

How an analytical database stopped me smoking:

A practical story with Exasol



Valerie Wiedemann
Sales Engineer



Christian Langmayr
Head of Community

Introduction

Christian Langmayr

- Works at Exasol for 6+ Year
- Learned about Exasol during his 10+ years at MicroStrategy
- Has many Bikes (10+) but all of them have 2 Wheels

Contact:

- Christian.Langmayr@exasol.com
- www.linkedin.com/in/christian-langmayr-3620982



Introduction

Valerie Wiedemann

- Sales Engineer at Exasol (3+ years)
- Fell in love with Exasol's technology at adidas

Contact:

- Valerie.wiedemann@exasol.com
- www.linkedin.com/in/valerie-wiedemann

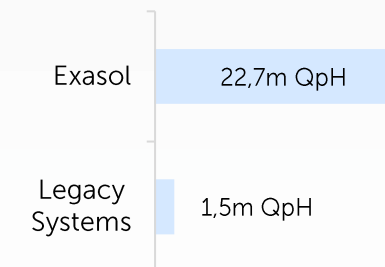


Exasol is the performance leader



Speed

Exasol delivers 100x faster query performance

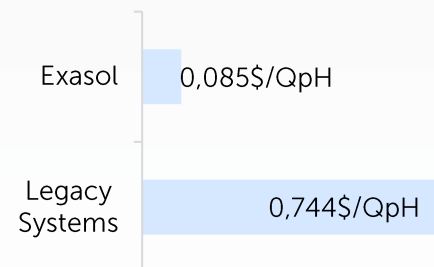


Results at 30TB scale factor



Ownership cost

Significantly lower than other data warehouse solutions

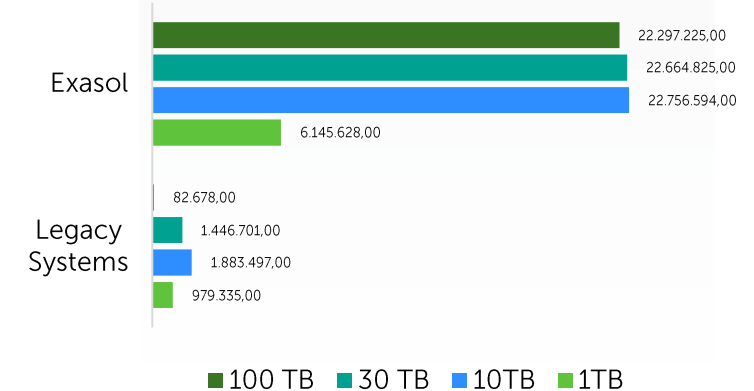


Results at 30TB scale factor



Data growth

Performance advantage grows with larger data volumes



Exasol - usecases



BI ACCELERATION

A quick addition to any warehouse that brings immediate performance benefits

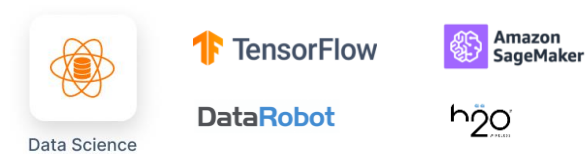
- No replacement of data warehouse
- No changes to integrations, queries, applications



DATA WAREHOUSE REPLACEMENT

A more involved project that brings more lasting benefits and staying power

- Accelerate BI reporting, real-time, and ad-hoc query requests as well as AI and ML projects
- Simplify your landscape and save money



ADVANCED SOLUTIONS

Once in the door, there's room for more valuable opportunities

- Data Consolidation
- Data science operationalization
- Hybrid cloud

Adidas

- Added as Performance Layer to survive a POC
- Fighting SAP Hana
- Basis for eCommerce and Reporting



Revolut

- Brought in by former Badoo Employee
- Solving the performance problem
- Kick it off with Community Edition
- Moving to Google Cloud
- From 1 TB to 500 TB



ARZ – Austrian Banking Center

- Added to the POC to fill an empty spot
- Kicking out the big ones
- Overtaking themselves
- Superior Support

