REAL-TIME BI & ANALYTICS

P. Anitha Sahaya Mercy[1], M. Pramila[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Annai Vailankanni College of Engineering, Kanyakumari.
mercy-cs@avce.edu.in
[2]Assistant Professor, Department of Computer Science and Engineering, Annai Vailankanni College of Engineering, Kanyakumari.
pramila-cs@avce.edu.in

ABSTRACT
In today's fast-paced world, traditional business intelligence (BI) that relies on historical data is no longer sufficient. Real-time BI and analytics (RTBI&A) emerge as a powerful solution, enabling organizations to analyze data as it arrives, offering near-instant insights for proactive decision-making. RTBI&A is a game-changer, providing a competitive edge in the data-driven landscape. It equips businesses with the ability to act on insights in real-time, shaping a future of proactive and informed decision-making. RTBI&A prioritizes analyzing data with minimal latency, allowing businesses to react to trends, issues, and opportunities as they unfold. By providing real-time insights, RTBI&A empowers businesses to optimize operations, personalize customer experiences, and mitigate risks proactively. Stream processing engines, in-memory computing, and real-time data visualization techniques are crucial for analyzing and presenting insights from high-velocity data streams. Organizations can leverage RTBI&A to achieve improved operational efficiency, increased productivity, reduced costs, enhanced customer satisfaction, and superior decision-making.

Keywords: Real-time Analytics, Streaming Analytics, Stream Processing.

INTRODUCTION
In the era of big data, where information flows like a raging river, businesses need insights faster than ever. Enter real-time business intelligence (RTBI) and analytics, the dynamic duo that analyzes data as it arrives, providing near-instantaneous knowledge to guide decisions. RTBI and analytics act like a crystal ball, offering a glimpse into the present to make informed choices about the future. By monitoring key metrics like production line efficiency, businesses can identify areas for improvement in real-time. Understanding customer behavior as it happens allows for more personalized interactions, leading to a more satisfying customer experience. Proactively identifying potential issues enables preventive actions to safeguard the business. The magic behind RTBI and analytics lies in its speed, utilizing techniques like streaming data ingestion and real-time processing to analyze massive amounts of data with minimal lag. This enables businesses to react instantly to changing market conditions, customer preferences, and internal operations.

REAL-TIME BI & ANALYTICS TECHNIQUES
Real-time BI and analytics (RTBI&A) leverage a combination of advanced techniques to analyze data with near-zero latency. Below is a breakdown of some key methods:

Data Ingestion and Processing:
- **Streaming Data Ingestion:** Specialized tools like Apache Kafka or Amazon Kinesis capture data from various sources, such as sensors, social media, and transactions, as it is generated. This is akin to a continuous firehose of data flowing into the system.
- **Stream Processing:** Engines like Apache Flink or Apache Spark Streaming cleanse, transform, and analyze incoming data streams on the fly. This process is comparable to a factory assembly line, where data is processed in real-time.

Analytical Techniques:
- **In-Memory Computing:** Data is stored in RAM for ultra-fast processing, allowing for complex calculations and real-time analytics without the delays associated with disk access. It's like having all the data instantly available at your fingertips for immediate analysis.
- **Time Series Analysis:** This technique analyzes data points collected at regular intervals to identify trends, seasonality, and anomalies. It is particularly useful for tracking metrics like website traffic, stock prices, or sensor readings over time. For example, it can help spot traffic spikes on a website and determine whether they are due to a seasonal trend or a successful campaign.
- **Machine Learning:** Techniques such as anomaly detection and real-time classification can be applied to identify unusual patterns or categorize data streams as they arrive. For instance, it can automatically flag fraudulent transactions or classify customer sentiment on social media in real-time.

Table 1: Key Aspects of Real-Time BI & Analytics

Feature	Description
Focus	Analyze live data and extract insights in real-time.
Data	Processes data streams as soon as they enter the system.
Benefits	- Enables informed decision-making within minutes. - Allows for faster response and adaptation to changing situations. - Identifies emerging trends and opportunities. - Helps prevent potential issues proactively.
Technology	Leverages tools such as: - Real-time streaming platforms - In-memory data storage - Interactive data visualization tools
Applications	- Customer service monitoring - Fraud detection - Social media analytics - Stock market analysis - IoT sensor data analysis

OPERATIONAL BUSINESS INTELLIGENCE

Operational BI facilitates bidirectional communication between operational systems and analytic applications. The insights and knowledge obtained from analytic systems are not only used for decision-making but also for improving business processes and adapting operational systems to better respond to changing market conditions.

