UNIVERSITY^{OF} BIRMINGHAM

In-Memory Computing –EXASOL Evaluation

1. Purpose

EXASOL (http://www.exasol.com/en/) provides an in-memory computing solution for data analytics. It combines inmemory, columnar storage and massively parallel technologies to support high performance, flexibility and scalability. In this report, we introduced what in-memory database is and the issues that require attention when considering in-memory computing. We focus our evaluation on the Exasol in-memory database solution in terms of its functionality, deployment solutions and license models. The report aims to give the audience a background understanding of the in-memory database (IMC) and a detail analysis of the Exasol solution.

2. In-Memory Computing?

2.1 What is In-memory Computing

In-memory computing (IMC) is an architecture-style where applications assume all the data required for processing is located in the main memory of their computing environment. In-memory computing (IMC) is a broader concept which includes in-memory database and in-memory analytics. An in-memory database (IMDB, also known as a main memory database or MMDB) is a database whose data is stored in main memory to facilitate faster response times. The source data are loaded into system memory in a compressed, non-relational format. In-memory analytics is an approach to querying data when it resides in a computer's random access memory (RAM), as opposed to querying data that is stored on physical disks.

What benefits can in-memory computing deliver? The main benefit is it improves the speed dramatically. For example, data processing consists three components: the processor to perform the calculations, the storage to store and manipulate data, and a system to transfer the data between the two. The current bottleneck is the latency storage, more specifically, it is the latency of hard disks. Processing power is not used to full capacity because the data to be processed is not retrieved fast enough from hard disks. Thus, IMC putting data into memory can dramatically speed up the process of data analysis. The diagram below shows a traditional data warehouse approach vs. IMC-enabled data warehouse approach. The shift from traditional, hard disk-enabled data warehouses to IMC-enabled warehouses implies a reduction in layers on the way from raw data to the results of data analysis.

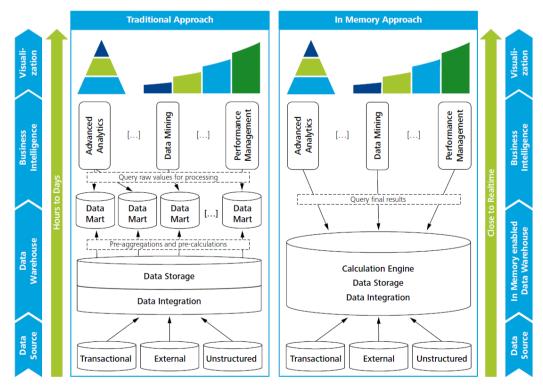


Figure 1. Traditional data warehouse approach vs. IMC-enabled data warehouse approach¹

Ref: In Memory Computing –EXASOL Evaluation 17/05/2016 Li Zhao, IT Innovation Specialist itinnovation@contacts.bham.ac.uk

Besides the main benefit of quicker speed, IMC can also deliver in process innovation, simplification and flexibility.

- Process innovation: The performance gain of quicker processing speed for data makes it possible to consider the real time data analysis, which could lead the innovative applications.
- Simplification: Due to the reductions in layers on the way from raw data to the results of data analysis, the simpler data models and a more unified interface can be designed.
- Flexibility: Integration of additional data sources and modification of the data analysis is more flexible because the reductions in layers.

2.2 In-Memory Computing Adoption

Although in-memory computing promises a massive gain in speed, it cannot be achieved without proper planning and strategic support. IMC cannot guarantee a competitive value out of the box. A great effort will need to be put into maintaining the data quality, identifying the data volumes, the respective analytical and transaction processes in the business.

Depending on the maturity of the business, there are a number of issues which will require attention when considering in-memory computing:

- Varying levels of technology maturity in the business will affect the adoption of IMC, complete renewal or updates of some technology will be necessary.
- Existing analytic solution may need to adapt to the new processes which IMC brings in.
- IMAC will improve the speed massively, as a result, it will lead to process innovation. Some real time process which was not possible will become service on offer.
- Lack of skills for IMC internally may be the reality in many businesses. Training internal staff and a complete IMC support plan will help in such situations.

To overcome these issues, it is recommended to consider the following measures:

- Identify the most easily achieved goals where IMC will benefit advanced analytics initiatives.
- Understand how IMC can link analytics with the transaction process.
- Look for initiatives that could benefit from real-time analytics enabled by IMC.
- Work with line of business specialists in domains to bring them into the business analytics team.
- Embracing IMC, however, plan and prepare how it can coexist with the existing analytic technologies.

