(IJPS) 2020, Vol. No. 10, Jul-Dec **e-ISSN: 2455-6270; p-ISSN: 2455-7455**

HARNESSING ADVANCED GPU ACCELERATED TECHNIQUES FOR AUGMENTED DATA ANALYSIS AND VISUALIZATION IN ENTERPRISE RESOURCE PLANNING (ERP)

Bhavesh Gondaliya
University of Mumbai, Mumbai, India

ABSTRACT

In the context of the fourth industrial revolution, the enterprise resource planning system (ERP) serves as a crucial nexus for integrating logistics systems, production facilities, innovative machinery, IoT devices, and other enterprise data sources. This paper presents an approach to augment the analytical capabilities of ERP-integrated tools by leveraging a multi-tenant GPU-enabled high-performance computing (HPC) environment. By harnessing corporate analytics alongside GPU-accelerated, in-memory processing of both structured and unstructured data, this approach enhances the efficiency and effectiveness of enterprise machine learning (ML) tasks. The proposed method advocates for data sharing in GPU memory through an open analytic platform, complementing existing ERP analytical functionalities, demonstrated with SAP S/4Hana as an example. This solution streamlines the workflow of data scientists dealing with ERP datasets, facilitating faster, higher-quality AI model development and simplifying data interaction through non-ERP visualization methods such as immersive learning with virtual or augmented reality (VR/AR).

INTRODUCTION

The enterprise resource planning system (ERP) is the cornerstone of intelligent manufacturing under the industry 4.0 framework. It aggregates and analyses data from various sources, including smart devices and cyber-physical systems. This data, originating from manufacturing execution systems (MES), client relationship management (CRM), supply chain management (SCM), and other enterprise applications, forms a crucial part of ERP data. Extracting valuable insights from this aggregated data enhances product quality and reduces manufacturing costs.

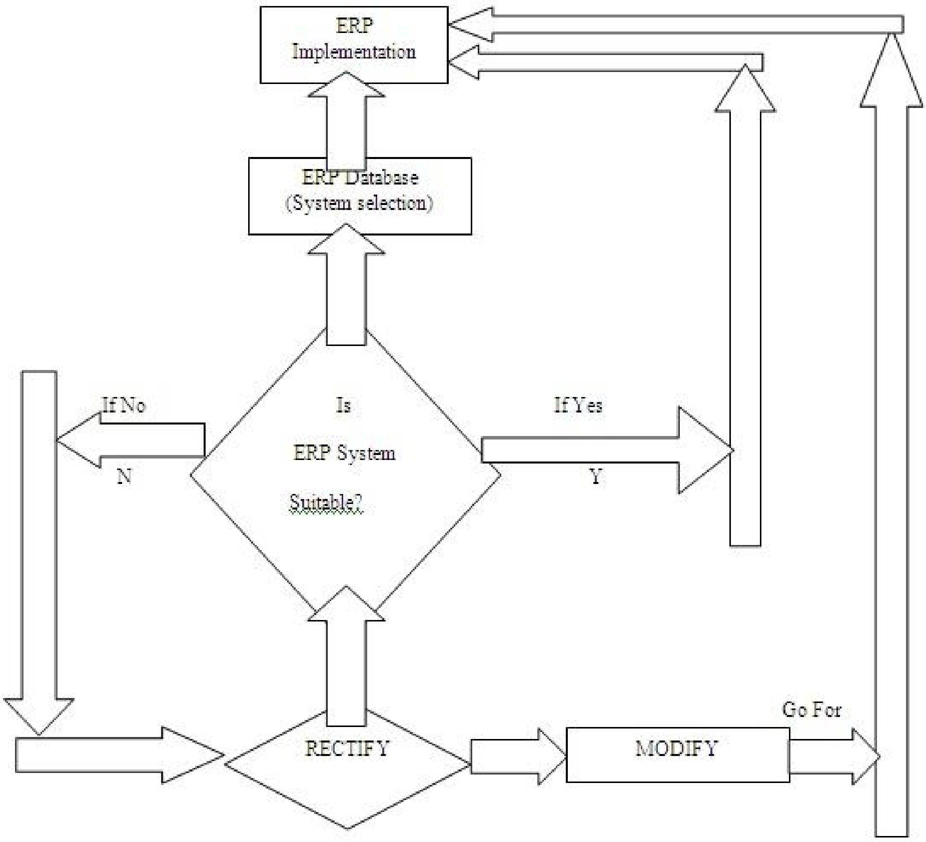

Fig 1: ERP System Implementation Process

This paper proposes a big data processing approach to analyze enterprise data within a high-performance computing environment and visualize it in a virtual reality (VR) environment.

