‌داده، چارچوب معماری، امنیت‌داده

۱. استادیار، هیئت‌علمی دانشکده رایانه، دانشگاه هوایی شهید ستاری، تهران، ایران (نویسنده مسئول) *kh.halili@ssau.ac.ir*
۲. دانشیار، دانشکده رایانه، دانشگاه هوایی شهید ستاری، تهران، ایران *mdeypir@ssau.ac.ir*

## جستارگشایی

کاربردگسترده و روزافزون فنّاوری اطلاعات و ارتباطات در همه عرصه‌های زندگی بشر، موجب تولید چشمگیر و خیره‌کننده داده از منابع مختلف شده و مفهوم جدیدی به نام کلان‌داده[1] را به وجود آورده است. کلان‌داده به مجموعه داده‌های حجیمی گفته می‌شود؛ که مدیریت، کنترل و پردازش آن‌ها، در یک زمان معقول و مورد انتظار، فراتر از توانایی ابزارهای نرم‌افزاری و پایگاه‌های داده سنتی است. آنچه در مواجهه با حجم زیاد اطلاعات جذابیت و اهمیت دارد؛ تحلیل و استخراج دانش از این اطلاعات است. امروزه داده و اطلاعات، یکی از مهم‌ترین دارایی‌های نامشهود، محرک نوآوری و ابزاری برای کسب مزیت رقابتی، در هر سازمان محسوب می‌شود. ارزش بالقوه نهفته در کلان‌داده برای شرکت‌های مختلف تجاری، صنعتی، علمی و سازمان‌های دولتی به حدی است؛ که برخی آن را نفت دوران جدید نامیده‌اند. مؤسسه تحقیقات بین‌المللی گارتنر[2] کلان‌داده را یکی از ۱۰ فنّاوری برتر در روند حیاتی فنّاوری‌های آینده معرفی نموده است.

کلان‌داده از مهم‌ترین مباحث روز دنیا و تحقیقات آینده است که در حوزه‌های مختلفی ازجمله پزشکی (ژنتیک و پیش‌بینی شیوع بیماری‌ها)، تحلیل‌های اقتصادی، مالی و تجاری (بورس)، هواشناسی (پیش‌بینی وضعیت آب‌وهوا)، شبیه‌سازی‌های پیچیده فیزیک، تحقیقات زیست‌شناسی و محیطی، جست‌وجوی اینترنت، تبلیغات، جاسوسی و به‌ویژه دستیابی به امنیت کاربرد دارد.

در بیشتر کشورها، دولت‌ها، بهره‌گیری و تحلیل کلان‌داده را در برنامه‌ریزی راهبردی برای دستیابی به اهداف سیاسی، اقتصادی، امنیتی، اجتماعی و سلامت، مورد توجه قرارداده‌اند. از منظر حاکمیتی، توسعه زیرساخت‌ها و ابزارهای تجزیه‌وتحلیل داده، زمینه‌ساز بهره‌برداری از مزایای کلان‌داده است. بااین‌حال، کلان‌داده در کنار قابلیت‌ها، مزایا و فرصت‌های فراوان و بالقوه خود، چالش‌های امنیتی زیادی نیز ایجاد نموده است.

چارچوب‌های معماری مختلف دارای اجزاء و ویژگی‌های مشابهی هستند؛ اما هرکدام بر اساس نیازمندهای خاصی توسعه یافته‌اند. ارائه چارچوب معماری علاوه بر آنکه موجب همگرایی و هماهنگی در نوع نگاه به مفاهیم و موضوعات یک حوزه می‌شود؛ مبنایی برای راهکارهای عملی پیاده‌سازی فراهم می‌سازد و هم‌راستایی دیدگاه‌ها و اقدامات مختلف را به همراه دارد.

بیان مسئله

در فنّاوری کلان‌داده، ویژگی‌هایی مانند حجم، سرعت، تنوع، کیفیت داده، دوره نگهداری اطلاعات، مصورسازی، اعتبار و هزینه‌های جمع‌آوری، ذخیره‌سازی، جست‌وجو، اشتراک و تحلیل داده‌ها، پارادایم

1. Big Data
2. www.gartner.com

نوینی در مدل‌های معماری داده- محور، ایجاد نموده است. در چارچوب معماری کلان‌داده، ضمن دسته‌بندی و لحاظ نمودن همه جنبه‌ها و فرایندها، یک ساختار منطقی برای بهره‌برداری از حجم انبوهی از داده‌های متنوع فراهم می گردد.

این فنّاوری نوظهور، در کنار مزایای فراوان، چالش‌ها و مخاطرات امنیتی گسترده‌ای ازجمله بهره‌برداری از داده‌ها برای جاسوسی، نقض حریم خصوصی، حملات سایبری، تهدیدات امنیت ملی و افزایش هزینه‌های دولت درزمینهٔ تأمین امنیت کلان‌داده‌ها را به همراه دارد.

در جمهوری اسلامی ایران، گسترش فعالیت‌ها در حوزه فنّاوری اطلاعات و ارتباطات با افزایش پهنای باند، توسعه دولت الکترونیک، حرکت اپراتورهای مخابراتی به نسل‌های 3G و 4G، فراگیر شدن تجهیزات هوشمند دیجیتال و گرایش به استفاده از شبکه‌های اجتماعی، ورود به حوزه کلان‌داده را نشان می‌دهد. از طرفی چالش‌های امنیتی موجود و بالقوه، درزمینهٔ پذیرش و ورود این فنّاوری به کشور، از حساسیت بالایی برخوردار است؛ که این مسئله، لزوم دستیابی به یک چارچوب معماری را مشخص می‌سازد.

بنابراین دغدغه اصلی شکل‌گیری این تحقیق، ارائه یک چارچوب معماری با در نظر گرفتن مؤلفه‌های اثرگذار و اثرپذیر در حوزه فنّاوری کلان‌داده و ارزیابی کارکرد آن در مقابله با چالش‌های امنیتی این حوزه است. در این مقاله، با بررسی و مقایسه تطبیقی چارچوب‌های معماری موجود، یک چارچوب معماری پیشنهاد شده و در ادامه پس از شناسایی چالش‌های فنّاورانه مدیریتی و امنیتی کلان‌داده، ارزیابی چارچوب معماری ارائه شده مورد توجه قرار گرفته است.

انجام پژوهش‌های علمی درزمینهٔ مفهوم شناسی روش‌ها، ابزارها و فنّاوری‌های مورد استفاده در مراحل جمع‌آوری، ذخیره‌سازی، جست‌وجو، اشتراک و تحلیل کلان‌داده‌ها، پیش‌نیاز برنامه‌ریزی برای حضور فعال و ورود به این عرصه است. ارائه یک چارچوب معماری در حوزه کلان‌داده، می‌تواند فرهنگ جدیدی را برای بهره‌برداری از زیرساخت، سرویس‌ها و تجزیه‌وتحلیل داده، ایجاد نماید و موجب ارتقای بینش و درک مدیران و ترغیب آن‌ها به توسعه فعالیت‌ها در این حوزه شود.

توجه ویژه آژانس‌های امنیتی کشورهای صنعتی، ازجمله ایالات‌متحده به جاسوسی گسترده از طریق برنامه‌ها و سخت‌افزارهای جمع‌آوری و تحلیل داده، مانند پریزم[1] و خبرچین بیکران[2]، نشانگر اهمیت بهره‌برداری از فرصت‌ها و ظرفیت‌های این فنّاوری است؛ بنابراین فقدان یک چارچوب معماری و لحاظ نمودن سازه‌ها و شاخص‌های امنیتی در آن، موجب غفلت از مخاطرات امنیتی شده و چالش‌ها و تهدیدات مرتبط با آن همچنان باقی خواهد ماند.

1. *PRISM*
2. *Boundless Informant*

اهداف تحقیق

هدف اصلی: ارائه یک چارچوب معماری برای کلان‌داده و بررسی تأثیر آن بر شاخص‌های امنیتی این حوزه

حال پرسش اینست: چارچوب معماری مناسب برای کلان‌داده چگونه است و چه تأثیری بر شاخص‌های امنیتی این حوزه دارد؟

در بسیاری از پژوهش‌های مرتبط با کلان‌داده، موضوعاتی مانند رشد سرسام‌آور حجم داده و روند فزاینده سرعت تولید و پردازش اطلاعات بررسی شده است (متی و کورنیک، ۲۰۱۲). برخی نیز با اشاره به سرمایه‌گذاری گسترده کشورها درزمینهٔ این فنّاوری، به لحاظ نمودن این فنّاوری در راهبردهای ملی کشورها پرداخته‌اند (جن و ژانگ، ۲۰۱۴). در سایر پژوهش‌ها مانند (آنوار، یعقوب و هاشم، ۲۰۱۵)، (فرانکس، ۲۰۱۲)، (هیلبرت، ۲۰۱۳) و (دوو و چن، ۲۰۱۳) نیز، مزایا و چالش‌های کلان‌داده مورد توجه قرار گرفته است. در برخی از تحقیقات نرم‌افزارهایی مانند هادوپ و اسپارک برای تحلیل و داده‌کاوی کلان‌داده بررسی شده است (ماهک، سجویین و دیپاردون، ۲۰۱۳). در تحقیق (مشارا، ۲۰۱۴) نیز بر لزوم توجه به استانداردها تأکید شده است.

