

BAB II

TINJAUAN PUSTAKA

Dasar-dasar penelitian sebelumnya yang menjadi tinjauan pustaka pada penelitian ini dirangkum dalam Tabel 2.1.

Penelitian ini menggunakan pendekatan dari beberapa pendekatan dari penelitian (Nawawi & Rubedo, 2021) Banyak Perguruan Tinggi yang belum memiliki sistem dan database khusus untuk mengelola data penelitian dan PKM. Hal ini menyebabkan dosen terkadang lupa terhadap hasil laporan penelitian dan PKM yang telah dilakukan. Dosen akan mendapatkan kendala dalam pengajuan Jabatan Fungsional maupun Akreditasi, sehingga mereka harus mencari-cari dan bahkan tidak menemukan karena tak ada catatan dan pengelolaan yang baik. Metode yang digunakan adalah deskriptif-action, yaitu mendeskripsikan keseluruhan sistem terlebih dahulu lalu menganalisis kekurangan dan kelemahan yang terjadi dan selanjutnya adalah melakukan tindakan disertai dengan solusi-solusi pemecahannya. Hasil dari penelitian tersebut diharapkan sistem akan memberikan kemudahan akses bagi dosen dan stakeholder lain yang berkepentingan di *International Women Universty* seperti pencarian, penyortiran, termasuk penyimpanan data.

Dampak negatif bagi sebuah Universitas bila tidak didukung oleh suatu sistem yang terintegrasi dalam melakukan pencarian data ijazah alumni. Pencarian data Ijazah alumni yang telah lulus untuk memenuhi kebutuhan informasi pemangku kepentingan (stakeholders) adalah salah satu hal yang sangat sulit dilakukan (Nasution et al., 2022). Di sisi lain, stakeholders terutama lembaga, instansi dan perusahaan yang akan merekrut para lulusan, membutuhkan waktu yang cepat untuk mengetahui dan menguji kebenaran Ijazah para lulusan. Dalam penelitian tersebut digunakan metode Grid Data (GD) adalah salah satu metode yang dapat digunakan dalam proses BI, dimana metode GD dapat melakukan analisis yang tepat dalam mengukur validitas data terutama data ijazah. Dengan menggunakan sistem informasi berbasis kecerdasan bisnis (*business intelligent*) memudahkan

proses integrasi data lulusan, sehingga tidak sulit untuk menghasilkan data secara optimal, efektif dan efisien.

Sistem pengelolaan dokumen terkadang hanya dijalankan pada lembaga induk secara terpusat namun tidak terlaksana dengan baik pada level low management. Hal ini menyebabkan kesulitan untuk mengakses data secara langsung. Audit internal atau akreditasi kelompok ini haruslah memiliki data yang dibutuhkan secara cepat dan tepat dan sangat sulit untuk mendapatkannya (Santoso, 2022). Metode yang digunakan yaitu teknik pengumpulan data menggunakan observasi, wawancara, penilaian pakar. Instrumen yang digunakan adalah lembar wawancara, ceklist observasi. Lembar validasi pakar dan angket. Model desain pengembangan yang digunakan adalah desain pengembangan Borg and Gall yang telah dimodifikasi. 5 kategori penilaian, pada penilaian > 4.2 dalam kategori Sangat baik, $> 3.4 - 4.2$ adalah Baik, $> 2.6 - 3.4$ adalah Cukup, $> 1.8 - 2.6$ dalam kategori Kurang, dan < 1.8 dalam kategori Sangat kurang. Hasil dari penelitian tersebut yaitu sistem database dari segi media, praktisi, dan pengguna sangat layak untuk dapat diterapkan.

Jika pengelolaan SDM di perguruan tinggi tidak menggunakan perangkat lunak HRIS maka Data SDM yang dimiliki perguruan tinggi menjadi tidak terpusat sehingga ada kemungkinan terjadi *redundancy* data dan Pengolahan data dan penyajian data membutuhkan waktu yang lama (Fahmi & Sar, 2018). Pendekatan pengembangan sistem menggunakan metode *prototyping*. Sedangkan metode perancangan yang digunakan yaitu perancangan sistem dan perancangan rinci, yang meliputi perancangan objek. Alat yang digunakan untuk mengukur implementasi sistem yang dibangun menggunakan metode kuisisioner. Menggunakan validitas isi (*content validity*) sebagai alat ukur dalam penelitian tersebut. Hasil dari penelitian tersebut yaitu, sistem ini sangat baik (45%), kemudahan di dalam penggunaannya (65% menyatakan sangat baik dan 35% menyatakan baik), disamping itu 56% menyatakan sangat baik terhadap desain interface, keakuratan informasi yang diberikan dalam sistem dinyatakan 72% baik, bahasa yang digunakan dalam aplikasi 55% sangat baik dan 45% baik, dan

terakhir mengenai tentang *error handling* sebanyak 32% menyatakan sangat baik, 68% yang menyatakan baik.