Wayne W. Eckerson, Research Director at The Data Warehouse Institute (TDWI), introduced a maturity model for BI in his paper [WAECK07], as presented in Figure 2.

This model is based on the concept of latency, which refers to the temporal delay between the initiation of an event and the moment when its effects become apparent. In the model, the red line represents the freshness of the data, indicating how current or outdated the data is. The blue line represents the latency of the decision-making process. There are three types of latency in a decision-making process:

- **Data Latency:** The time required to collect data from source systems, prepare it for analysis, and store it in data warehouses or data centers.
- **Analytic Latency:** The time needed to access and analyze the data, transform it into information, and apply business rules.
- **Decisional Latency:** The time required to review the analysis, make decisions, and implement actions.

In Wayne Eckerson's model, traditional business intelligence corresponds to the pre-natal and child phases, which are characterized by high latency in the decision-making process and low freshness of data.
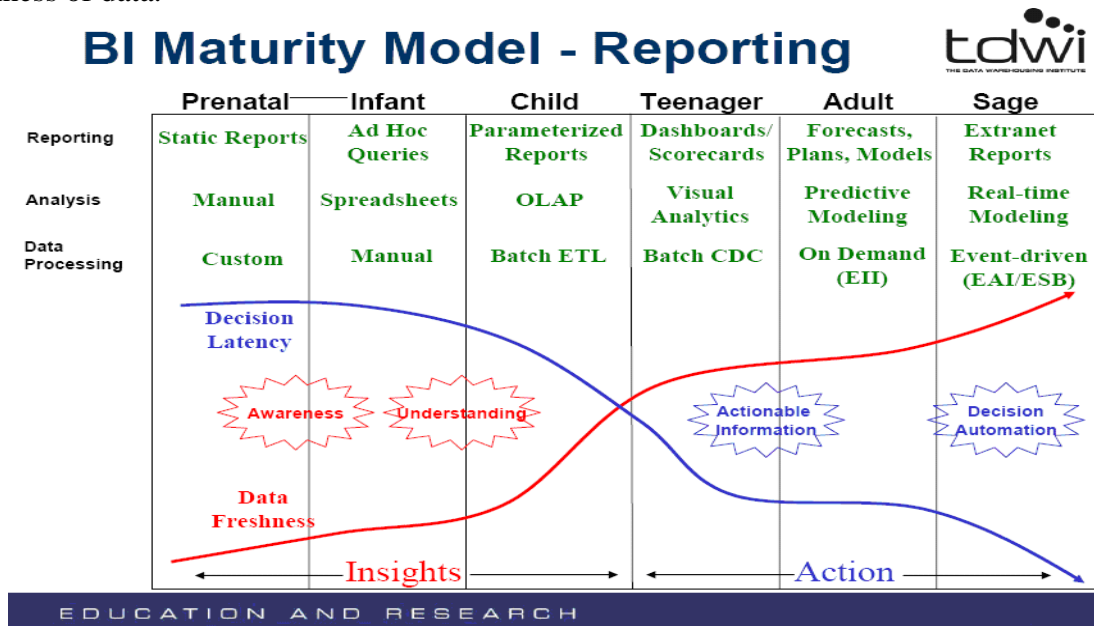

Fig.1:   Maturity   Model   for   Business   Intelligence
Source: Wayne W. Eckerson, TDWI

REAL-TIME BI & ANALYTICS SYSTEM

Real-time analytics architecture empowers businesses to collect, process, and analyze data in real time, enabling swift, data-driven decisions, deep operational insights, and a competitive edge in the industry. This advanced data processing architecture is crucial for harnessing the potential of real-time data, providing the tools necessary to thrive in today's fast-paced market.

To effectively process and analyze big data, businesses require a robust real-time analytics infrastructure. This infrastructure must rapidly collect, process, and store data from various log files in an easily accessible manner. Due to the high velocity of data, gathering log data from multiple servers can be complex. Once processed, the data must be enriched with reference and lookup data before being imported into the destination database. The Real-Time Analytics architecture supporting this process must be resilient, with load-balancing and fail-over capabilities to ensure uninterrupted performance.

The main elements of a typical real-time analytics architecture are as follows:

1. **Data Sources**
   The first step in real-time analytics is gathering data from numerous sources, including social media feeds, sensors, online logs, and mobile apps. These sources produce essential data that must be immediately consumed and analyzed.

