THE MODERN DATA PLATFORM

DATA LAKEHOUSE AND DATAOPS IN DATABRICKS – BY MATHIAS WEBER

TABLE OF CONTENTS

ABOUT ME

ABOUT ME

Name:

- Mathias Weber

Professional:

- Half a Decade of Analytics Consulting Experience

- Senior Delivery Architect at the largest consulting firm of European origin (source: Wikipedia)

- Currently focusing on the technical design of data platforms

ABOUT ME

Extracurricular Engagement:

- Applied Technology Scout, attended 120+ tech events

- Author of 140+ LinkedIn articles ([see LinkedIn](#))

- Author of two books:

  - [Europe's Sweet Spot for Software Architects](#)

  - [Other People's Software Endeavours](#)

APPLIED TECHNOLOGY SCOUTING

APPLIED TECHNOLOGY SCOUTING - THEORY

There are four main goals in technology scouting:

1. **Early identification** of technologies, technological trends and technological shocks.

2. **Raising awareness** of the threats and opportunities posed by technological developments.

3. **Promoting innovation** by combining technology reports with the assessment of business potential.

4. **Facilitating the procurement of external technologies** by enabling direct access to information sources through the network of technology scouts.
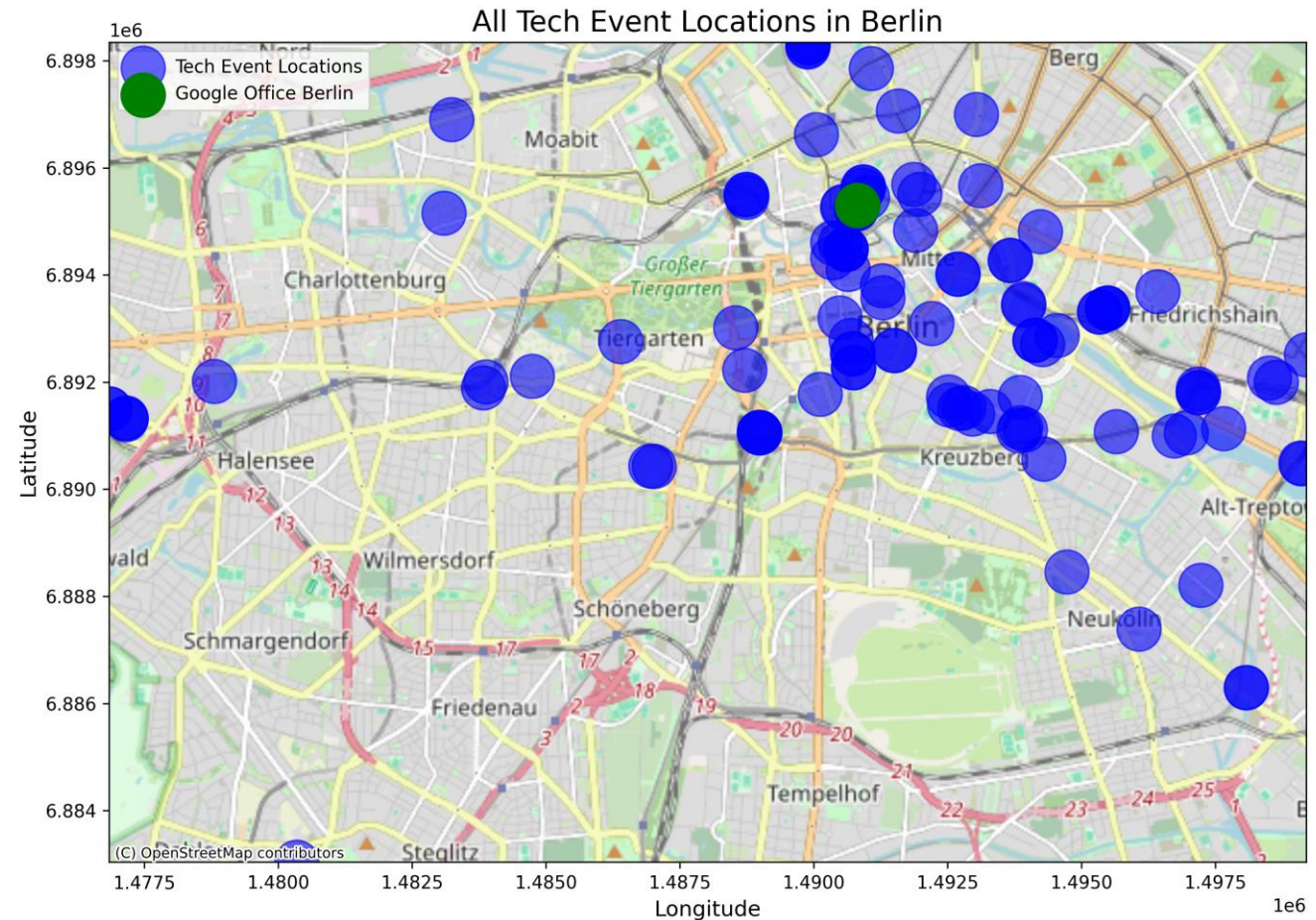
Source: Rohrbeck, 2010, „Harnessing a network of experts for competitive advantage: technology scouting in the ICT industry"