در مقاله پاندری و همکاران، با بررسی مقالات و تحقیقات انجام‌شده، اجزای زیست‌بوم حوزه کلان‌داده، استخراج و ارائه شده است. در این زیست‌بوم، عناصری مانند چارچوب معماری، چرخه حیات، فراساختار، ساختارهای داده‌ای و تحلیلی و مدل‌های مربوط به کلان‌داده، موردبررسی قرار گرفته است (پاندری و همکاران، ۲۰۱۶).

در مقاله مرت و همکاران، یک معماری مرجع منبع باز برای کلان‌داده مطابق شکل (۱) پیشنهاد شده است (مرت و همکاران، ۲۰۱۷). در این معماری اجزای مختلفی مانند سیستم فایل، منبع داده، پردازش، تحلیل داده و مصورسازی به‌عنوان اجزای اصلی معماری در نظر گرفته شده است. حکمرانی و امنیت نیز به‌صورت لایه‌های عمودی در کنار اجزای معماری قرار گرفته است.


**شکل ۱. چارچوب معماری پیشنهادی منبع باز کلان‌داده** (مرت و همکاران، ۲۰۱۷)

در مقاله اینوبلی و همکاران چارچوب‌های معماری هادوپ، اسپارک، استورم و فلینک از دیدگاه کارایی و جزئیات فنی مورد مطالعه و ارزیابی قرار گرفته است (اینوبلی، ۲۰۱۸). در این مقاله، برخی ویژگی‌های کلیدی مانند مدل و زبان برنامه‌نویسی، نوع منابع داده، ظرفیت پردازش، سازگاری و خطا در نظر گرفته شده است. برای مقایسه چارچوب‌های مختلف نیز از معیارهایی مانند زمان اجرا و میزان ترافیک مصرفی استفاده‌شده؛ اما به موضوع امنیت پرداخته نشده است.

در مقاله مورنو و همکاران یک معماری مرجع اولیه برای امنیت و حریم خصوصی در کلان‌داده ارائه شده است (مورنو و همکاران، ۲۰۱۸). این مدل معماری بر اساس زبان مدل‌سازی یکپارچه (UML) به الزامات امنیتی و حریم خصوصی در این حوزه پرداخته است. در این معماری به جایگاه امنیت در یک سیستم کلان‌داده و ارتباط آن با سایر سرویس‌ها پرداخته شده است. برتری مدل ارائه شده در این مقاله بیان جزئیات توسط UML است که امکان پیاده‌سازی آن را فراهم می‌کند.

در چارچوب‌های معماری کلان‌داده، جنبه‌ها، دیدگاه‌ها و اجزای آن در نظر گرفته می‌شود. با بررسی پژوهش‌های انجام‌شده در این زمینه، سه چارچوب مهم در این تحقیق مورد بررسی قرار گرفته است. چارچوب معماری اتحادیه مرکز داده آزاد توسط (عستس و موتی، ۲۰۱۳) ارائه شده است. شرکت مایکروسافت نیز که از پیشگامان معماری کلان‌داده است؛ یک چارچوب معماری مرجع برای زیست‌بوم کلان‌داده ارائه نموده است (لوین، ۲۰۱۳). چارچوب معماری مرجع تحلیل و بهینه‌سازی حرفه آی. بی. ام، نیز توسط (دمچنکو، یوری و انگو، ۲۰۱۳) مطرح شده است.

مروری بر تحقیقات منتشرشده نشان می‌دهد؛ اغلب معماری‌های مرجع برای کلان‌داده توسط شرکت‌های بزرگ مانند مایکروسافت، اوراکل، آمستردام، آی. بی. ام، موسسه ملی فنّاوری و استاندارد (NIST)

ارائه شده است. بااینکه این چارچوب‌های معماری نقاط مشترکی باهم دارند؛ اما از دیدگاه‌های مختلف به این مسئله پرداخته شده و روش استانداردی در این زمینه وجود ندارد.

## مفاهیم و مبانی نظری تحقیق

### مفهوم و ویژگی‌های کلان‌داده

تعریف کلان‌داده اولین بار توسط داگ لنی[1] در موسسه گارتنر در سال ۲۰۰۱ مطرح شد. طبق این تعریف، منظور از کلان‌داده اطلاعاتی با حجم زیاد است که با به‌کارگیری روش‌های نوین پردازش و ذخیره‌سازی استخراج شده و به‌عنوان یک سرمایه اطلاعاتی برای درک بهتر از دنیا و روند تصمیم‌گیری مورد استفاده قرار می‌گیرد. وی سه ویژگی حجم[2]، سرعت[3] و تنوع[4] را تحت عنوان $3Vs$ برای کلان‌داده در نظر گرفت. ویژگی‌های کلان‌داده به‌مرور زمان از حجم، سرعت و تنوع در تعریف اولیه، فراتر رفت و با اضافه شدن ۵ ویژگی دیگر تحت عنوان $8Vs$ شناخته می‌شود (ملاس، ۲۰۱۲). این ویژگی‌های جدید عبارتند از: ۱- صحت[5] (کیفیت داده و حفظ مشخصات آن)، ۲- نوسان[6] (نرخ تغییرات زمانی داده‌ها یا مدت‌زمان و دوره نگهداری اطلاعات)، ۳- مصورسازی[7] (به‌کارگیری گراف‌های تصویری و بصری سازی اطلاعات برای فهم روابط پیچیده بین آن‌ها)، ۴- اعتبار[8] (اطمینان از روایی اطلاعات علاوه بر صحت آن‌ها)، ۵- ارزش[9] (برآورد میزان هزینه جمع‌آوری پردازش و نگهداری اطلاعات در برابر حجم داده).

استانداردسازی کلان‌داده، کمک زیادی به شرکت‌ها و افراد مختلف برای ورود به این فنّاوری می‌کند. در این زمینه، اتحادیه امنیت ابری[10] (CSA) و مؤسسه ملی استاندارد و فنّاوری آمریکا[11] (NIST) اقدام به شناسایی و ایجاد استانداردهای لازم درزمینهٔ مشکلات امنیتی و حریم خصوصی نموده‌اند (مشارا، ۲۰۱۴).

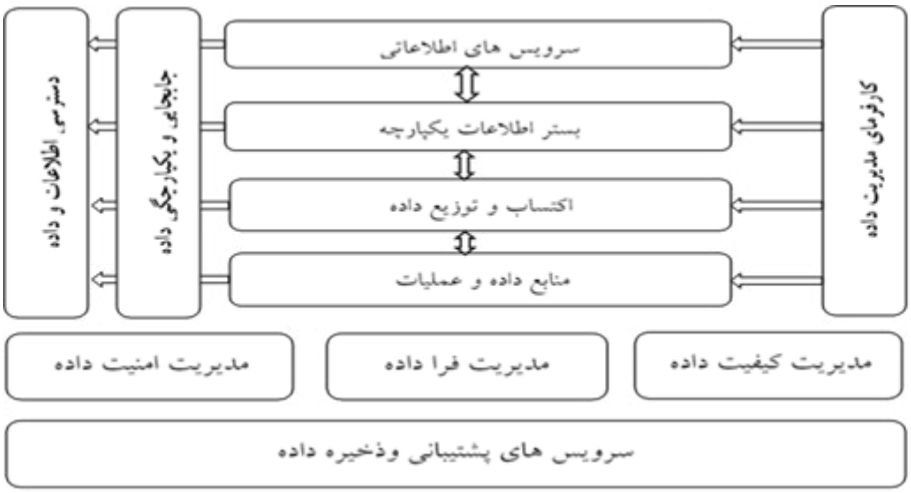
### چارچوب‌های معماری مهم برای کلان‌داده

چارچوب‌های معماری[12] قالب‌هایی برای ارائه اطلاعات در مورد دیدگاه‌ها و جنبه‌های یک سیستم (سازمان)، نمایش ارتباط بین اجزای آن و راهنمایی برای تکمیل و پیاده‌سازی عملی معماری است. استفاده از چارچوب معماری متناسب با موقعیت و وضعیت محیطی سازمان و با تکیه بر ظرفیت‌های داخلی، قابل توسعه و بومی‌سازی است (معینی و مرآتی، ۱۳۹۴). چارچوب‌های معماری معمولاً شامل

1. *Doug Laney*
2. *Volume*
3. *Velocity*
4. *Variety*
5. *Veracity*
6. *Volatity*
7. *Visualization*
8. *Validity*
9. *Value*
10. *Cloud Security Alliance*
11. *National Institute of Standards and Technology*
12. *Architecture Frameworks*

توصیه‌های اجرایی در مورد فرایند کلی معماری و روش‌های مدل‌سازی است. این چارچوب‌ها متناسب با نیازهای سازمان، محدودیت منابع، زمان و سازگاری محیطی در نظر گرفته می‌شوند؛ بنابراین مرحله انتخاب و طراحی چارچوب معماری، نقش بسزایی در چگونگی اجرای معماری دارد. در این بخش، سه نمونه از چارچوب‌های معماری مهم در حوزه کلان‌داده شناسایی و بررسی شده است.