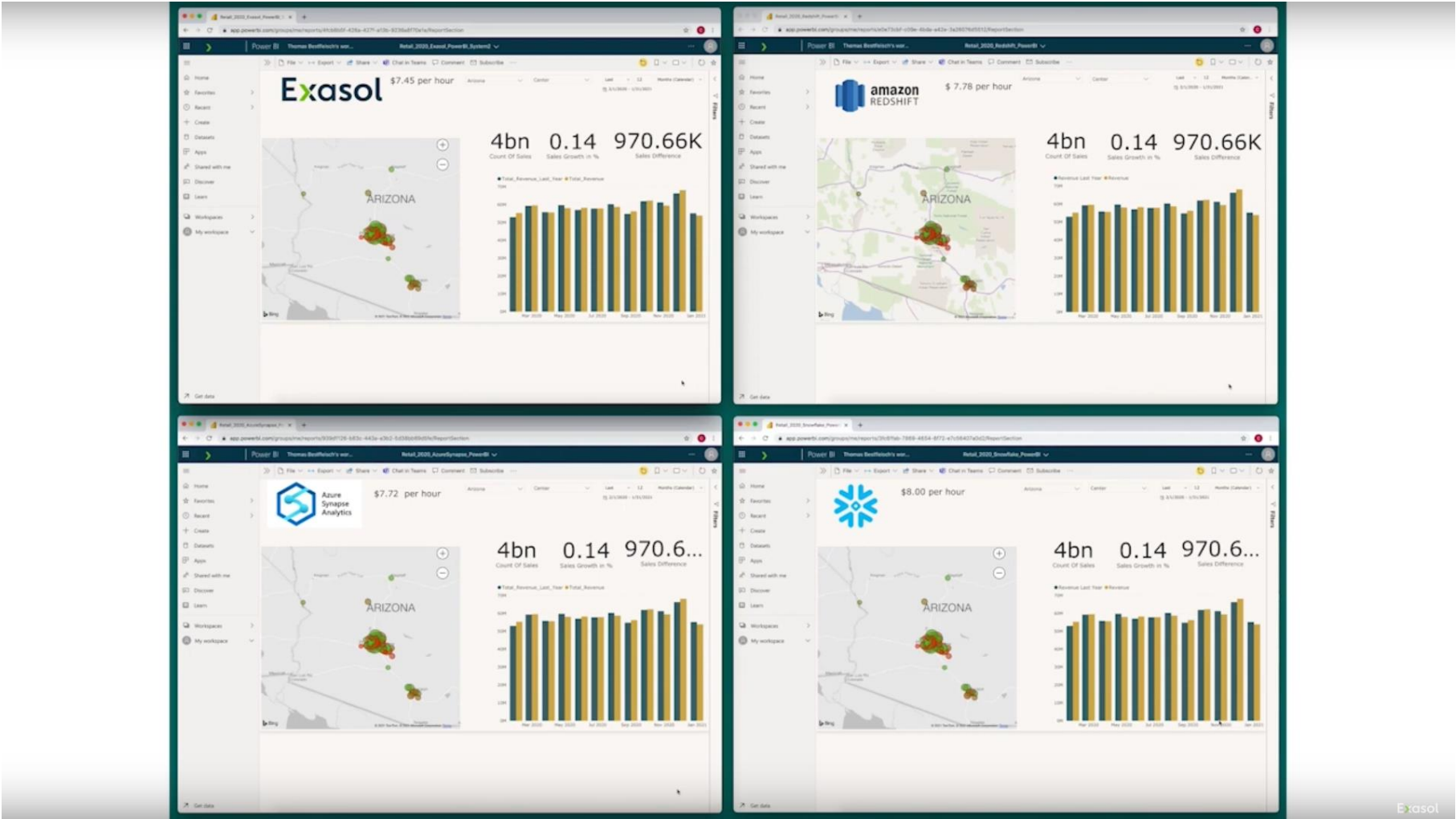


Hahn Air

- Beating MSSQL from 8 Mins to 6 Seconds on a smaller hardware
- IT kept fighting for 6 Month – „We can optimize Microsoft“

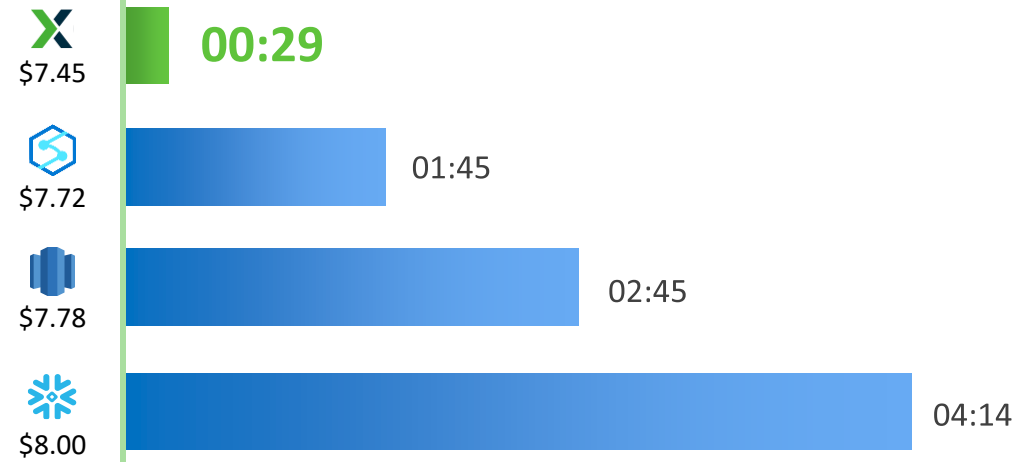
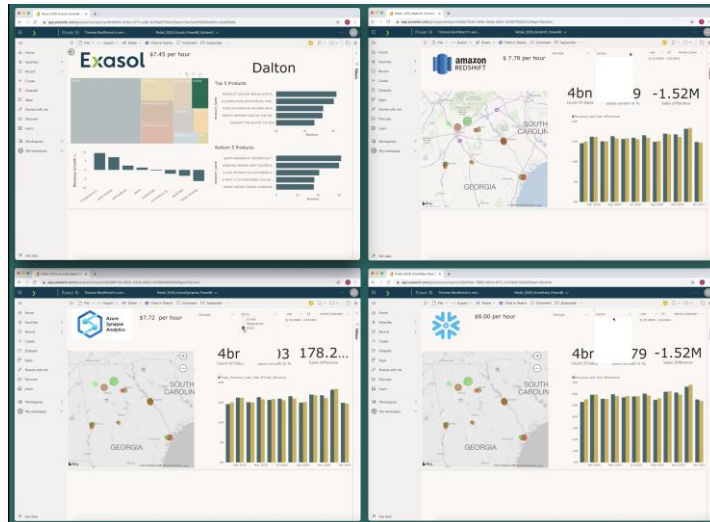


Exasol vs competition



Exasol Powers the World's Fastest Analytics

- ...or in head-to-head competition



We go head-to-head with key competitors in our [Built for Speed performance demo](#).

