<div align="center">**SUBJECT OF THE THESIS**</div>

University: Paul Sabatier University, Toulouse 3

Research Laboratory: IRIT Institut de Recherche en Informatique de Toulouse

Team: PYRAMID (Dynamic Query Optimization in Large-scale Distributed Environments; https://www.irit.fr/PYRAMIDE/)

Supervisor: Abdelkader Hameurlain (hameurlain@irit.fr)

Co-supervisor: Franck Morvan (morvan@irit.fr)

Doctoral School: MITT Mathématiques, Informatique, Télécommunications de Toulouse

PhD Subject: Predictive Query Optimization for Multi-tenant Cloud DBMSs

1. Scientific Context

In parallel and distributed large-scale environments (Cluster, Grid, Cloud), the Pyramid team addresses the main problems of query processing and optimization, targeting large volumes of data distributed in large scale. In cloud environments, users are often called tenants. A cloud DBMS shared by many tenants is called a multi-tenant DBMS. The resource consolidation in such a DBMS allows the tenants to only pay for the resources that they consume, while providing the opportunity for the provider to increase its economic gain. For this, a Service Level Agreement (SLA) is usually established between the provider and a tenant. However, in the current systems, the SLA is often defined by the provider, while the tenant should agree with it before using the service. In addition, only the availability objective is described in the SLA, but not the performance objective. In one of our previous work [8], an SLA negotiation framework was proposed for OLAP applications, in which the provider and the tenant define the performance objective together in a fair way. To demonstrate the feasibility and the advantage of this framework, we evaluated its impact on query optimization. We formally defined the problem by including the cost-efficiency aspect, we designed a cost model and improved two execution plan search methods to adapt to the new context, and we proposed a heuristic to solve the resource contention problem caused by concurrent queries of multiple tenants. We also conducted a performance evaluation to show that, our optimization approach (i.e., driven by the SLA) can be much more cost-effective than the traditional approach which always minimizes the query completion time.

2. PhD Subject: Predictive Query Optimization for Multi-tenant Cloud DBMSs

In the above work, we proposed a new criterion: the Unit Benefit Factor (UBF) which is the profit generated in a unit of time (by the execution of a query). For example, if a query lasts 2 seconds and it allows the provider to have 10 cents of profit, the UBF is then 5 cents / second. For each given query, the optimizer chooses the execution plan that maximizes this criterion. Obviously, this does not guarantee the maximum profit when considering all the queries of all tenants in a long term. Indeed, the workload of a multi-tenant DBMS varies over time and influences both the QoS (tenant side) and the economic cost (provider side). Some work proposes to build models in order to predict the future load [2, 4, 6, 9]. This prediction can help the optimizer to choose execution plans that improve both QoS and profitability in a long term

[1]. Taking into account this prediction (that becomes a new constraint) requires extending the cost model and revisiting the search strategy.

In this perspective, the candidate is expected to design and develop a query optimization method by taking into account the workload prediction. More precisely, she/he will: (i) study the related work [e.g., 2-9], (ii) propose a predictive query optimization method that maximizes the provider's long term profit while meeting the SLAs established with the tenants, and (iii) conduct an experimental study to evaluate and validate the proposed method.

3. *References*

[1] Abadi, D., et al. ; The Seattle Report on Database Research; SIGMOD Record, December 2019, Vol. 48, No. 4.

[2] Picado, J., Lang W., Thayer E.C.; Survivability of Cloud Databases - Factors and Prediction. SIGMOD '18: Proceedings of the 2018 International Conference on Management of Data. May 2018, p. 811-823.

[3] Pietri, I., Chronis, Y., and Ioannidis, Y.; Fairness in Dataflow Scheduling in the Cloud. Information Systems, Elsevier, Vol. 83, 2019, p. 118-125.

[4] Taft, R., El-Sayed, N., Serafini, M. , Lu, Y., Aboulnaga, A.I., Stonebraker, M., Mayerhofer, R., and Andrade, F. ; P-Store: An Elastic Database System with Predictive Provisioning. SIGMOD '18: Proceedings of the 2018 International Conference on Management of Data, May 2018, Pages 205-219.

[5] Tan, Z., and Babu, S. Tempo: robust and self-tuning resource management in multi-tenant parallel databases. Proceedings of the VLDB Endowment 9.10, 2016, p. 720-731.

[6] Viswanathan, L., Chandra, B., Lang, W., Ramachandra, K., Patel, JM., Kalhan, A., DeWitt, D. J., and Halverson, A.; Predictive Provisioning: Efficiently Anticipating Usage in Azure SQL Database. IEEE 33rd International Conference on Data Engineering (ICDE), 2017, p. 1111-1116.

[7] Wong, P., He, Z., Feng, Z., Xu, W., and Lo, E.; Thrifty: Offering Parallel Database as a Service using the Shared-Process Approach. SIGMOD Conference 2015, p. 1063-1068.

[8] Yin, S., Hameurlain, A., and Morvan, F.; SLA Definition for Multi-tenant DBMS and its Impact on Query Optimization. IEEE Transactions on Knowledge and Data Engineering, Vol. 30, N. 11, 2018, p. 2213-2226.

[9] Zhang, W., Zheng, N., Chen, Q., Yang, Y., Song, Z., Ma,T., Leng, J., and Guo, M.; URSA: Precise Capacity Planning and Fair Scheduling based on Low-level Statistics for Public Clouds. ICPP '20: 49th International Conference on Parallel Processing - ICPP. August 2020, p.1- 11.

4.   Requirements & Application

Requirements:

Distributed and Parallel Systems, Data Management Systems, Database Systems, Query Processing and Optimization, Cost Models, Cloud Systems, Programming Languages (e.g. C++, Java, Python).

The Application should include following documents (PDF format, see: http://www.edmitt.ups-tlse.fr/):

1- CV mentioning all your degrees

2- Motivation letter from the applicant explaining his/her choice of the proposed thesis subject

3- Recommendation letters

4- Details of your grades since you started higher education with ranking.

Applications in digital form (pdf) should be sent to: hameurlain@irit.fr

Application Deadline: March 31st, 2021

Start Date: October 1st, 2021.

Remuneration: the PhD Scholarship is approximatively 1768 Euros (Gross).

Once the candidate is selected by the Pyramid team, she/he will be auditioned by a committee.