APPLIED TECHNOLOGY SCOUTING
– MY APPROACH

"Dit is Berlin, wa?"

In English:

Mapping the Berlin Tech Scene
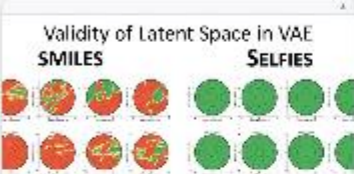
APPLIED TECHNOLOGY SCOUTING – MY APPROACH

Attendance and documentation of 120+ tech events (List of Articles)

APPLIED TECHNOLOGY SCOUTING – THE VISION

„My aim is to make the Berlin tech scene visible and tangible for others."

Mathias Weber

A BRIEF HISTORY OF DATA PLATFORMS

SOURCE: ARMBRUST, M., GHODSI, A., XIN, R., & ZAHARIA, M. (2021, JANUARY). LAKEHOUSE: A NEW GENERATION OF OPEN PLATFORMS THAT UNIFY DATA WAREHOUSING AND ADVANCED ANALYTICS. IN PROCEEDINGS OF CIDR (VOL. 8, P. 28).

A BRIEF HISTORY OF DATA PLATFORMS – GENERATION I

➤ On-Premises-Software

➤ ETL process: transforming structured data from operational databases into data warehouses

➤ Decision support and business intelligence

➤ Personal experience with Cubeware BI Suite
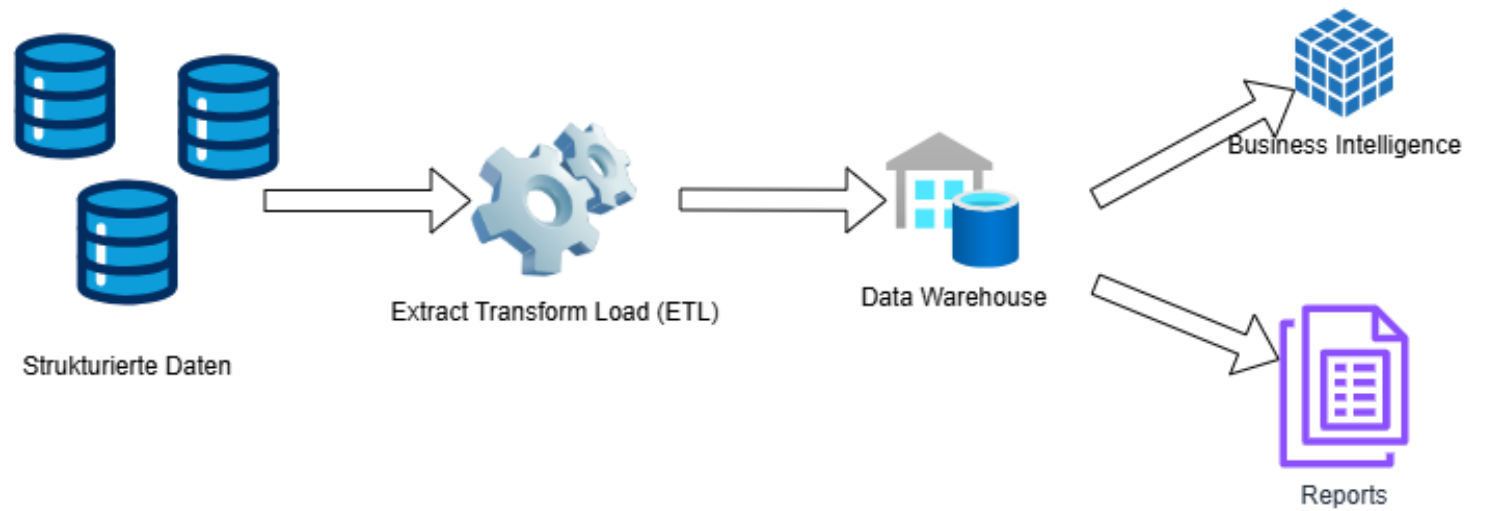
A BRIEF HISTORY OF DATA PLATFORMS – GENERATION II