3. EXASOL Solution

EXASOL (<u>http://www.exasol.com/en/</u>) is a German company based in Nuremberg. The company's key product is a column-store, in-memory, massively parallel processing (MPP) relational database management system (RDBMS) designed specifically for data warehousing and analytics applications. It can run on standard x86 hardware and uses standard SQL language (with optional extension). In the recent TPC-H benchmark ²for clustered, decision support databases, EXASOL holds the number one position for all volumes of data.

3.1 Key Features

The key features of EXASOL solution include:

- In-memory technology: large amounts of data processed in main memory dramatically faster the processing time
- Column-based storage and compression: reduces the number of I/O operations
- Massively parallel processing: queries are distributed across all nodes in each cluster
- High user concurrency: multiple users can simultaneously access and analyse large amounts of data
- Scalability: extend the system and increase performance by adding additional nodes
- Tuning-free database: self-tuning and optimizing system performance
- Industry-standard interfaces: easy connect to SQL-based BI and other data integration tool via ODBC, JDBC etc
- Advanced Analytics: user defined functions (UDF) allow in-database advanced analytics
- Big Data: MapReduce processing capabilities and Haddop integration service available

The diagram below shows the EXASOL solution plan in a visual way.

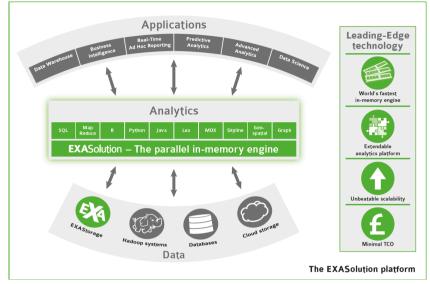


Figure 2. The EXASolution Platform³

3.2 EXASOL Architecture

3.2.1 Logical Architecture

EXASOL has a shared nothing architecture (SN architecture: a distributed computing architecture in which each node is independent and self-sufficient) to ensure high resilience, it runs on low cost commodity x86 server and can scale from a single note to hundreds of notes to support 1000's concurrent users and 100's TB data. Figure 3 shows the logical architecture for EXASOL. EXAOperation is a web based GUI which manages the databases, while the EXACluster OS and EXAStorage are the cluster and storage management software which are in charge of deployment and configuration. EXASOL supports ANSI 2008 SQL and the commonly sued Oracle SQL dialect. As figure 3 shows, the resilience and redundancy is provided by deploying additional "hot standby" servers in the cluster.

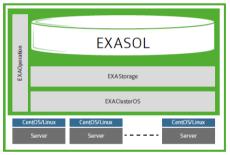


Figure 3. The EXASOL Logical Architecture⁴

3.2.2 Data Source Interfaces

EXASOL can be connected to various data sources. Currently, EXASOL supports three standardized interfaces:

- ODBC (Open Database Connectivity)
- JDBC (Java Database Connectivity)
- ADO.NET (Abstract Data Objects.Net)

In addition, the OLE DB Provider for ODBC Driver from Microsoft allows EXASolution to be accessed via an OLE DB interface.

The above protocols are mainly used for single-node connections, while in the multi-node connections, EXAloader – a bulk loader for ETL (Extract, Transform and Load) scenarios is recommended.

In order to be able to integrate EXASolution into in-house applications, EXASOL additionally provides a client SDK (Software Development Kit) for C/C++.

3.2.3 Query Process

In order to achieve maximum performance, EXASOL aims to utilize all the available hardware resources when answering queries. When EXASOL receives a SQL statement, the query is made available on all the machines in the cluster. On each machine, the SQL compiler generates an execution graph. The machines start executing and allocating resources fully in parallel. To guarantee the maximum independence, all the tables in EXASOL are

distributed horizontally across all machines in the cluster – this means that each machine holds a segment of the table's rows and typically only operates on its own part of most operations.

Many database systems use vastly different and incompatible dialects of standard SQL, so EXASOL pre-processes these commands in a two-level strategy:

- Compatibility features that are consistent with existing features in the database and that do not cause performance penalties are regularly added to EXASOL's version of SQL.
- A query pre-processor in the compiler enables ad-hoc adoption of the input language.⁵

In this way, EXASOL maintains the best compatible SQL engine.

3.2.4 Analytic Functions

EXASOL can be extended with EXAPowerlytics which includes comprehensive analytical functions. User can write UDFs using Lua, Python, Java or R to support in-database analytics. It also supports the integration of Hadoop clusters, the map-reduce algorithms and the processing of unstructured data.

3.3 Deployment Options

There are three options to deploy EXASOL:

- 1) Software-only solution: EXASOL runs on any standard x86 architecture. User can download the database and install it on the existing infrastructure and get started with the in-memory analytics.
- 2) Appliance: EXASOL can also be offered as an appliance solution. As a package, it includes the EXASOL solution, standard hardware from EXASOL's hardware and service partner plus service and support.
- Cloud: EXASOL is also available in the cloud. It offers cloud-based data analytics with the flexibility of payas-you-go. Currently EXASOL cloud solutions are available on EXACloud, Amazon Web Services, Microsoft Azure and Bigstep.

