How and Where Oracle 26ai Beats Python for Data Preparation

Dr. Holger Friedrich, Chief Executive Officer, sumIT AG

ANALYTICS AND DATA
TechCasts

Past & Future TechCasts:

July 23rd

Temporal Data Lineage and Impact Analysis in Oracle Database: Evolution to Oracle Graph Architecture with Time-Travel Capabilities

Presented by **Karin Patenge & Tomislav Zarkovic**

August 20th

How and Where Oracle 26ai Beats Python for Data Preparation

Presented by **Dr. Holger Freidrich**

September 17th

Graph Analytics with Oracle AI Database 26ai and Linkurious Ogma

Presented by **Leo Nicolle & Karin Patenge**

TechCast Archive

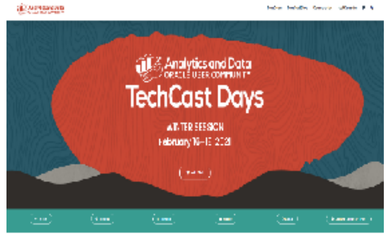
2026 | 2025 | 2024 | 2023 | 2022 | 2021

Date	Title	Presenter(s)	Replay	Download(s)
Jun 25	Our Favorite Features of OAC: March & May 2026 Releases	Gautam Pisharam & Carter Beaton	Video	Slides
Apr 30	Pythian's AI Journey with UMass: From AI Assistant to AI Agents!	Jon Mead & Ambili Thottathil, Pythian	Video	Slides
Apr 2	Build Practical AI Prototypes with Oracle Analytics (Even with a Small Team)	Dennis Gray & Adam Hume, Definian	Video	Slides
Mar 5	Fraud Stops with Oracle Database 26ai Vector Search + OML	Dr. Abigail Giles-Haigh	Video	Slides
Feb 5	Our Favorite Features of OAC: November 2025 & January 2026 Releases	Branden Pavol, Gautam Pisharam, & Carter Beaton	Video	Slides

Submit a topic to share at **https://andouc.org/techcasts/**

Analytics and Data ORACLE USER COMMUNITY

Let's Connect



Website
http://andouc.org/

Chat with the Experts
https://bit.ly/Join-ANDOUC-Slack

Watch Previous TechCasts
https://bit.ly/3qmGgHN

https://www.linkedin.com/company/analytics-and-data-oracle-user-community

Spatial + Graph SIG
bit.ly/Spatial-Graph-LinkedIn

@analyticsndataouc.bsky.social

ANALYTICS AND DATA TechCasts

Analytics and Data SUMMIT 2027

APRIL 13-15, 2027

SAVE THE DATE!
April 13-15, 2027

Call for Speakers is OPEN!!

https://andouc.org/analytics-and-data-summit-2027/

ANALYTICS AND DATA TechCasts

Agenda

- Introduction
- Use Case
- Examples & Demos
- Conclusions

sumIT AG

- Consulting- and implementation services in Switzerland for
  - Data Ware-/Lakehousing a.k.a. ‘modern data platforms'
  - Business Intelligence
  - Advanced Analytics
  - Data Integration/Engineering/Mesh
- Technical focus
  - Oracle on-premises
  - Oracle and others in the Cloud
- Part of Sequotech Group since 2025
- Our motto: Get Value From Data