➢ Growing and more diverse datasets (Volume & Variety)

➢ New data types: semi-structured and unstructured data

➢ Introduction of data lakes, ingestion of raw data

➢ New opportunities for data science and machine learning
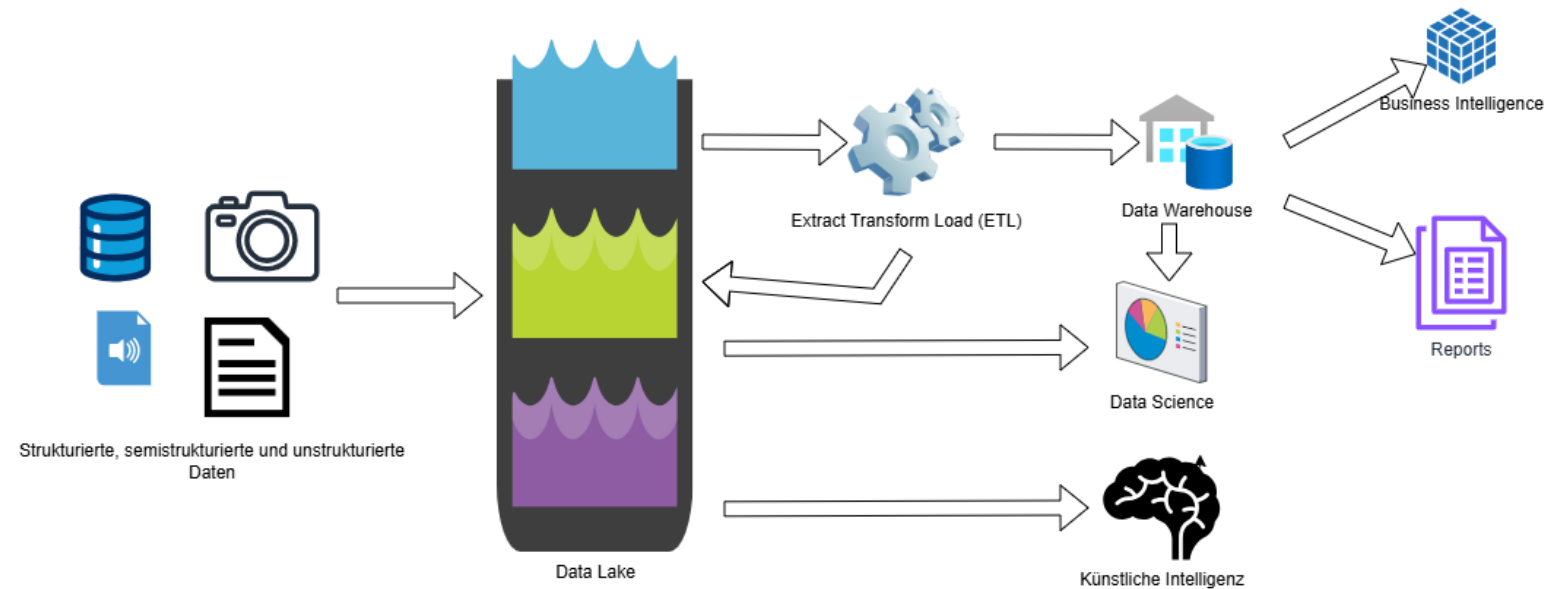
A BRIEF HISTORY OF DATA PLATFORMS – GENERATION III