Hourly costs shown, time in mm:ss

- Benefits:
 - Speed
 - Low Tco
 - Self tuning
 - Stops customers from smoking
-
- → HOW?







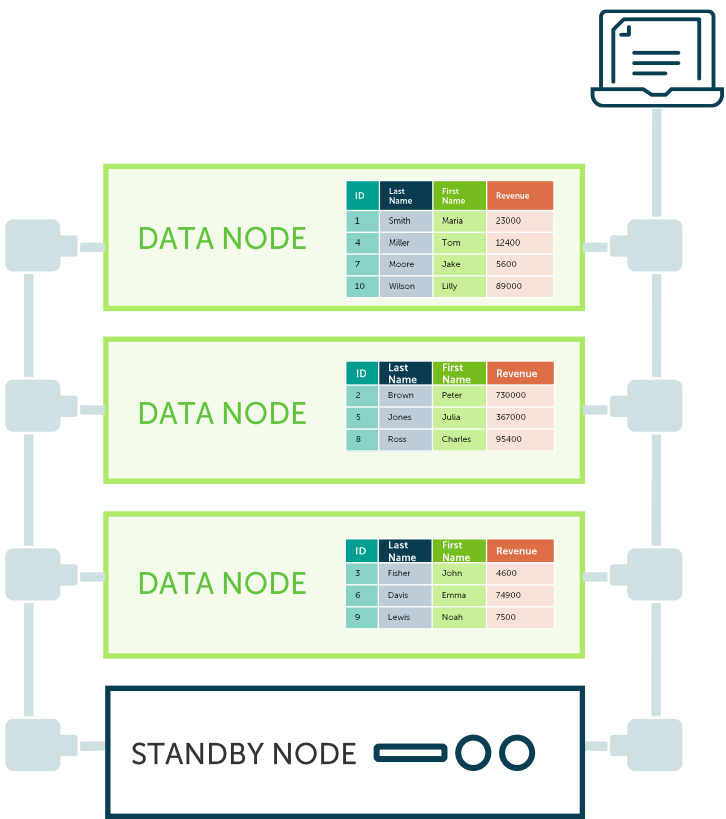




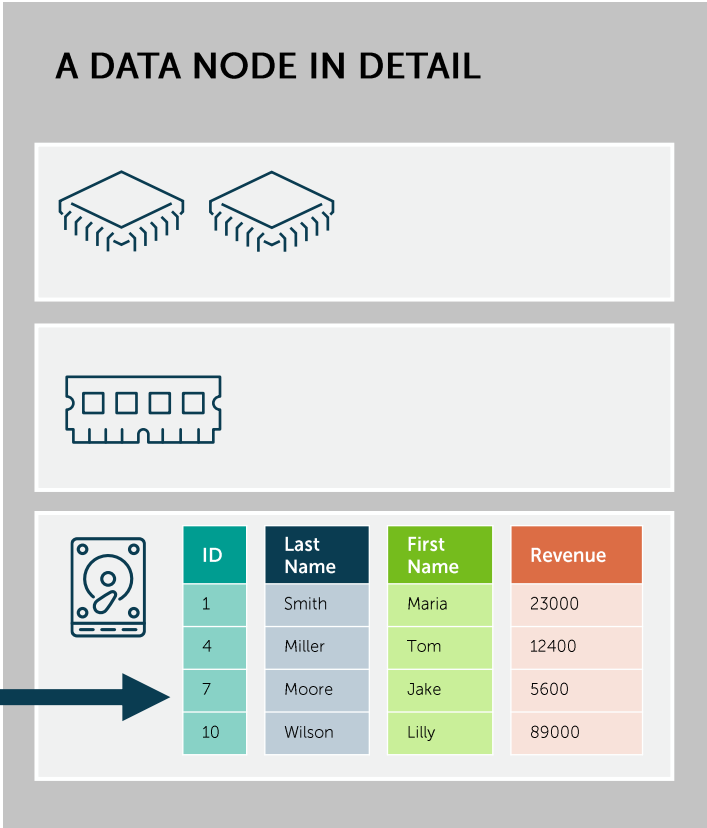
A full-page underwater photograph. In the upper half, a diver in a light-colored wetsuit and a large white scuba tank is swimming horizontally, holding a camera. In the lower half, a large sea turtle with a patterned shell is swimming over a vibrant, colorful coral reef. The water is clear and blue.