Pengaruh adanya Perubahan aturan Akreditasi Perguruan Tinggi (APT) 3.0 yang di berlakukan sejak April 2019 menyebabkan Beberapa sistem informasi khususnya SIMMawa, SIMHumas dan Kerjasama serta SIMAKU Mengalami kekosongan data Adanya kekosongan data dalam bentuk Laporan Kinerja Perguruan Tinggi (LKPT) Sebagai acuan dalam membuat Laporan Evaluasi Diri (LED). Sehingga dilakukan proses pengintegrasian fungsi bisnis, entitas data dan aplikasi yang mengacu pada referensi Zachman Framework (ZF) dengan menggunakan pangkalan data perguruan tinggi, data primer dan sekunder perguruan tinggi. Mendapatkan sebuah model alternatif pengelolaan data terintegrasi yang dapat dijadikan masukan untuk pengembang sistem informasi di lingkup perguruan tinggi untuk mengantisipasi kekosongan data tersebut (Primadewi & Hanafi, 2020).

Sistem Informasi Akademik Mahasiswa pada perguruan tinggi yang masih monoton dan lazim serta tidak memiliki *Novelty* (Kebaruan) menyebabkan Kurangnya *attractives* (daya tarik pengguna) pada sistem informasi yang dibangun (Henim & Sari, 2020). Pengukuran dilakukan dengan menggunakan *user experience questionnaire* (UEQ) dengan data *User Experience Questionnaire* (UEQ) 179 responden mahasiswa yang masih aktif, daya tarik 1,375, kejelasan 1,552, Efisiensi 1,354, Stimulasi 1,346, dan kebaruan 0,855. Dari hasil *user experience questionnaire* memberikan rekomendasi untuk perancangan *user experience system* informasi akademik yang baru dengan melakukan inovasi pada berbagai aspek.

Belum adanya system pengarsipan pada Perguruan Tinggi Universitas Catur Insan Cendekia Cirebon untuk mengelola sehingga menyebabkan data terkait penelitian dan pengabdian kepada masyarakat (LPPM) tidak di kelola dengan baik. Sehingga digunakan metode *Alphabetical filing system* dalam metode pengarsipannya, metode ini dituangkan pada bahasa pemrograman *PHP* dan Javascript. Kemudian data kerjasama, kegiatan penelitian dan pengabdian kepada masyarakat ini akan disimpan pada database *MYSQL*. Dengan demikian dapat

mempermudah lembaga penelitian dan pengabdian kepada masyarakat (LPPM) dalam mengarsipkan data kerjasama, kegiatan penelitian dan pengabdian kepada masyarakat yang dilakukan oleh dosen (Rully et al., 2020).

Data pendukung untuk sistem remunerasi diambil melalui pusat data namun ada juga yang diambil langsung dari masing-masing sistem pendukung sehingga menyebabkan jika terjadi perubahan data pada satu sistem akan timbul ketidakkonsistenan dan duplikasi data sistem di UPT TIK belum terdokumentasi (Aryani et al., 2022). Menggunakan pendekatan Zachman Framework, yang merupakan salah satu metode EAP dalam perancangan system dengan sistem pangkalan data dosen (PDD), sistem informasi akademik (SIAk), sistem aplikasi hukum dan tata laksana (Shakuntala), E-Kinerja, sistem penelitian dan pengabdian (SILIDIA) dan sistem absensi wajah. Hasil dari penelitian ini adalah gambar arsitektur aplikasi, arsitektur teknologi dan rancangan integrasi data sistem penunjang remunerasi yang diterjemahkan dari hasil pemetaan matrik Zachman yang berjumlah 36 sel yang terdiri dari 6 kolom dan 6 baris.

Data tidak dapat dengan mudah ditafsirkan dalam bentuk mentahnya, sehingga meningkatnya sejumlah sifat eksistensial entitas bisnis, termasuk manajemen organisasi, kapabilitas pasar, dan umpan balik konsumen (Science & Science, 2023). Dalam penelitian tersebut digunakan teknologi data pipeline untuk proses *extract-transform-load* (ETL) data dengan menggunakan data klien, konsumen, dan karyawan. Hasil dari penelitian tersebut teknologi data pipeline mampu mengatasi kendala bisnis yang ada, dan mendapatkan data yang diproses langsung di akhir siklus sinkronisasi data.