➤ Data Lake + Data Warehouse = Lakehouse

➤ Metadata, caching, and indexing layer on the data lake

➤ Management functions of a data warehouse on the data lake
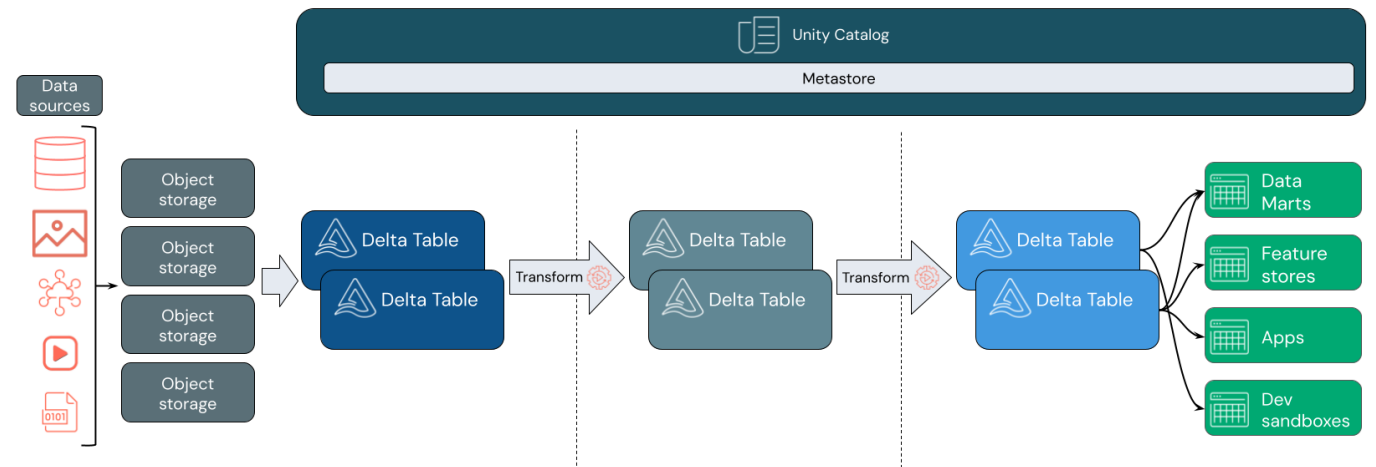
Source: https://learn.microsoft.com/de-de/azure/databricks/_static/images/lakehouse-architecture/lakehouse-diagram.png