In most cases, the software-only solution is suitable for trial and evaluation purpose because there is no hardware or much cost involved. For commercial adoptions, companies will need to consider their requirements and choose from appliance or cloud. Appliance implementation has the data in house and does not require big resources on the network bandwidth, however, it means a bigger investment at the project beginning and a long time lock-in for the services. While the cloud implementation is flexible and pay as you go, it potentially raises data boundaries, security issues and can cost more in the long run.

3.4 License Models

There are three different license models to choose depending on the requirements of the customers.

- 1) Perpetual License:
 - One off license fee based on the total RAM in the cluster: there is no restriction on data volume, number of nodes or the number of users.
 - Annual Support & Maintenance: Standard Support 08:00 –17:00 UK time or Enterprise Support (24 x 7)
- 2) Annual Subscription License:
 - Annual fee based on the total RAM in cluster includes standard support
- 3) OEM License EXASOL embedded in Partner solution:
 - ▶ License restricted to use with the named application
 - Perpetual or Subscription license options

4. Demo of EXASOL Solution

EXASOL offers two ways of trials: EXASOL cloud demo edition where users can easily access the sample database and test its functionalities, and the EXASOL community edition – a one-node image of the EXASOL in memory database which could be installed on virtual machines, In our experiment, both of the trials were tested. EXAplus - a standard user interface for querying EXASOL in-memory database was used in the experiment.

4.1 EXASOL Cloud Demo Editon

EXASOL cloud demo is hosted in the cloud. It provides users an environment to test data and scripts without the need to do hardware set-up and configuration.

4.1.1 Testing Environment Setup

To set up the EXASOL cloud demo system involves two steps:

1) Sign up to the EXASOL could system, EXASOL will email the demo account detail to the user.

2) Install the EXAplus (EXASOL SQL editor) on the windows PC. It is also possible to install the EXASOL drivers and connecting user's own BI front-end to EXASolution.

4.1.2 Sample Data Testing

EXASOL cloud edition provides sample database "TPC", which contains the test data similar to the TPC-H online transaction processing benchmark from the Transaction Processing Performance Council (TPC). In the experiment we will use the EXAplus client and the "TPC" sample data.

First, use EXAplus to connect to the EXASOL cloud edition with the demo account, then users will be able to see the EXAplus main screen. It has the database browser to navigate the database, the history browser for the past commands, the favourites browser to organise the favourite folders and the templates browser for various command references.

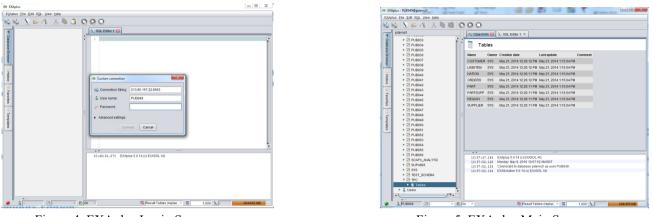
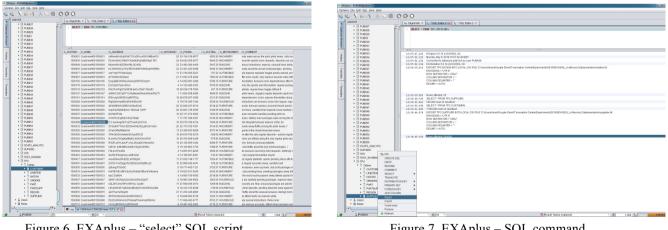
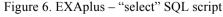


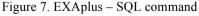
Figure 4. EXAplus Login Screen

Figure 5. EXAplus Main Screen

EXAplus supports the Oracle standard SQL language. Users can either write the SQL script or run the SQL command by right clicking the database. First, I run the SQL script to select all the customers and list them. There are 7,500,000 rows of all the customers and it takes 0.7 seconds to calculate the results. Second, I select the "Export" command to export the all suppliers information to a text file. The suppliers file size 73.7MB, and it only takes seconds to export the file. Consider the database is hosted in the cloud, the processing speed is very quick.







4.1.3 Findings

1) EXASOL cloud edition provides users an environment to test data and scripts without the need to do hardware setup and configuration.

2) Through the sample database "TPC" testing, the processing speed is fast based on the real processing time and the size of the data.

3) EXAplus is an easy-to-use SQL client which supports comprehensive SQL commands.