۱) چارچوب معماری اتحادیه مرکز داده آزاد[1] (*ODCA*): این چارچوب معماری از سه بخش اصلی سرویس‌های پشتیبانی و ذخیره‌سازی، مدیریت داده و بسترهای تبادل، دسترسی و توزیع یکپارچه، تشکیل‌شده است. در شکل (۲) ارتباط بین این اجزاء نشان داده شده است (عستس و موتی، ۲۰۱۳).



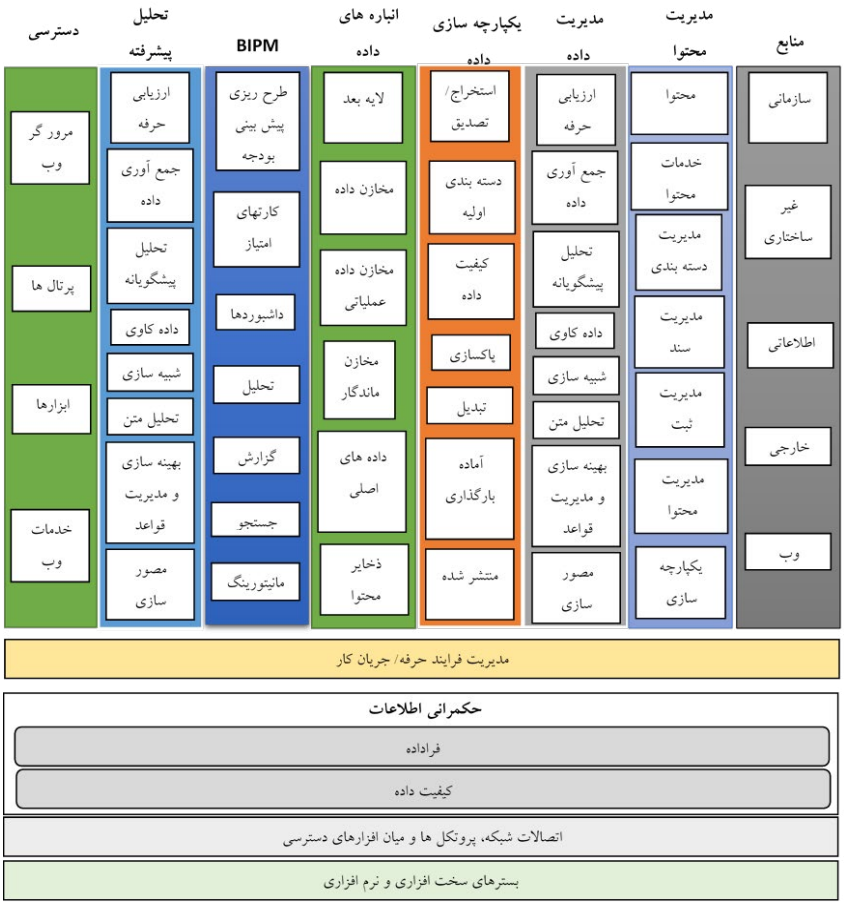

**شکل ۲. چارچوب معماری اتحادیه مرکز داده آزاد ( عستس و موتی ، ۲۰۱۳)**

۲) چارچوب معماری مرجع زیست‌بوم کلان‌داده (شرکت مایکروسافت): معماری مرجع زیست‌بوم کلان‌داده یک نمودار مبتنی بر داده سطح بالاست که در مورد جریان کلان‌داده و انتقال داده از جمع‌آوری تا استفاده را توصیف می‌کند. زیست‌بوم کلان‌داده شامل چهار بخش اصلی منابع، تبدیل، زیرساخت و کاربرد داده است. در این مدل امنیت و مدیریت به‌عنوان دو بخش مستقل در کنار سرویس‌ها و زیرساخت داده قرارگرفته‌اند. در شکل (۳) اجزای این معماری نشان داده شده است (لوین، ۲۰۱۳).

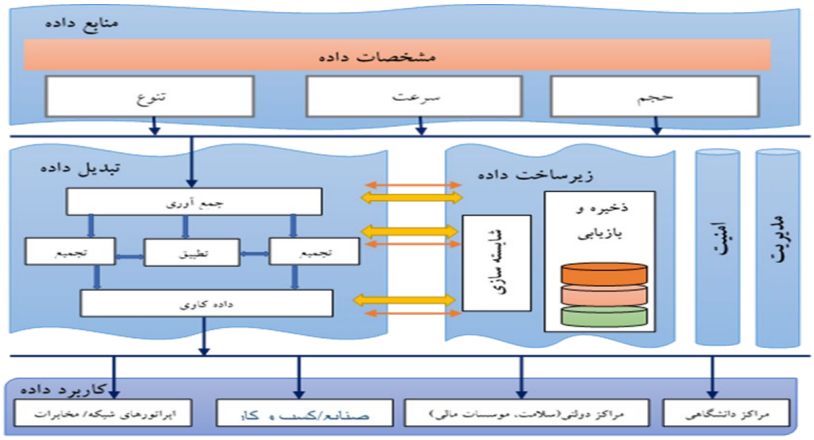
۳) معماری مرجع تحلیل و بهینه‌سازی حرفه *IBM*: در این مدل، ۸ لایه عمودی شامل دسترسی، تحلیل پیشرفته، *BIPM*، انباره‌های داده، یکپارچه‌سازی داده، مدیریت داده، مدیریت محتوا و منابع مشخص‌شده و پارامترهای مرتبط با هرکدام در داخل ستون مربوط به آن گنجانده‌شده است. چهار لایه افقی نیز شامل مدیریت فرایند حرفه/ جریان کار، حکمرانی اطلاعات (شامل مباحث مربوط به فراداده و کیفیت داده) و اتصالات شبکه، پروتکل‌ها و میان‌افزارهای دسترسی قرارگرفته‌اند. در شکل (۴) این مدل نشان

1. *Open Data Center Alliance (ODCA)*

داده شده است.




**شکل ۳. چارچوب معماری زیست‌بوم کلان‌داده مایکروسافت** (لوین، ۲۰۱۳)




**شکل ۴. چارچوب معماری مرجع تحلیل و بهینه‌سازی حرفه *IBM* ( دمچنکو، یوری و انگو، ۲۰۱۳)**

چند چارچوب معماری دیگر مانند *HPCC*، *TMF*، *LexisNexis* و *NIST* در پژوهش (دمچنکو، یوری و انگو، ۲۰۱۳) بررسی‌شده است. از دیدگاه دمچنکو، استفاده از سرویس‌های رایانش ابری، پارادایم جدیدی در امنیت کلان‌داده ایجاد کرده است؛ که با مکانیسم‌های سنتی مانند احراز هویت، مجاز شناسی و کنترل دسترسی متفاوت است و نیازمند رمزنگاری اجباری است.

روش تحقیق

تحقیق حاضر ازنظر هدف، کاربردی و ازنظر ماهیت و چگونگی گردآوری داده‌ها، توصیفی- پیمایشی است. همچنین از جنبه زمانی، مقطعی و ازنظر نوع داده‌ها، تحقیقی کمّی است. جامعه آماری مورد نظر در این پژوهش را خبرگان آشنا با معماری سازمانی، امنیت فضای سایبر و فنّاوری کلان‌داده به تعداد ۵۵ نفر تشکیل می‌دهند که از میان آن‌ها ۳۵ نفر به‌عنوان نمونه آماری انتخاب شدند. به‌منظور ارزیابی تأثیر چارچوب معماری ارائه شده بر شاخص‌های امنیتی کلان‌داده از یک پرسشنامه محقق ساخته شامل ۲۳ سؤال که به‌صورت طیف پنج سطحی لیکرت تدوین شده؛ استفاده شده است. برای سنجش پایایی و روایی تحقیق از نرم‌افزار اسمارت پی. آل. اس. استفاده شده است.