SAMPLE ARCHITECTURE WITH AZURE AND DATABRICKS

SAMPLE ARCHITECTURE WITH AZURE AND DATABRICKS

Source: https://learn.microsoft.com/en-us/azure/architecture/solution-ideas/articles/azure-databricks-modern-analytics-architecture
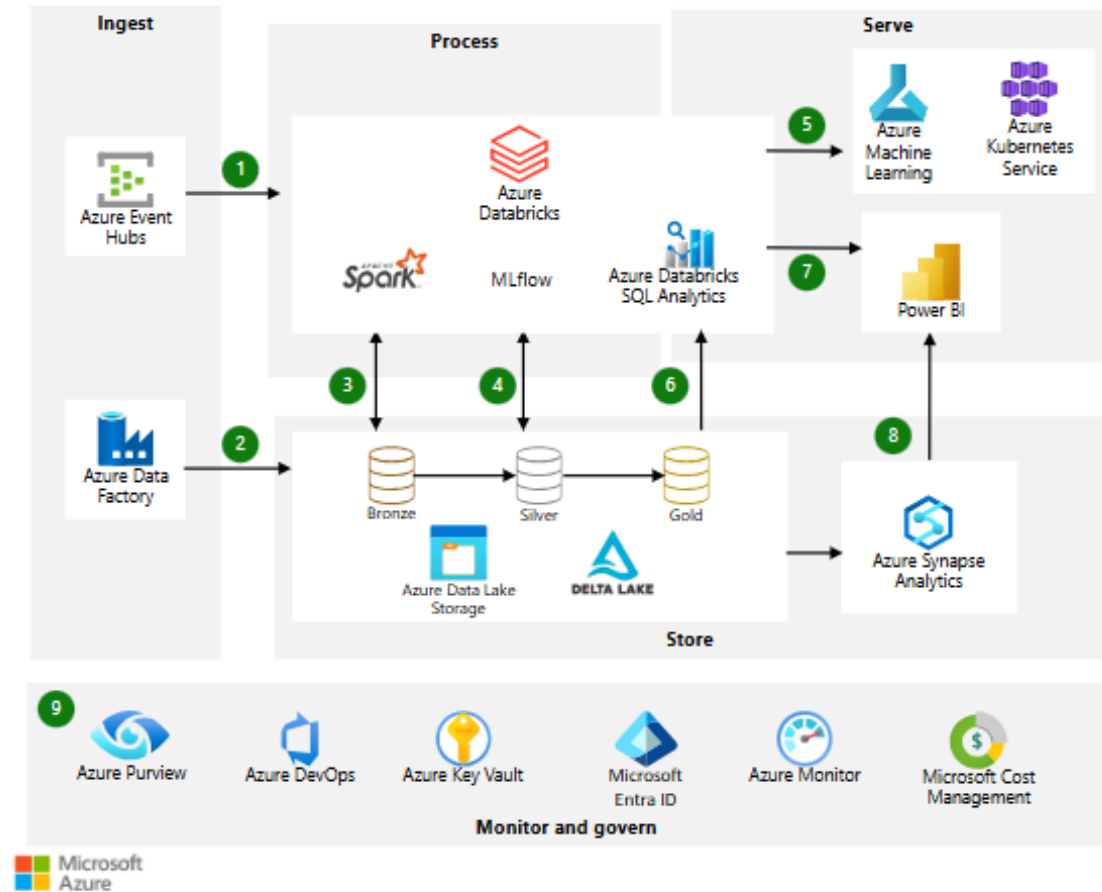
FROM UNVERSIONED JUPYTER NOTEBOOKS TO SUSTAINABLE DATAOPS