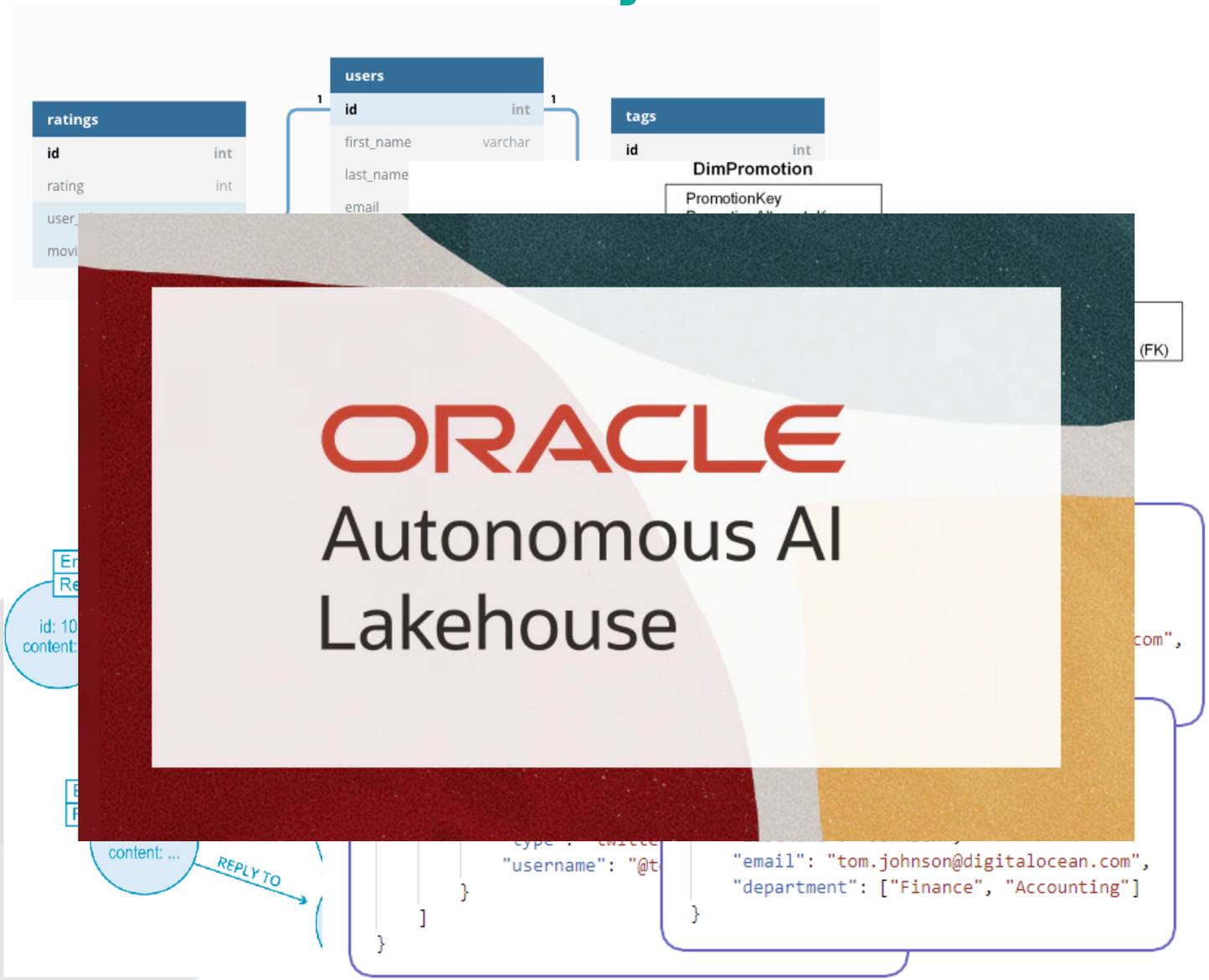
Holger Friedrich

- CS Diploma from Karlsruhe Institute of Technology (KIT)
- Ph.D. in Robotics and Machine Learning (KIT)
- >25 years of Oracle tech experience
- expert for
  - Data Integration
  - Data Ware-/Lakehousing,
  - Machine Learning, and
  - Business Intelligence/Analytics
- CEO sumIT AG
- www.sumit.ch