2. **Data Processing**
   Once gathered from multiple sources, data must be processed in real-time to extract insights. This process involves filtering, cleaning, and transformation to ensure the data is accurate and relevant for analysis.

3. **Data Storage**
   A vital element of the architecture for real-time analytics is data storage. It must be quick, scalable, and capable of handling the massive amounts of data generated by various sources. Examples of popular data storage options for real-time data analytics architecture include:

   o **In-Memory Databases**
     These databases store data in the system's main memory, allowing for rapid storage and retrieval. Their minimal latency and capacity to handle large amounts of data make them ideal for real-time analytics.

   o **Distributed File Systems**
     Distributed file systems, such as the Hadoop Distributed File System (HDFS), allow large volumes of data to be stored across a cluster of computers. As data volumes increase, storage capacity can easily be scaled up.

   o **NoSQL Databases**
     NoSQL databases are non-relational and can store substantial amounts of unstructured and semi-structured data. They are frequently used in real-time analytics applications and offer fast and flexible data processing.

Scalable, fast storage solutions are necessary for handling the large volumes of data produced by multiple sources, with data stored in a manner that allows for easy retrieval and analysis.

4. **Analytics Engine**
   The analytics engine is central to real-time analytics architecture, responsible for processing, storing, and analyzing data. The architecture must be designed to handle large data volumes and deliver real-time analytics capabilities. Businesses can build a scalable and robust analytics engine using cloud-based infrastructure and data platforms, providing a competitive advantage. The analytics engine enables businesses to make data-driven decisions and gain a competitive edge.

5. **Visualization**
   Making the insights generated by the analytics engine accessible to users is essential. Visualization tools, such as dashboards and reports, present findings in a way that is easy to understand and act upon.

These components work together seamlessly to enable organizations to gain real-time insights into their operations, improve decision-making, and secure a competitive advantage in their industry.
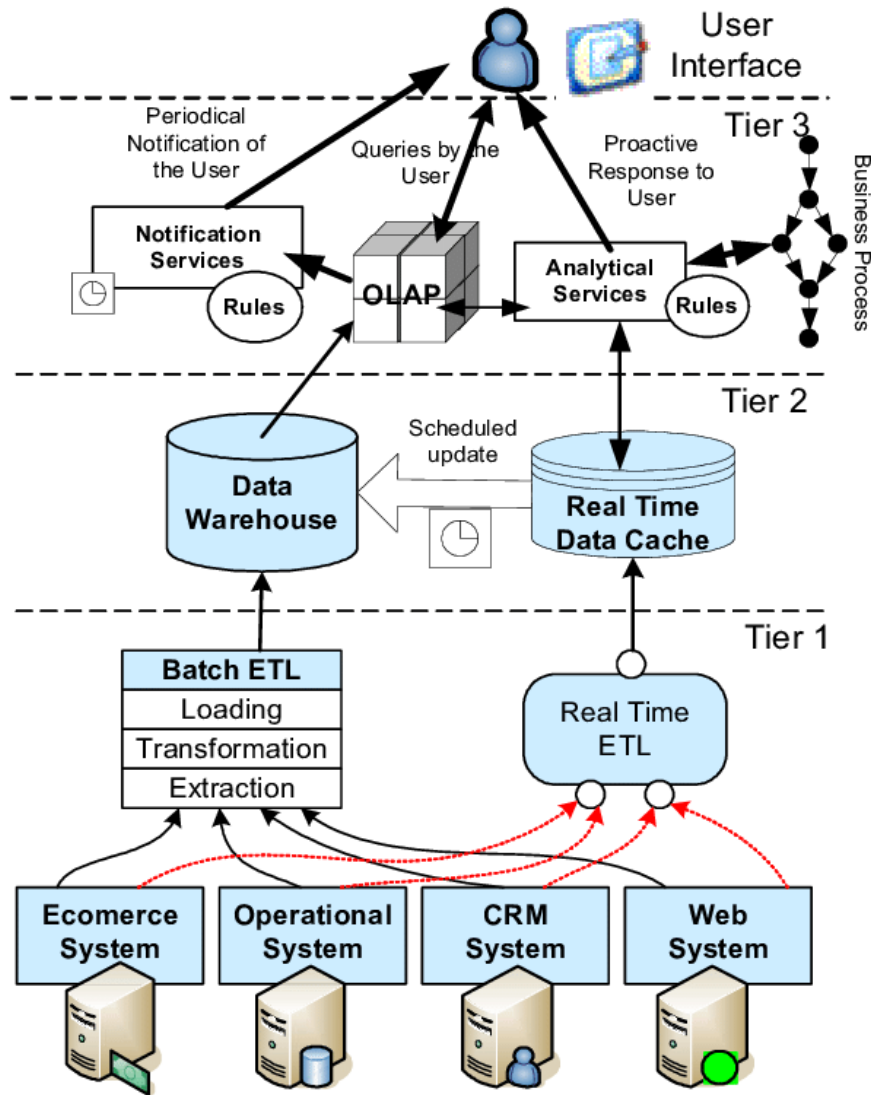
Fig. 2: Real-Time BI & Analytics System