Technical deep dive

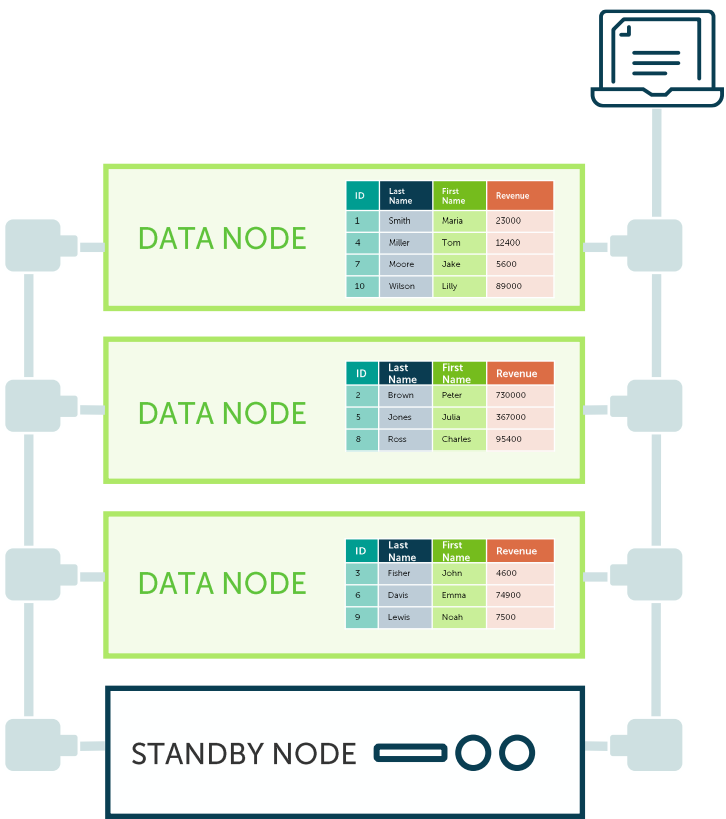
Pure Design: Exasol's Performance Architecture



ID	Last Name	First Name	Revenue
1	Smith	Maria	23000
2	Brown	Peter	730000
3	Fisher	John	4600
4	Miller	Tom	12400
5	Jones	Julia	367000
6	Davis	Emma	74900
7	Moore	Jake	5600
8	Ross	Charles	95400
9	Lewis	Noah	7500
10	Wilson	Lilly	89000



Pure Design: Exasol's Performance Architecture



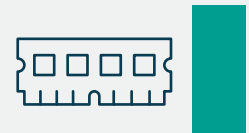
Access times

A DATA NODE IN DETAIL

10 ns



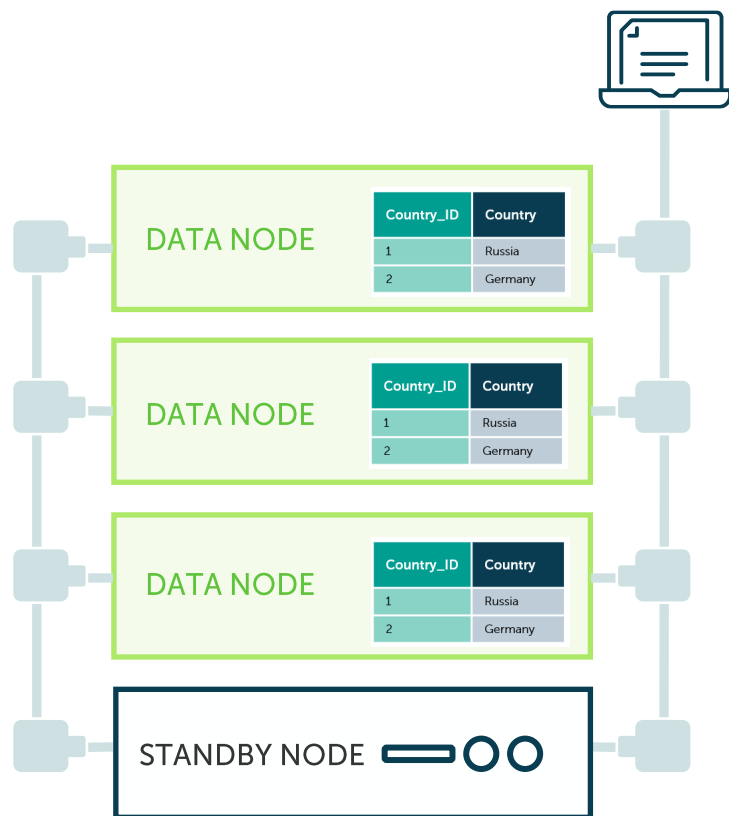
100 ns



10 000 000 ns



Replication of small tables



Country_ID	Country
1	Russia
2	Germany

For small tables: Replicate across all nodes to make joins local

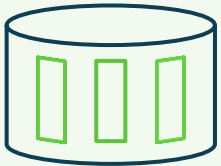
→ Replication border is a one-time setup parameter

`-soft_replicationborder_in_numrows=1000000`

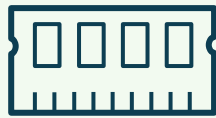
Exasol



Self-tuning



Columnar



In Memory



MPP

Ultimate flexibility: Decide once, deploy anywhere



ON PREMISES

- Single point of contact
- Commodity hardware (x86)
- Virtual environments like Vmware, Docker



PUBLIC CLOUD

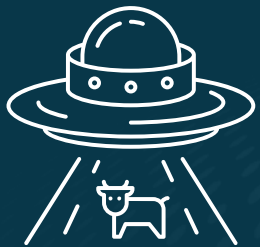
- Your cloud environment, your control
- Tight integration into cloud-ecosystem, e.g. S3 backup