Di era *Internet of Things* dan platform media sosial, sejumlah besar data digital dihasilkan oleh dan dikumpulkan dari banyak sumber, termasuk sensor, perangkat seluler, pelacak yang dapat dikenakan, dan kamera keamanan. Hal ini tentunya memberikan tantangan bagi pengembang untuk mencari solusi dalam kemampuan penyimpanan, pemrosesan, dan analisis saat ini (Belcastro et al., 2022). Dalam penelitian tersebut dilakukan analisis dan meninjau paradigma, bahasa, dan sistem paralel dan terdistribusi yang digunakan saat ini untuk menganalisis dan belajar dari Big Data pada komputer yang dapat diskalakan.

Hasil dari analisis tersebut memberikan pemahaman yang mendalam tentang properti paradigma pemrograman paralel utama (*MapReduce*, alur kerja, BSP, penyampaian pesan, dan seperti SQL) dan, melalui contoh pemrograman, dan menjelaskan sistem yang paling banyak digunakan untuk analisis Big Data (misalnya, *Hadoop*, *Spark*, dan *Storm*).

Data mentah susah untuk dikonsolidasikan atau disatukan sehingga berbagai aspek karakteristik eksistensial data mentah tidak dapat ditanyakan secara langsung karena tidak tersedia pada perusahaan tersebut, seperti kompetensi pasar. Data komoditas, data pelanggan, dan data karyawan di gunakan dalam proses *extract-transform-load* (ETL). Dengan menggunakan metode *extract-transform-load* (ETL) mampu mengatasi tantangan korporasi fasilitas cloud perantara yang dikenal sebagai data lake, di mana dengan menggunakan teknologi data pipeline dapat menerima data secara kemudian dimuat ke database untuk diproses di akhir siklus sinkronisasi data (Cottur & Gadad, 2020).

Arsitektur tanpa server pipeline *Extract-Transform-Load* (ETL) menyebabkan adanya hambatan pada muatan peristiwa aliran data dengan berbagai frekuensi, kecepatan dan ukuran *payload* (Pogiatzis & Samakovitis, 2021). Dalam penelitian tersebut dilakukan 92 eksperimen berparameter pada arsitektur AWS sederhana dengan menggunakan data tabular. Hasil penelitian tersebut didapatkan sebuah arsitektur yang mampu mencapai pemrosesan data yang konsisten waktu dari muatan peristiwa lebih dari 100 MB, dengan *throughput* 750 KB/dtk di empat frekuensi peristiwa.

Lamanya proses dalam memproses data pada suatu big data disebabkan oleh keanekaragaman jenis data dalam suatu big data. Data dalam suatu big data sangat bervariasi misalnya data media sosial, data transaksi keuangan (Oliviani et al., 2018). Pada penelitian tersebut melakukan analisis performa dari *Hadoop MapReduce* dan membandingkannya dengan *Apache Spark* yaitu platform yang dibuat untuk memproses suatu big data yang dikembangkan berdasarkan *Hadoop MapReduce* dengan peningkatan performa pemrosesan. Skenario yang digunakan adalah memproses *wordcount* suatu data dengan besar data yang berbeda yang bertujuan untuk menganalisis *response time* dan penggunaan *hardware* dari kedua

platform tersebut. Hasil pengujian disimpulkan bahwa penggunaan *Apache Spark* untuk memproses big data sangatlah tepat karena dapat menurunkan response time rata-rata 50% sampai 70% dari *Hadoop MapReduce*. Tetapi untuk mengoptimalkan kemampuannya dibutuhkan juga spesifikasi *hardware* yang mumpuni khususnya spesifikasi *memory* yang besar jika ingin mengoptimalkan kinerja dari *Apache Spark*.