4.2 EXASOL Community Edition:

In our experiment, the EXASOL community edition is installed on the Virtual PC and user can connect to the virtual server via the EXAplus. One thing extra from the cloud edition is that users with the administrator right can access the EXAoperation. EXAOperation is a critical part of the EXASOL solution, so we will have a look how it operates.

4.2.1 Testing Environment Setup

Setting up the testing environment involves the following steps:

- 1) Register with EXASOL to download the free trial single-node copy of the EXASOL in-memory database, EXASOL will send an email containing the user login details.
- 2) Set up the Oracle Virtual box version 5.0.18.
- 3) Install the single-node image of EXASOL on the virtual box.
- 4) Download and install EXAplus, use the login detail to connect to the EXASOL in-memory database.

4.2.2 EXAOperation Configuration

EXAOperation can be accessed through a web interface. After the login page, the first page will show the status of the databases. Admin can start or shutdown the instances as necessary. It is also easy to add or delete the database here. It is possible to add database on the same node or expand to the other nodes.

Image: Source Sourc	s.html
etting standel - Sudian V 1 Soldan Genez-Solian V 1 Norgenetion dutat 1 X 0 Google X etting standel - Sudian V 2 Soldan Genez-Solian V 1 Soldan Genez-Solian V 2 Solian Genez	
Subject as used. States Hangement States Hange	7 7
Excs EXASolution Instances Background RestoreBackup RestoreComment Schuldban Image: State	
Stat	•
operation Office DB Name Connection String DB RAM Space Usage Status Background RestoreBackup/RestoreComment Private Connection String DB RAM Space Usage Status Background RestoreBackup/RestoreComment "Network Interfaces "Network Interfaces "Network Interfaces "Network Interfaces "Network Interfaces"	
	v v v
uration LDAP Server URLs	
Vare * Connection Port #563	
* DB RAM (cite) #	
Nork Extra DB Parameters	
If and Control	

Figure 8. EXASolution – Status

Figure 9. EXASolution - Add Database

Ki Getting Started - Solution × Ki Solution Center - Solution × X EXAppendian cluster) × G Google

On the left navigation menu, there are two main categories – Services and Configuration. The services section provides the following functions:

- EXASolution: show the status of the instances, where admin can also add or delete database, as shown in figure 4 and 5.
- EXAStorage: manage the storage of database, storage can be added horizontally to the nodes or vertically to the single node, as shown in figure 6.
- EXAoperation: manage and prioritize the nodes, as figure 7.
- Monitoring: monitor the services and access to all the logs.
- Support: get the debug information which could be used to communicate with the EXASOL support.

• • • • • • • • • • • • • • •	Home - Welcome - EXASC X	X EXAppration: duster1 ×	Home - Welcome - EXASC × X EXAoperation: cluster1 ×	
EXES EXES	← → C fi 🔒 batps://	192.168.56.102/cluster1/storage/index.html		
Add Universe Units Service Universe Nodes Nodes Network Access Management	Content dat experts Services EXASolution EXASolution BXASorage EXASorage Support Support Software Nodes Network	EXAStorage Index Size Redum: Type Disk Labels Prio Recovery Size 0 00000 Table Gills Data 403.storage EXASolo_pervisitent 10 n/a OHL/HE 0 Add Volume Image: Add Remote Volume Image: Add Remote Volume Image: Add Remote Volume Image: Add Volume Image: Add Remote Volume Image: Sum State Disk Image: Add Volume Image: Sum State Disk State Disk Image: Add Volume Image: Sum State Disk State Disk Disk Image: Add Volume Image: Sum State Disk Disk	Services EXASolution EXASolution EXASorage EXAstrage Configuration Software Nodes Network	

Figure 10. EXAStorage

Figure 11. EXAOperation

The configuration section includes the following functions:

- Software: it controls the version, license, drivers and UDF libraries, see figure 8. Software function is a very important part. On one hand, it connects to the data sources through JDBC drivers (see figure 9). Admin can install and configure many data sources as required. On the other hand, it manages the configuration of UDF libraries (see figure 10). UDF libraries were named to EXAPowerlytics in the real versions. As long as the UDF repositories is loaded in the configuration, users can write UDFs using Python, Java or R to support indatabase analytics.
- Nodes: it manages the cluster and disk configuration, see figure 11.
- Network: network configuration for the clusters, may add additional private or public network as necessary.
- Access Management: manage the user rights, assign roles and reset the passwords.