## یافته‌های تحقیق

چارچوب معماری پیشنهادی

به‌منظور دستیابی به یک چارچوب معماری جامع و یکپارچه، لازم است تمام مؤلفه‌های اثرگذار و اثرپذیر در حوزه فنّاوری کلان‌داده شناسایی گردد. در این تحقیق، سه چارچوب معماری مهم برای کلان‌داده مورد توجه قرار گرفته است. اجزای اصلی این چارچوب‌ها در جدول (۱) آمده است.

**جدول ۱. اجزای اصلی سه چارچوب معماری مورد نظر**

| چارچوب معماری | اجزای اصلی |
|---|---|
| چارچوب معماری ODCA | سرویس‌های پشتیبانی و ذخیره‌سازی، مدیریت کیفیت داده، مدیریت فراداده، مدیریت امنیت‌داده، بسترهای تبادل، دسترسی و توزیع یکپارچه |
| چارچوب معماری مایکروسافت | منابع داده، تبدیل داده، زیرساخت داده، کاربرد داده، مدیریت،امنیت |
| چارچوب معماری IBM | بستر سخت‌افزاری و نرم‌افزاری، اتصالات شبکه و پروتکل‌ها، حکمرانی اطلاعات، مدیریت فرایند، منابع، مدیریت محتوا، مدیریت داده، یکپارچه‌سازی داده، دسترسی |


در چارچوب‌های معماری، اغلب از مدل‌های لایه‌ای استفاده می‌شود که هر لایه دارای عناصر خاصی است. نحوه همبندی، اولویت‌بخشی و ارتباط میان لایه‌ها از اهمیت زیادی برخوردار است. با بررسی چارچوب‌های مطرح شده، پس از برگزاری جلسات خبرگی، در این تحقیق اجزای اصلی چارچوب معماری پیشنهادی بر اساس مدل لایه‌ای مطابق شکل (۵) در نظر گرفته شد.

شکل ۵. لایه‌های چارچوب معماری پیشنهادی

این چارچوب معماری، شامل هفت لایه زیرساخت سخت‌افزاری و ارتباطی، بستر محاسباتی، تحلیل داده، سرویس‌ها و خدمات و لایه کاربرد است. دو لایه مقررات، قوانین و استانداردها به‌صورت عمودی در کنار پنج لایه دیگر در نظر گرفته شده است و لایه امنیت نیز همه لایه‌ها و جنبه‌ها را در برمی‌گیرد. با توجه به تولید و تبادل روزافزون داده‌ها از منابع متنوع، به‌ویژه شبکه‌های اجتماعی و تبعات سوء امنیتی اجتماعی و فرهنگی ناشی از بهره‌برداری از این اطلاعات در کشور، مسئله امنیت در سه بخش

داده، خدمات و زیرساخت از اولویت و اهمیت خاصی برخوردار است که با قراردادن آن در لایه محاطی چارچوب معماری مورد تأکید قرار گرفته است.

نقش و جایگاه امنیت کلان‌داده در چارچوب موردنظر و قراردادن آن در لایه بیرونی ساختار لایه‌ای، نشانگر این دیدگاه است که شناسایی و پاسخگویی به حملات و تهدیدات سایبری برای مصون ماندن لایه‌های داخلی امری ضروری است. توجه به جنبه‌های قانونی و استاندارد نیز در این معماری باید در تمام لایه‌ها و فرایندها لحاظ شود. در ادامه ویژگی‌های لایه‌های معماری به‌صورت کلی بررسی شده است.

در لایه زیرساخت، منابع داده و تجهیزات ارتباطی قرار دارند. داده‌ها از طریق حسگرهای محیطی، شبکه وب جهانی، شبکه‌های اجتماعی، صنایع و سازمان‌های تجاری تولید شده و از طریق تجهیزات ارتباطی انتقال می‌یابند. وجود مراکز ملی داده از الزاماتی است که در این بخش باید مورد توجه قرار گیرد. این کار توسط بعضی از سازمان‌ها انجام‌شده است؛ برای مثال در پایگاه داده علوم زمین، اطلاعات مربوط به اطلس زمین‌شناسی، معادن، نفت و گاز، ژئوفیزیک و غیره قرار دارد. مرکز ملی داده‌های سلامت نیز از طرح‌های کلان ملی در این زمینه است. اطلاعات و محتوای ملی تولیدشده در حوزه‌های مختلف کلان‌داده باید به شبکه ملی اطلاعات به‌عنوان زیرساخت ارتباطی فضای مجازی کشور مرتبط شود؛ تا امنیت لازم برای آن فراهم گردد. این شبکه دارای مزایایی مانند سرعت دسترسی بالا، عرضه محتوای سالم و مناسب و حفظ امنیت ملی زمینه‌ساز ایجاد و تقویت محتوای بومی و ملی در کشور است.

در لایه بستر محاسباتی قالب‌های مختلف داده‌های ساختاریافته و غیر ساختاریافته که در پایگاه‌های داده قرار دارند؛ با استفاده از نرم‌افزارهای برنامه‌نویسی پردازش می‌شوند؛ تا در لایه‌های بالاتر مورداستفاده قرار گیرند. در این لایه باید به خاطر تنوع داده‌های مرتبط با کلان‌داده، از پایگاه داده مشابه هادوپ و اسپارک در کشور استفاده شود.

در لایه تحلیل داده[1] داده‌های حجیم جمع‌آوری‌شده پس از دسته‌بندی، به‌منظور استخراج اطلاعات مفید برای تصمیم‌گیری مورد تجزیه‌وتحلیل قرار می‌گیرند. این لایه در حوزه‌های مختلفی مانند اقتصاد، بازاریابی، پزشکی و مهندسی کاربرد دارد. در معماری پیشنهادی، این لایه شامل سه بخش داده‌کاوی، استخراج دانش و مصورسازی در نظر گرفته شده است.

در بخش داده‌کاوی[2]، داده‌های جمع‌آوری‌شده در لایه بستر محاسباتی که در پایگاه داده خاص کلان‌داده ذخیره‌شده‌اند؛ به‌منظور استخراج اطلاعات نهان یا الگوها و روابط مشخص مورد استفاده قرار می‌گیرند. داده‌کاوی بر مدل‌سازی و کشف دانش از داده‌ها، برای اهداف قابل پیش‌بینی و نه صرفاً توصیفی متمرکز است. استفاده از داده‌کاوی در بخش دولتی اهمیت زیادی در مباحث مربوط به تأمین امنیت

1. *Data analysis*
2. *Data Mining*

ملی دارد و به‌عنوان ابزاری برای کشف جرم، کلاه‌برداری و ردیابی فعالیت‌ها استفاده می‌شود. در بخش خصوصی نیز داده‌کاوی در اموری مانند بانکداری، بیمه، بهداشت و بازاریابی کاربرد دارد. در این بخش با فنّاوری خوشه‌بندی، یادگیری ماشینی، هوش مصنوعی و مباحث ریاضی و آماری مرتبط با داده مواجه هستیم. در بخش استخراج دانش، داده‌های اولیه پس از یکپارچه‌سازی، فیلترینگ و اعتبارسنجی به داده‌های مفهومی و دانشی تبدیل می‌شوند. در بخش مصورسازی، اطلاعات ناشی از بخش استخراج دانش دسته‌بندی‌شده و به اطلاعات قابل‌فهم برای کاربران تبدیل می‌شوند. هرچه این اطلاعات شفاف‌تر و جذاب‌تر باشند و به‌صورت آنلاین در اختیار کاربران قرار گیرند؛ تصمیمات بهتر و دقیق‌تری اتخاذ می‌شود.

در لایه سرویس و خدمات با استفاده از رایانش ابری سرویس‌های مختلف آنلاین و آفلاین به کاربران ارائه می‌شود. این لایه شامل خدمات رایانش ابری مانند داده به‌عنوان سرویس[1] (DaaS)، نرم‌افزار به‌عنوان سرویس[2] (SaaS)، پلتفرم به‌عنوان سرویس[3] (PaaS) و یا زیرساخت به‌عنوان سرویس[4] (IaaS) است.