EXACLOUD

- Full database-as-a-service
- Dedicated optimized hardware
- Fully managed 24x7
- Secure Site2Site VPN

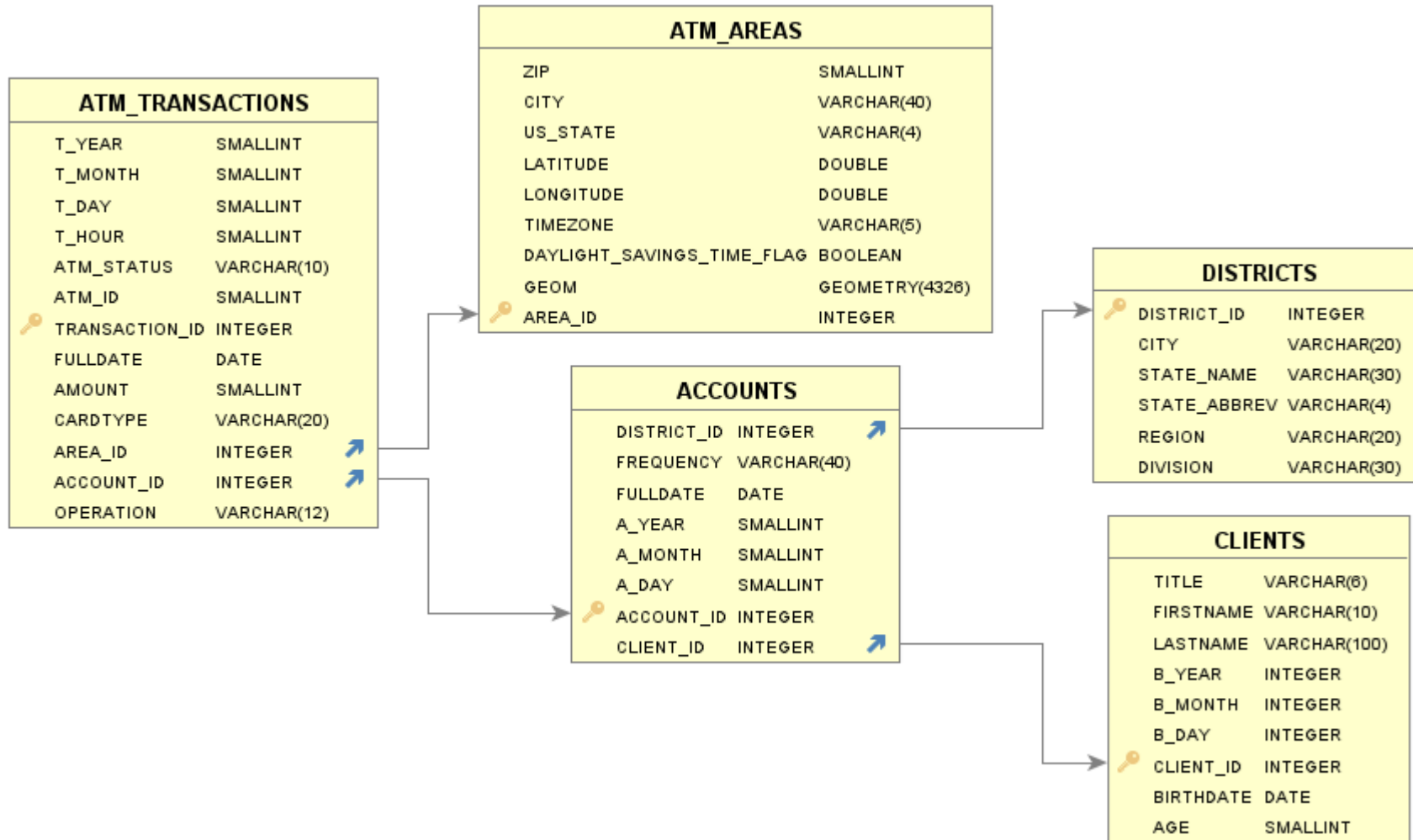
Everyone can
promise
awesome
things



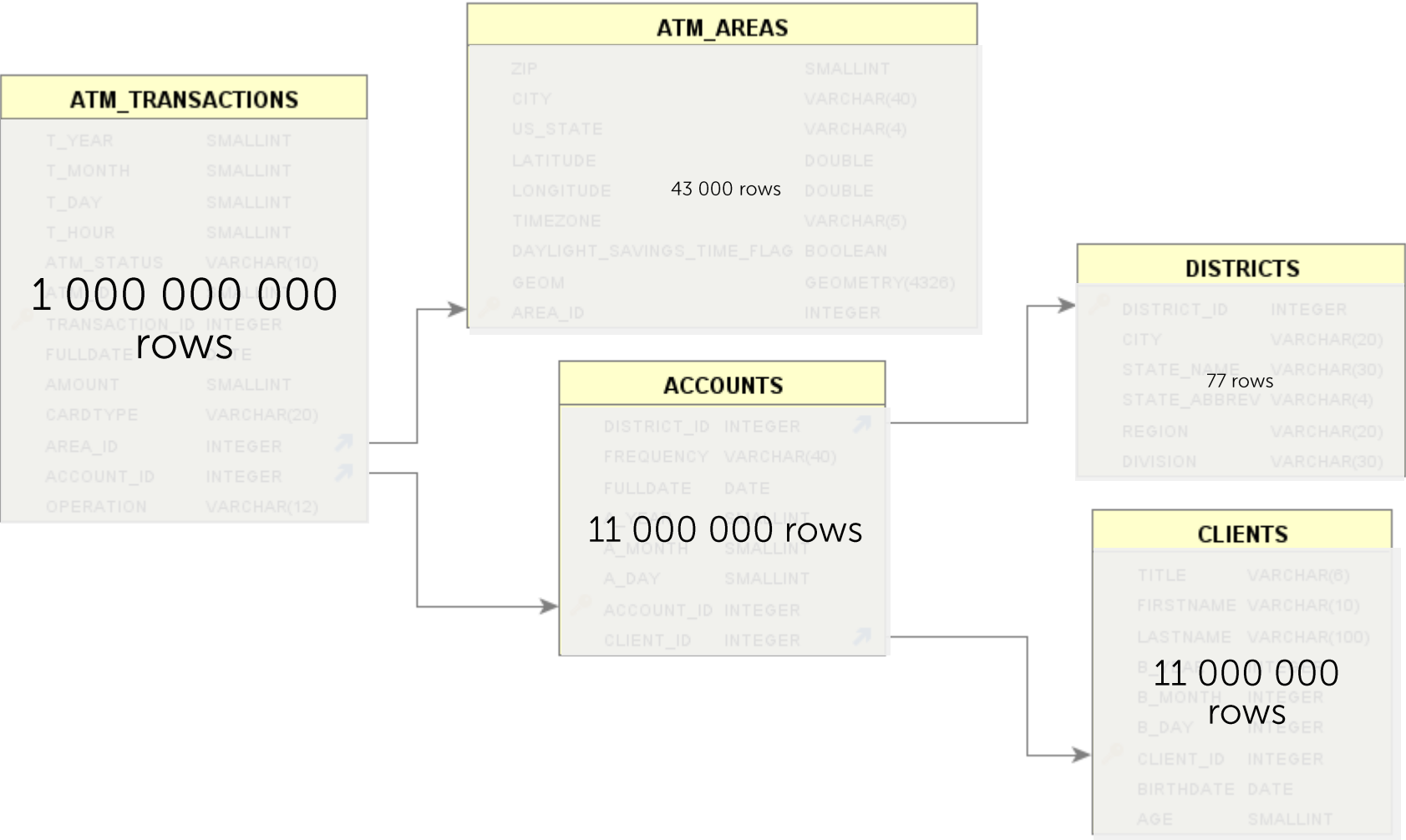
but:



Dataset: Banking



Dataset: Banking



Plan for today

- Load data from Oracle
- Run queries on newly loaded data
- Have a look at Exasol's automatic tuning

From zero to a tuned dataset

- Creating tables
- Load data
- Distributing data
- [Migrate primary + foreign keys]
- Done – start querying data



Indices

- Created automatically
- Maintained for DML statements (INSERT, UPDATE and DELETE)
- If not used for more than 5 weeks, index gets dropped

Tuning an Exasol

- Define a replication border
- Set primary and foreign keys
- Distribute large tables via distribution key

That's it 😎

What We Believe

DATA SCIENCE SHOULD BE DEMOCRATIZED

- Modeling from data in the warehouse is just the beginning
- Data engineers deploy models directly in the database as UDFs for maximum performance
- Analysts use those functions without needing to know the technical tools or steps

HOW?

- Decide on your use cases
- Find the list of script languages and packages which you want to use in Exasol
- Perform user-defined operations on big datasets embedded in SQL
- **Bring algorithms to the data!**

The Benefits of Running Data Science in Exasol



Save time, optimise processes & easily operationalise models otherwise neglected.

By bringing the algorithm to the data you **avoid data transfer** out of the database which can become an obstacle for leveraging data science & ML models in production.



Harness flexibility and work in an agile way, don't be limited by a specific language or tool.

Exasol UDF's are **very flexible** and allow multiple programming languages (Lua, Python, R, Java) and the addition of new packages



Achieve better quality models & insights faster by leveraging the blazing fast speed of the Exasol engine.

UDF & BucketFS allows you to **run algorithms easily distributed** to all the nodes of the cluster with automatic distribution of UDF's for pre-processing.



Avoid risk and practice better data security & governance by keeping everything in a single store.

Because the data never leaves the database, you benefit of the **security mechanism** of the database.