Setting Started - Solution 🗴 🔀 Solution Center - S	intuition X X EXAmenation cluster	/ 🕅 Home - Welco		
C fi & https://192.168.56.102/clu		. < → C ń	(k berps://192.168.56.102/clus	ter1/software.html
		EXAS	0	
ASOL			of data experts.	
Escellent data experts.		duster1		
		Services	Versions	License UDF Libraries JDBC Drivers Legal
vices Versions	License UDF Libraries JDBC Drivers Legal	EXASolutio EXAStorag		
CASolution CAStorage EXACluster CAoperation	05 5.0.13 (+ 05 patchlevel 9)	EXAoperati Monitoring	n MYSQL JTDSSYL POSTGR	com.mysql.jdbc.Driver jdbc:mysql: preinstalled Version 5.1.18 BASE net.sourceforge.jtds.jdbc.Driver jdbc:jtds:sybase: preinstalled ESOL or opostarsa0.Driver idbc:osotoresol: oreinstalled Version 9.1-901
onitoring EXASoluti	ion 5.0.13-S	Support	JTDSMS EXASolu	ISOL net.sourceforge.itds.idbc.Driver idbc:itds:salserver: preinstalled
	ate	Configuratio	Choose File No fi	
oftware Fi odes	te Choose File No file chosen	Nodes Network		he
etwork ccess Management		Access Mar	agement 🌍 Add 🤤	Delete 🐉 Edit 👷 Grant
🔅 Define Upda	ate URL 🔝 List Available Updates			
Software Databa:	Type Software Name EXASolution 5.0.13-S Delete			
1	Figura 12 Softwara	Fig	roll Sol	fuvora IDDC Drivera
]	Figure 12. Software	Figu	are 13. Sot	ftware - JDBC Drivers
]	Figure 12. Software	Figu	/ 🔣 Home - Welcome - EXAS: 🗙	X EXApperation: cluster1 ×
]	Figure 12. Software	Figu	/ 🔣 Home - Welcome - EXAS: 🗙	
]	Figure 12. Software	Figu	Ki Home - Welcome - EXAS: × ← ⇒ C' ff	X EXApperation: cluster1 ×
∕ X Home - Welcome - EXASC >		Figu	/ 🔣 Home - Welcome - EXAS: 🗙	X EXApperation: cluster1 ×
Home - Welcome - EXASC	X EXApperation: cluster1 ×	Figu	Ki Home - Welcome - EXAS: × ← ⇒ C' ff	X EXApperation: cluster1 ×
Home - Welcome - EXAS		Figu	Ki Home - Welcome - EXAS: × ← ⇒ C' ff	X EXApperation: cluster1 ×
✓ Home - Welcome - EXASI > ← → C ↑ ▲ ▲ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	X EXApperation: cluster1 ×	Figu	Image: Website - BKAC N Image: Websit	X Dolgenson Gutteri x
Home - Welcome - EXASC	X EXApperation: cluster1 ×	Figu	M Hone - Weckere - BAS ← → C fit B Length // EXECUTE Executed dat exerts Barry Loss Barry Loss EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction	X Despenden dutte:
★ Home - Weicome - EXAS: > ★ → C ↑ ▲ ▲ Latps.	X EXApperation: cluster1 ×	Figu	M Hone-Wednere DAT × ★ → C A R Butter. Control Response of the regrets. Control Response of the regre	X Despendin dutre: Despendin dutre: Despendin dutre: Despendin dutre: Despendin dutre: Despendin Disk Configuration Ndcs Ndc Male Info State(Op. MAL Addresses
✓ Home - Welcome - EXASI > ← → C' ↑ ଢ bttp5:	X EXApperation: cluster1 ×	Figu	M Hone - Weckere - BAS ← → C fit B Length // EXECUTE Executed dat exerts Barry Loss Barry Loss EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction EXAStruction	X Desperation Guted: * 192168-55.102/chatteri/modes.html Claster Nodes Default Disk Configuration Nodes
✓ Home - Welcome - EXASI > ← → C' ↑ ଢ bttp5:	X EXAoperation: cluster1 x //192.168.56.102/cluster1/+/AddScriptExtension.html	Figu	Monet - Woodene - DAL Image: A standard - DAL <	X Despenden dated: X Des
★ Home - Welcome - EXAS >> ★ → C fi	X EXApperation: cluster1 ×	Figu	Minore-Woodener UM Image: An analysis Image: Analysis	X Despendion dute: x 192.168.56.102/cluster/.modes.html Clinator Nodes Default Disk Configuration Nodes Nodes Extension Addresses Addresses Addresses
/ M Home - Welcome - EXAST > ← → C ↑ ⊗ LattpS:	X EXAoperation: cluster1 x //192.168.56.102/cluster1/+/AddScriptExtension.html	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	X Desperation duted: X Desperation duted: X Desperation duted: X Desperation Clinator Nodes Clinator Nodes Default Disk Carliguration Nodes
Image: Home - Welcome - EXAS Image: Home - Home - Exas Image: Home - Home - Exas Image: Home - H	Add UDF Library	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	
★ Home - Welcome - EXAS > ★ → C ★ ● ▲ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ C ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ ★ K ★ K K	X EXAoperation: cluster1 x //192.168.56.102/cluster1/+/AddScriptExtension.html	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	Dedgement on date: X Dedgement on date: X Default Disk Configuration Nodes Default Disk Configuration Nodes Default Disk Configuration Nodes Default Disk Configuration Nodes Default Disk Configuration Default Disk Config
X Home - Welcome - EXAS ★ → C A EXASOL Excelent data experts	Add UDF Library	Figu	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	License Rodes State Rodes License Rodes State Rodes License Rodes State Rodes License Rodes
★ Home - Welcome - EXAS 1) ★ → C ↑ @ bttpS ★ → C ↑	X EXAoperation: cluster1 x //192.168.56.102/cluster1/+/AddScriptExtension.html	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	License Nodes Indee Infe Exactly Consult Relation Addresses
K Home - Welcome - EXAS K → C A Image: Constraint of the second secon	K EXAoperation: cluster1 x X X	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	License Rodes License
★ → C ↑ € ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	Add UDF Library	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	
K Home - Welcome - EXAS K → C A Image: Constraint of the second secon	Add UDF Library * Name * Repository URL * Language Python * Description HTTP proxy Java	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	Character Nodes Default Disk Configuration Default Disk Confi
★ Home - Welcome - EXAS >> ★ → C ↑	Add UDF Library * Name * Repository URL * Language Python * Description HTTP proxy Java HTTPS proxy	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	Character Nodes Default Disk Configuration Default Disk Confi
★ Home - Welcome - EXAS ★ → C ↑	Add UDF Library * Name * Repository URL * Language Python * Description HTTP proxy Java	Fig	Minore-Wooder-BAL N Image: An analysis Image: Analysis Image: Analysis Image: Analysis	Character Nodes Default Disk Configuration Default Disk Confi