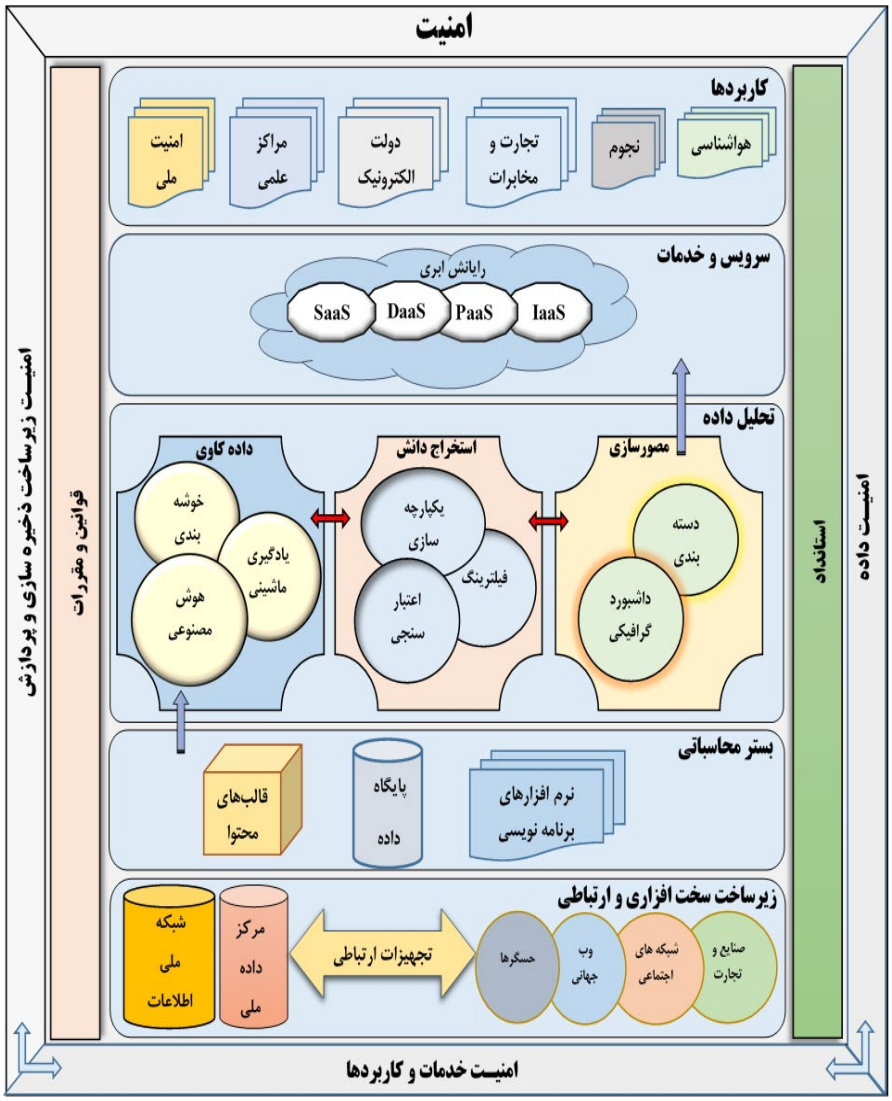
در لایه کاربرد، کاربردهای مختلف کلان‌داده در تحقیقات علمی و دانشگاهی، نجوم، شبکه‌های اجتماعی، تجارت و بورس، سلامت، هواشناسی و غیره عرضه می‌شود.

لایه مقررات و قوانین برای احاطه بر حوزه کلان‌داده و رفع مشکلات حقوقی توسط دولت به کار گرفته می‌شود. این لایه به‌صورت عمودی در تمام لایه‌ها قرار دارد و مسائل حقوقی و قانون‌گذاری در تمام حوزه‌های کلان‌داده را در برمی‌گیرد.

در لایه استاندارد با تدوین استانداردهای مشخص به فراگیری خدمات کلان‌داده پرداخته می‌شود. این لایه نیز به‌صورت عمودی با لایه‌های دیگر قرار می‌گیرد.

با توجه به مطالب مطرح شده در مورد هر لایه، اجزاء چارچوب معماری پیشنهادی در شکل (۶) ارائه شده است.

1. *Data as a Service*
2. *Software as a Service*
3. *Platform as a Service*
4. *Infrastructure as a Service*




**شکل ۶. چارچوب معماری پیشنهادی برای کلان‌داده**

لایه امنیت در چارچوب معماری

چارچوب‌های معماری بررسی‌شده، متناسب با موقعیت و وضعیت سازمان بهره‌بردار طراحی شده‌اند و اطلاعات جزئی آن‌ها منتشر نشده است. در چارچوب معماری اتحادیه مرکز داده (ODCA)، مدیریت امنیت‌داده به‌عنوان یک بخش در کنار مدیریت فراداده و مدیریت کیفیت داده آمده است؛ اما امنیت

خدمات و زیرساخت و وضعیت حقوقی و قوانین در آن مشخص نشده است. در چارچوب معماری مرجع زیست‌بوم کلان‌داده مایکروسافت، امنیت به‌عنوان یک لایه عمودی در کنار لایه زیرساخت داده و تبدیل داده قرار گرفته و محیط بر همه فرایندها ازجمله لایه کاربرد داده‌ها نیست. در چارچوب معماری مرجع تحلیل و بهینه‌سازی حرفه ($IBM$)، ازنظر اجزاء نسبت به چارچوب‌های قبلی جامع‌تر است؛ اما به مسئله امنیت کلان‌داده در آن پرداخته نشده است.

در چارچوب معماری پیشنهادی برای کلان‌داده، با توجه به حساسیت چالش‌های امنیتی موجود و بالقوه، درزمینهٔ به‌کارگیری این فنّاوری در سازمان‌های مختلف دولتی و خصوصی، تعیین شاخص‌های امنیتی کلان‌داده و نوع ارتباط آن با چارچوب معماری مورد نظر باید به شکل مناسبی مدنظر قرار گیرد. در ایران، انحصار داده‌های در حد کلان‌داده در اختیار دولت است و دولت در این حوزه به‌عنوان بازیگر اصلی، نقش نظارت و رگولاتوری دارد؛ بنابراین در طراحی چارچوب معماری، علاوه بر جنبه‌هایی مانند اجزاء فرایندها، وضعیت و ظرفیت‌های داخلی، باید دیدگاه‌های مختلف مرتبط با معماری مانند برنامه‌ریزی، مالکیت، خدمات و کاربردهای داده، در سطح ملی موردتوجه قرار گیرد

امنیت کلان‌داده از مهم‌ترین موضوعات مطرح در کلان‌داده است که تضمین‌کننده صحت دسترسی، کنترل دسترسی، مجاز شناسی، احراز هویت و سایر مکانیسم‌های امنیتی است. در مدل پیشنهادی امنیت را در تمام مراحل مدیریت داده از منابع داده تا کاربردهای آن باید در نظر بگیریم. لذا امنیت بر تمام لایه‌های هفتگانه محاط شده است.

بر اساس چارچوب معماری پیشنهادی، امنیت کلان‌داده شامل سه بخش اصلی امنیت‌داده، امنیت خدمات و کاربردها و امنیت زیرساخت ذخیره‌سازی و پردازش در نظر گرفته شد که ویژگی‌های هرکدام در ادامه آمده است.

امنیت‌داده: مکانیسم‌های امنیت سنتی مانند محرمانگی، یکپارچگی، دسترس‌پذیری، احراز هویت، عدم انکار و مجاز شناسی در داده‌های کلان نیز باید مورد توجه قرار گیرد. در این حالت به علت حجم زیاد، سرعت و تنوع داده، نیاز به الگوریتم‌هایی داریم که علاوه بر قابلیت اطمینان و یکپارچگی، از سرعت کافی در ثبت اطلاعات، اصلاح و یا حذف آن‌ها برخوردار باشند. برای بهبود کارایی، داده‌ها بین گره‌های مختلف تقسیم‌شده و به اشتراک گذاشته می‌شوند. انتقال خودکار داده‌ها بین گره‌های مختلف شناسایی محل قرارگیری داده در هرلحظه و تعداد گره‌های موجود را دشوار می‌سازد؛ بنابراین استفاده از مدل‌های قدیمی امنیت متمرکز، جوابگو نیست.

امنیت خدمات و کاربردها: برای صرفه‌جویی در هزینه سازمان، منابع داده را می‌توان بین برنامه‌ها، داده‌ها و کاربران مختلف، به اشتراک گذاشت. این کار هزینه و بهره‌وری سیستم را برای ارائه خدمات بالاتر می‌برد.