Oracle ACE Director

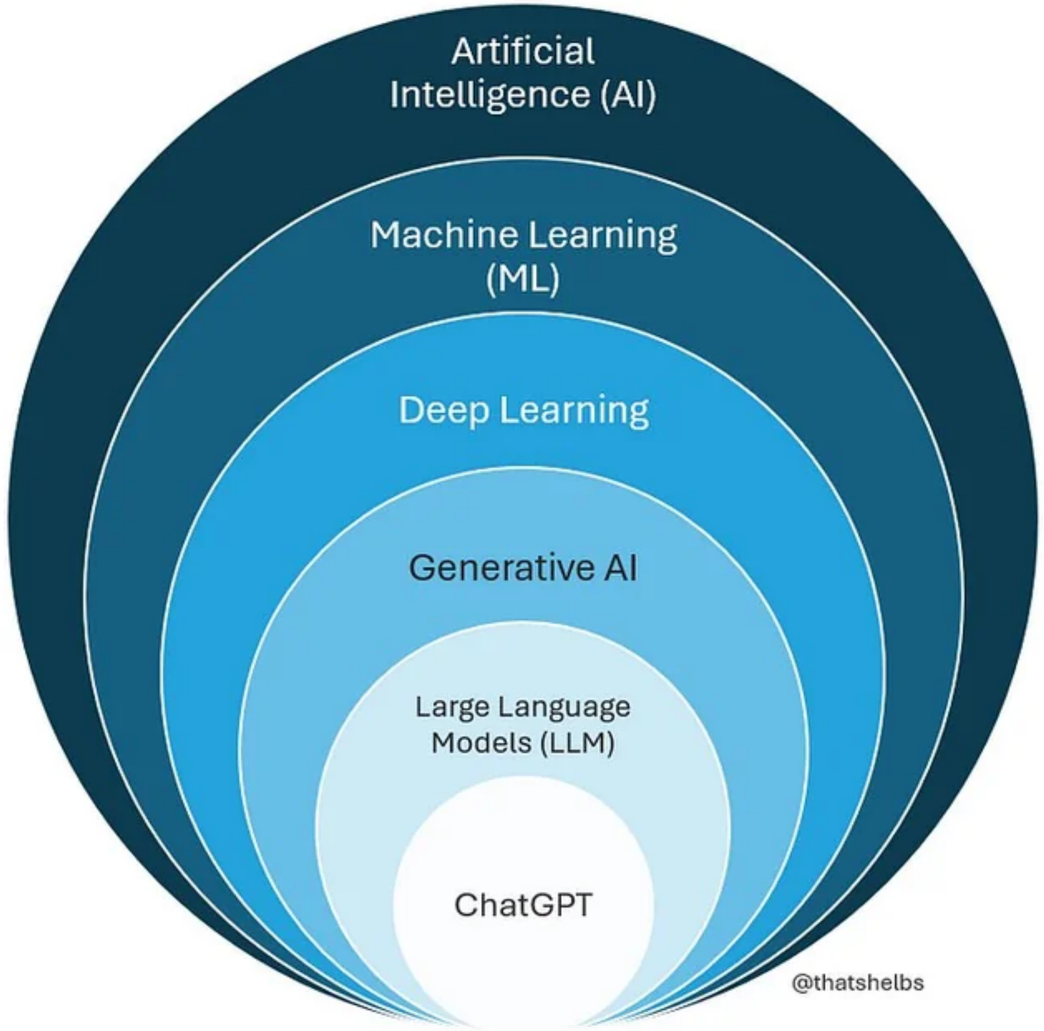
Today's Data World

- Many different kinds of data in databases
  - relational
  - dimensional
  - graph
  - documents
  - …
- Lots of data in
  - files
    - csv
    - json
    - parquet
  - open table formats, e.g. Iceberg



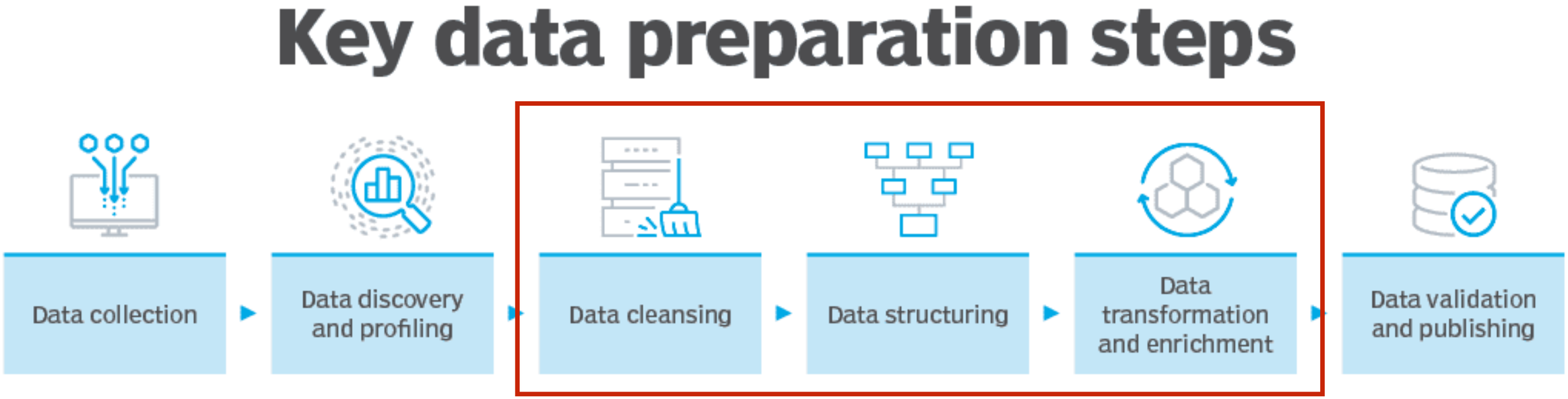
Today's AI World

- Many AI use-cases on the same data
  - Generative AI
  - RAG
  - Agentic AI
  - classic Machine Learning
    - churn
    - classification
    - prediction
    - …



All use cases require performant data preparation!