Pengolahan data secara langsung dengan ukuran data yang besar dengan menggunakan metode dan alat konvensional sangat sulit dan tidak efisien (Aminudin & Cahyono, 2019). Dalam penelitian tersebut dilakukan pengujian system dengan dua skenario yaitu penggunaan jumlah node dan ukuran dataset yang berbeda oleh apache spark saat melakukan komputasi terdistribusi dengan menggunakan algoritma deep learning. Pengujian sistem dengan kedua skenario tersebut dilakukan menggunakan alur pengujian yang telah ditentukan agar menghasilkan data yang akurat dan valid untuk keperluan pengambilan kesimpulan. Data yang digunakan yaitu dua buah jenis dataset yang telah disediakan oleh database National Climate Data Center (NCDC) yaitu Storm Event Database (SED) dan Global Surface Summary of the Day (GSSOD). Berdasarkan hasil pengujian nilai running dan nilai throughput bahwa Apache Spark mampu memproses data dengan skala yang besar di dalam lingkungan komputasi parallel dengan membagi beberapa node yang berbeda. Kemudian berdasarkan dari pengujian parameter yang telah diujikan dengan menggunakan parameter yang didapatkan dari tools Benchmark Hibench nilai pengujian parameter bersifat linear artinya hasil nilai parameter yang diujikan didasarkan atas jumlah node dan jumlah besarnya data. Ada dua parameter yang ketika diujikan tidak bersifat linear yaitu Max Memory Usage dan Max CPU Usage hal ini dikarenakan pengiriman data antar node di dalam RDD yang tidak beraturan serta penggunaan salah satu tools Sparkling Water yang digunakan untuk integrasi di dalam lingkungan terdistribusi dan tools machine learning H2O.

Merancang arsitektur sebagai tolak ukur untuk membandingkan dampak kinerja penggunaan Apache Beam pada tiga kerangka kerja streaming: Apache Spark Streaming, Apache Flink, dan Apache Apex. Hal dilakukan untuk

mengimplementasikan system yang sudah ada ke system yang baru karena memerlukan penyesuaian aplikasi yang ada ke antarmuka baru. Pengukuran dapat diperluas sehubungan dengan berbagai aspek seperti jumlah sistem yang dipelajari atau kompleksitas kueri serta perilaku penskalaan, paralelisme, atau toleransi kesalahan. Selain itu, versi Apache Beam yang akan datang dan lapisan abstraksi lainnya dapat dibandingkan dengan hasil yang disajikan untuk melengkapi ikhtisar awal yang disajikan (Hesse et al., 2019).

Dalam penelitian (Rahutomo & Pardamean, 2021) “Data Engineering Pipeline to Analyse Jakarta’s Air Quality during COVID-19-Caused Lockdown Periods” Untuk menganalisis kualitas udara Jakarta setelah melewati lockdown, diperlukan *pipeline data engineering* . Dengan mengakuisisi data deret waktu dari openaq.com, sistem database time-series dikembangkan dengan pemrograman Python bahasa dan perpustakaan fundamentalnya yaitu Pandas, NumPy, SQLite.

Membangun arsitektur data lake untuk melakukan pengawasan terhadap berita-berita online yang beredar (Thenata, 2020). Hasil dari arsitektur yang dibangun yaitu menggabungkan dan menyeragamkan struktur data berita online dari beberapa saluran berita online dan kemudian mengalirkannya secara real-time untuk mengisi data lake. Hasil penggunaan arsitektur data lake untuk berita online akan disimpan pada MongoDB yang berfungsi sebagai database untuk menyimpan semua data baik untuk jangka pendek maupun jangka panjang. Terakhir, data lake ini akan menjadi sarana untuk menampung, menyelami, dan menganalisis data berita online yang beredar.

Melakukan komputasi untuk menghitung nilai rata-rata, maksimum dan minimum pada data streaming. Tujuannya untuk mengembangkan platform pengolahan data streaming IoT secara terdistribusi menggunakan Spark Streaming. Hal ini dilakukan karena arsitektur yang berbasis centralized akan menyebabkan delay yang tinggi dalam penyediaan layanan. Setelah hasil pengolahan didapatkan, hasilnya dikirim ke sistem eksternal, yaitu Apache Kafka dan Apache (Rosianti et al., 2020).

Dari beberapa penelitian di atas, penulis melihat masih banyak perguruan tinggi dan universitas maupun perusahaan besar yang belum memiliki sebuah

sistem yang baik dalam mengelolah data secara baik, baik dari segi penyimpanan data, kecepatan dan kenyamanan dalam penggunaan sebuah sistem. Ada beberapa system yang sudah dibangun namun belum berjalan secara optimal. Dalam penelitian ini penulis akan membangun suatu infrastruktur data terpadu pada kampus Akademi Teknik Biak yang menggunakan teknologi apache spark untuk proses extract-transform-load (ETL) karena kecepatannya dalam mengelolah data secara real-time dan mampu mengolah data dalam jumlah besar.