FROM UNVERSIONED JUPYTER NOTEBOOKS …

Humble Beginnings
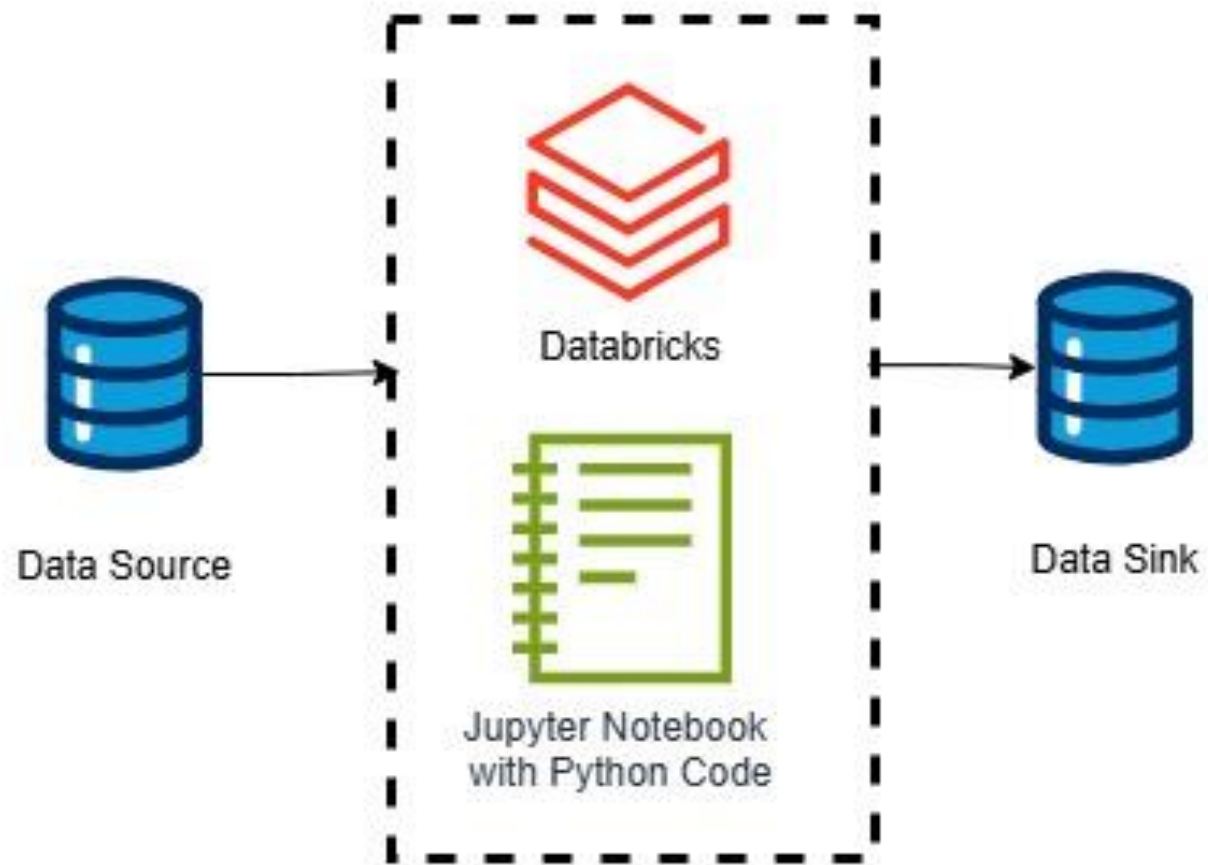
… ISSUES, ISSUES, ISSUES …

How do I design new data analytics applications or extend existing ones with new features?

How and with what do I develop?

How do I build my applications?

How do I configure the environment in which my applications run?

How do I test my applications?

How do I release new versions?

How do I ship code to production?

How do I ensure the smooth operation of my application?

How do I monitor my application?

How do I do all of this in a continuous, iterative, and sustainable manner?

… SOLUTIONS, SOLUTIONS, SOLUTIONS …

- Plan
- Develop
- Build
- Environment Management
- Test
- Release
- Deploy
- Operate
- Monitor

The overarching answer is DataOps.

… EXAMPLES, EXAMPLES, EXAMPLES …

Plan > Agile Sprints in Jira

Develop > Local Development in VS Code

Build > CI-Workflow in GitHub

Environment Management > IaC with Azure ARM

Test > Testing Pyramid with Pytest

Release > Version Control with Git

Deploy > CD-Workflow in GitHub

Operate > Orchestration of Workloads within Azure Data Factory

Monitor > Traces, Metrics and Logs with LogAnalytics

… TO SUSTAINABLE DATAOPS

*"DataOps (short for data operations) is a **data management practice** that makes building, testing, deploying, and managing data products and data apps the same **as it is for software products**. It combines technologies and processes to improve trust in data and reduce your company's data products' time to value."*

Source: www.dataops.live/what-is-dataops

BONUS CONTENT

TOP 5 TECH EVENTS IN 2024

Linkedin
Artikel Personen E-Learning Jobs Spiele

My Humble Recap: Top 5 Tech Events in 2024

Mathias Weber
Data Architect and Author of the Book "Other People's Software Endeavours"
Veröffentlicht: 19. Dez. 2024

+ Folgen

Last week, I powered through *four* tech events in just one week. Why, you ask? Well, I had a mission: to hit my 2024 goal of attending at least **50 tech events**— and, let's be real, to flex my judgmental muscles and handpick the crème de la crème, the top 10% (aka my top 5).

If you're curious about my ongoing tech event scouting saga, I shamelessly recommend my book, *"Europe's Sweet Spot for Software Architects"*. But enough with the teasers—let's dive right in, shall we?

Let's kick things off from the bottom and climb our way to the very top—because the best is always worth the wait.

