

Diterima 25 November 2016, diterima 14 Maret 2017, tanggal publikasi 29 Maret 2017, tanggal versi saat ini 17 Mei 2017.

Pengenal Objek Digital 10.1109/ACCESS.2017.2689040

Analisis Data IoT Besar: Arsitektur, Peluang, dan Terbuka Tantangan Penelitian

MOHSEN MARJANI¹, FARIZA NASARUDDIN², ABDULAH GANI¹ (Anggota Senior, IEEE), AHMAD KARIM³ DAN , IBRAHIM ABAKER TARGIO HASHEM¹, AISHA SIDDIQA¹ ,

IBRAR YAQOOB¹ 1Jurusan

Sistem dan Teknologi Komputer, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Malaya, Kuala Lumpur, Malaysia

2Jurusan Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Malaya, Kuala Lumpur 50603 , Malaysia

3Departemen Teknologi Informasi, Universitas Bahaiddin Zakariya, Multan, Punjab, Pakistan Penulis koresponden: Fariza Nasaruddin

(fariza@um.edu.my); Abdullah Gani (abdullah@um.edu.my)

Pekerjaan ini didukung oleh Kementerian Pendidikan Tinggi Malaysia di bawah University Malaya Research Grant (UMRG)

Projek/Program UM.0000168/HRU.RP.IT dan RP029D-14AET.

ABSTRAK Sejumlah besar data telah dihasilkan, sejak dekade terakhir seiring dengan meningkatnya miniaturisasi perangkat Internet of things (IoT). Namun, data tersebut tidak berguna tanpa kekuatan analitik.

Banyak solusi data besar, IoT, dan analitik telah memungkinkan orang memperoleh wawasan berharga tentang data besar yang dihasilkan oleh perangkat IoT. Namun, solusi ini masih dalam tahap awal, dan domain tersebut tidak memiliki survei yang komprehensif. Makalah ini menyelidiki upaya penelitian mutakhir yang diarahkan pada analitik data IoT besar. Hubungan antara analitik data besar dan IoT dijelaskan. Selain itu, makalah ini menambah nilai dengan mengusulkan arsitektur baru untuk analitik data IoT yang besar. Selanjutnya, jenis, metode, dan teknologi analitik data IoT besar untuk penambangan data besar dibahas. Banyak kasus penggunaan penting juga disajikan.

Beberapa peluang yang dibawa oleh data analytics dalam paradigma IoT kemudian dibahas. Terakhir, tantangan penelitian terbuka, seperti privasi, penambangan data besar, visualisasi, dan integrasi, disajikan sebagai arah penelitian masa depan.

ISTILAH INDEKS Data besar, Internet of Things, analisis data, komputasi terdistribusi, kota pintar.

I. PENDAHULUAN

Perkembangan big data dan Internet of things (IoT) dengan cepat mempercepat dan mempengaruhi semua bidang teknologi dan bisnis dengan meningkatkan manfaat bagi organisasi dan individu. Pertumbuhan data yang dihasilkan melalui IoT telah memainkan peran utama dalam lanskap data besar. Big data dapat dikategorikan menurut tiga aspek: (a) volume, (b) variasi, dan (c) kecepatan [1]. Kategori ini pertama kali diperkenalkan oleh Gartner untuk menggambarkan elemen tantangan data besar [2]. Peluang besar disajikan oleh kemampuan untuk menganalisis dan memanfaatkan data IoT dalam jumlah besar, termasuk aplikasi di kota pintar, sistem transportasi dan jaringan pintar, pengukur energi pintar, dan perangkat pemantauan perawatan kesehatan pasien jarak jauh.

Popularitas IoT yang meluas telah membuat analisis data besar menjadi menantang karena pemrosesan dan pengumpulan data melalui berbagai sensor di lingkungan IoT.

Laporan International Data Corporation (IDC) menunjukkan bahwa pasar big data akan mencapai lebih dari US\$125 miliar

pada tahun 2019 [3]. Analisis data besar IoT dapat didefinisikan sebagai langkah-langkah di mana berbagai data IoT diperiksa [4] untuk mengungkapkan tren, pola yang tidak terlihat, korelasi tersembunyi, dan informasi baru [5]. Perusahaan dan individu dapat mengambil manfaat dari menganalisis sejumlah besar data dan mengelola sejumlah besar informasi yang dapat mempengaruhi bisnis [6]. Oleh karena itu, analitik data besar IoT bertujuan untuk membantu asosiasi bisnis dan organisasi lain untuk mencapai pemahaman yang lebih baik tentang data, dan dengan demikian, membuat keputusan yang efisien dan terinformasi dengan baik.

Analisis data besar memungkinkan penambang data dan ilmuwan untuk menganalisis sejumlah besar data tidak terstruktur yang dapat dimanfaatkan menggunakan alat tradisional [5]. Selain itu, analitik data besar bertujuan untuk segera mengekstrak informasi yang berpengetahuan luas menggunakan teknik penambangan data yang membantu dalam membuat prediksi, mengidentifikasi tren terkini, menemukan informasi tersembunyi, dan membuat keputusan [7].

Teknik dalam penambangan data digunakan secara luas untuk metode khusus masalah dan analitik data umum.

Oleh karena itu, metode statistik dan pembelajaran mesin adalah

dimanfaatkan. Data IoT berbeda dari data besar normal yang dikumpulkan melalui sistem dalam hal karakteristik karena berbagai sensor dan objek yang terlibat selama pengumpulan data, yang meliputi heterogenitas, kebisingan, variasi, dan pertumbuhan yang cepat. Statistik [8] menunjukkan bahwa jumlah sensor akan meningkat sebesar 1 triliun pada tahun 2030. Peningkatan ini akan mempengaruhi pertumbuhan big data. Memperkenalkan analitik data dan IoT ke dalam data besar membutuhkan sumber daya yang besar, dan IoT memiliki kemampuan untuk menawarkan solusi yang sangat baik. Sumber daya yang tepat dan aplikasi intensif dari platform disediakan oleh layanan IoT untuk komunikasi yang efektif di antara berbagai aplikasi yang digunakan. Proses tersebut cocok untuk memenuhi persyaratan aplikasi IoT, dan dapat mengurangi beberapa tantangan di masa depan analitik data besar. Penggabungan teknologi ini meningkatkan kemungkinan penerapan IoT ke arah yang lebih baik. Selain itu, menerapkan IoT dan solusi integrasi data besar dapat membantu mengatasi masalah pada penyimpanan, pemrosesan, analisis data, dan alat visualisasi. Itu juga dapat membantu dalam meningkatkan kolaborasi dan komunikasi di antara berbagai objek di kota pintar [9]. Area aplikasi, seperti lingkungan ekologi cerdas, lalu lintas cerdas, jaringan cerdas, bangunan cerdas, dan manajemen cerdas logistik, dapat memanfaatkan pengaturan yang disebutkan di atas. Banyak penelitian tentang data besar berfokus pada manajemen data besar; khususnya, analitik data besar telah disurvei [10], [11]. Namun, survei ini berfokus pada data besar IoT dalam konteks analitik dari sejumlah besar data. Kontribusi dari survei ini adalah sebagai berikut. a) Upaya penelitian mutakhir yang dilakukan dalam hal analitik data besar diselidiki. b) Sebuah arsitektur untuk analitik data IoT besar diusulkan. c) Beberapa peluang yang belum pernah terjadi sebelumnya yang dibawa oleh analitik data di domain IoT diperkenalkan. d) Kasus penggunaan yang kredibel disajikan. e) Tantangan penelitian yang masih harus ditangani diidentifikasi dan didiskusikan.

Kontribusi ini disajikan dari Bagian 3 sampai 6. Kesimpulan disediakan di Bagian 7.

II. IKHTISAR IoT DAN DATA BESAR

Ikhtisar teknologi IoT dan data besar diberikan sebelum diskusi.

A. IoT

IoT menawarkan platform untuk sensor dan perangkat untuk berkomunikasi secara mulus dalam lingkungan yang cerdas dan memungkinkan berbagi informasi di seluruh platform dengan cara yang nyaman. Adaptasi terbaru dari teknologi nirkabel yang berbeda menempatkan IoT sebagai teknologi revolusioner berikutnya dengan memanfaatkan peluang penuh yang ditawarkan oleh teknologi Internet. IoT telah menyaksikan adopsi baru-baru ini di kota-kota pintar dengan minat dalam mengembangkan sistem cerdas, seperti kantor pintar, ritel pintar, pertanian pintar, air pintar, transportasi pintar, perawatan kesehatan pintar, dan energi pintar [12], [13]. IoT telah muncul sebagai tren baru dalam beberapa tahun terakhir, di mana perangkat seluler, fasilitas transportasi,

fasilitas umum, dan peralatan rumah tangga semuanya dapat digunakan sebagai peralatan akuisisi data di IoT. Semua peralatan elektronik di sekitar untuk memudahkan operasi kehidupan sehari-hari, seperti jam tangan, mesin penjual otomatis, alarm darurat, dan pintu garasi, serta peralatan rumah tangga, seperti lemari es, oven microwave, AC, dan pemanas air terhubung ke jaringan IoT dan dapat dikendalikan dari jarak jauh.

Ciuffo [14] menyatakan bahwa perangkat ini "berbicara" satu sama lain dan ke perangkat pengontrol pusat. Perangkat tersebut ditempatkan di daerah yang berbeda dapat mengumpulkan berbagai jenis data, seperti data geografis, astronomi, lingkungan, dan logistik.

Sejumlah besar perangkat komunikasi dalam paradigma IoT disematkan ke perangkat sensor di dunia nyata. Perangkat pengumpul data merasakan data dan mengirimkan data ini menggunakan perangkat komunikasi tertanam. Kontinum perangkat dan objek saling berhubungan melalui berbagai solusi komunikasi, seperti Bluetooth, WiFi, ZigBee, dan GSM. Perangkat komunikasi ini mengirimkan data dan menerima perintah dari perangkat yang dikendalikan dari jarak jauh, yang memungkinkan integrasi langsung dengan dunia fisik melalui sistem berbasis komputer untuk meningkatkan standar hidup.

Lebih dari 50 miliar perangkat mulai dari smartphone, laptop, sensor, dan konsol game diantisipasi untuk terhubung ke Internet melalui beberapa jaringan akses heterogen yang diaktifkan oleh teknologi, seperti identifikasi frekuensi radio (RFID) dan jaringan sensor nirkabel. Referensi [15] menyebutkan bahwa IoT dapat dikenali dalam tiga paradigma: berorientasi internet, sensor, dan pengetahuan [16]. Adaptasi terbaru dari teknologi nirkabel yang berbeda menempatkan IoT sebagai teknologi revolusioner berikutnya dengan memanfaatkan peluang penuh yang ditawarkan oleh teknologi Internet.

B. DATA BESAR

Volume data yang dihasilkan oleh sensor, perangkat, media sosial, aplikasi perawatan kesehatan, sensor suhu, dan berbagai aplikasi perangkat lunak dan perangkat digital lainnya yang terus-menerus menghasilkan sejumlah besar data terstruktur, tidak terstruktur, atau semi terstruktur meningkat pesat.

Pembangkitan data yang masif ini menghasilkan "big data" [17]. Sistem basis data tradisional tidak efisien saat menyimpan, memproses, dan menganalisis jumlah data atau data besar yang berkembang pesat [18]. Istilah "data besar" telah digunakan dalam literatur sebelumnya tetapi relatif baru dalam bisnis dan TI [19]. Contoh studi terkait data besar adalah inovasi, persaingan, dan produktivitas berikutnya; McKinsey Global Institute [20] mendefinisikan big data sebagai ukuran kumpulan data yang merupakan alat sistem basis data yang lebih baik daripada alat biasa untuk menangkap, menyimpan, memproses, dan menganalisis data tersebut [18]. Studi "The Digital Universe" [21] melabeli teknologi data besar sebagai generasi baru teknologi dan arsitektur yang bertujuan untuk mengambil nilai dari volume besar data dengan berbagai format dengan memungkinkan pengambilan, penemuan, dan analisis berkecepatan tinggi. Penelitian sebelumnya juga mengkarakterisasi big data menjadi tiga aspek: (a) sumber data, (b) analisis data, dan (c) penyajian hasil analisis.

Definisi ini menggunakan model 3V (volume, variasi, kecepatan) yang diusulkan oleh Beyer [2]. Model tersebut menyoroti tren e-commerce dalam manajemen data yang menghadapi tantangan untuk mengelola volume atau ukuran data, variasi atau sumber data yang berbeda, dan kecepatan atau kecepatan pembuatan data. Beberapa penelitian menyatakan volume sebagai karakteristik utama dari big data tanpa memberikan definisi yang murni [22]. Namun, peneliti lain memperkenalkan karakteristik tambahan untuk data besar, seperti kejujuran, nilai, variabilitas, dan kompleksitas [23], [24]. Model 3V, atau turunannya, adalah deskripsi paling umum dari istilah "data besar".

