RICK VAN DER LANS

LEAN DATA ARCHITECTURES TO MINIMIZE DATA COPYING

From Data-by-copying to Data-on-demand

ONLINE LIVE STREAMING

DECEMBER 14, 2022

ABOUT THIS SEMINAR

There was a time when people would visit a record store to buy a copy of an album to listen to at home. There was also a time when people went to a video store to rent a DVD to have a copy at home to watch the movie. Not anymore, music and video are streamed. People no longer listen to or watch copies. Music-by-copying has been replaced by music-on-demand and video-by-copying by video-on-demand.
Unfortunately, the world of data has remained unchanged. Data is still copied, in fact several times before it is even consumed. Many organizations are lightyears away from Data architectures that support data-on-demand.

Most data Architectures are duplication-heavy. So much data is duplicated multiple times. For example, data about a specific customer can be stored in a transactional system, a staging area, a Data Warehouse, several Data Marts, and in a Data Lake. Even within one database, data can be stored multiple times to support different data consumers. Additionally, redundant copies of the data are stored in development and test environments. But business users also copy data. They may have copied data from central databases to private files and spreadsheets. Also, data infrastructures currently consist of Data Lakes, Data Hubs, Data Warehouses, and Data Marts. And all these systems contain overlapping data.

In addition to these intra-organizational forms of data copying, massive inter-organizational copying takes place. When organizations exchange data with each other, the receiving organizations store the data in their own systems, creating even more copies of the data.
It is time for lean data Architectures that minimize copying of data. The advantages of this are manyfold, such as the Architecture is more flexible, improves productivity and maintenance, lowers low data latency enabling real or near real data-on-demand solutions, and less error-prone.

A lean Architecture relies more heavily on the strength and performance of technology. The entire workload is not distributed across several Data Marts or a Data Lake with a Data Warehouse, but must be processed by a smaller number of databases. The good news is that new technology is available that can handle these bigger workloads. Especially fast analytical databases and scalable Cloud platforms make lean Architectures a reality.

In the old days, several reasons existed to create data copies. But database performance, Cloud technology, and network speed have improved enormously, often making copying of data unnecessary. Unfortunately, new data Architectures are still being designed in which data is stored redundantly. Architects think too casually about copying data and storing it redundantly. Copying data has many drawbacks and challenges:

• Higher data latency
• Complex data synchronization
• More complex data security and data privacy enforcement
• Higher development and maintenance costs
• Higher technology costs
• More complex database administration
• More complex metadata administration
• Reduced data quality

Redundant data is introduced too easily, and this unrestrained duplication must stop. Lean data Architectures aim to reduce the time data is copied.
During this seminar, Rick van der Lans explains how to design a lean data Architecture and which solutions and technologies are available to develop one. Design guidelines for zero-copy and single-copy data Architectures and a comparison with duplication-heavy Architectures are discussed. How to minimize intra -and inter- organization copying is discussed. The impact on existing Data Warehouse, Data Lake, and Data Hub Architectures are presented. A complete picture of designing lean data Architectures in real-life projects is given.

WHO SHOULD ATTEND

- Business Intelligence Specialists
- Data Analysts
- Data Warehouse Designers
- Business Analysts
- Data Scientists
- Technology Planners
- Technical Architects
- Enterprise Architect
- IT Consultant
- IT Strategists
- Systems Analysts
- Database Developers
- Database Administrators
- Solutions Architects
- Data Architects
- IT Managers

WHAT YOU WILL LEARN

- How to design lean data integration Architectures using examples
- What the real drawbacks are of creating too many copies of the data are, including higher data latency, complex data synchronization, more complex data security and privacy, and higher development and maintenance costs
- How new database, integration, and Cloud technology can help to design lean data Architectures that contain less copied data
- What the effect is of applying data minimization to Data Warehouse and Data Lake Architectures
- How to design the data in single-copy solutions
- What the 1:1+ approach for data Architectures means
- How to replace managed-file-transfer solutions by data-on-demand solutions, and how to reduce inter-organizational data flows
- How to design data Architectures from the perspective of data processing specifications and not data stores

*O*UTLINE

1. Unrestrained Copying of Data

- Examples of intra-organization data copies
- Examples of inter-organization data copies
- Copying data in new data Architectures, such as Data Lakes and Data Hubs
- What is data minimization?
- From data-by-delivery to data-on-demand
- Risks and drawbacks of copying and duplicating data

2. Justifying Lean Data Architectures

- Business advantages of lean data Architectures, such as improved time-to-market, support for (near) real-time data consumers (internal and external), improved conformance to data security and privacy, and improved data quality
- Technical advantages of lean data Architectures, such as simplification of development, management and operation of synchronization programs, less complex database and metadata administration