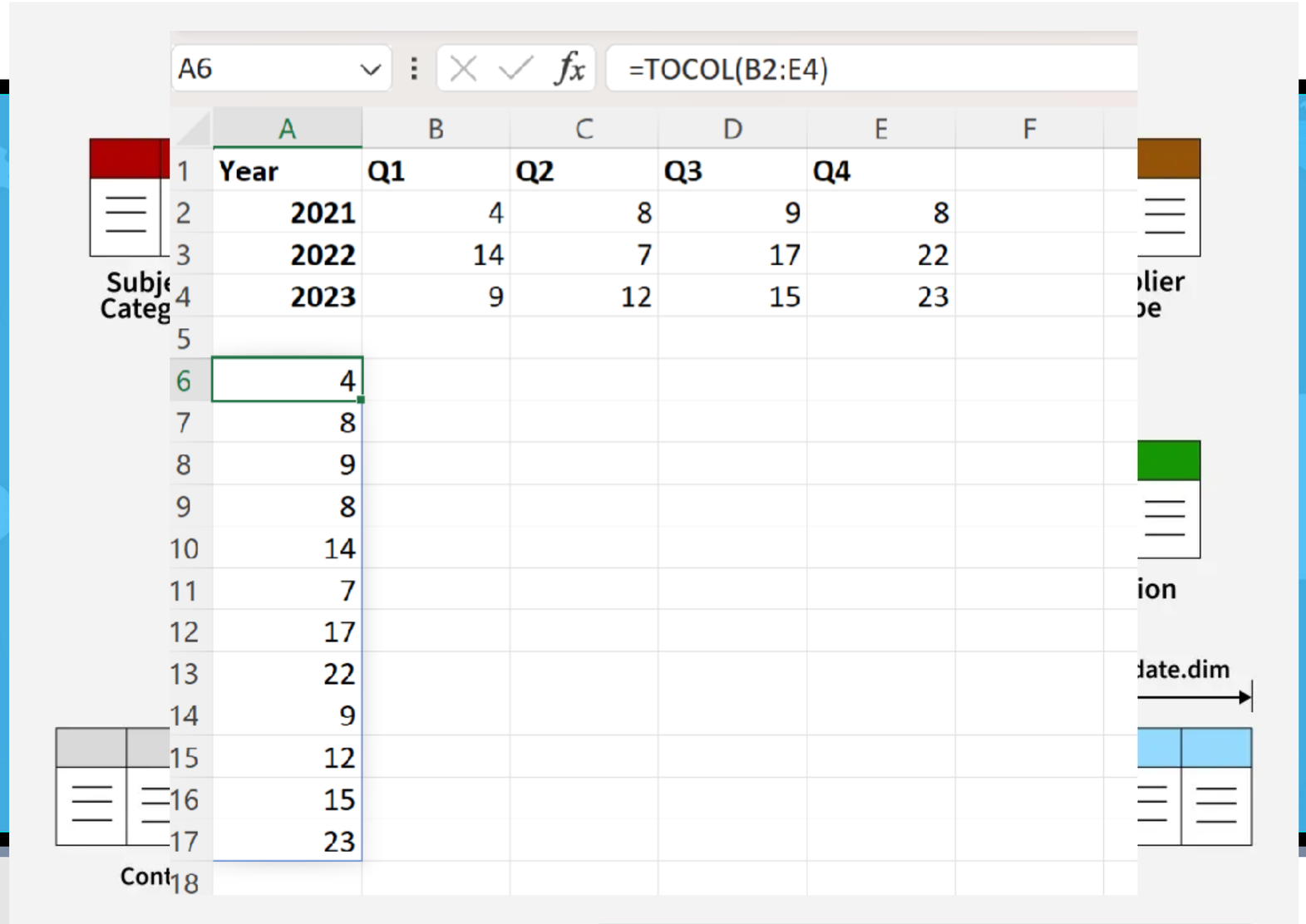
Data Preparation Steps



ICONS FROM LEFT: PRIYANKA GUPTA/GETTY IMAGES, TIM_IURII/GETTY IMAGES, ENIS AKSOY/GETTY IMAGES, FINGERMEDIUM/GETTY IMAGES, ENOTMAKS/GETTY IMAGES, BROWNDOGSTUDIOS/GETTY IMAGES