AKU AKU AKU. ANALISIS DATA BESAR

Analisis data besar melibatkan proses pencarian basis data, penambangan, dan analisis data yang didedikasikan untuk meningkatkan kinerja perusahaan [25].

Analisis data besar adalah proses pemeriksaan kumpulan data besar yang berisi berbagai tipe data [4] untuk mengungkapkan pola yang tidak terlihat, korelasi tersembunyi, tren pasar, preferensi pelanggan, dan informasi bisnis berguna lainnya [5]. Kemampuan untuk menganalisis data dalam jumlah besar dapat membantu organisasi menangani informasi yang cukup besar yang dapat mempengaruhi bisnis [6]. Oleh karena itu, tujuan utama analitik data besar adalah membantu asosiasi bisnis untuk meningkatkan pemahaman tentang data, dan dengan demikian, membuat keputusan yang efisien dan terinformasi dengan baik. Analisis data besar memungkinkan penambang data dan ilmuwan untuk menganalisis sejumlah besar data yang mungkin tidak dapat dimanfaatkan menggunakan alat tradisional [5].

Analitik data besar memerlukan teknologi dan alat yang dapat mengubah sejumlah besar data terstruktur, tidak terstruktur, dan semi-terstruktur menjadi format data dan metadata yang lebih mudah dipahami untuk proses analitik. Algoritme yang digunakan dalam alat analisis ini harus menemukan pola, tren, dan korelasi pada berbagai cakrawala waktu dalam data [26]. Setelah menganalisis data, alat ini memvisualisasikan temuan dalam tabel, grafik, dan bagan spasial untuk pengambilan keputusan yang efisien. Dengan demikian, analisis data besar merupakan tantangan serius bagi banyak aplikasi karena kompleksitas data dan skalabilitas algoritma yang mendasari yang mendukung proses tersebut [27].

Talia (2013) menyoroti bahwa memperoleh informasi yang berguna dari analisis data besar adalah masalah penting yang memerlukan algoritme dan teknik analitis yang dapat diskalakan untuk mengembalikan hasil yang tepat waktu, sedangkan teknik dan algoritme saat ini tidak efisien untuk menangani analisis data besar. Oleh karena itu, infrastruktur besar dan aplikasi tambahan diperlukan untuk mendukung paralelisme data. Selain itu, sumber data, seperti aliran data berkecepatan tinggi yang diterima dari sumber data yang berbeda, memiliki format yang berbeda, yang membuat integrasi berbagai sumber untuk solusi analitik menjadi penting [28]. Oleh karena itu, tantangannya difokuskan pada kinerja algoritma saat ini yang digunakan dalam analisis data besar, yang tidak meningkat secara linier dengan peningkatan pesat dalam sumber daya komputasi [19].

Proses analitik data besar menghabiskan banyak waktu untuk memberikan umpan balik dan pedoman kepada pengguna, sedangkan hanya beberapa alat [29] yang dapat memproses kumpulan data besar dalam arti dari data. Selain itu, analitik besar-besaran

jumlah waktu pemrosesan yang wajar. Sebaliknya, sebagian besar alat yang tersisa menggunakan metode coba-coba yang rumit untuk menangani sejumlah besar kumpulan data dan heterogenitas data [30].

Ada sistem analitik data besar. Misalnya, Lingkungan Analisis Data Eksplorasi [31] adalah sistem analisis visual data besar yang digunakan untuk menganalisis simulasi sistem bumi yang kompleks dengan sejumlah besar data.

A. SISTEM ANALISIS YANG ADA

Jenis analitik yang berbeda digunakan sesuai dengan kebutuhan aplikasi IoT [32]. Jenis analitik ini dibahas dalam subbagian ini di bawah kategori analitik real-time, off-line, level memori, intelijen bisnis (BI), dan level masif. Selain itu, perbandingan berdasarkan jenis analitik dan levelnya disajikan pada Tabel 1.

Analisis waktu nyata biasanya dilakukan pada data yang dikumpulkan dari sensor. Dalam situasi ini, data berubah secara konstan, dan teknik analisis data yang cepat diperlukan untuk mendapatkan hasil analisis dalam waktu singkat. Akibatnya, dua arsitektur yang ada telah diusulkan untuk analisis real-time: cluster pemrosesan paralel menggunakan database relasional tradisional dan platform komputasi berbasis memori [33]. Greenplum [34] dan Hana [35] adalah contoh arsitektur analitik real-time.

Analitik offline digunakan ketika respons cepat tidak diperlukan [32]. Misalnya, banyak perusahaan Internet menggunakan arsitektur analitik off-line berbasis Hadoop untuk mengurangi biaya konversi format data [36]. Analitik semacam itu meningkatkan efisiensi akuisisi data. SCRIBE [37], Kafka [38], Time Tunnel [39], dan Chukwa [40] adalah contoh arsitektur yang melakukan analitik off-line dan dapat memenuhi tuntutan akuisisi data.

Analisis tingkat memori diterapkan ketika ukuran data lebih kecil dari memori cluster [32]. Sampai saat ini, memori cluster telah mencapai tingkat terabyte (TB) [41]. Oleh karena itu, beberapa teknologi database internal diperlukan untuk meningkatkan efisiensi analitis. Analisis tingkat memori cocok untuk melakukan analisis waktu nyata. MongoDB [42] adalah contoh dari arsitektur ini.

Analisis BI diadopsi ketika ukuran data lebih besar dari tingkat memori, tetapi dalam kasus ini, data dapat diimpor ke lingkungan analisis BI [43]. Analisis BI saat ini mendukung data tingkat TB [32]. Apalagi, BI dapat membantu menemukan peluang bisnis strategis dari membanjirnya data. Selain itu, analitik BI memungkinkan interpretasi volume data yang mudah. Mengidentifikasi peluang baru dan menerapkan strategi yang efektif memberikan keunggulan pasar yang kompetitif dan stabilitas jangka panjang.

Analisis besar-besaran diterapkan ketika ukuran data lebih besar dari seluruh kapasitas produk analisis BI dan database tradisional [44]. Analisis besar-besaran menggunakan sistem file terdistribusi Hadoop untuk penyimpanan data dan pemetaan/pengurangan untuk analisis data. Analisis besar-besaran membantu menciptakan fondasi bisnis dan meningkatkan daya saing pasar dengan mengekstraksi nilai yang

TABEL 1. Perbandingan berbagai jenis analitik dan levelnya.

Analytic Types/Level	Specified Use	Existing Architectures/Tools	Advantages/Category
Real time[33]	To analyze the large amounts of data generated by the sensors	+Greenplum +HANA	+Parallel processing clusters using traditional databases memory based computing platforms
Offline [36]	To use for the Applications where there is no high requirements on response time	+Scribe + Kafka +Timetunnel +Chukwa	+Efficient Data acquisition +Reduce the cost of data format conversion
Memory level [41]	To use where the total data volume is smaller than the maximum Memory of the cluster	+MongoDB	+Real time
Business intelligence level [43]	To use when the data scale surpasses the memory level	+Data analysis plans.	+Both offline and Online
Massive level [44]	To use when data scale is totally surpassed the capacity of business intelligence products and traditional databases	+MapReduce	+Mostly belong to Offline

memperoleh data akurat yang memanfaatkan risiko yang terlibat dalam membuat keputusan bisnis apa pun. Selain itu, analitik masih menyediakan layanan secara efektif.

B. HUBUNGAN ANTARA IoT DAN ANALISIS DATA BESAR

Analisis data besar dengan cepat muncul sebagai inisiatif utama IoT untuk meningkatkan pengambilan keputusan. Salah satu fitur IoT yang paling menonjol adalah analisis informasinya tentang "hal-hal yang terhubung". Analisis data besar di IoT memerlukan pemrosesan data dalam jumlah besar dengan cepat dan penyimpanan data di berbagai teknologi penyimpanan. Mengingat bahwa sebagian besar data tidak terstruktur dikumpulkan langsung dari "hal-hal" yang mendukung web, implementasi data besar akan mengharuskan dilakukannya analisis secepat kilat dengan kueri besar untuk memungkinkan organisasi memperoleh wawasan cepat, membuat keputusan cepat, dan berinteraksi dengan orang-orang dan pihak lain. perangkat. Interkoneksi perangkat penginderaan dan penggerak memberikan kemampuan untuk berbagi informasi di seluruh platform melalui arsitektur terpadu dan mengembangkan gambaran operasi umum untuk memungkinkan aplikasi inovatif.

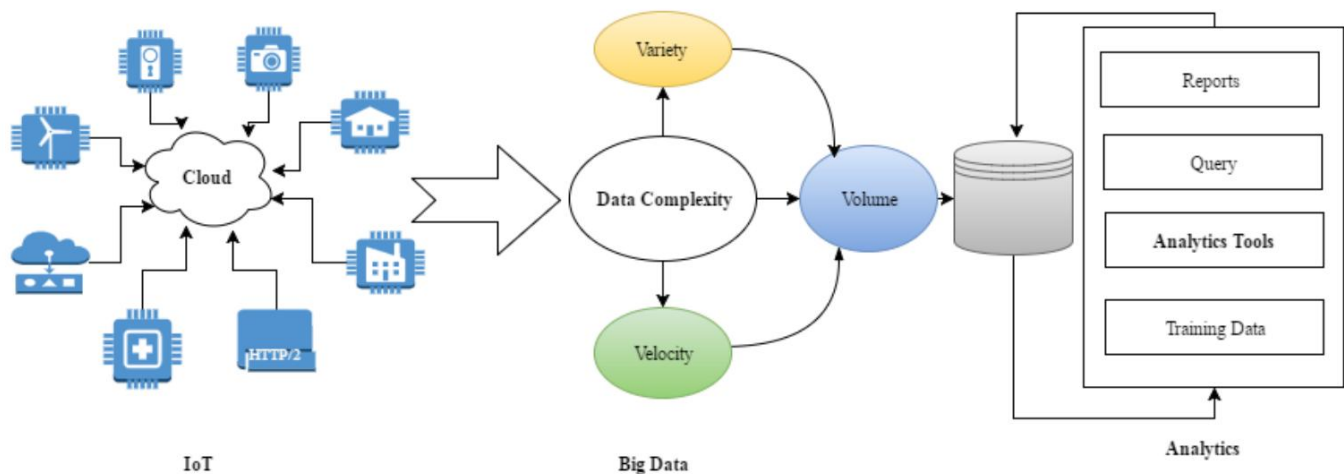
Kebutuhan untuk mengadopsi data besar dalam aplikasi IoT sangat menarik. Kedua teknologi ini sudah diakui di bidang IT dan bisnis. Walaupun

perkembangan big data sudah tertinggal, teknologi ini saling bergantung dan harus dikembangkan bersama.

Secara umum, penyebaran IoT meningkatkan jumlah data dalam jumlah dan kategori; karenanya, menawarkan peluang untuk aplikasi dan pengembangan analitik data besar.

Selain itu, penerapan teknologi big data di IoT mempercepat kemajuan penelitian dan model bisnis IoT.

Hubungan antara IoT dan data besar, yang ditunjukkan pada Gambar 1, dapat dibagi menjadi tiga langkah untuk memungkinkan pengelolaan data IoT. Langkah pertama terdiri dari pengelolaan sumber data IoT, di mana perangkat sensor yang terhubung menggunakan aplikasi untuk berinteraksi satu sama lain. Misalnya, interaksi perangkat seperti kamera CCTV, lampu lalu lintas pintar, dan perangkat rumah pintar, menghasilkan sejumlah besar sumber data dengan format yang berbeda. Data ini dapat disimpan dalam penyimpanan komoditas berbiaya rendah di cloud. Pada langkah kedua, data yang dihasilkan disebut "data besar", yang didasarkan pada volume, kecepatan, dan variasinya. Data dalam jumlah besar ini disimpan dalam file data besar di database toleransi kesalahan terdistribusi bersama. Langkah terakhir menerapkan alat analitik seperti MapReduce, Spark, Splunk, dan Skytree yang dapat menganalisis kumpulan data IoT besar yang tersimpan. Empat tingkat analitik dimulai dari data pelatihan, kemudian beralih ke alat analitik, kueri, dan laporan.



GAMBAR 1. Hubungan antara IoT dan analitik data besar.