Tabel 2.1 Tinjauan Pustaka

Nama & Tahun	Permasalahan	Efek Dari Permasalahan	Teknologi Atau Metode	Kontribusi
(Fahmi & Sar, 2018)	- Jika pengelolaan SDM di perguruan tinggi tidak menggunakan perangkat lunak HRIS	- Data SDM yang dimiliki perguruan tinggi menjadi tidak terpusat sehingga ada kemungkinan terjadi redundancy data. - Pengolahan data dan penyajian data membutuhkan waktu yang lama.	Metode yang digunakan pada penelitian ini adalah metode terapan (action research). Pendekatan pengembangan sistem menggunakan metode prototyping. Sedangkan metode perancangan yang digunakan yaitu perancangan sistem dan perancangan rinci, yang meliputi perancangan objek.	Hasil penelitian ini disajikan secara deskriptif dengan rincian sebagai berikut sistem ini sangat baik (45%), kemudahan di dalam penggunaannya (65% menyatakan sangat baik dan 35% menyatakan baik), disamping itu 56% menyatakan sangat baik terhadap desain interface, keakuratan informasi yang diberikan dalam sistem dinyatakan 72% baik, bahasa yang digunakan dalam aplikasi 55% sangat baik dan 45% baik, dan terakhir mengenai tentang error handling sebanyak 32% menyatakan sangat baik, 68% yang menyatakan baik.
(Olivandi et al., 2018)	- Keanekaragaman jenis data dalam suatu big data, data dalam suatu big data sangat bervariasi misalnya data media sosial, data transaksi keuangan	- Lamanya proses dalam memproses data	Menganalisis performa dari Hadoop MapReduce dan membandingkannya dengan Apache Spark yaitu platform yang dibuat untuk memproses suatu big data yang dikembangkan berdasarkan	Hasil pengujian disimpulkan bahwa penggunaan Apache Spark untuk memproses big data sangatlah tepat karena dapat menurunkan response time rata-rata 50% sampai 70% dari Hadoop MapReduce. Tetapi untuk mengoptimalkan kemampuannya dibutuhkan juga

			<p>Hadoop MapReduce dengan peningkatan performa pemrosesan. Skenario yang digunakan adalah memproses wordcount suatu data dengan besar data yang berbeda yang bertujuan untuk menganalisis response time dan penggunaan hardware dari kedua platform tersebut</p>	<p>spesifikasi hardware yang mumpuni khususnya spesifikasi memory yang besar jika ingin mengoptimalkan kinerja dari Apache Spark</p>
(Aminudin & Cahyono, 2019)	<p>Pengolahan data secara langsung dengan ukuran data yang besar dengan menggunakan metode dan alat konvensional</p>	<p>- Sangat sulit dan tidak efisien.</p>	<p>-Melakukan pengujian system dengan 2 skenario yaitu penggunaan jumlah node dan ukuran dataset yang berbeda oleh apache spark saat melakukan komputasi terdistribusi dengan menggunakan algoritma deep learning. -Pengujian sistem dengan kedua skenario tersebut dilakukan menggunakan alur pengujian yang telah ditentukan agar menghasilkan data yang akurat dan valid untuk keperluan pengambilan kesimpulan.</p>	<p>- Berdasarkan dari pengujian nilai running dan nilai throughput bahwa Apache Spark mampu memproses data dengan skala yang besar di dalam lingkungan komputasi parallel dengan membagi beberapa node yang berbeda. - Berdasarkan dari pengujian parameter yang telah diujikan dengan menggunakan parameter yang didapatkan dari tools Benchmark Hibench nilai pengujian parameter bersifat linear artinya hasil nilai parameter yang diujikan didasarkan atas jumlah node dan jumlah besarnya data. - Ada dua parameter yang ketika diujikan tidak bersifat linear yaitu Max Memory Usage dan Max CPU Usage hal ini dikarenakan pengiriman data antar</p>