استفاده از سامانه‌های پردازش مانند هادوپ، مپ ردیوس و اسپارک از این روش استفاده می‌کنند. کلان‌داده برای امنیت سیستم نیز استفاده می‌شود. با استفاده از تجزیه‌وتحلیل داده‌ها، مانیتورینگ رویدادها و مدیریت اطلاعات امنیتی، نفوذپذیری به سیستم، کنترل می‌شود. چند نمونه از کاربردهای کلان‌داده در امنیت سیستم عبارت‌اند از: سامانه‌های مدیریت اطلاعات امنیتی و رویدادها (SIEM)[1]، سامانه‌های پیشگیری و تشخیص نفوذ (IPS/IDS)، سامانه‌های هشدار زودهنگام[2] (EWS) (فرانکس، ۲۰۱۲).

امنیت زیرساخت ذخیره‌سازی و پردازش: در کلان‌داده از فایل سامانه‌های توزیع‌شده برای ذخیره‌سازی جریان کارهای مربوط به کلان‌داده استفاده می‌شود. در این حالت نیز مباحث امنیتی باید مورد توجه قرار گیرد. در محرمانگی از رمزنگاری و کدینگ اطلاعات، استفاده می‌شود. برای جامعیت داده از روش‌های درهم‌سازی و امضای دیجیتال استفاده می‌شود و با استفاده از مکانیسم‌های کنترل دسترسی و احراز هویت، کاربر مجاز و میزان دسترسی او مشخص می‌شود.

تأثیر چارچوب معماری پیشنهادی بر امنیت کلان‌داده

در چارچوب‌های معماری معرفی‌شده، امنیت به‌عنوان بخشی از چارچوب در نظر گرفته شده است. درحالی‌که امنیت تنها یک مکانیسم نیست؛ بلکه فرایندی است که باید در تمامی مراحل پیاده‌سازی و مؤلفه‌های یک چارچوب معماری دخالت داشته باشد. در نمونه چارچوب‌های خاص امنیتی برای کلان‌داده نیز، صرفاً جنبه‌های فنی امنیت‌داده یا امنیت زیرساخت و سرویس‌ها مدنظر قرار گرفته و تنها به مسائل فنی و جزءنگر پرداخته شده است.

در چارچوب معماری پیشنهادی، امنیت کلان‌داده یک لایه از چارچوب معماری محسوب می‌شود؛ باوجوداین، با توجه به نوع پیکربندی و جایگاه آن در این چارچوب، می‌توان با فرض کردن آن، به‌عنوان متغیر وابسته تأثیر بقیه مؤلفه‌های معماری به‌عنوان متغیر مستقل بر آن را در یک مدل مناسب مورد ارزیابی قرارداد.

علاوه بر آن، باید توجه داشت که مباحث مرتبط با کلان‌داده از جنس فنی است و در برخی از منابع، برای ارزیابی معماری از روش‌هایی مانند $FAAM$[3] به‌منظور ارزیابی معماری سامانه‌های اطلاعاتی و $SAAM$[4] به‌منظور ارزیابی معماری برنامه‌های کاربردی استفاده شده است؛ اما موضوعاتی مانند طراحی چارچوب معماری در سطح کلان، نیازمند نظرسنجی از کارشناسان و خبرگان آشنا به مسائل فنی و مدیریتی است؛ بنابراین در این تحقیق، به‌منظور اطمینان از صحت ارتباط بین اجزاء انتخاب‌شده برای چارچوب معماری پیشنهادی، تأثیر چارچوب معماری پیشنهادی بر امنیت کلان‌داده مورد ارزیابی

1. *Security information and event management*
2. *Early Warning Systems*
3. *Family Architecture Assessment method*
4. *Scenario-Based Software Architecture Evaluation Methods*

قرار گرفته است. بدین منظور از یک پرسشنامه محقق ساخته استفاده شده است. در این پرسشنامه، شاخص‌های مرتبط با سازه‌ها (لایه‌های) چارچوب معماری و امنیت کلان‌داده مطابق جدول (۲) در نظر گرفته شد. برای تجزیه‌وتحلیل کمّی داده‌های دریافت شده توسط پرسشنامه از جامعه آماری، از نرم‌افزار SmartPLS استفاده شده است.

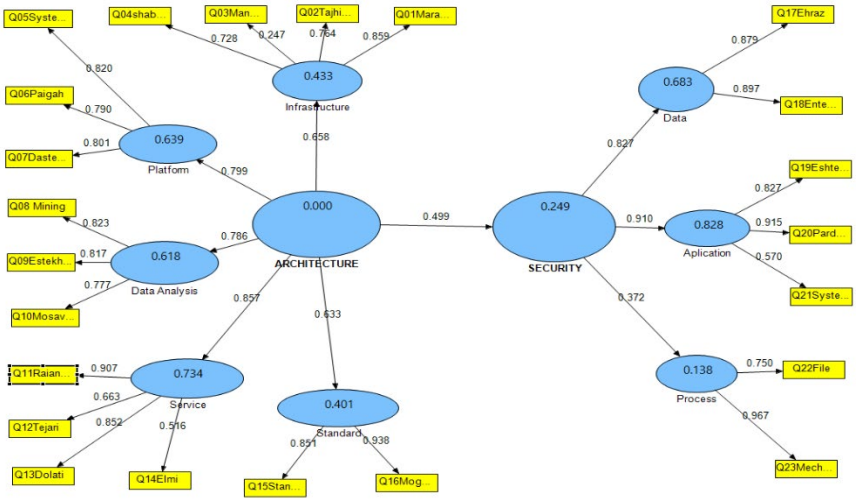
آزمون برازش مدل، شامل سه مرحله است. در مرحله اول، برای برازش مدل کلی، پایایی و روایی مورد ارزیابی قرار گرفته است. در مرحله دوم، برازش مدل ساختاری یعنی روابط میان متغیرهای مکنون بررسی شده و درنهایت در مرحله سوم برازش مدل کلی بررسی شده است.

**جدول ۲. ابعاد، سازه‌ها و شاخص‌های مدل پژوهش**

| ابعاد | سازه‌ها | شاخص‌ها | نشانگر در مدل |
|---|---|---|---|
| چارچوب معماری (ARCHITECTURE) | زیرساخت سخت‌افزاری و ارتباطی Infrastructure | ایجاد مراکز داده ملی | Marakez |
| | | استفاده از تجهیزات سخت‌افزاری و ارتباطی | Tajhizat |
| | | گسترش منابع جمع‌آوری داده | Manabe |
| | | راه‌اندازی شبکه ملی اطلاعات | Shabakemeli |
| | بستر محاسباتی Platform | استفاده از نرم‌افزار و سیستم‌عامل بومی | Systemamel |
| | | استفاده از پایگاه داده داخلی | Paigah |
| | | فنّاوری دسته‌بندی قالب‌های محتوایی داده | Dastebandi |
| | تحلیل داده Data Analysis | فن‌های داده‌کاوی | Mining |
| | | استخراج دانش | Estekhraj |
| | | مصورسازی | Mosavar |
| | خدمات و کاربردها Service | سرویس‌های رایانش ابری | Raianesh |
| | | کاربردهای تجاری (بورس، اپراتورهای مخابراتی و...) | Tejari |
| | | کاربردهای دولتی (امنیت ملی، دولت الکترونیک و...) | Dolati |

| ابعاد | سازه‌ها | شاخص‌ها | نشانگر در مدل |
|---|---|---|---|
| | | کاربردهای علمی، هواشناسی و اجتماعی | Elmi |
| | مقررات و استاندارد Standard | استانداردسازی | Standard |
| | | مقررات قانونی و حقوقی | Mogharart |
| امنیت کلان‌داده (SECURITY) | امنیت‌داده Data | سرویس‌های احراز هویت، کنترل دسترسی | Ehraz |
| | | فن‌های انتقال خودکار داده بین گره‌های مختلف و اشتراک آن | Enteghal |
| | امنیت کاربرد Application | اشتراک منابع داده بین برنامه‌ها، داده‌ها و کاربران مختلف | Eshterak |
| | | ایجاد سامانه‌های پردازش بومی مشابه هادوپ و ... | Pardazesh |
| | | ایجاد سامانه‌های پیشگیری و تشخیص نفوذ و هشدار | System |
| | امنیت زیرساخت و پردازش Process | ایجاد فایل سامانه‌های توزیع‌شده | File |
| | | مکانیسم‌های سخت‌افزاری و نرم‌افزاری کنترل دسترسی | Mechanism |

در مرحله اول برازش چارچوب معماری ارائه شده، برای پایایی شاخص‌ها از ضرایب بار عاملی، آلفای کرونباخ، پایایی ترکیبی و برای سنجش روایی، از روایی همگرا (ضرایب *AVE*) و روایی واگرا (ماتریس بار عاملی متقابل) استفاده شده است. با ترسیم مدل و اجرای فرمان *Calculate/PLS Algorithm* شکل (۷) حاصل می‌شود. در این حالت اعداد نوشته شده روی پیکان‌های متصل به شاخص‌ها ضرایب بار عاملی را نشان می‌دهند که این مقادیر باید بالای ۰/۴ باشند.