C. METODE ANALISIS DATA BESAR

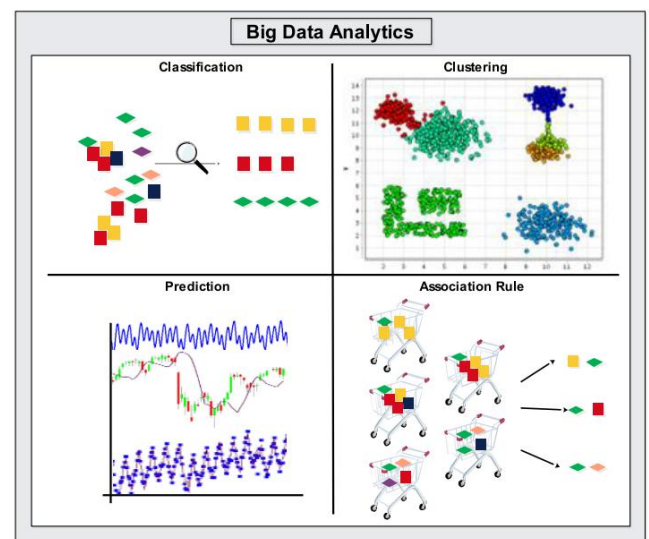
Analisis data besar bertujuan untuk segera mengekstrak informasi yang dapat membantu dalam membuat prediksi, mengidentifikasi tren terkini, menemukan informasi tersembunyi, dan akhirnya, membuat keputusan [7]. Teknik penambangan data digunakan secara luas untuk metode khusus masalah dan analitik data umum. Oleh karena itu, metode statistik dan pembelajaran mesin digunakan. Evolusi data besar juga mengubah persyaratan analitik. Meskipun persyaratan untuk mekanisme yang efisien terletak pada semua aspek pengelolaan data besar [30], seperti pengambilan, penyimpanan, prapemrosesan, dan analisis; untuk diskusi kami, analitik data besar membutuhkan kecepatan pemrosesan yang sama atau lebih cepat daripada analitik data tradisional dengan biaya minimum untuk data volume tinggi, kecepatan tinggi, dan variasi tinggi [45].

Berbagai solusi tersedia untuk analitik data besar, dan kemajuan dalam mengembangkan dan meningkatkan solusi ini terus dicapai agar sesuai dengan tren data besar baru. Penambangan data memainkan peran penting dalam analitik, dan sebagian besar teknik dikembangkan menggunakan algoritma penambangan data sesuai dengan skenario tertentu.

Pengetahuan tentang opsi analitik data besar yang tersedia sangat penting ketika mengevaluasi dan memilih pendekatan yang tepat untuk pengambilan keputusan. Pada bagian ini, kami menyajikan beberapa metode yang dapat diterapkan untuk beberapa studi kasus big data.

Beberapa metode analitik ini efisien untuk analitik data IoT besar. Kumpulan data berukuran besar dan beragam berkontribusi lebih banyak dalam wawasan data besar. Namun, keyakinan ini tidak selalu valid karena lebih banyak data mungkin memiliki lebih banyak ambiguitas dan kelainan [7].

Kami menyajikan metode analitik data besar di bawah klasifikasi, pengelompokan, penambangan aturan asosiasi, dan kategori prediksi. Gambar 2 menggambarkan dan meringkas masing-masing kategori ini. Setiap kategori adalah fungsi penambangan data dan melibatkan banyak metode dan algoritme untuk memenuhi persyaratan ekstraksi dan analisis informasi. Misalnya, jaringan Bayesian,



GAMBAR 2. Ikhtisar metode analisis data besar.

support vector machine (SVM), dan k-nearest neighbor (KNN) menawarkan metode klasifikasi. Demikian pula, partisi, pengelompokan hierarkis, dan kejadian tersebar luas dalam pengelompokan. Penambangan dan prediksi aturan asosiasi terdiri dari metode yang signifikan.

Klasifikasi adalah pendekatan pembelajaran terawasi yang menggunakan pengetahuan sebelumnya sebagai data pelatihan untuk mengklasifikasikan objek data ke dalam kelompok [46]. Kategori yang telah ditentukan sebelumnya diberikan ke suatu objek, dan dengan demikian, tujuan memprediksi grup atau kelas untuk suatu objek tercapai (lihat Gambar 2). Menemukan pola yang tidak diketahui atau tersembunyi lebih menantang untuk data IoT. Selanjutnya, mengekstraksi informasi berharga dari kumpulan data besar untuk meningkatkan pengambilan keputusan adalah tugas penting. Jaringan Bayesian adalah metode klasifikasi yang menawarkan interpretabilitas model. Jaringan Bayesian lebih efisien untuk menganalisis struktur data kompleks yang diungkapkan melalui data besar

daripada format data terstruktur tradisional. Jaringan ini diarahkan grafik asiklik, di mana node adalah variabel acak dan tepi menunjukkan ketergantungan bersyarat [47]. Nave, selektif nave, semi-nave Bayes, dan Bayes multi-jaring adalah kategori yang diusulkan untuk klasifikasi [48].

Menganalisis pola data dan membuat grup dilakukan secara efisien menggunakan SVM, yang juga merupakan pendekatan klasifikasi untuk analitik data besar. SVM menggunakan teori pembelajaran statistik untuk menganalisis pola data dan membuat grup. Beberapa aplikasi klasifikasi SVM dalam analitik data besar antara lain klasifikasi teks [49], pencocokan pola [50], diagnosis kesehatan [51], dan perdagangan. Demikian pula, KNN biasanya dirancang untuk menyediakan mekanisme yang efisien untuk menemukan pola tersembunyi dari kumpulan data besar, sehingga objek yang diambil mirip dengan kategori yang telah ditentukan sebelumnya [52]. Menggunakan kasus lebih lanjut meningkatkan algoritma KNN untuk aplikasi dalam deteksi anomali [53], data berdimensi tinggi [54], dan eksperimen ilmiah [55]. Klasifikasi memiliki ekstensi lain sambil mengadopsi sejumlah besar kecerdasan buatan dan teknik penambangan data. Akibatnya, klasifikasi adalah salah satu teknik penambangan data yang tersebar luas untuk analitik data besar.

Clustering adalah teknik penambangan data lain yang digunakan sebagai metode analitik data besar. Berlawanan dengan klasifikasi, pengelompokan menggunakan pendekatan pembelajaran tanpa pengawasan dan menciptakan kelompok untuk objek tertentu berdasarkan ciri khasnya yang bermakna [56]. Seperti yang telah kami sajikan pada Gambar 2 bahwa mengelompokkan sejumlah besar objek dalam bentuk cluster membuat manipulasi data menjadi sederhana. Metode yang terkenal digunakan untuk pengelompokan adalah pengelompokan hirarkis dan representasi untuk pengambilan keputusan yang lebih baik. Pendekatan pengelompokan hirarkis terus menggabungkan kluster kecil objek data untuk membentuk pohon hirarkis dan membuat kluster aglomerasi. Cluster divisive dibuat dengan cara yang berlawanan dengan membagi satu cluster yang berisi semua objek data menjadi cluster yang lebih kecil yang sesuai [57].

Analisis pasar dan pengambilan keputusan bisnis adalah aplikasi paling signifikan dari analisis data besar. Proses penambangan aturan asosiasi melibatkan pengidentifikasian hubungan yang menarik di antara objek, peristiwa, atau entitas lain yang berbeda untuk menganalisis tren pasar, perilaku pembelian konsumen, dan prediksi permintaan produk (lihat Gambar 2). Penambangan aturan asosiasi [58] berfokus pada identifikasi dan pembuatan aturan berdasarkan frekuensi kemunculan untuk data numerik dan non numerik. Pemrosesan data dilakukan dalam dua cara di bawah aturan asosiasi. Pertama, pemrosesan data sekuensial menggunakan algoritma berbasis apriori, seperti MSPS [59] dan LAPIN SPAM [60], untuk mengidentifikasi asosiasi interaksi. Pendekatan pemrosesan data penting lainnya di bawah aturan asosiasi adalah analisis urutan temporal, yang menggunakan algoritme untuk menganalisis pola peristiwa dalam data berkelanjutan.

Analisis prediktif menggunakan data historis, yang dikenal sebagai data pelatihan, untuk menentukan hasil sebagai tren atau perilaku dalam data. SVM dan algoritma logika fuzzy digunakan untuk mengidentifikasi hubungan antara variabel independen dan dependen dan untuk mendapatkan kurva regresi untuk prediksi, seperti untuk bencana alam. Selanjutnya, prediksi pembelian pelanggan dan tren media sosial dianalisis melalui analisis prediktif [61] (lihat Tabel 2). Dalam kasus analitik data besar, persyaratan pemrosesan dimodifikasi sesuai dengan sifat dan volume data. Akses data yang cepat dan metode penambangan untuk data terstruktur dan tidak terstruktur menjadi perhatian utama terkait analitik data besar. Selain itu, representasi data merupakan persyaratan signifikan dalam analitik data besar. Analisis deret waktu mengurangi dimensi tinggi yang terkait dengan data besar dan meningkatkan representasi untuk pengambilan keputusan yang lebih baik. Penelitian terkait representasi time series meliputi ARMA [62], bitmaps [63], dan fungsi wavelet [64].

Metode analitik data besar yang dibahas di bagian ini diadopsi secara luas di banyak area aplikasi data besar,

TABEL 2. Aplikasi penambangan data besar untuk IoT.

Method	Applications									
	Disaster management	Healthcare	Medical Imaging	Human Genetics	Market Analysis	Industry	Speech Recognition	Bioinformatics	NLP	Social Network Analysis
Classification [46]	-	-	✓	-	-	✓	✓	-	✓	-
Clustering [57]	-	✓	✓	✓	✓	✓	-	✓	-	✓
Association rule[58, 65]	-	✓	-	-	✓	✓	-	✓	-	-
Prediction [61]	✓	-	-	-	✓	-	-	-	-	-
Time Series [62] [63] [64]	✓	-	✓	-	-	-	✓	-	-	✓

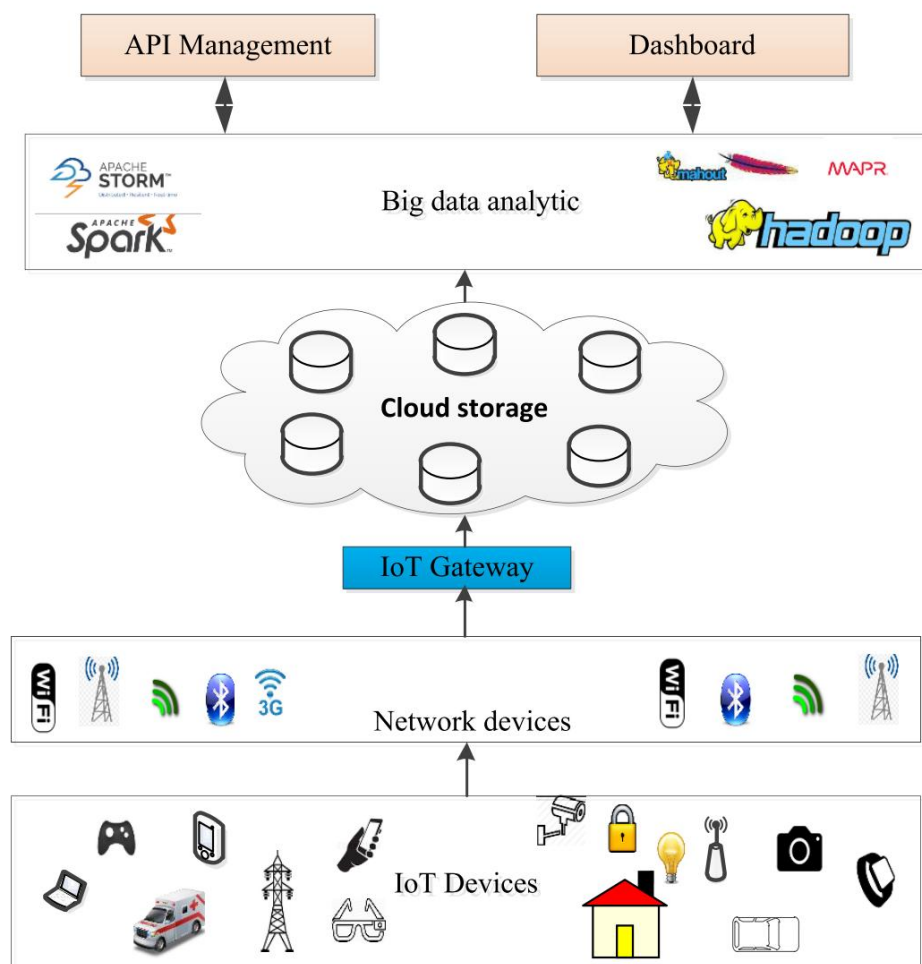
✓ has support
- not obvious

seperti manajemen bencana, kesehatan, bisnis, industri, dan e-governance. Pada Tabel 2, kami menyajikan aplikasi area fungsi penambangan data besar yang diuraikan di bagian ini, 'X' digunakan untuk menunjukkan dukungan untuk aplikasi sedangkan '-' menunjukkan bahwa tidak jelas apakah metode mendukung untuk aplikasi atau tidak. Khususnya, Tabel 2 menunjukkan bahwa metode klasifikasi cocok untuk pencitraan medis, industri, pengenalan suara, pemrosesan bahasa alami, dan e-governance. Metode analitik data berbasis aturan pengelompokan dan asosiasi berlaku untuk industri dan e-governance dan diadopsi dengan baik dalam perawatan kesehatan, e-commerce, dan bioinformatika. Analisis prediktif berguna untuk prediksi bencana dan pasar, sedangkan analisis deret waktu digunakan dalam peramalan bencana, medis, pencitraan, pengenalan ucapan, analisis jaringan sosial, dan pemerintahan elektronik.