3. New Technologies Enabling Lean Data Architectures

- Analytical database servers and their distributed, share-based Architecture
- Translytical database servers: combining transactions and analysis
- Cloud technology offers the required scalability and centralization of data
- Data virtualization enables reduction of redundant data
- Messaging and streaming technology

4. Design guidelines for Lean Data Architectures

- Differences between zero-copy solutions (real data-on-demand) and single-copy solutions (near real data-on-demand)
- Valid reasons for copying data, such as source does not keep track of history, availability level of source is too low, and extracting data from source is too expensive
- Use the 1:1+ approach for table design with single-copy solutions
- Extended copies contain data not stored by the source; reasons may be the need for artificial data, additional metadata and auditability
- The difference between technical and functional data corrections
- Trust the performance of database servers
- Keeping track of data history only once
- Copy when needed, but not by default

5. Minimizing Inter-organization of Copying Data

- Replacing managed file transfer by data-on-demand across organizations
- Challenges: Extra infrastructure needed at source, more unpredictable workloads, service-level agreements
- Accessing geographically dispersed data sources
- Maximizing performance of distributed queries by centralizing data in the cloud
- What can we learn from video-streaming services, such as Netflix and Amazon Prime?

OUTLINE

6. Transforming Current Data Architectures to Lean Architectures

• From traditional Data Warehouse Architectures to logical Data Warehouse Architectures
• From physical Data Lake with zones and tiers to virtual Data Lakes
• From Data Lakehouses to logical Data Lakehouses
• From Data Fabrics to logical Data Fabrics
• The impact of lean Data Architectures on data privacy aspects

7. Closing remarks

• General recommendations for designing lean data Architectures
• 'Netflixing' your data

SPEAKER

Rick van der Lans is a highly-respected independent analyst, consultant, author, and internationally acclaimed lecturer specializing in data warehousing, business intelligence, big data, and database technology.

He has presented countless seminars, webinars, and keynotes at industry-leading conferences. He also helps clients worldwide to design their data warehouse, big data, and business intelligence architectures and solutions and assists them with selecting the right products. He has been influential in introducing the new logical data warehouse architecture worldwide which helps organizations to develop more agile business intelligence systems.

Over the years, Rick has written hundreds of articles and blogs for newspapers and websites and has authored many educational and popular white papers for a long list of vendors. He was the author of the first available book on SQL, entitled including Introduction to SQL, which has been translated into several languages with more than 100,000 copies sold. More recently, he published his book **Data Virtualization for Business Intelligence Systems**.

He presents seminars, keynotes, and in-house sessions on Big data and analytics, data virtualization, the logical data warehouse, data warehousing and business intelligence.

INFORMATION

PARTICIPATION FEE

€ 650

The fee includes all seminar documentation.

SEMINAR TIMETABLE

9.30 am - 1.00 pm
2.00 pm - 5.00 pm

HOW TO REGISTER

You must send the registration form with the receipt of the payment to:
info@technologytransfer.it

TECHNOLOGY TRANSFER S.r.l.
Piazza Cavour, 3 - 00193 Rome (Italy)

PAYMENT

Wire transfer to:
Technology Transfer S.r.l.
Banca: Credit Agricole
Agenzia 1 di Roma
IBAN Code:
IT 03 W 06230 03202 000057031348
BIC/SWIFT: CRPPIT2P546

GENERAL CONDITIONS

DISCOUNT

The participants who will register 30 days before the seminar are entitled to a 5% discount.

If a company registers 5 participants to the same seminar, it will pay only for 4.

Those who benefit of this discount are not entitled to other discounts for the same seminar.

CANCELLATION POLICY

A full refund is given for any cancellation received more than 15 days before the seminar starts. Cancellations less than 15 days prior the event are liable for 50% of the fee. Cancellations less than one week prior to the event date will be liable for the full fee.

CANCELLATION LIABILITY

In the case of cancellation of an event for any reason, Technology Transfer's liability is limited to the return of the registration fee only.

RICK VAN DER LANS
LEAN DATA ARCHITECTURES TO MINIMIZE DATA COPYING

December 14, 2022

Registration fee:
€ 650

If registered participants are unable to attend, or in case of cancellation of the seminar, the general conditions mentioned before are applicable.

first name ....

surname ....

job title ....

organisation ....

address ....

postcode ....

city ....

country ....

telephone ....

fax ....

e-mail ....

Stamp and signature

*Send your registration form
with the receipt of the payment to:*
Technology Transfer S.r.l.
*Piazza Cavour, 3 - 00193 Rome (Italy)
Tel. +39-06-6832227 - Fax +39-06-6871102
info@technologytransfer.it
www.technologytransfer.it*