MOVING ERP DATA PROCESSING TO GPU

ERP vendors offer different analytical capabilities based on underlying database architectures. Major vendors like Oracle and SAP utilize column-oriented database formats to accelerate ERP in-database analytics. In-memory column-based data processing offers high performance compared to traditional row-oriented storage methods. However, ERP-specific in-database machine learning processes, including model inference and training, are predominantly CPU-based, which may need to be more efficient, especially during training.

This paper introduces ERP data extraction followed by GPU-accelerated, in-memory data processing in column-based Apache Arrow format, using SAP S/4Hana as an example.

SAP S/4HANA ML AND ANALYTICAL SOLUTIONS

SAP Leonardo Machine Learning and analytics are integral to the SAP Leonardo platform, alongside services like the Internet of Things (IoT) and Blockchain. The Leonardo Machine

Learning Foundation offers a wide range of machine learning algorithms and services on the cloud. However, utilizing APIs for data processing with cloud-based ML services may raise security concerns for sensitive corporate data.

On-premise analytical capabilities include the Predictive Analytics Library (PAL) for in-database analytics housed within the HANA Application Function Library (AFL). Additionally, SAP Predictive Analytics, as a standalone application, extends PAL functionalities on-premise.

The Automated Predictive Library (APL) enables the development of predictive modelling processes and automates tasks such as feature engineering, selection, sampling, and cross-validation.

The SAP HANA Python Client API for Machine Learning allows for launching ML algorithms within a Python environment alongside HANA DataFrame without data movement to the client. Similar functionality is available through the R API in the latest SAP HANA release.

SAP solutions for machine learning primarily utilize in-database and standalone analytics without GPU acceleration or cloud-based solutions, which may need to be improved for private data cases. Even with R and Python, massive data processing typically remains in the database rather than on the client side.

With SAP HANA External Machine Learning (EML), it is possible to utilize the TensorFlow platform for ML to avoid processing sensitive corporate data in the cloud. TensorFlow Model Server (TMS) is an external service with GPU acceleration in this scenario.

SAP DATA EXTRACTION FOR GPU PROCESSING

Figure 2 illustrates the interaction of components with data extraction flows depicted by blue arrows and accelerated data processing by red arrows.

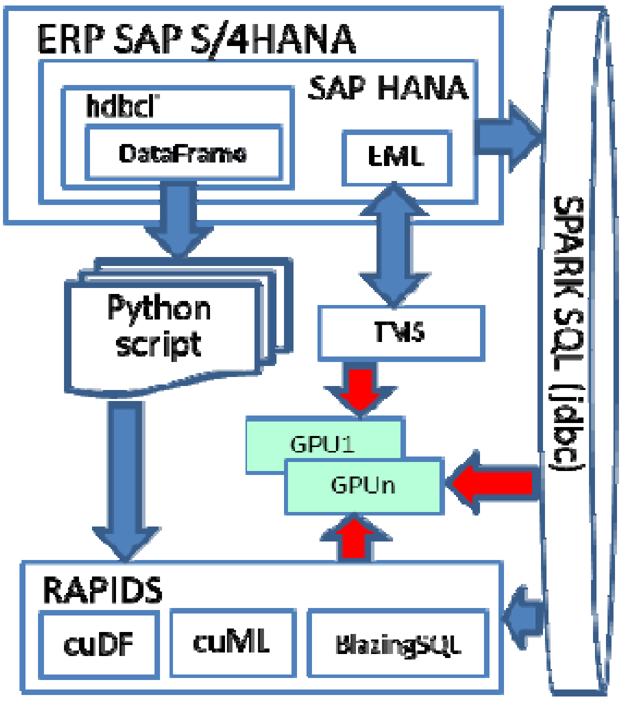

Fig. 2. The components interaction with data extraction flows (blue arrows) and in GPU processing (red arrows)