D. ARSITEKTUR IoT UNTUK ANALISIS DATA BESAR

Konsep arsitektur IoT memiliki beberapa definisi berdasarkan: pada abstraksi dan identifikasi domain IoT. Ini menawarkan model referensi yang mendefinisikan hubungan di antara berbagai IoT

vertikal, seperti, lalu lintas pintar, rumah pintar, transportasi pintar, dan kesehatan pintar. Arsitektur untuk data besar analytics menawarkan desain untuk abstraksi data. Lebih-lebih lagi, standar ini menyediakan arsitektur referensi yang membangun pada model referensi. Banyak arsitektur IoT ditemukan dalam literatur [13], [66], [67]. Misalnya, [13] ditawarkan arsitektur IoT dengan komputasi awan di pusatnya dan a model interaksi ujung ke ujung di antara berbagai pemangku kepentingan dalam kerangka IoT cloud-centric untuk perbandingan yang lebih baik dengan arsitektur IoT yang diusulkan. Arsitektur ini tercapai dengan penginderaan di mana-mana tanpa batas, analisis data, dan representasi informasi dengan IoT sebagai arsitektur pemersatu. Namun, arsitektur saat ini berfokus pada IoT berkaitan dengan untuk komunikasi. Sepengetahuan kami, arsitektur yang kami usulkan, yang mengintegrasikan IoT dan analitik data besar, belum telah dipelajari dalam literatur saat ini. Gambar 3 mengilustrasikan Arsitektur IoT dan analitik data besar. Pada gambar ini, lapisan sensor berisi semua perangkat sensor dan objek, yang terhubung melalui jaringan nirkabel. Komunikasi jaringan tanpa kabel ini dapat berupa RFID, WiFi, ultra wideband, ZigBee, dan Bluetooth. Gerbang IoT memungkinkan



GAMBAR 3. Arsitektur IoT dan analitik data besar.

komunikasi Internet dan berbagai web. Lapisan atas menyangkut analitik data besar, di mana sejumlah besar data yang diterima dari sensor disimpan di cloud dan diakses melalui aplikasi analitik data besar. Aplikasi ini berisi manajemen API dan dasbor untuk membantu interaksi dengan mesin pemroses.

Pendekatan berbasis meta-model baru untuk mengintegrasikan objek arsitektur IoT diusulkan. Konsep ini semi otomatis digabungkan ke dalam lingkungan arsitektur perusahaan digital holistik. Tujuan utamanya adalah untuk memberikan dukungan keputusan yang memadai untuk bisnis yang kompleks, manajemen arsitektur dengan pengembangan sistem penilaian, dan lingkungan TI. Dengan demikian, keputusan arsitektur untuk IoT terkait erat dengan implementasi kode untuk memungkinkan pengguna memahami integrasi manajemen arsitektur perusahaan dengan IoT.

IV. GUNAKAN KASUS

Bagian ini menyajikan sejumlah kasus penggunaan untuk analitik data IoT besar. Meskipun kasus penggunaan relevan dengan aplikasi IoT, pilihannya telah dipandu untuk yang paling umum digunakan dalam aplikasi IoT dan untuk jumlah data yang dapat dihasilkan untuk analitik.

A. PENGUKURAN CERDAS

Smart metering adalah salah satu kasus penggunaan aplikasi IoT yang menghasilkan data dalam jumlah besar dari berbagai sumber, seperti smart grid, level tangki, dan aliran air, dan penghitungan stok silo, di mana pemrosesan membutuhkan waktu lama bahkan pada perangkat khusus dan khusus. mesin yang kuat [68]. Smart meter adalah perangkat yang secara elektronik mencatat konsumsi data energi listrik antara meteran dan sistem kontrol. Mengumpulkan dan menganalisis data smart meter di lingkungan IoT membantu pengambil keputusan dalam memprediksi konsumsi listrik. Selain itu, analitik smart meter juga dapat digunakan untuk memperkirakan permintaan guna mencegah krisis dan memenuhi tujuan strategis melalui rencana penetapan harga tertentu.

Dengan demikian, perusahaan utilitas harus mampu mengelola data volume tinggi dan analitik canggih yang dirancang untuk mengubah data menjadi wawasan yang dapat ditindaklanjuti.

B. TRANSPORTASI CERDAS

Sistem transportasi pintar adalah use case berbasis IoT yang bertujuan untuk mendukung konsep kota pintar. Sistem transportasi pintar bermaksud untuk menerapkan teknologi komunikasi yang canggih dan canggih untuk pengelolaan kota pintar.

Sistem transportasi tradisional, yang didasarkan pada pemrosesan citra, dipengaruhi oleh kondisi cuaca, seperti hujan lebat dan kabut tebal. Akibatnya, gambar yang diambil mungkin tidak terlihat jelas. Desain sistem e-plate [69] menggunakan teknologi RFID memberikan solusi yang baik untuk pemantauan cerdas, pelacakan, dan identifikasi kendaraan.

Selain itu, memperkenalkan IoT ke dalam teknologi kendaraan akan memungkinkan manajemen kemacetan lalu lintas untuk menunjukkan kinerja yang jauh lebih baik daripada infrastruktur yang ada. Teknologi teknologi ini dapat meningkatkan sistem lalu lintas yang ada di mana kendaraan

dapat berkomunikasi secara efektif satu sama lain secara sistematis tanpa campur tangan manusia.

Sistem dan sensor navigasi satelit juga dapat diterapkan di truk, kapal, dan pesawat terbang secara real time. Rute kendaraan ini dapat dioptimalkan dengan menggunakan sebagian besar data publik yang tersedia, seperti kemacetan lalu lintas, kondisi jalan, alamat pengiriman, kondisi cuaca, dan lokasi stasiun pengisian ulang. Misalnya, dalam kasus perubahan alamat runtime, informasi yang diperbarui (rute, biaya) dapat dioptimalkan, dihitung ulang, dan diteruskan ke driver secara real time. Sensor yang dimasukkan ke dalam kendaraan ini juga dapat memberikan informasi waktu nyata untuk mengukur kesehatan mesin, menentukan apakah peralatan memerlukan perawatan, dan memprediksi kesalahan [70].

C. RANTAI PASOKAN CERDAS

Teknologi sensor tertanam dapat berkomunikasi dua arah dan menyediakan aksesibilitas jarak jauh ke lebih dari 1 juta elevator di seluruh dunia [71]. Data yang diambil digunakan oleh teknisi di dalam dan di luar lokasi untuk menjalankan diagnostik dan opsi perbaikan untuk membuat keputusan yang tepat, yang menghasilkan peningkatan waktu kerja alat berat dan peningkatan layanan pelanggan. Pada akhirnya, analitik data IoT yang besar memungkinkan rantai pasokan untuk mengeksekusi keputusan dan mengontrol lingkungan eksternal. Peralatan pabrik yang mendukung IoT akan dapat berkomunikasi dalam parameter data (yaitu, penggunaan mesin, suhu) dan mengoptimalkan kinerja dengan mengubah pengaturan peralatan atau alur kerja proses [72]. Visibilitas dalam perjalanan adalah kasus penggunaan lain yang akan memainkan peran penting dalam rantai pasokan masa depan dengan adanya infrastruktur IoT. Teknologi utama yang digunakan oleh visibilitas transit adalah RFID dan Sistem Pemosisian Global (GPS) berbasis cloud, yang menyediakan lokasi, identitas, dan informasi pelacakan lainnya. Data ini akan menjadi tulang punggung rantai pasokan yang didukung oleh teknologi IoT. Informasi yang dikumpulkan oleh peralatan akan memberikan visibilitas rinci dari item yang dikirim dari produsen ke pengecer. Data yang dikumpulkan melalui teknologi RFID dan GPS akan memungkinkan manajer rantai pasokan untuk meningkatkan pengiriman otomatis dan informasi pengiriman yang akurat dengan memprediksi waktu kedatangan. Demikian pula, manajer akan dapat memantau informasi lain, seperti kontrol suhu, yang dapat mempengaruhi kualitas produk dalam perjalanan.

D. PERTANIAN CERDAS

Pertanian pintar adalah kasus penggunaan yang bermanfaat dalam analitik data IoT besar. Sensor adalah aktor dalam kasus penggunaan pertanian cerdas. Mereka dipasang di lapangan untuk mendapatkan data tingkat kelembaban tanah, diameter batang tanaman, kondisi iklim mikro, dan tingkat kelembaban, serta untuk prakiraan cuaca. Sensor mengirimkan data yang diperoleh menggunakan jaringan dan perangkat komunikasi. Data ini melewati gateway IoT dan Internet untuk mencapai lapisan analitik yang ditunjukkan pada Tabel 1. Lapisan analitik memproses data yang diperoleh dari jaringan sensor untuk mengeluarkan perintah. Kontrol iklim otomatis sesuai dengan persyaratan panen, irigasi tepat waktu dan terkontrol, dan kontrol kelembaban untuk pencegahan jamur adalah

TABEL 3. Perbandingan kasus penggunaan analitik data besar IoT.

Use cases	Benefits	IoT devices	Data source	Big data analytics applications
Smart metering [68]	Predict electricity consumption	Sensors	Text	Hadoop
Smart transportation[69] [70]	Improve existing traffic system by which vehicles can effectively communicate with one another in a systematic manner without human intervention	Sensors, cameras	Text, video, audio	Hadoop, Spark, Hive
Smart supply chains [71] [72]	Allow a supply chain to execute decisions and control the external environment	Sensors, mobile devices	Text, image	Hadoop
Smart agriculture [12, 13]	Obtain moisture level of soil, trunk diameter of plants, microclimate condition and humidity level; forecast weather	Sensors	Text, image	Hadoop
Smart grid [73, 74] [75] [76]	Improves reliability, safety, and efficiency, along with real-time control and monitoring	Sensors	Text	Hadoop
Smart traffic [77]	Detect the presence of vehicles, bikers, and pedestrians	Cameras	Video, image	Hadoop, Spark

contoh tindakan yang dilakukan berdasarkan analitik data besar rekomendasi.

E. SMART GRID

Jaringan pintar adalah generasi baru jaringan listrik di mana mengelola dan mendistribusikan listrik antara pemasok dan konsumen ditingkatkan menggunakan teknologi komunikasi dua arah dan kemampuan komputasi untuk meningkatkan keandalan, keamanan, efisiensi dengan kontrol waktu nyata, dan pemantauan [73], [74]. Salah satu tantangan utama dalam sistem tenaga adalah mengintegrasikan energi terbarukan dan desentralisasi. Listrik sistem membutuhkan jaringan cerdas untuk mengelola perilaku volatil dari sumber daya energi terdistribusi (DER) [75]. Namun, sebagian besar sistem energi harus mengikuti undang-undang pemerintah dan peraturan, serta mempertimbangkan analisis bisnis dan potensi kendala hukum [76]. Sensor dan perangkat grid secara terus menerus dan cepat menghasilkan data yang terkait dengan loop kontrol dan perlindungan serta memerlukan pemrosesan dan analitik real-time bersama dengan interaksi mesin-ke-mesin (M2M) atau manusia-ke-mesin (HMI) untuk mengeluarkan perintah kontrol ke sistem. Namun, sistem harus memenuhi visualisasi dan persyaratan pelaporan.

F. SISTEM LAMPU LALU LINTAS CERDAS

Sistem lampu lalu lintas pintar terdiri dari node yang secara lokal berinteraksi dengan sensor dan perangkat IoT untuk mendeteksi keberadaan kendaraan, pengendara sepeda, dan pejalan kaki. Node ini berkomunikasi dengan lampu lalu lintas tetangga untuk mengukur kecepatan dan jarak mendekati sarana transportasi dan mengelola sinyal lalu lintas hijau [77]. Data IoT dikumpulkan menggunakan sistem memerlukan pemrosesan analitik waktu-nyata untuk melakukan yang diperlukan tugas, seperti mengubah siklus waktu sesuai dengan lalu lintas kondisi, mengirimkan sinyal informatif ke node tetangga, dan mendeteksi kendaraan yang mendekat yang menggunakan sensor IoT dan perangkat untuk mencegah antrian panjang atau kecelakaan. Apalagi pintar sistem lampu lalu lintas dapat mengirim data IoT yang dikumpulkan ke cloud

penyimpanan untuk analisis lebih lanjut. Tabel 3 menyajikan kasus penggunaan Analisis data besar IoT.