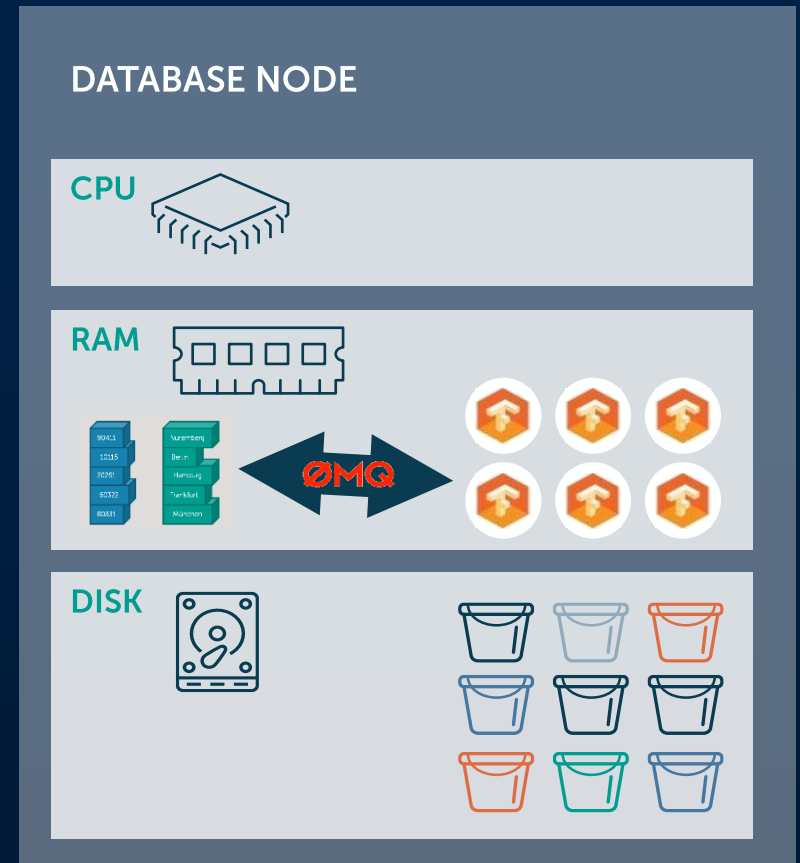
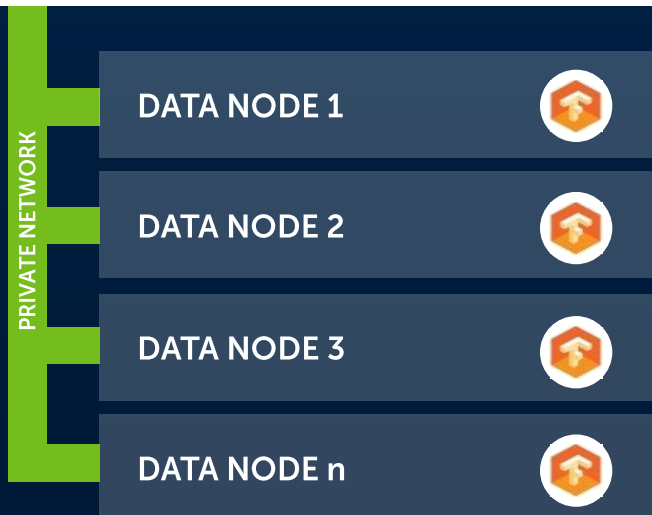
Data Science with Python

Develop python UDFs to run In-Database Analytics in parallel on very big datasets

```

1 --Execute the predict udf on large data
2 select predict_linear_regression(S1,TARGET)
3 from
4 "VERY_BIG_TABLE"
5 group by iproc(), -- the node number for data locality
6 mod (rownum, 32) --32 parallel processes per server
7 ;

```



Start operationalizing data science!



Private Cloud, Public Cloud, Community Edition

- <https://www.exasol.com/test-it-now/>
- <https://cloudtools.exasol.com>



Get Resources

- [Exasol Documentation](#)
- [Exasol on Github](#)



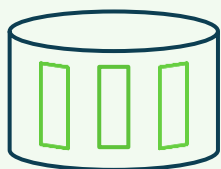
Start Experiencing

- [Data Science Examples](#)
- [Online Trainings](#)

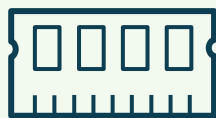
Fraud prediction



Exasol



Columnar



In Memory



MPP

Wrap up:

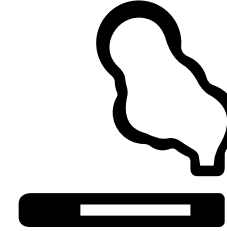
What sets Exasol apart?

- Automatic memory management of hot and cold data which maximizes in-memory resources
- A massively parallel architecture that enables high performance even as you grow
- Complete choice of deployment - you're free to choose the hardware or cloud, so this keeps the price affordable

How does Exasol stop customers from smoking?



ORACLE

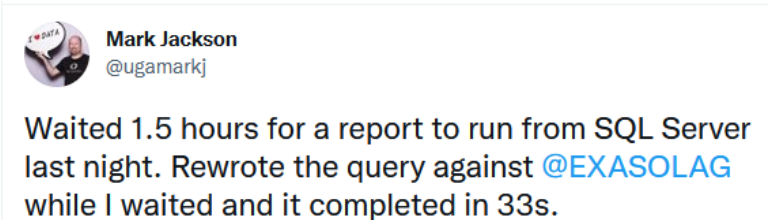


VS.