Previous versions of R/3 ERP, based on Advanced Business Application Programming (ABAP) language, allowed data loading and extraction through the SAP NetWeaver Integration platform. Data could be obtained directly from R/3 and S/4 (in the case of NetWeaver) via the SAP Remote Function Call (RFC) protocol using standard SAP connectors or libraries such as pyrfc and node-rfc.

SAP S/4Hana utilizes the integration features of the in-memory SAP Hana database. When the performance of CPU-based ML inference and training and in-memory processing is inadequate, data extraction for subsequent GPU memory processing becomes necessary. Several methods for data extraction are proposed:

1. ODBC or JDBC can connect directly to Hana from external applications and extract the required data.

2. Leveraging SAP Smart Data Access (SDA) to create virtual tables connected to remote data sources, allowing for data movement to external sources.

3. Utilizing Spark SQL connection to Hana tables via JDBC to fetch and store data on HDFS in the Arrow-compatible Parquet format, followed by processing in the GPU-aware Blazing SQL engine.

(IJPS) 2020, Vol. No. 10, Jul-Dec **e-ISSN: 2455-6270; p-ISSN: 2455-7455**

To initiate data transfer from the Hana side, the SAP HANA Spark controller deployed in the Hadoop ecosystem and the Spark SQL adapter (as an SDA plugin) on the Hana side is recommended. Alternatively, SAP Hana Vora can facilitate the initiation of bidirectional data transfer. Data analysis in Python is possible using the Python client API, enabling data transfer from HANA DataFrame to Python Pandas DataFrame and subsequently to cuDF for GPU memory processing.

DATA PROCESSING IN ARROW PLATFORM

Upon completion of the extraction process and with GPU access to ERP data, accelerated data processing can commence. Figure 3 illustrates the subset of ERP data loaded into GPU memory.

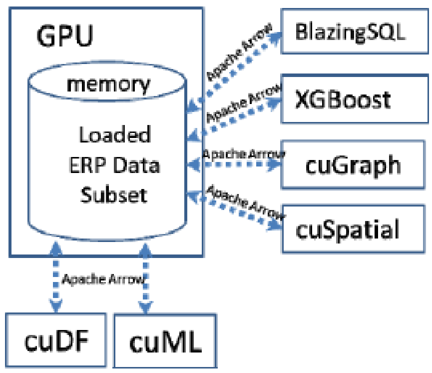

Fig. 3. The loaded in GPU memory subset of ERP data and sharing among the RAPIDS and other libraries

The NVIDIA RAPIDS open data science framework processes ERP data in GPU memory. The framework's libraries facilitate data sharing in the particular Apache Arrow format, allowing data exchange without actual data movement. The data loaded into GPU memory are referenced within the same GPU card. Remote Direct Memory Address (RDMA) is utilized with GPUDirect technology for Quadro and Tesla GPUs to reference data in other GPU unit memory and avoid data movement. Python-based libraries such as cuDF for GPU DataFrame, cuML for GPU-aware ML algorithms, and BlazingSQL for GPU-conscious SQL engine enable enterprise data analysis on the GPU after extraction from the HANA database.

Within the HANA SQL engine, the GPU-accelerated workload of the TensorFlow ML Platform is initiated and accessed through the SAP HANA External Machine Learning (EML) library. The Spark XGBoost application is also utilized for GPU data processing, enabling Gradient Boosting accessible by Spark within this framework.

DATA VISUALIZATION

Exploratory Data Analysis (EDA) is employed to investigate and visualise data sets statistically. While traditional methods for data visualization exist in Python, VR or AR environments offer immersive learning experiences for better data understanding. Various services for data visualization in VR are available.