Seperti yang ditunjukkan pada Tabel 3, sebagian besar kasus penggunaan terkait dengan: teknologi komunikasi M2M dan mengurangi peran interaksi Manusia. Namun, teknologi menggunakan prediksi metode dan teknik pengambilan keputusan untuk meningkatkan kontrol, pemantauan, dan kinerja waktu nyata. Data tekstual adalah di antara tipe data umum yang dihasilkan oleh perangkat IoT, yang sebagian besar adalah sensor dan kamera. Data berbasis teks adalah cocok untuk analisis oleh sistem file terdistribusi, seperti: hadop.

V. PELUANG

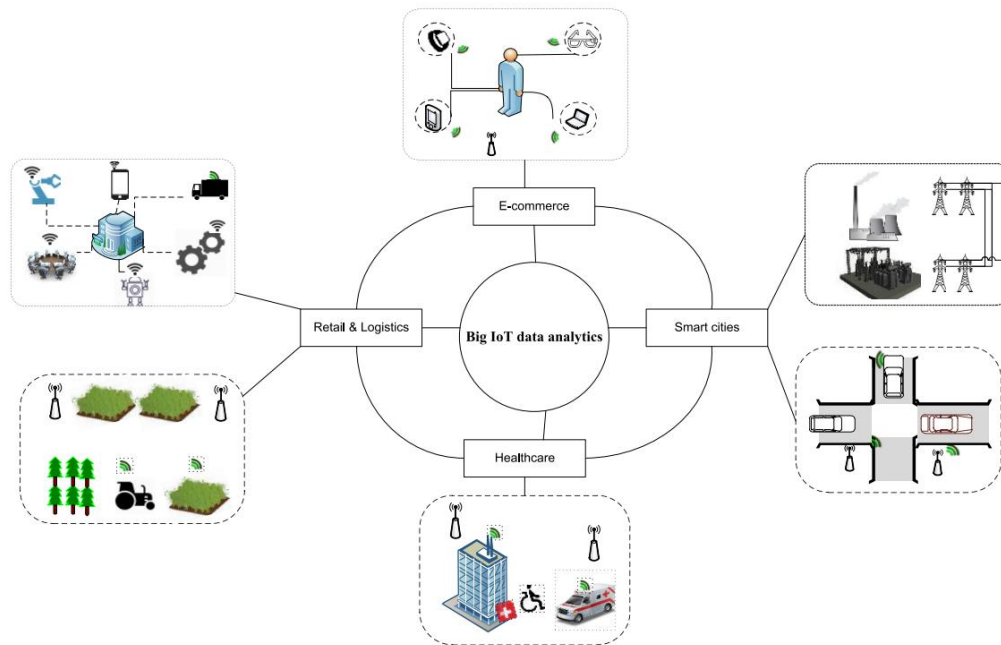
IoT saat ini dianggap sebagai salah satu transisi paling mendalam dalam teknologi. IoT saat ini menyediakan beberapa peluang analitik data untuk analitik data besar. Gambar 4 menunjukkan contoh kasus penggunaan dan peluang yang dibahas dalam Bagian 4 dan 5.

A. E-COMMERCE

Analisis data IoT besar menawarkan alat yang dirancang dengan baik untuk diproses data besar waktu nyata, yang menghasilkan hasil tepat waktu untuk pengambilan keputusan. Data IoT besar menunjukkan heterogenitas, meningkat volume, dan fitur pemrosesan data waktu nyata. Konvergensi big data dengan IoT membawa tantangan baru dan peluang untuk membangun lingkungan yang cerdas. Analisis data IoT besar memiliki aplikasi luas di hampir setiap industri. Namun, area keberhasilan utama analitik adalah di e-niaga, pertumbuhan pendapatan, peningkatan ukuran pelanggan, akurasi hasil perkiraan penjualan, pengoptimalan produk, risiko manajemen, dan meningkatkan segmentasi pelanggan.

B. KOTA PINTAR

Data besar yang dikumpulkan dari kota pintar menawarkan peluang baru di mana keuntungan efisiensi dapat dicapai melalui platform/infrastruktur analitik yang sesuai untuk menganalisis besar



GAMBAR 4. Contoh kasus penggunaan dan peluang untuk arsitektur analitik data IoT besar.

data IoT. Berbagai perangkat terhubung ke Internet dalam lingkungan yang cerdas dan berbagi informasi. Selain itu, biaya penyimpanan data telah berkurang secara dramatis setelah penemuan teknologi komputasi awan. Kemampuan analisis telah membuat lompatan besar. Dengan demikian, peran big data dalam smart city berpotensi mentransformasi setiap sektor perekonomian suatu bangsa.

Hadoop dengan manajer sumber daya YARN telah menawarkan kemajuan terbaru dalam teknologi data besar untuk mendukung dan menangani banyak beban kerja, pemrosesan waktu nyata, dan penyerapan data streaming.

C. ECERAN DAN LOGISTIK

IoT diharapkan memainkan peran kunci sebagai teknologi baru di bidang ritel dan logistik. Dalam logistik, RFID melacak kontainer, palet, dan peti. Selain itu, mempertimbangkan kemajuan teknologi IoT yang dapat memfasilitasi pengecer dengan memberikan beberapa manfaat. Namun, perangkat IoT menghasilkan data dalam jumlah besar setiap hari. Dengan demikian, analitik data yang kuat memungkinkan perusahaan untuk mendapatkan wawasan dari sejumlah besar data yang dihasilkan melalui teknologi IoT.

Menerapkan analitik data ke kumpulan data logistik dapat meningkatkan pengalaman pengiriman pelanggan. Selain itu, perusahaan ritel dapat memperoleh keuntungan tambahan dengan menganalisis data pelanggan, yang dapat memprediksi tren dan permintaan barang. Dengan melihat data pelanggan, mengoptimalkan rencana harga dan promosi musiman dapat direncanakan secara efisien untuk memaksimalkan keuntungan.

D. PERAWATAN KESEHATAN

Beberapa tahun terakhir telah menyaksikan pertumbuhan luar biasa dalam perangkat pemantauan kesehatan pintar. Perangkat ini menghasilkan sejumlah besar data. Jadi, menerapkan analitik data ke data

dikumpulkan dari monitor janin, elektrokardiogram, monitor suhu, atau monitor kadar glukosa darah dapat membantu spesialis kesehatan menilai kondisi fisik pasien secara efisien. Selain itu, analitik data memungkinkan profesional kesehatan untuk mendiagnosis penyakit serius pada tahap awal untuk membantu menyelamatkan nyawa. Selain itu, analitik data meningkatkan kualitas perawatan klinis dan memastikan keselamatan pasien.

Selain itu, profil dokter dapat ditinjau dengan melihat riwayat perawatan pasien, yang dapat meningkatkan kepuasan, akuisisi, dan retensi pelanggan.

VI. TANTANGAN TERBUKA DAN ARAH MASA DEPAN

IoT dan analitik data besar telah diterima secara luas oleh banyak organisasi. Namun, teknologi ini masih dalam tahap awal. Beberapa tantangan penelitian yang ada belum ditangani. Bagian ini menyajikan beberapa tantangan di bidang analitik data IoT besar.

A. PRIVASI

Masalah privasi muncul ketika sistem dikompromikan untuk menyimpulkan atau memulihkan informasi pribadi menggunakan alat analitik data besar, meskipun data dihasilkan dari pengguna anonim. Dengan menjamurnya teknologi analitik data besar yang digunakan dalam data IoT besar, masalah privasi telah menjadi masalah inti dalam domain penambangan data. Akibatnya, kebanyakan orang enggan untuk mengandalkan sistem ini, yang tidak memberikan kondisi perjanjian tingkat layanan (SLA) yang solid terkait pencurian atau penyalahgunaan informasi pribadi pengguna. Faktanya, informasi sensitif pengguna harus diamankan dan dilindungi dari gangguan eksternal. Meskipun identifikasi sementara, anonimitas, dan enkripsi menyediakan beberapa cara untuk menegakkan privasi data,

keputusan harus dibuat berkaitan dengan faktor etika, seperti apa yang digunakan, bagaimana menggunakan, dan mengapa menggunakan data IoT besar yang dihasilkan [7].

Risiko keamanan lain yang terkait dengan data IoT adalah keragaman jenis perangkat yang digunakan dan sifat data yang dihasilkan, seperti perangkat mentah, tipe data, dan protokol komunikasi. Perangkat ini dapat memiliki ukuran dan bentuk yang berbeda di luar jaringan dan dirancang untuk berkomunikasi dengan aplikasi kooperatif. Jadi, untuk mengotentikasi perangkat ini, sistem IoT harus menetapkan sistem identifikasi yang tidak dapat disangkal untuk setiap perangkat. Selain itu, perusahaan harus memelihara meta-repositori dari perangkat yang terhubung ini untuk tujuan audit. Arsitektur IoT yang heterogen ini baru bagi para profesional keamanan, dan dengan demikian, menghasilkan peningkatan risiko keamanan. Akibatnya, setiap serangan dalam skenario ini membahayakan keamanan sistem dan memutuskan perangkat yang saling berhubungan.

Dalam konteks data IoT yang besar, keamanan dan privasi adalah tantangan utama dalam memproses dan menyimpan data dalam jumlah besar. Selain itu, untuk melakukan operasi penting dan menyimpan data pribadi, sistem ini sangat bergantung pada layanan dan infrastruktur pihak ketiga. Oleh karena itu, pertumbuhan eksponensial dalam kecepatan data menyebabkan kesulitan dalam mengamankan setiap bagian dari data penting. Seperti yang telah dibahas sebelumnya, solusi keamanan yang ada (Karim, 2016 #86) tidak lagi berlaku untuk menyediakan keamanan lengkap dalam skenario data IoT besar. Algoritme yang ada tidak dirancang untuk pengamatan data yang dinamis, dan dengan demikian, tidak diterapkan secara efektif. Solusi keamanan data lama dirancang khusus untuk kumpulan data statis, sedangkan persyaratan data saat ini berubah secara dinamis (Lafuente, 2015). Jadi, menerapkan solusi keamanan ini sulit untuk meningkatkan data secara dinamis. Selain itu, masalah legislatif dan peraturan harus dipertimbangkan saat menandatangani SLA.

Berkenaan dengan data yang dihasilkan melalui IoT, masalah keamanan berikut dapat muncul [78]: (a) pembaruan tepat waktu - kesulitan dalam menjaga sistem tetap mutakhir, (b) manajemen insiden - mengidentifikasi pola lalu lintas yang mencurigakan di antara yang sah dan kemungkinan kegagalan untuk menangkap insiden yang tidak dapat diidentifikasi, (c) interoperabilitas - prosedur khusus pemilik dan vendor akan menimbulkan kesulitan dalam menemukan yang tersembunyi

atau serangan zero day, (d) dan konvergensi protokol - meskipun IPv6 saat ini kompatibel dengan spesifikasi terbaru, protokol ini belum sepenuhnya digunakan. Oleh karena itu, penerapan aturan keamanan melalui IPv4 mungkin tidak berlaku untuk melindungi IPv6.

Saat ini, tidak ada jawaban yang dapat mengatasi tantangan ini dan mengelola keamanan dan privasi perangkat yang saling terhubung.

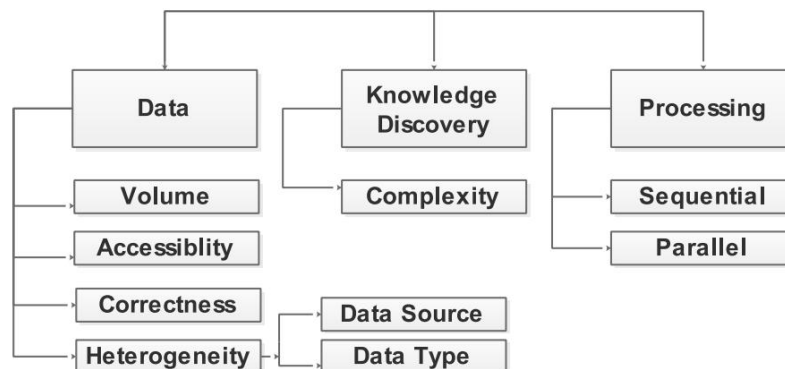
Namun, panduan berikut dapat mengatasi kesulitan tersebut. (a) Pertama, ekosistem terbuka sejati dengan API standar diperlukan untuk menghindari masalah interoperabilitas dan keandalan. (b) Kedua, perangkat harus terlindungi dengan baik saat berkomunikasi dengan rekan-rekan. (c) Ketiga, perangkat harus di-hardcode dengan praktik keamanan terbaik untuk melindungi dari ancaman keamanan dan privasi umum.

B. PERTAMBAHAN DATA

Metode data mining memberikan solusi prediktif atau deskriptif yang efisien dan paling pas untuk data besar yang juga dapat digeneralisasi untuk data baru [45]. Evolusi data IoT besar dan platform komputasi awan telah membawa tantangan eksplorasi data dan ekstraksi informasi [79]. Namun, untuk arsitektur data IoT besar secara keseluruhan, Gambar 5 menyajikan tantangan utama yang terkait dengan pemrosesan dan penambahan data.