				node di dalam RDD yang tidak beraturan serta penggunaan salah satu tools Sparkling Water yang digunakan untuk integrasi di dalam lingkungan terdistribusi dan tools machine learning H2O.
(Hesse et al., 2019)	Memindahkan implementasi system yang sudah digunakan ke system yang baru	Memerlukan penyesuaian aplikasi yang ada ke antarmuka baru	Arsitektur sebagai tolak ukur baru untuk membandingkan dampak kinerja penggunaan Apache Beam pada tiga kerangka kerja streaming: Apache Spark Streaming, Apache Flink, dan Apache Apex.	Penggunaan Apache Beam untuk aplikasi streaming yang diperiksa menyebabkan variasi waktu eksekusi kueri yang tinggi dengan perlambatan hingga faktor 58 dibandingkan dengan kueri yang dikembangkan tanpa lapisan abstraksi
(Primadewi & Hanafi, 2020)	- Pengaruh adanya Perubahan aturan Akreditasi Perguruan Tinggi (APT) 3.0 yang di berlakukan sejak April 2019	- Beberapa sistem informasi khususnya SIMMawa, SIMHumas dan Kerjasama serta SIMAKU Mengalami kekosongan data - Adanya kekosongan data dalam bentuk Laporan Kinerja Perguruan Tinggi (LKPT) Sebagai acuan dalam membuat Laporan Evaluasi Diri (LED)	Proses pengintegrasian fungsi bisnis, entitas data dan aplikasi mengacu pada referensi Zachman Framework (ZF)	Mendapatkan sebuah model alternatif pengelolaan data terintegrasi yang dapat dijadikan masukan untuk pengembang sistem informasi di lingkup perguruan tinggi untuk mengantisipasi kekosongan data tersebut
(Henim & Sari, 2020)	Sistem Informasi Akademik Mahasiswa pada perguruan tinggi yang masih monoton dan lazim serta tidak memiliki <i>Novelty</i> (Kebaruan)	Kurangnya <i>attractives</i> (daya tarik pengguna) pada sistem informasi yang dibangun	Pengukuran dilakukan dengan menggunakan User Experience Questionnaire (UEQ).	Memberikan rekomendasi untuk perancangan <i>user experience system</i> informasi akademik yang baru dengan melakukan inovasi pada berbagai aspek

(Rully et al., 2020)	Belum adanya system pengarsipan pada Perguruan Tinggi Universitas Catur Insan Cendekia Cirebon untuk mengelola	Data terkait Penelitian dan Pengabdian kepada masyarakat (LPPM) tidak di Kelola dengan baik	<ul style="list-style-type: none"> - Menggunakan metode Alphabetical filing system dalam metode pengarsipannya, metode ini dituangkan pada bahasa pemrograman PHP dan Javascript. - Kemudian data kerjasama, kegiatan penelitian dan pengabdian kepada masyarakat ini akan disimpan pada database MYSQL, software yang digunakan yaitu sublime text, XAMPP, Browser dan aplikasi pembuatan perancangan UML, dan hardware yang digunakan adalah notebook Lenovo G41 dengan processor AMD A8 APU Radeon 2.2 GHz. 	Mempermudah Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) dalam mengarsipkan data kerjasama, kegiatan penelitian dan pengabdian kepada masyarakat yang dilakukan oleh dosen.
(Thenata, 2020)	- Berita online menarik perhatian pembaca dengan memberikan kemudahan dan kecepatan dalam menyajikan berita dari berbagai bidang.	- Kebutuhan masyarakat untuk mengkonsumsi berita dalam berbagai referensi (variety) dapat mempengaruhi kehidupan masyarakat	- Merancang arsitektur data lake	- Data lake ini akan menjadi sarana untuk menampung, menyelami, dan menganalisis data berita online yang beredar
(Cottur & Gadad, 2020)	-Data mentah susah untuk dikonsolidasikan/disatukan	- Berbagai aspek karakteristik eksistensial data mentah tidak	- Menggunakan metode ETL	- Mengatasi tantangan korporasi fasilitas cloud perantara yang dikenal sebagai data

		dapat ditanyakan secara langsung karena tidak tersedia pada perusahaan tersebut, seperti kompetensi pasar		lake, di mana membawa teknologi perpipaan data dan langsung menerima data dibersihkan dan kemudian dimuat ke gudang untuk data yang diproses di akhir siklus sinkronisasi data. - Mewakili pengambilan terus menerus dari data yang dihasilkan atau data yang diubah selama periode tertentu seperti dua minggu atau satu
(Rosianti et al., 2020)	- Arsitektur yang berbasis centralized	- Akan menyebabkan delay yang tinggi dalam penyediaan layanan	- Mengembangkan platform pengolahan data streaming IoT secara terdistribusi menggunakan Spark Streaming	- Pengujian fault tolerance menunjukkan 6 node memproses lebih cepat dengan waktu pemrosesan 60 detik. - Spark Streaming juga menunjukkan kemampuan penggunaan resource terdistribusi secara efisien dengan menunjukkan penggunaan CPU dan memori di semua node tidak memiliki perbedaan yang signifikan. -Rata-rata perbedaan penggunaan memori dan CPU pada masing-masing node adalah 3,6% dan 2,04%.
(Nawawi & Rubedo, 2021)	- Banyak Perguruan Tinggi yang belum memiliki sistem dan database khusus untuk mengelola data penelitian dan PKM.	- Dosen terkadang lupa terhadap hasil laporan penelitian dan PKM yang telah dilakukan. - Dosen akan mendapatkan kendala dalam pengajuan Jabatan Fungsional maupun Akreditasi, sehingga mereka harus mencari-cari	- Metode yang digunakan adalah deskriptif-action, yaitu mendeskripsikan keseluruhan sistem terlebih dahulu lalu menganalisis kekurangan dan kelemahan yang terjadi dan selanjutnya adalah melakukan tindakan disertai dengan	- Diharapkan sistem akan memberikan kemudahan akses bagi dosen dan stakeholder lain yang berkepentingan di <i>International Women Universty</i> seperti pencarian, penyortiran, termasuk penyimpanan data