This work preferred an open specification approach for VR visualization via the browser. The WebVR API enables the use of devices like Oculus Rift, Oculus Go, Daydream, and Google Cardboard. A React JavaScript library-based application and D3 visualization components were created in the testing environment. The VR-Viz component was utilized for high-level aggregation. Data transferred from SAP HANA via Spark SQL or direct JDBC connection was transformed into JavaScript Object Notation (JSON) format through Spark program logic or HANA SQL JSON functions. The result was visualized in the VR environment as Bar Charts, Scatter Plots, various Maps, and Surface Plots.

CONCLUSION

This study utilised the SAP HANA database to investigate methods for accessing data stored in SAP S/4HANA, the top-level aggregation and analytical enterprise chain. We proposed extracting valuable insights from aggregated data flows and analyzing the results on the GPU using a set of open frameworks and libraries deployed on-premise or in a private cloud. Additionally, we proposed a method for visualizing enterprise data within a VR environment, utilizing JSON data conversion for visualization.

REFERENCES

[1] Martins S., Varela M.L.R., Machado J., "Development of a system for supporting industrial management", 2020, Lecture Notes in Mechanical Engineering, pp. 209-215.

[2] Ghobakhloo M., "The future of manufacturing industry: a strategic roadmap toward Industry 4.0", Journal of Manufacturing Technology Management, vol. 29, Issue 6, pp. 910-936, October 2018

[3] Skrzeszewska M., Patalas-Maliszewska J., "Assessing the effectiveness of using the MES in manufacturing enterprises in the context of industry 4.0", 2020, Advances in Intelligent Systems and Computing 1004, pp.49-56

[4] Trif S.-M., Dutu C., Tuleu D.-L., "Linking CRM capabilities to business performance: A comparison within markets and between products", 2019, Management and Marketing 14(3), pp. 292-303

[5] Tang Y., Liu Y., "Information Management System and Supply Chain Management (SCM)", 2020, Advances in Intelligent Systems and Computing 928, pp. 1421-1426

[6] Bag S., Wood L.C., Xu L., Dhamija P., Kayikci Y., "Big data analytics as an operational excellence approach to enhance sustainable supply chain performance", 2020, Resources, Conservation and Recycling 153,104559

[7] Cui Y., Kara S., Chan K.C., "Manufacturing big data ecosystem: A systematic literature review", 2020, Robotics and Computer-Integrated Manufacturing 62,101861

[8] Seera N.K., Taruna S., "Leveraging mapreduce with column-oriented stores: Study of solutions and benefits", 2018, Advances in Intelligent Systems and Computing 654, pp. 39-46

[9] Faerber F., Kemper A., Larson P.-Г., Neumann T., Pavlo A., "Main memory database systems", 2017, Foundations and Trends in Databases 8(1-2), pp. 1-130

[10] Romeo L., Loncarski J., Paolanti M., Mancini A., Frontoni E., "Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0", 2020, Expert Systems with Applications 140,112869

[11] Hong S., Choi W., Jeong W.-K., "GPU in-memory processing using spark for iterative computation", 2017, Proceedings - 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGRID 2017 7973686, pp. 31-41

[12] Peltenburg J., van Straten J., Brobbel M., Hofstee H.P., Al-Ars Z., "Supporting Columnar In-memory Formats on FPGA: The Hardware Design of Fletcher for Apache Arrow", 2019, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11444 LNCS, pp. 32- 47

[13] Fleig C., Augenstein D., Maedche A., "Process mining for business process standardization in ERP implementation projects вЂ“ An SAP S/4 HANA case study from manufacturing",2018, CEUR Workshop Proceedings 2196, pp. 149-155

[14] Shraideh M., Drieschner C., Betzwieser B., Utesch M., Krcmar H., "Using a project-based learning approach for teaching emerging technologies: An example of a practical course for introducing SAP Leonardo and SAP HANA",2018, IEEE Global Engineering Education Conference, pp. 2047-2051

[15] Oktaş K., "Analytics powered by cloud infrastructure", 2020, Advances in Intelligent Systems and Computing 1029, pp. 386-392