**شکل ۷. برازش مدل اندازه‌گیری**

در برازش اولیه مدل دیده می‌شود شاخص گسترش منابع جمع‌آوری در سازه زیرساخت دارای ضریب بار عاملی کمتر از ۰/۴ است و باید حذف شود. با استفاده از بخش *Report html* مقادیر آلفای کرونباخ پایایی ترکیبی و روایی همگرایی (*AVE*) محاسبه‌شده که در جدول (۳) این مقادیر آمده است. در این جدول مقدار آلفای کرونباخ برای همه مؤلفه‌ها به‌جز مؤلفه امنیت کاربرد، بالاتر از ۰/۷ است؛ بنابراین لازم است در مدل، اصلاحاتی انجام شود برای این کار مؤلفه‌ای که دارای کمترین ضریب بار عاملی است را حذف می‌کنیم. با توجه به مقادیر شکل (۷) مؤلفه سامانه‌های پیشگیری و تشخیص نفوذ دارای ضریب بار عاملی برابر ۰/۵۷۰ است که نسبت به دو مؤلفه دیگر کمتر است. پس از حذف آن و اجرای برنامه، مقدار ضریب آلفای کرونباخ برای مؤلفه کاربرد به‌جای ۰/۶۶۵۸ در جدول (۳) برابر ۰/۸۲۹۰ به دست می‌آید. مقادیر پایایی ترکیبی[1] محاسبه‌شده در جدول نیز همگی بالای ۰/۷ هستند.

**جدول ۳. آلفای کرونباخ، پایایی ترکیبی و روایی همگرایی**

| | روایی همگرا | پایایی ترکیبی | آلفای کرونباخ |
|---|---|---|---|
| چارچوب معماری | ۰/۳٤۱ | ۰/۸۸٤ | ۰/۸۵۸ |
| امنیت کاربرد | ۰/۶۱۵ | ۰/۸۲۲ | ۰/۶۶۵ |
| امنیت‌داده | ۰/۷۸۹ | ۰/۸۸۲ | ۰/۷۳۳ |
| تحلیل داده | ۰/۶٤۹ | ۰/۸٤۷ | ۰/۷۳۳ |

1. *Composite Reliability*

| | روایی همگرا | پایایی ترکیبی | آلفای کرونباخ |
|---|---|---|---|
| زیرساخت سخت‌افزاری و پردازش | ۰/۶۲۶ | ۰/۸۳۲ | ۰/۷۰۸ |
| بستر محاسباتی | ۰/۶۴۵ | ۰/۸۴۵ | ۰/۷۳۴ |
| امنیت زیرساخت و پردازش | ۰/۷۴۸ | ۰/۸۵۴ | ۰/۷۱۴ |
| امنیت کلان‌داده | ۰/۳۹۸ | ۰/۸۰۲ | ۰/۷۰۸ |
| خدمات و کاربردها | ۰/۵۶۳ | ۰/۸۳۱ | ۰/۷۲۵ |
| مقررات و استاندارد | ۰/۸۰۲ | ۰/۸۸۹ | ۰/۷۶۲ |

برای بررسی روایی همگرا از ضرایب *AVE* نشان داده‌شده در جدول (۴) استفاده می‌شود این معیار، میزان همبستگی هر سازه با سؤالات (شاخص‌های) خود را نشان می‌دهد. طبق روش فرونل و لارکر (داوری و رضازاده، ۱۳۹۳) این مقادیر باید بالای ۰/۵ باشند. با توجه به جدول (۴) دیده می‌شود که همه این مقادیر به‌جز مقادیر متغیرهای پنهان مرتبه دوم یعنی چارچوب معماری (*ARCHITECTURE*) و امنیت کلان‌داده (SECURITY) همگی بالای ۵/. هستند. برای محاسبه *AVE* مربوط به این متغیر، میانگین ضرایب بار عاملی که روی پیکان متصل به این متغیر نشان داده‌شده است را محاسبه می‌کنیم که مقادیر ۰/۵۶۵ و ۰/۵۵۰ به ترتیب برای مؤلفه‌های چارچوب معماری و امنیت کلان‌داده حاصل می‌شود. این مسئله روایی همگرایی مناسب برای مدل را نشان می‌دهد

روایی واگرایی میزان همبستگی یک سازه با شاخص‌هایش در مقابل همبستگی آن سازه با سایر سازه‌ها را نشان می‌دهد. برای محاسبه آن، از بخش *Latent Variable Correlation* نرم‌افزار *SmartPLS* استفاده می‌شود. برای این کار قطر اصلی به‌صورت $\sqrt{AVE}$ مربوط به سازه‌ها که در جدول (۴) آمده است؛ محاسبه می‌شود. همان‌طور که در این جدول دیده می‌شود؛ اکثر مقادیر از قطر اصلی کمتر از این مقدار هستند که این مسئله روایی واگرایی مناسب را نشان می‌دهد.

مرحله دوم برازش مدل ساختاری است که در آن تنها روابط متغیرهای پنهان بررسی می‌شود و با متغیر آشکار (سؤالات) نقشی ندارند. برای این کار از Z یا همان *t-values* با اجرای فرمان *Bootstrapping* استفاده می‌شود. با توجه به اینکه مقادیر محاسبه‌شده به‌جز متغیر سیستم پردازش بومی همگی از ۱/۹۶ بیشتر هستند لذا می‌توان گفت تمام سؤالات و روابط میان متغیرها در سطح اطمینان ۹۵ درصد معنادار هستند.

مرحله سوم برازش مدل کلی است؛ برای این کار از معیار GoF[1] که توسط تننهاوس[2] و همکاران ابداع شد استفاده شده است. این معیار از رابطه زیر محاسبه می‌شود.

1. *Goodness of Fit*
2. *Tenenhaus et al*

$$GOF = \sqrt{\overline{Communalities} \times \overline{R^2}} = \sqrt{۰.٦٤١٤ \times ۰.٥١٣٨} = ۰.٥٧٤۰$$

با توجه به اینکه GoF از ۰/۳۶ بزرگ‌تر است می‌توان نتیجه گرفت مدل از برازش کلی قوی، مناسب و قابل‌تأییدی برخوردار است.

**جدول ۴. نتایج روایی واگرایی سازه‌ها**

| | چارچوب معماری | امنیت کاربرد | امنیت داده | تحلیل داده | زیرساخت سخت‌افزاری | بستر محاسباتی | امنیت پردازش | امنیت کلان‌داده | خدمات و کاربرد | مقررات و استاندارد |
|---|---|---|---|---|---|---|---|---|---|---|
| چارچوب معماری | ۰/۷۵۲ | | | | | | | | | |
| امنیت کاربرد | ۰/۲۷۳ | ۰/۷۸٤ | | | | | | | | |
| امنیت داده | ۰/۱۹۵ | ۰/۵۷۷ | ۰/۸۸۸ | | | | | | | |
| تحلیل داده | ۰/۷۸٤ | ۰/۲۰۹ | ۰/۱۷۷ | ۰/۸۰۶ | | | | | | |
| زیرساخت سخت‌افزاری | ۰/۶۵٤ | ۰/۲۳۶ | ۰/۲۱۳ | ۰/۲۹۸ | ۰/۷۰۱ | | | | | |
| بستر محاسباتی | ۰/۷۹۸ | ۰/۰۸۳ | ۰/۱۱۷ | ۰/۶۸۳ | ۰/۳۲ | ۰/۸۰۳ | | | | |
| امنیت پردازش | ۰/۷۹۶ | ۰/۱٤۷ | ۰/۰۳۶ | ۰/۵۱۲ | ۰/۵۱۱ | ۰/۶۲۲ | ۰/۸۷ | | | |
| امنیت کلان‌داده | ۰/٤۹۱ | ۰/۸۶۲ | ۰/۸۲۵ | ۰/۳۸۰ | ۰/۳۵۱ | ۰/۳۱۲ | ۰/۳۷ | ۰/۷٤۱ | | |
| خدمات و کاربرد | ۰/۸۵۶ | ۰/۲۳۵ | ۰/۰۵٤ | ۰/٤۷۷ | ۰/۵۲ | ۰/۶٤ | ۰/۹۰۷ | ۰/٤۱٤ | ۰/۷۵ | |
| مقررات و استاندارد | ۰/۶۳۵ | ۰/۳۱٤ | ۰/۲۶۰ | ۰/٤۶۱ | ۰/٤۰۷ | ۰/۲٤۰ | ۰/۳۰٤ | ۰/٤۲۹ | ۰/٤۵ | ۰/۸۹۶ |