		dan beberapa bahkan tidak menemukan karena tak ada catatan dan pengelolaan yang baik.	solusi-solusi pemecahannya.	
(Fauzi et al., 2021)	- Dalam masa Pemilihan Umum di Indonesia, masih ada berita-berita yang disampaikan secara berpihak atau tidak aktual	- Banyak berita yang memiliki unsur keberpihakan dalam memberikan informasi	- Metode yang digunakan untuk mengklasifikasi sentimen sebuah data (pada kasus ini adalah teks berita) yaitu Naïve Bayes Classifier. - Menggunakan tabel Confusion Matrix sebagai tabel evaluasi dari hasil klasifikasi sentimen berita, dengan menghitung nilai evaluasi seperti akurasi, presisi, recall, dan F-Measure.	- Berdasarkan berbagai pengujian dan skenario yang telah dilakukan, nilai evaluasi terbaik dihasilkan pada pengujian menggunakan K-Fold Cross Validation dengan nilai K=10. Pada pecahan (fold) ke-8 menghasilkan nilai akurasi sebesar 100%, presisi sebesar 100%, recall sebesar 100%, dan F-Measure sebesar 100%.
(Pogiatzis & Samakovitis, 2021)	- Arsitektur tanpa server pipeline Extract-Transform-Load (ETL)	- Adanya hambatan pada muatan peristiwa aliran data dengan berbagai frekuensi, kecepatan dan ukuran payload.	- Menjalankan 92 eksperimen berparameter pada arsitektur AWS sederhana	- Hasil kami menunjukkan bahwa arsitektur referensi kami dapat mencapai pemrosesan data yang konsisten waktu dari muatan peristiwa lebih dari 100 MB, dengan throughput 750 KB/dtk di empat frekuensi peristiwa.
(Fauzi et al., 2021)	- Media berita pada saat pemilihan umum masih ada berita-berita yang disampaikan secara sepihak atau tidak actual.	- Hal tersebut pun membuat masyarakat masih berpandangan bahwa banyak berita yang memiliki unsur keberpihakan dalam memberikan informasi	- Metode yang digunakan untuk mengklasifikasi sentimen sebuah data (pada kasus ini adalah teks berita) yaitu Naïve Bayes Classifier. - Menggunakan tabel Confusion Matrix sebagai	- Berdasarkan berbagai pengujian dan skenario yang telah dilakukan, nilai evaluasi terbaik dihasilkan pada pengujian menggunakan K-Fold Cross Validation dengan nilai K=10. Pada pecahan (fold) ke-8 menghasilkan nilai akurasi sebesar 100%, presisi sebesar 100%, recall sebesar

			tabel evaluasi dari hasil klasifikasi sentimen berita, dengan menghitung nilai evaluasi seperti akurasi, presisi, recall, dan F-Measure	100%, dan F-Measure sebesar 100%.
(Santoso, 2022)	- Sistem pengelolaan dokumen terkadang hanya dijalankan pada lembaga induk secara terpusat namun tidak terlaksana dengan baik pada level low management.	- Kesulitan untuk mengakses data secara langsung. - Audit intenat atau akreditasi kelompok ini haruslah memiliki data yang dibutuhkan secara cepat dan tepat dan sangat sulit untuk mendapatkannya.	- Teknik pengumpulan data menggunakan observasi, wawancara, penilaian pakar. Instrumen yang digunakan adalah lembar wawancara, ceklisch observasi. lembar validasi pakar dan angket.	- Sistem database dari segi media, praktisi, dan pengguna sangat layak untuk dapat diterapkan.
(Nasution et al., 2022)	- Dampak negative bagi sebuah Universitas bila tidak didukung oleh suatu system yang terintegrasi dalam melakukan pencarian data ijazah alumni.	- Pencarian data Ijazah alumni yang telah lulus untuk memenuhi kebutuhan informasi pemangku kepentingan (stakeholders) adalah salah satu hal yang sangat sulit dilakukan Di sisi lain, stakeholders terutama lembaga, instansi dan perusahaan yang akan merekrut para lulusan, membutuhkan waktu yang cepat untuk mengetahui dan menguji kebenaran Ijazah para lulusan.	- Metode Grid Data (GD) adalah salah satu metode yang dapat digunakan dalam proses BI, dimana Metode GD dapat melakukan analisis yang tepat dalam mengukur validitas data terutama data ijazah	- Dengan menggunakan sistem informasi berbasis kecerdasan bisnis (business intelligent) memudahkan proses integrasi data lulusan, sehingga tidak sulit untuk menghasilkan data secara optimal, efektif dan efisien.
(Aryani et al., 2022)	- Data pendukung untuk sistem remunerasi diambil	- Menyebabkan jika terjadi perubahan data	- Menggunakan pendekatan Zachman	Hasil dari penelitian ini adalah gambar arsitektur aplikasi, arsitektur