Baca/tulis data yang lengkap: Kualitas volume tinggi, kecepatan tinggi, dan keragaman tinggi dari data IoT besar menantang eksplorasi, integrasi, komunikasi heterogen, dan proses ekstraksi. Ukuran dan heterogenitas data memaksakan persyaratan penambahan data baru, dan keragaman sumber data juga menimbulkan tantangan [80]–[82]. Selanjutnya, dibandingkan dengan kumpulan data kecil, kumpulan data besar terdiri lebih banyak kelainan dan ambiguitas yang memerlukan langkah-langkah pra-proses tambahan, seperti pembersihan, pengurangan, dan transmisi [23], [83]. Masalah lain terletak pada ekstraksi informasi yang tepat dan berpengetahuan dari volume besar data yang beragam. Akibatnya, memperoleh informasi yang akurat dari data yang kompleks memerlukan analisis properti data dan menemukan hubungan antara titik data yang berbeda.

Para peneliti telah memperkenalkan model pemrograman paralel dan sekuensial dan mengusulkan algoritma yang berbeda untuk meminimalkan waktu respons kueri saat menangani data besar.



GAMBAR 5. Masalah penambahan data besar di IoT.

Selain itu, para peneliti telah memilih algoritma penambangan data yang ada dengan cara yang berbeda untuk (a) meningkatkan penemuan pengetahuan sumber tunggal, (b) menerapkan metode penambangan data untuk platform multi-sumber, dan (c) mempelajari dan menganalisis metode penambangan data dinamis dan mengalirkan data [84]. Oleh karena itu, algoritma paralel k-means [85] dan metode penambangan aturan asosiasi paralel [65] diperkenalkan. Namun, kebutuhan untuk merancang algoritma tetap untuk menyediakan kompatibilitas dengan arsitektur paralel terbaru. Selain itu, masalah sinkronisasi dapat terjadi dalam komputasi paralel, sementara informasi dipertukarkan dalam metode penambangan data yang berbeda. Kemacetan metode penambangan data ini telah menjadi masalah terbuka dalam analisis data IoT besar yang harus ditangani arsitektur IoT.

C. VISUALISASI

Visualisasi adalah entitas penting dalam analitik data besar, terutama ketika berhadapan dengan sistem IoT di mana data dihasilkan secara besar-besaran. Selain itu, melakukan visualisasi data sulit dilakukan karena ukuran besar dan dimensi data besar yang tinggi. Situasi ini menunjukkan tren yang mendasari dan gambaran lengkap dari data yang diurai. Oleh karena itu, analitik dan visualisasi data besar harus bekerja dengan lancar untuk mendapatkan hasil terbaik dari aplikasi IoT dalam data besar. Namun, visualisasi dalam kasus data yang heterogen dan beragam (tidak terstruktur, terstruktur, dan semi terstruktur) adalah tugas yang menantang. Merancang solusi visualisasi yang kompatibel dengan kerangka kerja pengindeksan data besar yang canggih adalah tugas yang sulit. Demikian pula, waktu respons adalah faktor yang diinginkan dalam analitik data IoT yang besar. Akibatnya, arsitektur komputasi awan yang didukung dengan fasilitas GUI yang kaya dapat digunakan untuk mendapatkan wawasan yang lebih baik tentang tren data IoT yang besar [86].

Metode pengurangan dimensi yang berbeda telah diperkenalkan sebagai hasil dari data IoT besar yang kompleks dan berdimensi tinggi [87], [88]. Namun, metode ini tidak cocok untuk semua jenis data yang disajikan. Demikian pula, ketika dimensi butuh halus divisualisasikan secara efektif, kemungkinan untuk mengidentifikasi korelasi, pola, dan garis luar yang dapat diamati tinggi [89]. Selain itu, data harus disimpan secara lokal untuk mendapatkan informasi yang dapat digunakan secara efisien karena keterbatasan daya dan lebar pita. Selain itu, perangkat lunak visualisasi harus dijalankan dengan konsep lokalitas referensi untuk mencapai hasil yang efisien di lingkungan IoT. Mengingat jumlah data IoT besar meningkat dengan cepat, persyaratan paralelisasi yang sangat besar adalah tugas yang menantang dalam visualisasi. Jadi, untuk menguraikan masalah menjadi tugas-tugas independen yang dapat dikelola untuk menegakkan eksekusi query secara bersamaan merupakan tantangan untuk algoritma visualisasi paralel [90].

Saat ini, sebagian besar alat visualisasi data besar yang digunakan untuk IoT menunjukkan hasil kinerja yang buruk dalam hal fungsionalitas, skalabilitas, dan waktu respons. Untuk memberikan visualisasi sadar-ketidakpastian yang efektif selama proses analitik visual, menghindari ketidakpastian menimbulkan tantangan yang cukup besar [32]. Selanjutnya, beberapa isu penting dibahas [91], seperti (a) gangguan visual - sebagian besar objek kumpulan data terkait erat satu sama lain, dan dengan demikian, pengguna dapat melihat hasil yang berbeda dari jenis yang sama; (b) informasi

loss - menerapkan metode reduksi ke kumpulan data yang terlihat dapat menyebabkan hilangnya informasi; (c) pengamatan gambar besar - alat visualisasi data memiliki masalah yang melekat sehubungan dengan rasio aspek, resolusi rancangan, dan batas persepsi fisik; (d) gambar yang sering berubah - pengguna tidak akan melihat perubahan data yang cepat dalam output; dan (e) persyaratan kinerja tinggi - persyaratan kinerja tinggi diberlakukan karena data dihasilkan secara dinamis dalam lingkungan IoT. Selain itu, metode yang didukung oleh analitik canggih memungkinkan grafik interaktif di laptop, desktop, atau perangkat seluler, seperti smartphone dan tablet [92].

Analisis waktu nyata adalah pertimbangan lain yang disorot dalam arsitektur IoT. Beberapa pedoman visualisasi dalam data besar disajikan [93], seperti (a) kesadaran data, yaitu, keahlian domain yang sesuai, (b) kualitas data - pembersihan data menggunakan manajemen informasi atau kebijakan tata kelola data, (c) hasil yang berarti - data pengelompokan digunakan untuk memberikan abstraksi tingkat tinggi sehingga visibilitas kelompok data yang lebih kecil dimungkinkan, dan (d) outlier harus dihapus dari data atau diperlakukan sebagai entitas yang terpisah. Referensi [94] menyarankan bahwa visualisasi harus mematuhi pedoman berikut: (a) sistem harus memberikan perhatian khusus pada metadata, (b) perangkat lunak visualisasi harus interaktif dan harus memerlukan keterlibatan pengguna maksimum, dan (c) alat harus dibangun berdasarkan pada sifat dinamis dari data yang dihasilkan.

D. INTEGRASI

Integrasi mengacu pada memiliki pandangan seragam yang berbeda untuk tika. Integrasi data menyediakan tampilan tunggal dari data yang datang dari sumber yang berbeda dan menggabungkan tampilan data [95]. Integrasi data mencakup semua proses yang terlibat dalam pengumpulan data dari sumber yang berbeda, serta dalam menyimpan dan menyediakan data dengan tampilan terpadu. Untuk setiap momen, berbagai bentuk data terus-menerus dihasilkan oleh media sosial, IoT, dan pendekatan komunikasi dan telekomunikasi lainnya. Data yang dihasilkan dapat dikategorikan ke dalam tiga kelompok: (a) data terstruktur, seperti data yang disimpan dalam sistem database tradisional, termasuk tabel dengan baris dan kolom; (b) semi-terstruktur, seperti file HTML, XML, dan Json; dan (c) data tidak terstruktur, seperti video, audio, dan gambar. Data yang baik menawarkan informasi yang baik; namun, hubungan ini hanya dicapai melalui integrasi data [96]. Mengintegrasikan beragam tipe data adalah tugas yang kompleks dalam menggabungkan sistem atau aplikasi yang berbeda [97]. Tumpang tindih data yang sama, meningkatkan kinerja dan skalabilitas, dan memungkinkan akses data waktu nyata adalah beberapa tantangan yang terkait dengan integrasi data yang harus diatasi di masa mendatang.

Tantangan lainnya adalah menyesuaikan struktur pada data semi terstruktur dan tidak terstruktur sebelum mengintegrasikan dan menganalisis jenis data ini [98]. Informasi, seperti entitas dan hubungan, dapat diekstraksi dari data tekstual dengan menggunakan teknologi yang tersedia di era penambangan teks, pembelajaran mesin, pemrosesan alami, dan ekstraksi informasi. Namun, teknologi baru harus dikembangkan untuk mengekstrak gambar, video, dan informasi lain dari non-teks lainnya

format data tidak terstruktur [98]. Penambahan teks diharapkan dapat dilakukan dengan menerapkan beberapa ekstraktor khusus pada teks yang sama. Oleh karena itu, mengelola dan mengintegrasikan hasil ekstraksi yang berbeda dari sumber data tertentu memerlukan teknik lain [99].

VII. KESIMPULAN

Tingkat pertumbuhan produksi data telah meningkat secara drastis selama beberapa tahun terakhir dengan menjamurnya perangkat pintar dan sensor. Interaksi antara IoT dan data besar saat ini berada pada tahap di mana pemrosesan, transformasi, dan analisis data dalam jumlah besar pada frekuensi tinggi diperlukan. Kami melakukan survei ini dalam konteks analitik data IoT yang besar. Pertama, kami menjelajahi solusi analitik terbaru. Hubungan antara analitik data besar dan IoT juga dibahas.

Selain itu, kami mengusulkan arsitektur untuk analitik data IoT besar. Selanjutnya, jenis, metode, dan teknologi analitik data besar untuk penambahan data besar disajikan. Beberapa kasus penggunaan yang kredibel juga disediakan. Selain itu, kami menjelajahi domain dengan membahas berbagai peluang yang dibawa oleh analisis data dalam paradigma IoT. Beberapa tantangan penelitian terbuka dibahas sebagai arah penelitian masa depan. Akhirnya, kami menyimpulkan bahwa solusi analitik data IoT besar yang ada masih dalam tahap awal pengembangan. Di masa depan, solusi analitik real-time yang dapat memberikan wawasan cepat akan diperlukan.