THE PROPOSED MODEL

This paper proposes a Real-Time Business Intelligence (RTBI) solution for the utilities industry, taking into account the type of data source. For real-time data, such as weather information, sensor and meter states, and production and consumption data, we propose the use of an external real-time data cache. This cache will contain only the tables with real-time data, while non-real-time data will be extracted, transformed, and loaded directly into the traditional data warehouse.

By using an external real-time data cache, we can eliminate performance issues associated with integrating real-time data into a data warehouse. Additionally, this approach helps resolve other challenges, such as internal inconsistency and data latency.

With all real-time activity managed on the external cache database, the data warehouse is not subjected to any additional load, thereby solving scalability and query performance problems.

The connection between the real-time data cache and the data warehouse is managed through scheduled updates and real-time updates when specific conditions are met. For example, production and consumption data may be updated hourly, whereas alerts triggered by measurable factors or equipment malfunctions need to be updated instantly in the data warehouse.

To create advanced business analytics with a just-in-time information merging solution, we can seamlessly merge real-time data from the external real-time data cache with historical data from the data warehouse.

RESULTS

An ideal business intelligence system provides utilities with the right information in the right format at the right time, enabling them to make optimal business decisions. Given the time-sensitive nature of decision-making in this industry, utilities require Real-Time Business Intelligence (RTBI) solutions to enhance customer experiences and operational efficiencies. Our vision of real-time business intelligence for the utilities industry involves using an external real-time data cache in conjunction with a traditional data warehouse. This approach aims to eliminate performance issues and address challenges such as internal inconsistency, scalability, and data latency. Although various technologies exist for implementing real-time business intelligence, significant challenges remain in realizing this vision.

CONCLUSION

Real-time business intelligence will soon be accessible to everyone within an enterprise and will be embedded in many business systems. While numerous technologies are available to implement this vision, significant challenges must still be overcome. We have identified key challenges, such as automated analytics, semantics-based information fusion, and process automation, and presented examples that demonstrate the feasibility of our vision. We believe that technologies like intelligent data analysis, soft computing, and ontologies will play a pivotal role in the development of RTBI.

REFERENCES

1. Wayne W. Eckerson, Performance Dashboards: Measuring, Monitoring, and Managing Your Business, John Wiley & Sons, 07.10.2010 - 336 pp.
2. B.S. Sahay, Jayanthi Ranjan, "Real time business intelligence in supply chain analytics", Information Management & Computer Security, Vol. 16 Iss: 1, pp.28 – 48, 2008.
3. Purnendu Kumar, "Smart Grid…the next generation of electrical transmission and distribution system", 04.09.2012.
4. Judith. R. Davis „Right Time Business Intelligence Optimizing the Business Decision Cycle", research report published by BEYE Research, December 2005.
5. Azvine, B., Cui, Z. and Nauck, D. (2005), "Towards real-time business intelligence", BT Technology Journal, Vol. 23 No. 3, pp. 214-25.
6. Business Intelligence for the Enterprise, By Mike Biere, Publisher: Prentice Hall PTR, Pub Date: June 04, 2003, ISBN: 0-13-141303-1, 240 pages.
7. Justin Kern, "Utilities Surge Ahead with BI, Analytics Despite Immaturity and Integration Problems", 20.09.2012.

8. Colin White, "Now is the Right Time for Real-Time BI", Information Management Magazine, September, 2004.

9. B Azvine, Z Cui, D D Nauck, B Majeed, "Real Time Business Intelligence for the Adaptive Enterprise," cec-eee, pp.29, The 8th IEEE International Conference on ECommerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services (CEC/EEE'06), 2006.

10. Daniela Ioana Sandu, "Operational and real-time Business Intelligence", Informatica Economică Journal 3(47)/2008.