	melalui pusat data namun ada juga yang diambil langsung dari masing-masing sistem pendukung	pada satu sistem, akan timbul ketidakkonsistenan dan duplikasi data sistem di UPT TIK belum terdokumentasi dengan baik	Framework, yang merupakan salah satu metode EAP dalam perancangan sistem.	teknologi dan rancangan integrasi data sistem penunjang remunerasi yang diterjemahkan dari hasil pemetaan matrik Zachman yang berjumlah 36 sel yang terdiri dari 6 kolom dan 6 baris.
(Belcastro et al., 2022)	Di era Internet of Things dan platform media sosial, sejumlah besar data digital dihasilkan oleh dan dikumpulkan dari banyak sumber, termasuk sensor, perangkat seluler, pelacak yang dapat dikenakan, dan kamera keamanan	- Memberikan tantangan bagi pengembang untuk mencari solusi dalam kemampuan penyimpanan, pemrosesan, dan analisis saat ini	- Menganalisis dan meninjau paradigma, bahasa, dan sistem paralel dan terdistribusi yang digunakan saat ini untuk menganalisis dan belajar dari Big Data pada komputer yang dapat diskalakan. - Analisis mendalam tentang properti paradigma pemrograman paralel utama (MapReduce, alur kerja, BSP, penyampaian pesan, dan seperti SQL) dan, melalui contoh pemrograman, kami menjelaskan sistem yang paling banyak digunakan untuk analisis Big Data (misalnya, Hadoop, Spark, dan Storm).	- Analisis mendalam tentang properti paradigma pemrograman paralel utama (MapReduce, alur kerja, BSP, penyampaian pesan, dan seperti SQL) dan, melalui contoh pemrograman, kami menjelaskan sistem yang paling banyak digunakan untuk analisis Big Data (misalnya, Hadoop, Spark, dan Storm).
(Science & Science, 2023)	- Data tidak dapat dengan mudah ditafsirkan dalam bentuk mentahnya	- Meningkatnya sejumlah sifat eksistensial entitas bisnis, termasuk manajemen organisasi,	- ETL, atau mengekstrak, mengubah, dan memuat, adalah metode yang digunakan	- Mengatasi kendala bisnis yang ada untuk teknologi data pipeline dan mendapatkan data yang diproses langsung di akhir siklus sinkronisasi data

		kapabilitas pasar, dan umpan balik konsumen.		
(Heman & Bambang Purnomosidi, 2023)	<ul style="list-style-type: none"> - Belum ada suatu sistem terpadu antara kampus Akademi Teknik Biak dan Lembaga Layanan Pendidikan Tinggi Wilayah XIV Provinsi Papua 	<ul style="list-style-type: none"> - Kesulitan dalam memperoleh data yang akurat antara Lembaga Layanan Pendidikan Tinggi Wilayah XIV dengan perguruan tinggi yang ada di provinsi Papua. - Kesulitan dalam memperoleh data yang cukup besar dan real time. 	<ul style="list-style-type: none"> - Merancang suatu arsitektur data pipelines - Menggunakan <i>Apache Spark</i> untuk proses implementasi <i>extract transform load</i> (ETL) data. 	<ul style="list-style-type: none"> - Teknologi <i>Apache Spark</i> pada arsitektur data pipelines sistem informasi akademi teknik biak, dapat menjadi acuan untuk membangun suatu sistem terpadu antara Lembaga Layanan Pendidikan Tinggi Wilayah XIV dengan kampus Akademi Teknik Biak. - Kebutuhan akan data yang cukup besar dapat terpenuhi.