Figure 14. Software - add UDF Library

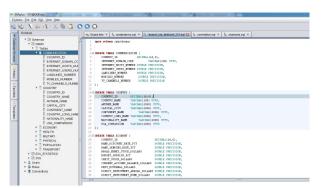
Figure 15. Nodes

Although the Community Edition is only one-node image of the EXASOL in-memory database, through the use of EXAOperation, we can see that the management of EXASOL is very straight forward and convenient. It has good horizontal and vertical expansion in terms of nodes and storage, and it supports a wide range of data source connectors and the data analytics methods.

4.2.3 Sample Data Testing

In the testing of EXASOL community version, the sample database - CIA World Factbook is used. The sample data was loaded from the local machine instead of the virtual EXASOL image. EXASOL virtual image doesn't provide any user interface for loading data directly, because loading from the local PC will achieve the best speed. EXAplus was installed on the local windows PC.

Before loading the data, a schema was created and the structure of the databases was defined. There are two options to load the data, either loading the data by right click the database name and select "import" to load the data, or through the SQL command "IMPORT". Figure 12 showed using the SQL script for defining the database structure and loading the data from the local disk. Because the time constraint of the experiment and the limited knowledge of SQL, only some simple data analytics were tried on the data. Figure 13 showed the analytic result of how many TV channels in each country and how many people per channel.



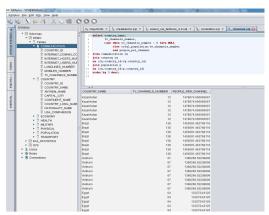


Figure 16. Create Sample Database

Figure 17. Data Analytics Sample Result

4.2.4 Findings

The findings of the experiment for the EXASOL community edition are as follows:

- 1) The management of EXASOL is through EXAOperation a web interface, which is easily accessible via web browser and straightforward to use.
- 2) EXASOL solution has good horizontal and vertical expansion. In terms of the upgrade options, users can either choose to extend more nodes to increase the storage and processing power, or just upgrade the storage and processing power on the existing nodes.
- 3) Through EXAOperation, user can easily configure a wide range of data connectors including ODBC, JDBC, and PostgreSQL etc.
- 4) EXASOL can be extended with EXAPowerlytic to include comprehensive analytical functions. To configure EXAPowerlytic, user can add UDF libraries in the EXAOperation. Currently, EXAPowerlytic support Java, Python, R and Lua.
- 5) EXAplus a SQL client offered by EXASOL fully supports the standard SQL, is easy to set up and use.
- 6) No performance issues detected, possibly due to the small size of the sample data, it would need further testing with bigger data set.
- 7) Single-node virtual PC was used in the testing, so massively parallel processing (MMP) was not tested in the scenario.