5TH PLACE

screenshot taken from: www.starknet.io/en

Event Report (#74): Starknet Germany - Basics & Advantages of Account Abstraction

Link to Event Report: **Event Report (#74): Starknet Germany - Basics & Advantages of Account Abstraction**
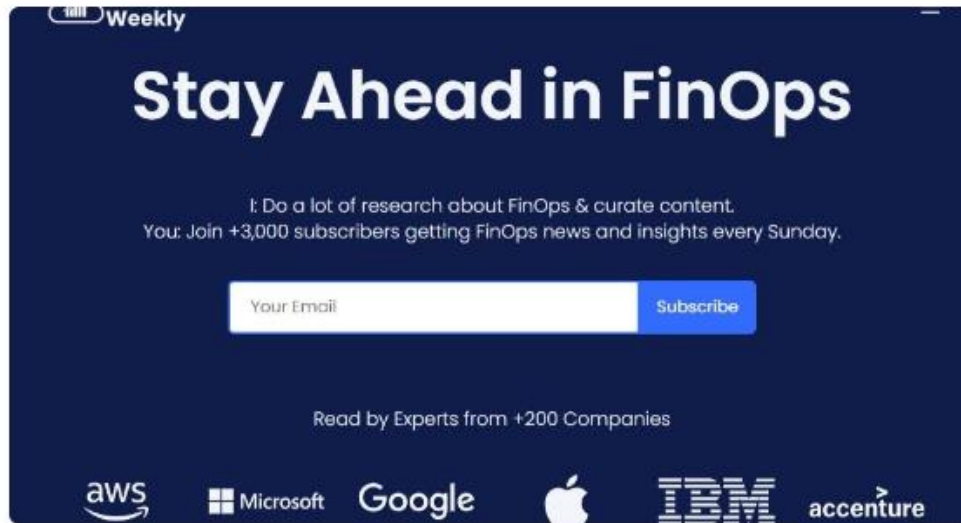
Objective Reason for Placement:

Web3 is very much alive, and crypto's riding a bull run. If you're not figuring out how to get your app on the blockchain, are you even keeping up?

Subjective Reason for Placement:

Hosting your event next to a standalone McDonald's? Now that's better than having it your way.

4TH PLACE

Link to Event Report: **Event Report (#120): Berlin FinOps and Cloud Cost Optimization No. 10**

Objective Reason for Placement:

After AI gave us those flirty eyes in 2024, it'll probably break our hearts—and our wallets—in 2025, especially when everyone jumps on the bandwagon without a proper FinOps plan, sending compute costs through the roof.

Subjective Reason for Placement:

The eerie calm of Berlin Adlershof's commercial area was pure bliss for my soul, worn down by the relentless noise of the inner city.

3RD PLACE

Event Report (#114): Games Ground 2024 Day 2

Link to Event Report: **Event Report (#114): Games Ground 2024 Day 2**

Objective Reason for Placement:

If we want the upcoming industrial metaverses to be more fun and less snooze-fest, it's probably worth taking a page from the playbook of video game development and how game servers are run.

Subjective Reason for Placement:

Those apocalyptic art installations just hit differently—like a beautiful chaos you can't look away from.

2ND PLACE

Event Report (#115):
Innovation Jam Tech
Conference 2024 at Gdansk
Olivia Centre

Link to Event Report: **Event Report (#115): Innovation Jam Tech Conference 2024 at Gdansk Olivia Centre**

Objective Reason for Placement:

AI inference when interacting with a smart device shouldn't test the customer's patience. At some point, we might need to craft custom chips designed specifically for the task—because lag isn't a good look.