## نتیجه‌گیری

نتایج تجزیه‌وتحلیل بررسی تأثیر چارچوب معماری بر شاخص‌های امنیت کلان‌داده، نشانگر آن است که در این پژوهش، شاخص‌های بستر محاسباتی، خدمات و کاربردها، مقررات و استاندارد مربوط به چارچوب معماری مورد تائید خبرگان قرار گرفت و تنها شاخص حذف‌شده گسترش منابع جمع‌آوری داده، مربوط به سازه زیرساخت سخت‌افزاری و ارتباطی است. دلیل این مسئله می‌تواند نگرانی جامعه آماری از تأمین امنیت منابع جمع‌آوری داده باشد. همچنین سازه‌های داده، کاربرد و زیرساخت و پردازش

ازنظر پرسش‌شوندگان برای بعد امنیت کلان‌داده مناسب تشخیص داده شد و تنها شاخص امنیت سامانه‌های پیشگیری و تشخیص نفوذ و هشدار برای داشتن آلفای کرونباخ بالای ۰/۷ قابل‌حذف تشخیص داده شد. این مسئله نیز می‌تواند بیانگر ماهیت فرامرزی تهدیدات سایبری و دیدگاه جامعه آماری در مورد ضرورت تغییر در نحوه برخورد با این تهدیدات از طریق سامانه‌های پیشگیری و تشخیص نفوذ سنتی باشد. مدل ازنظر ساختاری و برازش کلی نیز از مقادیر مناسب و قابل قبولی برخوردار است.

درزمینهٔ چارچوب معماری مربوط به کلان‌داده در کشور، پژوهش مشابهی انجام نشده است. در این مقاله پس از بررسی مفاهیم نظری کلان‌داده و چالش‌های مدیریتی، فنّاورانه و امنیتی مرتبط با آن، با محور قراردادن امنیت کلان‌داده در سطح کلان برای سازمان‌های دولتی و خصوصی، مؤلفه‌ها و شاخص‌های یک چارچوب معماری ارائه شده است.

چارچوب معماری حاضر تلفیقی از چارچوب‌های مورد مطالعه است که سعی شده است علاوه بر داشتن جنبه‌های موجود در معماری‌های مطرح‌شده به شکل مناسب‌تر و خوش‌تعریف تری ارائه شود. مقایسه میان چارچوب معماری پیشنهادی با چارچوب‌های معماری موجود برای کلان‌داده می‌تواند از طریق روش‌هایی مانند *AHP* مورد بررسی قرار گیرد. این کار مستلزم شناسایی جزئیات فنی سایر معماری‌های ارائه‌شده و خبرگان مسلط به جزئیات این معماری‌هاست و می‌تواند در تحقیقات آینده بررسی شود.

توجه به جنبه‌های قانونی و استاندارد در معماری پیشنهادی و لحاظ کردن امنیت در بخشه‌ای داده، زیرساخت، خدمات و کاربردهای به‌منظور محاط بودن آن بر تمامی اجزاء و فرایندها و توانایی پاسخگویی در برابر طیف وسیعی از حملات و تهدیدات سایبری در کشور است (نوع نگرش به امنیت در کشورها متفاوت است. مقوله امنیت فضای سایبر از مباحث چالشی مرتبط با امنیت ملی در کشور است که در مبحث کلان‌داده نیز باید به آن توجه ویژه‌ای داشت)

بومی‌سازی صنایع و فنّاوری‌های مربوط به کلان‌داده در افزایش بهره‌وری و توسعه خدمات نقش سازنده‌ای دارد و زیربنای خودکفایی، عدم وابستگی و تأمین امنیت ملی است. با توجه به چالش‌ها و تهدیدات امنیتی درزمینهٔ پذیرش و ورود فنّاوری کلان‌داده به کشور، تبیین سازه‌ها و شاخص‌های امنیت کلان‌داده و بررسی میزان تأثیر بقیه سازه‌ها و شاخص‌های چارچوب معماری بر آن به‌طور کمّی در این پژوهش، مورد تأکید قرار گرفته است. ارزیابی معیارها و شاخص‌های کارایی معماری پیشنهادی، از قبیل سرعت پاسخ‌گویی[1]، توان عملیاتی[2] و مقیاس‌پذیری[3] با افزایش حجم داده‌ها و تعداد درخواست‌ها می‌تواند در تحقیقات آتی مورد بررسی قرار گیرد.

1. *Response Time*
2. *Throughput*
3. *Scalability*

## منابع

- داوری، علی، رضازاده، آرش (۱۳۹۳). **مدل‌سازی معادلات ساختاری با نرم‌افزار PLS**. تهران، انتشارات جهاد دانشگاهی
- معینی، علی. مرآتی، احسان. (۱۳۹۴) **تدوین روش توسعه چارچوب معماری سازمانی: مطالعه پدیدارشناسی تفسیری**. مدیریت فنّاوری اطلاعات، دوره ۷ شماره ۱: ۱۶۲-۱۴۳.

- Chen, C. P, & Zhang, C. Y. (2014). **Data-intensive applications, challenges, techniques and technologies: *A survey on* Big Data**. *Information Sciences*, 275, 314-347.
- Chen, J. Chen, Y. Du, X. Li, C. Lu, J. Zhao, S. & Zhou, X. (2013). **Big data challenge: *a data management perspective***. *Frontiers of Computer Science*, 7(2): 157-164.
- Chen, M. Mao, S. & Liu, Y. (2014). **Big data: *A survey.*** *Mobile Networks and Applications*, 19(2): 171-209.
- Demchenko, Y. Ngo, C. & Membrey, P. (2013). ***Architecture framework and components for the big data ecosystem***. *Journal of System and Network Engineering*, 1-31.
- Depardon, B. Le Mahec, G. & Séguin, C. (2013). ***Analysis of six distributed file systems***. *Agence Nationale de la Recherche*, 1-45.
- EMC. (2012). **Big Data, Bigger Digital Shadows and Biggest Growth inhe Far East** *(IDC Digital Universe Study)*.
- Estes, M. Moty, F. (2013). **Master Usage Model، *Information as a Service*** *Rev.* 1.0. *Open Data Center Alliance Inc.*
- *European Big Data Value cPPP* - **Strategic Research and Innovation Agenda**. (2014). *European Big Data Value Strategic Research & Innovation Agenda*.
- *Franks*, B. (2012). ***Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics*** *(Vol. 49). John Wiley & Sons.*
- *Hashem, I. A. T. Yaqoob, I. Anuar, N. B. Mokhtar, S. Gani, A. & Khan, S. U. (2015).* ***The rise of "big data" on cloud computing: Review and open research issues***.

*Information Systems*, 47: 98-115.

- Hilbert, M. (2013). **Big data for development: From information-to knowledge societies**.
- Inoubli, W. Aridhi, S. Mezni, (2018). ***An experimental survey on big data frameworks,*** Future Generation Computer Systems.
- ITU-T Technology Watch Report. (2013). **Big Data: Big today, normal tomorrow.** Switzerland, Geneva.
- Levin, O. (2013). **Big Data Ecosystem Reference Architecture Microsoft Corporation**. Microsoft Corporation.
- Matti, M. O. N. A. & Kvernvik, T. (2012). ***Applying big-data technologies to network architecture***. Ericsson Review, 284: 23-31.
- Mert, O. Kerem, K. et al (2017). **Big-Data Analytics Architecture for Businesses: a comprehensive review on new open-source big-data tools**. Cambridge Service Alliance.
- Mills, S. e. a. (2012). **Demystifying Big Data: *A practical guide to transforming the business of Government, Technical report.*** Washington, D.C: TechAmerica Foundation.
- Moreno J. Manuel A. et al (2018). **Towards a Security Reference Architecture for Big Data,** International Journal of Scientific & Engineering Research Volume 8, Issue 10.
- Mishra, S. (2014). **Survey of Big Data Architecture and Framework from the Industry.** NIST Big Data Public Working Group.
- Pandey, N. Jena, P. et al. (2016). ***A Review on Big Data Architecture,*** International Journal of Trend in Research and Development, Volume 3(6).