5. Other Alternatives

In-memory computing is a growing subset of the database management system, there are more and more providers emerging in this field. Besides from the specialist provider, for example EXASOL, traditional database management system providers, including SAP and Oracle, also provide in-memory computing solutions. Here we will have a short overview of these two solutions. When choosing in-memory computing solution, as discussed in section 2.2, it will be worthwhile to look at the market and consider the business maturity for IMC.

5.1 SAP HANA

SAP HANA is an in-memory, column-oriented, relational database management system developed by SAP. SAP HANA connects to existing applications using open standards and offers a choice of ways to build web-based applications and applications based on micro services. As illustrated in the figure 18, SAP HANA converges database, advanced data processing, data integration, and application platform services in-memory to process transactions; analytics; text analysis; and spatial, streaming, and machine data processing in one system.



Figure 18. Database and Application Platform Convergence with SAP HANA⁶

SAP HANA can be deployed in the cloud or on premise. For on-premise installations, SAP HANA supports performance-optimized deployment of hardware appliances for SAP partners, with an option to build custom hardware based on a tailored data centre model. SAP HANA also supports virtualization software and hardware logical partitioning. In the cloud, SAP HANA is available as a comprehensive infrastructure combined with managed services, consumable through the SAP HANA Cloud Platform and the SAP HANA Enterprise Cloud service or other third-party cloud services, such as Amazon Web Services, IBM SoftLayer, and Microsoft Azure.

5.2 Oracle Database In-Memory

Oracle Database In-Memory accelerates analytic queries by orders of magnitude, enabling real-time business decisions. Oracle Database In-Memory accelerates both data warehouses and mixed workload OLTP databases, and is deployed under any existing application that is compatible with Oracle Database 12c. No application changes are required. Database In-Memory uses Oracle's mature scale-up, scale-out, and storage-tiering technologies to run any size workload⁷. In principle, Oracle Database In-Memory follows the key application design principles for maximizing the performance:

- Process data in the database, not in the application: push the computation down into the database since processing the data hosted in the database in-memory is much more efficient.
- Process data in sets, not row by row: the cost of database entry and exit and amortized by the number of rows processed, try to give the database enough data to process on each invocation.
- Use representative optimizer statistics: Plan differences can make a huge difference to the performance of a query, especially when in-memory access paths can provide orders of magnitude faster performance.
- Use parallel SQL when possible: Parallelism is essential to maximize performance, using all available CPU cores for in-memory processing.

It is worth to note that Oracle Database in-memory is not included in the Oracle database licenses, additional licenses will be required in order to use Oracle Database in-memory.

6. Conclusion

In this paper, we started by reviewing the definition of in-memory computing and the potential issues before adopting IMC. In-memory computing (IMC) is an architecture-style where applications assume all the data required for processing are located in the main memory of their computing environment. The key benefit of IMC is it can help improve the speed dramatically. IMC can also deliver in process innovation, simplification and flexibility. However, IMC cannot guarantee a competitive value out of the box. A great effort will need to be put into maintaining the data quality, identifying the data volumes, the respective analytical and transaction processes in the business.

Then we described a high level overview of the EXASOL in-memory database solution.

- EXASOL is a column-store, in-memory, massively parallel processing (MPP) relational database management system (RDBMS) designed specifically for data warehousing and analytics applications. EXASOL holds the number one position for all volumes of data in the recent TPC-H benchmark for clustered, decision support databases.
- EXASOL has a shared nothing architecture to ensure high resilience. EXASOL includes three major components. EXAOperation is a web based GUI which manages the databases, while the EXACluster OS and EXAStorage are the cluster and storage management software which are in charge of deployment and configuration.
- EXASOL can be connected to various data sources, including ODBC, JDBC, ADO.NET, etc. While in the multi-node connections, EXAloader a bulk loader for ETL can be used.
- EXASOL can be extended with EXAPowerlytics. User can write UDFs using Lua, Python, Java or R to support in-database analytics. It also supports the integration of Hadoop clusters, the map-reduce algorithms and the processing of unstructured data.
- EXASOL provides three options for deployment: software-only solution, appliance and cloud.
- Depending on the requirements of the customers, there are three different license models to choose: perpetual license, annual subscription license and OEM license.

In the experiment, we tested the EXASOL cloud demo edition and the EXASOL community edition with EXAplus - a standard user interface for querying EXASOL in-memory database.

• In the cloud demo edition testing, we used the sample database "TPC" which contains the test data similar to the TPC-H online transaction processing benchmark. We found the processing speed is fast based on the real processing time and the size of the data, and the EXAplus is an easy-to-use SQL client which supports comprehensive SQL commands.