Exasol

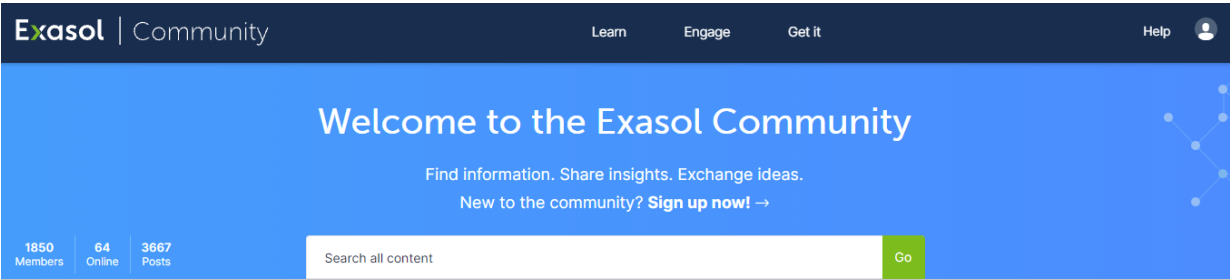


Could it even give you more time for your family?



Christian - community

- Join the Exasol community to discuss with peers
- Offer people to try out what we did in the session on your own by creating a trial account on our system
- Support us on GitHub



Community Activity Recents Start a topic Getting started in the community

Installing Exasol Community on a Mac M1 with UTM
07-10-2021 • Posted in Discussion Forum
by Isacy • Padawan
58 views 1 comment 0 likes

verify TLS certificate?
06-10-2021 • Posted in Discussion Forum
by drumcircle • Rising Star
124 views 3 comments 0 likes

MySQL Connection Problem "Communications link failure" ✓ Resolved!
06-10-2021 • Posted in Discussion Forum
by thschu • Just Joined
178 views 5 comments 0 likes

Events

October 2021

Sun	Mon	Tue	Wed	Thu	Fri	Sat
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30



Exasol is happy to share the latest and greatest from our open source community with the world. Below, you can see a list of our Github projects and a link to the latest release.

NOTE: Inclusion in this list does not guarantee that the project is officially supported by Exasol. Please view the Readme of the various projects to see if it is officially supported or not.

Exasol on Docker

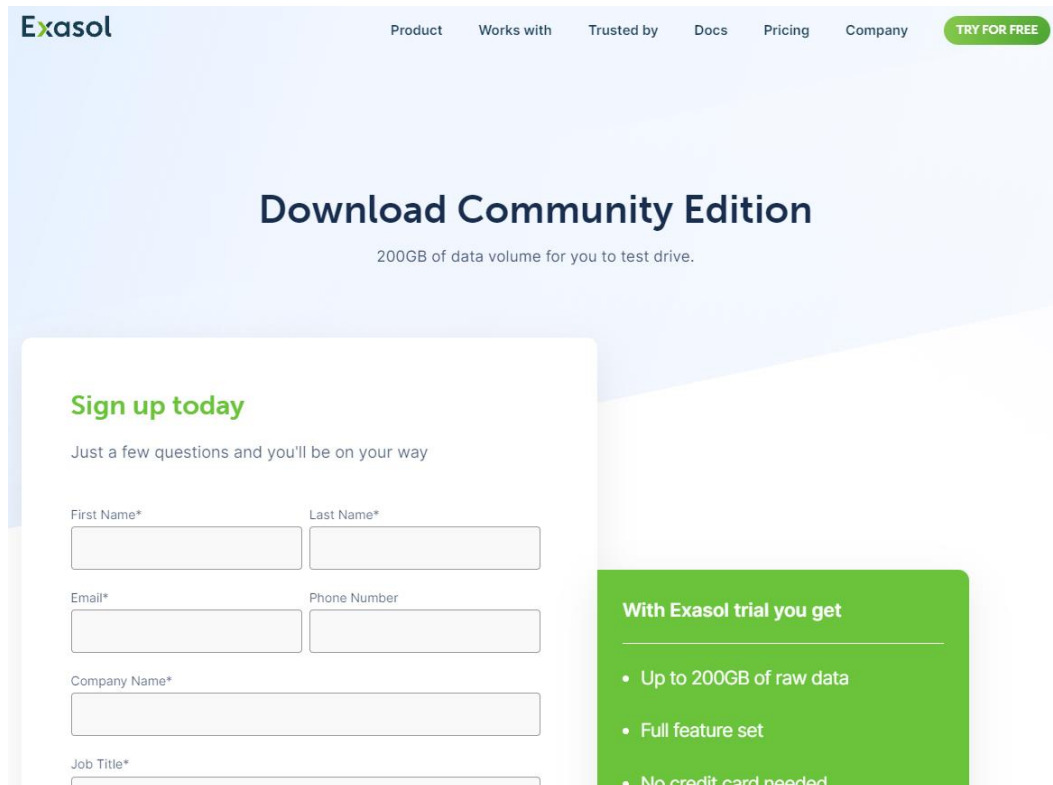
Project Name	Description	Download
Docker DB	Dockerized version of the Exasol Database for testing purposes	https://github.com/exasol/docker-db/releases

Virtual Schemas

Project Name	Description	Download

How can you try Exasol?

- Community edition
 - Download ISO/VM (200GB Raw)
 - Docker (20GB Raw)
 - Cloud (1TB – 30Days)



Exasol

Product Works with Trusted by Docs Pricing Company TRY FOR FREE

Download Community Edition

200GB of data volume for you to test drive.

Sign up today

Just a few questions and you'll be on your way

First Name* Last Name*

Email* Phone Number

Company Name*

Job Title*

With Exasol trial you get

- Up to 200GB of raw data
- Full feature set
- No credit card needed

- Proof of concept
 - Your data, our support



Q&A