Subjective Reason for Placement:

Nothing beats the view of Gdańsk Bay from the 34th floor—it's pure wellness for the eyes.
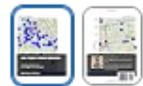
AND THE WINNER IS …
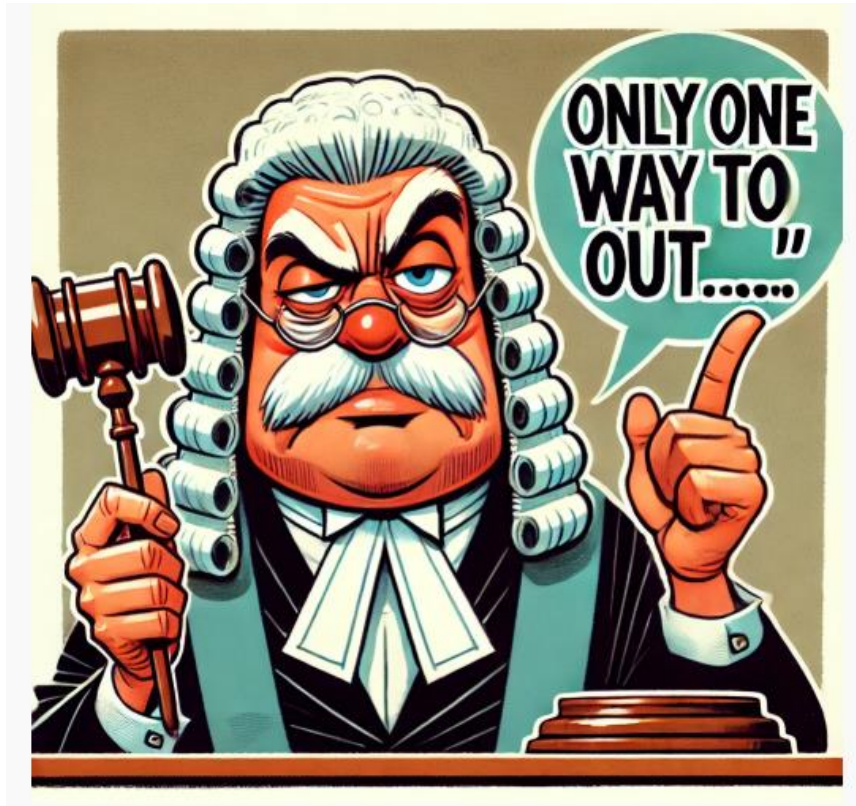
The only way to find out is to read the rest of the article yourself. Here is the link:

TOP 5 Tech Events 2024

A SMALL BUT IMPORTANT REQUEST

BUY MY BOOK!!!

Other People's Software Endeavours: The Fool's Final Journey & The Sage's First Conclusion
Taschenbuch – 8. Januar 2025

Englisch Ausgabe | von Mathias Weber (Autor)

Alle Formate und Editionen anzeigen

This book is for everyone who's tired of the same old routine and craving real adventures in Berlin's event scene.

Our young protagonist dives headfirst into 100 different events—ranging from secret backroom gatherings to spectacular mega-events—and discovers that networking can be more than just boring business card exchanges.

Here, you'll learn why stepping out of your comfort zone and striking up conversations with complete strangers is worth it. This book proves that events aren't just for industry nerds—anyone can draw inspiration, steal ideas, and forge genuine connections. Whether you're a programmer or have no clue what you're doing, grab this book and let Berlin's tech scene shake you up in the best way possible!

Ein Problem mit diesem Produkt melden

Seitenzahl der Print-Ausgabe	Sprache	Erscheinungstermin	Abmessungen	ISBN-13
136 Seiten	Englisch	8. Januar 2025	15.24 x 0.86 x 22.86 cm	979-8306296913

Alle Details anzeigen

Other People's Software Endeavours

The Fool's Final Journey & The Sage's First Conclusion

Mathias Weber

Leseprobe lesen

Making Business Connections That Count: The Gimmick-free Guide to Authentic Online Relationships with...
von Michal Stawicki
★★★★☆ 66 | 8,02 € prime | Taschenbuch
Jetzt einkaufen ›

Gesponsert

JUST KIDDING—NOW, HERE'S MY ACTUAL REQUEST.

SUPPORT THIS:

3 STEPS TO SUCCESS
STEP 1: SUPPORT **THE AGE OF DEVOPS COMMUNITY** IN BERLIN
STEP 2: ???
STEP 3: EVERYONE MOVES FROM SILICON VALLEY TO BERLIN

THANK YOU!

WITH BEST REGARDS,
MATHIAS WEBER