- In the community edition testing, we tested the EXAOperation web interface and used the sample database (CIA World Factbook), we found out:
 - EXAOperation as a web interface to manage the EXASOL, is easily accessible and straightforward to use. It supports horizontal and vertical expansion users can either choose to extend more nodes to increase the storage and processing power, or just upgrade the storage and processing power on the existing nodes. It can be configured for a wide range of data connectors. It can also be extended with EXAPowerlytic to include comprehensive analytical functions
 - Loading the sample database and processing the data is easy and no performance issues identified. However, possibly due to the small size of the sample data, it would need further testing with bigger data set.
 - Single-node virtual PC was used in the testing, so massively parallel processing (MMP) was not tested in the scenario.

Based on our evaluation, we found out that in-memory computing is a growing subset of the database management system, there are more and more providers emerging in this field. Besides from the specialist provider, for example EXASOL, traditional database management system providers are also keen in providing in-memory computing solutions. So in the last part of the paper, we gave a very brief introduction to the SAP HANA and Oracle in-memory database solution. The obvious advantage is that these traditional database management system providers offer in-memory computing on top of their ERP solutions, which could be easily adopted if customers are already using their ERP solutions.

7. Recommendations

We recommend that in-memory computing is the future of computing, but it requires a clear strategy for all steps from evaluation to implementation. This will include identifying the opportunities of utilization of IMC, business case development, management of implementation as well as learning and change management. Depending on the business model and varying levels of technology maturity, it is recommended to:

- Identify the most easily achieved goals where IMC will benefit advanced analytics initiatives.
- Understand how IMC can link analytics with the transaction process.
- Look for initiatives that could benefit from real-time analytics enabled by IMC.
- Work with line of business specialists in domains to bring them into the business analytics team.
- Embracing IMC, however, plan and prepare how it can coexist with the existing analytic technologies.

In-memory computing could be used in the university both within enterprise online transaction processing (OLTP) systems and with data warehouses for real time analytics.

- Enterprise OLTP system: enterprise OLTP systems typically include a mixture of both OLTP transactions and periodic analytic reporting. In-memory computing could potentially replace the analytic indexes in these systems and significant speedups the operation. The university is currently comparing various options for the New Core project, in-memory computing should be one of the key elements to consider.
- Data warehouse system: in-memory computing can significantly improve the performance of analytics and reporting on data that can be accommodated within the in-memory column store. At the university, for instance, it would be possible to have a real time dashboard of Clearing data, entries on the open day, the latest student recruitment data and so on.

Based on the testing of EXASOL solution, we recommend that

- EXASOL was easy to set up and configure. The web interface for configuring EXASOL is easily accessible and the functions are comprehensive. The processing speed of the cloud demo data is quick and satisfying.
- EXASOL provides different data connections to applications and support in-database analytics using Lua, Python, Java and R etc. It also supports the integration of Hadoop clusters, the map-reduce algorithms and the processing of unstructured data. It is flexible and scalable in terms of implementation.

• Due to the size of the test data on the community edition and the test environment was on a virtual PC, the processing performance will need further testing with real data and equipment.

8. References

² TPC-H benchmark, "TPC-H is an ad-hoc, decision support benchmark.", url at <u>http://www.tpc.org/tpch/</u> [Accessed on 21/4/2016]

³ EXASOL Introduction & Overview, <u>http://www.exasol.com/en/</u> [Accessed on 22/04/2016]

⁶ SAP Datasheet, <u>http://sapassets.edgesuite.net/sapcom/docs/2016/04/ac1e84d4-697c-0010-82c7-eda71af511fa.pdf</u> [accessed on 9/5/2016]

⁷ When to Use Oracle Database In-Memory, <u>http://www.oracle.com/technetwork/database/in-memory/overview/twp-dbim-usage-2441076.html</u> [accessed on 12/5/2016]

⁸ Magic Quadrant for Operational Database Management Systems, <u>https://www.gartner.com/doc/reprints?id=1-2PMFPEN&ct=151013</u> [accessed on 4/5/2016]

¹ Deloitte, "In-Memory Computing Technology – the holy grail of analytics?" url at <u>http://www2.deloitte.com/content/dam/Deloitte/de/Documents/technology-media-</u> <u>telecommunications/TMT_Studie_In_Memory_Computing.pdf</u> [Accessed on 20/04/2016]

⁴ EXASOL Business White Paper, <u>http://www.exasol.com/en/</u> [Accessed on 29/04/2016]

⁵ EXASOL Technical White Paper, <u>http://www.exasol.com/en/</u> [Accessed on 27/04/2016]