REFERENSI

- [1] P. Tiainen, "Peluang baru di bidang teknik elektro sebagai akibat dari munculnya Internet of Things," *Tech. Rep.*, AaltoDoc, Aalto Univ., 2016.
- [2] M. Beyer, "Gartner mengatakan memecahkan tantangan 'Big Data' melibatkan lebih dari sekadar mengelola volume data," *Tech. Rep.*, AaltoDoc, Aalto Univ., 2011.
- [3] J. Gantz dan D. Reinsel, "Mengambil nilai dari kekacauan," *IDC View*, vol. 1142, hlm. 1–12, Juni 2011.
- [4] R. Mital, J. Coughlin, dan M. Canaday, "Menggunakan teknologi data besar dan analitik untuk memprediksi anomali sensor," di *Proc. Adv. Pilihan Maui. Pengawasan Luar Angkasa. teknologi. Kon.*, Sep. 2014, hal. 84.
- [5] N. Golchha, "Big data-revolusi informasi," *Int. J. Adv. Res.*, vol. 1, tidak. 12, hlm. 791–794, 2015.
- [6] P. Russom, *Analisis Data Besar*. TDWI, Kuartal ke-4, 2011.
- [7] C.-W. Tsai, "Analisis data besar: Survei," *J. Big Data*, vol. 2, tidak. 1, hlm. 1–32, 2015.
- [8] M. Chen, *Teknologi Terkait dalam Big Data*. Heidelberg, Jerman: Springer, 2014, hlm. 11–18.
- [9] Z. Khan, A. Anjum, dan SL Kiani, "Analisis data besar berbasis cloud untuk kota masa depan yang cerdas," di *Proc. IEEE/ACM 6th Int. Kon. Komputasi Awan Utilitas. (UCC)*, Desember 2013, hlm. 381–386.
- [10] P. Russom, *Analisis Data Besar*. TDWI, Kuartal ke-4, 2011, hlm. 1–35.
- [11] S. LaValle, E. Lesser, R. Shockley, MS Hopkins, dan N. Kruschwitz, "Big data, analitik, dan jalur dari wawasan ke nilai," *MIT Sloan Manag. Pdt.*, vol. 52, tidak. 2, hal. 21, 2011.
- [12] E. Al Nuaimi *et al.*, "Aplikasi data besar ke kota pintar," *J. Internet Aplikasi Layanan*, vol. 6, hal. 25 Desember 2015.
- [13] J. Gubbi, R. Buyya, S. Marusic, dan M. Palaniswami, "Internet of Things (IoT): Sebuah visi, elemen arsitektur, dan arah masa depan," *Future Generat. Hitung. Sistem*, vol. 29, tidak. 7, hlm. 1645–1660, 2013.
- [14] CA Ciufo. (2014). *Peralatan Industri Berbicara tentang IoT? Lebih baik mendapatkan Gateway (Perangkat)*. [On line]. Tersedia: <http://eecatalog.com/caciufio/2014/07/15/iot-gateway-adlink/> [15] L. Atzori, A. Iera, dan G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, tidak. 15, hlm. 2787–2805, 2010.
- [16] H.-C. Hsieh dan C.-H. Lai, "Arsitektur Internet of Things berdasarkan PLC terintegrasi dan jaringan komunikasi 3G," di *Proc. IEEE 17th Int. Kon. Distribusi Paralel. sistem (ICPADS)*, Desember 2011, hlm. 853–856.
- [17] K. Kambatta, "Tren dalam analitik data besar," *J. Parallel Distrib. Komputasi*, jilid 74, tidak. 7, hlm. 2561–2573, 2014.
- [18] J. Manyika, *Big Data: The Next Frontier for Innovation, Competition, and Produktifitas*. McKinsey Global Inst. Rep., 2011.
- [19] IAT Hashem *et al.*, "Munculnya 'data besar' pada komputasi awan: Tinjau dan buka masalah penelitian," *Inf. Sistem*, vol. 47, hlm. 98–115, Januari 2015.
- [20] WB Ali, "Pemolisian cerdas berbasis data besar: pengiriman mobil patroli berbasis data besar," *J. Geotech. terjemahan Ind.*, vol. 1, tidak. 2, hlm. 1–16, 2016.
- [21] J. Gantz dan D. Reinsel, "Alam semesta digital pada tahun 2020: Big data, bayangan digital yang lebih besar, dan pertumbuhan terbesar di timur jauh," IDC iView: IDC Analyze the Future, Tech. Wakil, 2012.
- [22] V. Borkar, MJ Carey, dan C. Li, "Di dalam manajemen Big Data: Ogres, bawang, atau parfaits?" di *Proc. Int. 15 Kon. Memperluas Teknologi Basis Data*, 2012, hlm. 3–14.
- [23] A. Gani, "Sebuah survei tentang teknik pengindeksan untuk data besar: Taksonomi dan evaluasi kinerja," *Knowl. Inf. Sistem*, vol. 46, tidak. 2, hlm. 241–284, 2016.
- [24] A. Paul, "Pencarian dan pengindeksan video dengan agen penguatan untuk layanan multimedia interaktif," *ACM Trans. Menanamkan. Hitung. Sistem*, vol. 12, tidak. 2, hlm. 1–16, 2013.
- [25] O. Kwon dan NBL Shin, "Manajemen kualitas data, pengalaman penggunaan data, dan niat akuisisi analitik data besar," *Int. J. Inf. Kelola*, vol. 34, tidak. 3, hlm. 387–394, 2014.
- [26] S. Oswal dan S. Koul, "Analisis dan visualisasi data besar pada perangkat seluler," di *Proc. Nat. Kon. New Horizons IT-NCNHIT*, 2013, hlm. 223.
- [27] L. Candela dan DPC Pagano, "Mengelola big data melalui infrastruktur data hybrid," *ERICIM News*, vol. 89, hlm. 37–38, Juni 2012.
- [28] MD Assuncao, RN Calheiros, S. Bianchi, MAS Netto, dan R. Buyya. (2013). "Komputasi data besar dan awan: Tantangan, solusi, dan arah masa depan." [Online]. Tersedia: <https://arxiv.org/abs/1312.4722> [29] D. Singh dan CK Reddy, "Sebuah survei pada platform untuk analitik data besar," *J. Data Besar*, vol. 2, tidak. 1, hal. 1, 2014.
- [30] A. Siddiq, "Sebuah survei manajemen data besar: Taksonomi dan state-of-the-art," *J. Netw. Hitung. Aplikasi*, vol. 71, hlm. 151–166, Agustus 2016.
- [31] CA Steed, "Analisis visual data besar untuk analisis simulasi sistem bumi eksplorasi," *Comput. Geosci.*, vol. 61, hlm. 71–82, Maret 2013.
- [32] CLP Chen dan C.-Y. Zhang, "Aplikasi, tantangan, teknik, dan teknologi intensif data: Survei tentang data besar," *Inf. Sci.*, vol. 275, hlm. 314–347, Agustus 2014.
- [33] MW Pfaffl, "Model matematika baru untuk kuantifikasi relatif secara real-time RT-PCR," *Nucl. Asam Res.*, vol. 29, tidak. 9, hal. e45, 2001.
- [34] FM Waas, "Di luar pergudangan data konvensional-pemrosesan data paralel besar-besaran dengan database Greenplum," dalam *Lokakarya Internasional tentang Intelijen Bisnis untuk Perusahaan Real-Time*. Berlin, Jerman: Springer, 2008.
- [35] F. Färber, SK Cha, J. Primsch, C. Bornhövd, S. Sigg, dan W. Lehner, "database SAP HANA: Manajemen data untuk aplikasi bisnis modern," *ACM SIGMOD Rec.*, vol. 40, tidak. 4, hlm. 45–51, Desember 2011.
- [36] M. Cheng, "Mu kontrol kursor berbasis ritme: Analisis offline," *Clin. Neurofis.*, vol. 115, tidak. 4, hlm. 745–751, 2004.
- [37] M. Castro, P. Druschel, AM Kermarrec, dan AIT Rowstron, "Scribe: Sebuah infrastruktur multicast tingkat aplikasi skala besar dan terdesentralisasi," *IEEE J. Sel. Komune Area*, vol. 20, tidak. 8, hlm. 1489–1499, Oktober 2002.
- [38] J. Kreps, N. Narkhede, dan J. Rao, "KAFKA: Sistem pesan terdistribusi untuk pemrosesan log," di *Proc. NetDB*, 2011, hlm. 1–7.
- [39] H. Notsu, Y. Okada, M. Akaishi, dan K. Nijima, "Terowongan waktu: Alat analisis visual untuk data numerik deret waktu dan perluasannya ke koordinat paralel," di *Proc. Int. Kon. Hitung. Grafik., Gambar. melihat (CGIV)*, Juli 2005, hlm. 167–172.
- [40] A. Rabkin dan RH Katz, "Chukwa: Sistem pengumpulan log skala besar yang andal," di *Proc. LISA*, November 2010, hlm. 1–15.
- [41] S. Hong dan H. Kim, "Sebuah model analitis untuk arsitektur GPU dengan tingkat memori dan kesadaran paralelisme tingkat benang," *ACM SIGARCH Comput. Arsitek.*, vol. 37, tidak. 3, hlm. 152–163, Juni 2009.
- [42] K. Chodorow, *MongoDB: Panduan Definitif*. Newton, MA, AS: O'Reilly Media, Inc, 2014.
- [43] Z. Jourdan dan Rkte Rainer Marshall, "Kecerdasan bisnis: Analisis literatur 1," *Inf. sistem Kelola*, vol. 25, tidak. 2, hlm. 121–131, 2008.