[16] May N., Lehner W., Shahul H.P., Chowdhuri S., Goel A., "SAP HANA - From relational OLAP database to big data infrastructure", 2015, EDBT 2015 - 18th International Conference on Extending Database Technology, Proceedings, pp. 581-592

[17] Han B., "Teaching ERP programming using SAP ABAP/4", 2006, Enterprise Systems Education in the 21st Century, pp. 261-282

[18] Zhou K., Zhao L., Design of equipment management information system based on SAP NetWeaver, 2013, Tobacco Science and Technology (8), pp. 15-20

[19] Selmeci A., Orosz T., "SAP remote communications", 2012, SACI 2012 - 7th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings pp. 303-309

[20] Lawrence R., Brandsberg E., Lee R., "Next generation JDBC database drivers for performance, transparent caching, load balancing, and scaleout", 2017, Proceedings of the ACM Symposium on Applied Computing Part F128005, pp. 915-918

[21] Sharma K., Marjit U., Biswas U., "Efficiently processing and storing library linked data using apache spark and parquet", 2018, Information Technology and Libraries 37(3), pp. 29-49

[22] Aluko V., Sakr S., "Big SQL systems: an experimental evaluation", 2019, Cluster Computing 22(4), pp. 1347-1377

[23] Lentner G., "Shared memory high throughput computing with Apache arrow", 2019, ACM International Conference Proceeding Series 3335197,

[24] Peltenburg J., van Straten J., Brobbel M., Hofstee H.P., Al-Ars Z., "Supporting Columnar In-memory Formats on FPGA: The Hardware Design of Fletcher for Apache Arrow", 2019, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11444 LNCS, pp. 32- 47

[25] Dojchinovski D., Gusev M., Zdraveski V., "Efficiently Running SQL Queries on GPU", 2018, 2018 26th Telecommunications Forum, TELFOR 2018 - Proceedings 8611821

[26] Shehab E., Algergawy A., Sarhan A., "Accelerating relational database operations using both CPU and GPU co-processor", 2017, Computers and Electrical Engineering 57, pp. 69-80

[27] Sitaridi E.A., Ross K.A., "GPU-accelerated string matching for database applications", 2016, VLDB Journal 25(5), pp. 719-740

[28] Sisyukov A.N., Yulmetova O.S., Kuznecov V.A., "GPU accelerated industrial data analysis in private cloud environment",2019, Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2019, 8656751, pp. 348-352

[29] Ohno Y., Morishima S., Matsutani H., "Accelerating spark RDD operations with local and remote GPU devices", 2016, Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS 0,7823823, pp. 791-799

[30] Hu T., Song T., "Research on XGboost academic forecasting and analysis modelling", 2019, Journal of Physics: Conference Series 1324(1),012091

[31] Makransky G., Borre-Gude S., Mayer R.E., "Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments", 2019, Journal of Computer Assisted Learning 35(6), pp. 691-707

[32] Beck D. "Special Issue: Augmented and Virtual Reality in Education: Immersive Learning Research", 2019, Journal of Educational Computing Research 57(7), pp. 1619-1625

[33] Aebersold M., Rasmussen J., Mulrenin T., "Virtual Everest: Immersive Virtual Reality Can Improve the Simulation Experience", 2020, Clinical Simulation in Nursing 38, pp. 1-4

[34] Wang X., Guo C., Yuen D.A., Luo G., "GeoVReality: A computational interactive virtual reality visualization framework and workflow for geophysical research", 2020, Physics of the Earth and Planetary Interiors 298,106312

[35] Azraff Bin Rozmi M.D., Thirunavukkarasu G.S., Jamei E., Stojcevski A., Horan B., "Role of immersive visualization tools in renewable energy system development", 2019, Renewable and Sustainable Energy Reviews 115,109363

[36] Tudjarov B., Mitrev R., "Web-based VR environment for simulation and visualization of construction manipulator motion", 2019, ACM International Conference Proceeding Series

[37] Vincke, S., Hernandez, R.D.L., Bassier, M., Vergauwen, M., "Immersive visualisation of construction site point cloud data, meshes and bim models in a vr environment using a gaming engine", 2019, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 42(5/W2), pp. 77-83