- [44] A. Bifet, G. Holmes, R. Kirkby, dan B. Pfahringer, "MOA: Analisis online besar-besaran," *J. Mach. Mempelajari. Res.*, vol. 11, hlm. 1601–1604, Mei 2010.
- [45] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, dan CAC Coello, "Sebuah survei algoritma evolusioner multiobjektif untuk penambahan data: Bagian I," *IEEE Trans. Evolusi Komputer.*, vol. 18, tidak. 1, hlm. 4–19, Februari 2014.
- [46] V. Estivill-Castro, "Mengapa begitu banyak algoritma pengelompokan: Kertas posisi," *ACM SIGKDD Explorations Newslett.*, vol. 4, tidak. 1, hlm. 65–75, 2002.
- [47] C. Bielza dan P. Larrañaga, "Pengklasifikasi jaringan Bayesian Diskrit: Sebuah survei," *ACM Comput. Survei*, vol. 47, tidak. 1, hal. 5 Juli 2014.
- [48] F. Chen *et al.*, "Data mining untuk Internet of Things: Tinjauan dan tantangan literatur," *Int. J. Distribusi. Jaringan Sensor*, vol. 12 Agustus 2015.
- [49] R. Luss dan A. d'Aspremont, "Memprediksi abnormal return dari berita menggunakan klasifikasi teks," *Quant. Keuangan*, vol. 15, tidak. 6, hal. 999–1012, 2015.
- [50] P. Melin dan O. Castillo, "Sebuah tinjauan tentang aplikasi logika fuzzy tipe-2 dalam pengelompokan, klasifikasi dan pengenalan pola," *Appl. Komputer lunak.*, vol. 21, hlm. 568–577, Agustus 2014.
- [51] A. Soualhi, K. Medjaher, dan N. Zerhouni, "Memantau kesehatan berdasarkan transformasi Hilbert-Huang, mendukung mesin vektor, dan regresi," *IEEE Trans. alat musik Meas.*, vol. 64, tidak. 1, hlm. 52–62, Januari 2015.
- [52] DT Larose, *Algoritma K-Nearest Neighbor*. 2005, hlm. 90–106.
- [53] M.-Y. Su, "Sistem deteksi anomali waktu nyata untuk serangan Denial-of-Service oleh pengklasifikasi K-nearest-neighbor berbobot," *Expert Syst. Aplikasi.*, vol. 38, tidak. 4, hlm. 3492–3498, 2011.
- [54] M. Muja dan DG Lowe, "Algoritma tetangga terdekat yang dapat diskalakan untuk data dimensi tinggi," *IEEE Trans. Pola Anal. Mach. Intel.*, vol. 36, tidak. 11, hlm. 2227–2240, November 2014.
- [55] C. Hu, "Metode data-driven berdasarkan optimasi gerombolan partikel dan regresi tetangga K-terdekat untuk memperkirakan kapasitas baterai lithium-ion," *Appl. Energi*, vol. 129, hlm. 49–55, Sep. 2014.
- [56] K. Srivastava, "Data mining menggunakan algoritma pengelompokan aglomeratif hierarkis dalam lingkungan komputasi awan terdistribusi," *Int. J. Hitung. Teori Eng.*, vol. 5, tidak. 3, hal. 520, 2013.
- [57] P. Berkhin, *Sebuah Survei Teknik Data Mining Clustering, dalam Pengelompokan Data Multidimensi*. Berlin, Jerman: Springer, 2006, hlm. 25–71.
- [58] A. Gosain dan M. Bhugra, "Sebuah survei komprehensif aturan asosiasi pada data kuantitatif dalam penambahan data," di *Proc. IEEE Conf. Inf. komuni. Technol.*, April 2013, hlm. 1003–1008.
- [59] M. Fitzwater, "Penambahan yang efisien dari pola sekuensial maksimal menggunakan beberapa sampel," di *Proc. SIAM Int. Kon. Penambahan Data*, 2005, hal. 1.
- [60] Z. Yang dan M. Kitsuregawa, "LAPIN-SPAM: Algoritma yang ditingkatkan untuk menambang pola sekuensial," di *Proc. 21th Int. Kon. Data Eng. Bengkel kerja (ICDEW)*, April 2005, hlm. 1222.
- [61] A. Gandomi dan M. Haider, "Beyond the hype: Big data concept, methods, and analytics," *Int. J. Inf. Kelola.*, vol. 35, tidak. 2, hlm. 137–144, 2015.
- [62] K. Kalpakis dan DV Gada Puttagunta, "Ukuran jarak untuk pengelompokan seri waktu ARIMA yang efektif," di *Proc. IEEE Int. Kon. Data Mining (ICDM)*, Desember 2001, hlm. 273–280.
- [63] N. Kumar, *Time-Series Bitmaps: Alat Visualisasi Praktis untuk Bekerja dengan Database Deret Waktu Besar*. Philadelphia, PA, AS: SIAM, 2005.
- [64] D. Ryan, *Penemuan Kinerja Tinggi dalam Seri Waktu: Teknik dan Studi kasus*. Berlin, Jerman: Springer, 2013.
- [65] X. Wu dan S. Zhang, "Sintesis aturan frekuensi tinggi dari sumber data yang berbeda," *IEEE Trans. tahu. Data Eng.*, vol. 15, tidak. 2, hlm. 353–367, Maret 2003.
- [66] R. Duan, X. Chen, dan T. Xing, "Sebuah arsitektur QoS untuk IOT," di *Proc. Int. Kon. 4th Int. Kon. Cyber, Fisika. Perkumpulan Hitung. (iThings/CPSCoM)*, Oktober 2011, hlm. 717–720.
- [67] Y. Zhang, *Arsitektur Berbasis ICN untuk IoT*. Kontribusi IRTF, Oktober 2013.
- [68] S. Darby, "Pengukuran cerdas: Apa potensi keterlibatan rumah tangga?" *Building Res. Inf.*, jilid. 38, tidak. 5, hlm. 442–457, 2010.
- [69] TA Rahman dan SKA Rahim, "nomor plat kendaraan RFID (e-plate) untuk sistem pelacakan dan manajemen" di *Proc. Int. Kon. Distribusi Paralel. sistem (ICPADS)*, Desember 2013, hlm. 611–616.
- [70] J. Sherly dan D. Somasundareswari, "Sistem transportasi pintar berbasis Internet of Things," *Int. Res. J. Eng. Teknologi.*, vol. 2, tidak. 7, 2015.
- [71] N. Tohamy, "Apa yang perlu Anda ketahui tentang Internet of Things," *Solusi Rantai Pasokan MHD*, vol. 45, tidak. 3, hal. 32, 2015.
- [72] C. Kecil. (2015). *Lima Cara Internet of Things Akan Menguntungkan Rantai Pasokan*. [On line]. Tersedia: <http://www.gartner.com/smarterwithgartner/five-ways-the-internet-of-things-will-benefit-the-supply-chain-2/> [73] Y. Yan, Y. Qian, H. Sharif, dan D. Tipper, "A survei infrastruktur komunikasi jaringan pintar: Motivasi, persyaratan, dan tantangan," *IEEE Commun. Survei Tuts.*, vol. 15, tidak. 1, hlm. 5–20, Februari 2013.
- [74] S. Bera, S. Misra, dan JJPC Rodrigues, "Aplikasi komputasi awan untuk jaringan pintar: Survei," *IEEE Trans. Distribusi Paralel. Sistem*, vol. 26, tidak. 5, hlm. 1477–1494, Mei 2015.
- [75] T. Dethlefs, *Deskripsi Layanan Energi untuk Kemampuan Sumber Daya Energi Terdistribusi*. Berlin, Jerman: Springer, 2015.
- [76] C. Neureiter, "Sebuah pendekatan berbasis standar untuk pemodelan domain spesifik dari arsitektur sistem jaringan pintar," di *Proc. 11th Int. Kon. sistem sistem Ind. (SoSE)*, Kongsberg, Norwegia, Juni 2016, hlm. 1–6.
- [77] F. Bonomi, "Komputasi kabut dan perannya dalam Internet of Things," di *Proc. Edisi ke-1 Lokakarya MCC Mobile Cloud Comput.*, 2012, hlm. 1–6.
- [78] K. Steinklauber. (2014). *Perlindungan Data di Internet of Things*. [On line]. Tersedia: <https://securityintelligence.com/data-protection-in-the-internet-of-things>
- [79] T. Hu, H. Chen, L. Huang, dan X. Zhu, "Sebuah survei penambahan data massal berdasarkan komputasi awan," di *Proc. Anti-Pemalsuan, Aman. Identifikat.*, Agustus 2012, hlm. 1–4.
- [80] Y. Sun, "Menambang pengetahuan dari data yang saling berhubungan: Pendekatan analisis jaringan informasi yang heterogen," *Proc. Wakaf VLDB*, vol. 5, tidak. 12, hlm. 2022–2023, 2012.
- [81] M. Chen, LT Yang, T. Kwon, L. Zhou, dan M. Jo, "Perencanaan perjalanan untuk komunikasi agen hemat energi dalam jaringan sensor nirkabel," *IEEE Trans. weh. Teknologi.*, vol. 60, tidak. 7, hlm. 3290–3299, September 2011.
- [82] D. Zhang, "Sebuah taksonomi teknologi agen untuk lingkungan komputasi di mana-mana," *Trans. Informasi Internet Sistem*, vol. 6, tidak. 2, hlm. 547–565, 2012.
- [83] M. Chen, VCM Leung, dan S. Mao, "Fusi terkontrol terarah dalam jaringan sensor nirkabel," *Mobile Netw. Aplikasi.*, vol. 14, tidak. 2, hlm. 220–229, April 2009.
- [84] X. Wu, X. Zhu, G.-Q. Wu, dan W. Ding, "Data mining dengan data besar," *IEEE Trans. tahu. Data Eng.*, vol. 26, tidak. 1, hlm. 97–107, Januari 2014.
- [85] K. Su, "Sebuah kerangka logis untuk mengidentifikasi kualitas pengetahuan dari sumber data yang berbeda," *Sistem Pendukung Keputusan.*, vol. 42, tidak. 3, hlm. 1673–1683, 2006.
- [86] L. Wang, G. Wang, dan CA Alexander, "Big data dan visualisasi: Metode, tantangan dan kemajuan teknologi," *Digit. Teknologi.*, vol. 1, tidak. 1, hlm. 33–38, 2015.
- [87] AT Azar dan AE Hassani, "Pengurangan dimensi data besar medis menggunakan pengklasifikasi neural-fuzzy," *Soft Comput.*, vol. 19, tidak. 4, hlm. 1115–1127, 2015.
- [88] VL Popov dan M. Heß, *Metode Pengurangan Dimensi dalam Kontak Mekanika dan Gesekan*. Springer, 2015.
- [89] C. Donalek *et al.*, "Visualisasi data imersif dan kolaboratif menggunakan platform realitas virtual," di *Proc. IEEE Int. Kon. Big Data (USA Big Data)*, Oktober 2014, hlm. 609–614.
- [90] H. Childs *et al.*, "Tantangan penelitian untuk perangkat lunak visualisasi," *Komputer*, vol. 46, tidak. 5, hlm. 34–42, Mei 2013.
- [91] EY Gorodov dan VV Gubarev, "Tinjauan analitik metode visualisasi data dalam aplikasi ke data besar," *J. Elect. Hitung. Ind.*, vol. 2013, Oktober 2013, Seni. tidak. 969458.
- [92] Pusat. (2013). *Visualisasi Big Data: Mengubah Data Besar Menjadi Wawasan Besar [Sementara Makalah]*. [On line]. Tersedia: <https://future.transport.nsw.gov.au/wp-content/uploads/2016/02/big-data-visualization-turning-big-data-into-big-insights.pdf> [93] (2013). *Lima Tantangan Big Data dan Cara Mengatasinya Dengan Visual Analytics*. [On line]. Tersedia: <https://www.sas.com/resources/asset/five-big-data-challenges-article.pdf> [94] P. Simon, *Organisasi Visual: Visualisasi Data, Data Besar, dan Pencarian untuk Keputusan yang Lebih Baik*. Hoboken, NJ, AS: Wiley, 2014.
- [95] BB Ahamed, T. Ramkumar, dan S. Hariharan, "Kemajuan integrasi data dalam sumber data besar menggunakan afinitas pemetaan," di *Proc. 7th Int. Kon. Adv. lunak Ind. aplikasi (ASEA)*, Desember 2014, hlm. 16–21.
- [96] J. Liu dan X. Zhang, "Integrasi data dalam dokumen XML fuzzy," *Inf. Sci.*, vol. 280, hlm. 82–97, Maret 2014.

- [97] A. Ma'ayan *et al.*, "Lean integrasi data besar dalam biologi sistem dan sistem farmakologi," *Trends Pharmacol. Sci.*, vol. 35, tidak. 9, hlm. 450–460, 2014.
- [98] PBD Agrawal *et al.*, "Tantangan dan peluang dengan data besar: Buku putih komunitas yang dikembangkan oleh para peneliti terkemuka di seluruh Amerika Serikat," Konsorsium Komunitas Komputasi, Buku Putih, 2012.
- [99] R. Agrawal *et al.*, "Laporan claremont tentang penelitian basis data," *Komunikasi nikasi*, vol. 52, tidak. 6, hlm. 56–65, 2009.



MOHSEN MARJANI menerima gelar BE dalam matematika di Iran, pada tahun 2003, dan Master gelar Informasi dalam komputasi multimedia dari Universitas Multimedia, Malaysia, pada tahun 2011. Dia saat ini sedang mengejar gelar Ph.D. gelar dengan Departemen Sistem Komputer dan Teknologi, Universitas Malaya, dan anggota dari Proyek Penelitian Berdampak Tinggi (Mobile Cloud Komputasi: Arsitektur pusat data) yang sepenuhnya didanai oleh Kementerian Tinggi Malaysia

Pendidikan. Dia telah mengajar selama lebih dari sepuluh tahun di berbagai institusi di Iran.



FARIZA NASARUDDIN menerima B.Sc. gelar dalam ilmu komputer dan M.Sc. gelar dalam MIS dari Northern Illinois University, AS, dan Ph.D. gelar dari Universitas Malaya. Dia bekerja di industri sebagai Analis Sistem untuk sepuluh tahun sebelum bergabung dengan dunia akademis pada tahun 1997. Dia saat ini adalah Dosen Senior di Departemen Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas

Melayu. Dia menjadi terlibat dalam multi-disiplin penelitian tetapi fokus utamanya adalah database, sistem informasi, dan data ilmu pengetahuan.



ABDULLAH GANI (SM'12) menerima gelar sarjana dan magister dari Universitas Hull, Hull, Inggris, dan Ph.D. gelar dari The University of Sheffield, UK Beliau memiliki pengalaman mengajar yang luas dengan berbagai institusi pendidikan lokal dan luar negeri—sekolah, perguruan tinggi pengajar, kementerian pendidikan, dan universitas. Dia adalah saat ini menjadi Profesor di Departemen Sistem dan Teknologi Komputer, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Malaya, Malaysia. Dia

adalah Direktur Center for Mobile Cloud Computing Research, yang berfokus pada penelitian berdampak tinggi, a Visiting Professor dengan King Saud University, Saudi Arabia, dan juga an Profesor Ajun dengan Institut Teknologi Informasi COMSATS, Islamabad, Pakistan. Ketertarikannya pada penelitian dimulai pada tahun 1983 ketika dia dipilih untuk mengikuti Kursus Penelitian Ilmiah di RECSAM oleh Kementerian Pendidikan, Malaysia. Dia telah menerbitkan lebih dari 100 akademik makalah dalam konferensi dan jurnal terhormat. Dia aktif mengawasi banyak siswa di semua tingkat studi—sarjana, magister, dan Ph.D. Ketertarikannya penelitian meliputi sistem yang terorganisir sendiri, pembelajaran penguatan, dan jaringan terkait nirkabel. Dia terlibat dalam komputasi awan seluler dengan Hibah Penelitian Berdampak Tinggi sebesar USD 500.000 (RM 1,5 juta) untuk periode 2011 hingga 2016.



AHMAD KARIM menerima gelar MS (Hons.) dalam CS dari Universitas Bahauddin Zakariya, Multan, Pakistan, dan Ph.D. gelar dalam keamanan komputer dari Universitas Malaya, Malaysia. Dia saat ini adalah Dosen dengan Jurusan Teknologi Informasi, Universitas Bahaud di Zakariya. Bidang penelitiannya termasuk deteksi botnet seluler, keamanan komputer, jaringan komputer, jaringan yang ditentukan perangkat lunak, analitik data besar, dan Internet of Things.

Ia menerima Sertifikasi Internasional Cisco (CCNA, CCNP, dan CCAl).



IBRAHIM ABAKER TARGIO HASHEM diterima gelar BE dalam ilmu komputer di Sudan pada tahun 2007, dan gelar MS dalam komputasi di Malaysia pada tahun 2012. Ia menerima gelar Ph.D. gelar dalam ilmu komputer dari Universitas Malaya, Kuala Lumpur, Malaysia. Dia adalah seorang Tutor dengan Akademi CISCO, Universitas Malaya. Utamanya kepentingan penelitian termasuk data besar, komputasi awan, dan komputasi terdistribusi dan jaringan. Dia menerima sertifikat profesional dari CISCO (CCNP, CCNA, dan Keamanan CCNA) dan Grup APMG (PRINCE2 Yayasan, Yayasan ITIL v3, dan Yayasan OBASHI).



AISHA SIDDIQA menerima gelar B.Sc. gelar dalam teknologi informasi dari Bahauddin Zakariya University, Multan, Pakistan, dan M.Sc. derajat dalam ilmu komputer dari COMSATS Institute of Information Technology, Islamabad, Pak istan. Dia saat ini sedang mengejar gelar Ph.D. derajat dengan Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Malaya, dimana dia juga seorang Peneliti. Bidang penelitiannya adalah algoritma dan struktur data, penyimpanan data besar, analitik data besar, dan pengindeksan.



IBRAR YAQOOB menerima gelar BS (Hons.) dalam teknologi informasi dari Universitas Punjab, Gujranwala, Pakistan, pada tahun 2012. Ia telah mengejar gelar Ph.D. gelar dalam ilmu komputer dengan Universitas Malaya, Malaysia, sejak 2013. Dia juga Bright Spark Asisten Riset Program. Dia telah menerbitkan sejumlah artikel penelitian di jurnal dan majalah internasional yang direferensikan. Nya banyak artikel penelitian paling banyak diunduh di jurnal teratas.

Minat penelitiannya meliputi data besar, cloud seluler, Internet of Things, komputasi awan, dan jaringan nirkabel.

...