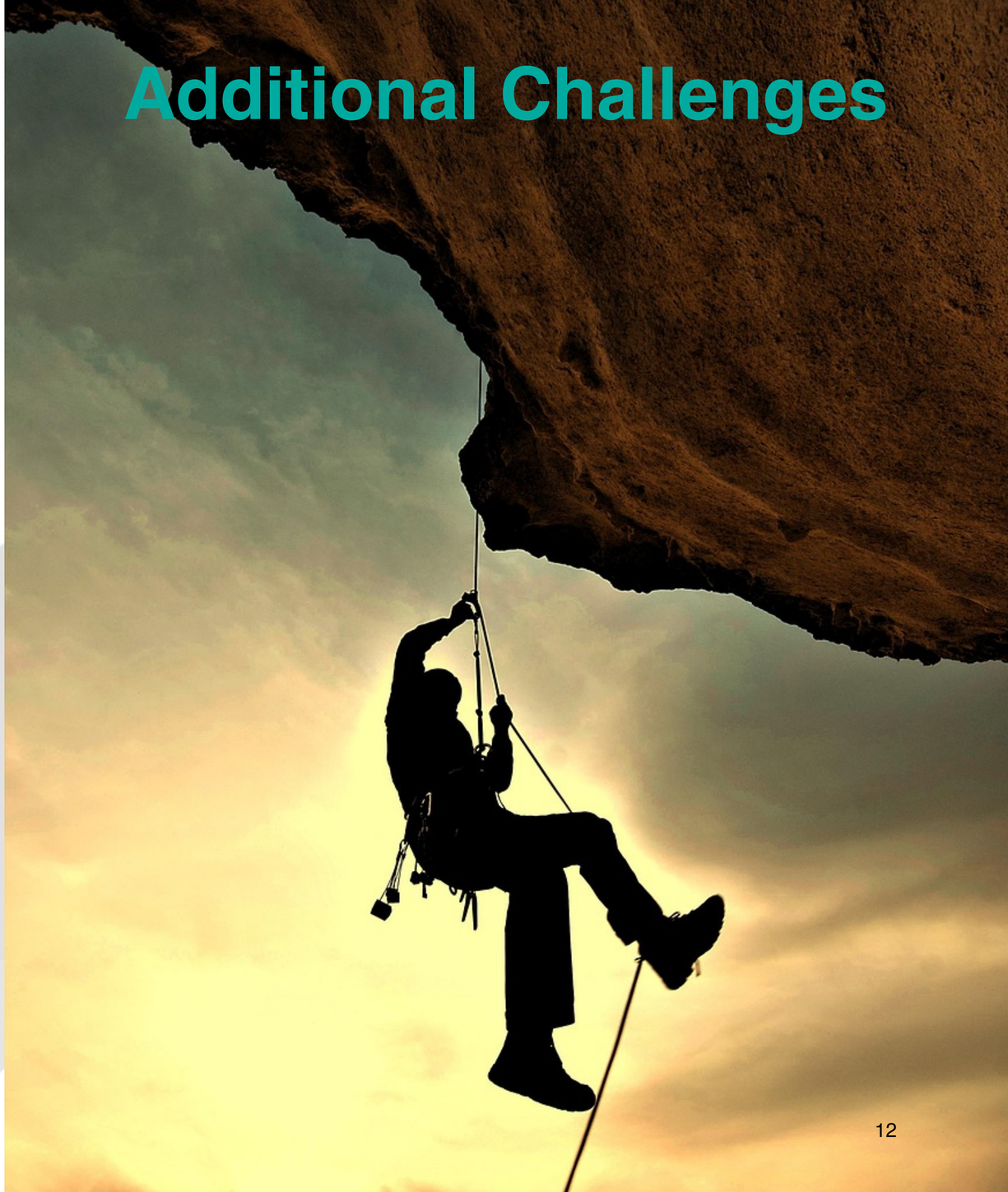
Challenging Data Preparation Actions

- Data Quality
- Binning & Grouping
- Tokenisation
- Standardisation
- Denormalisation
- Flattening
- …



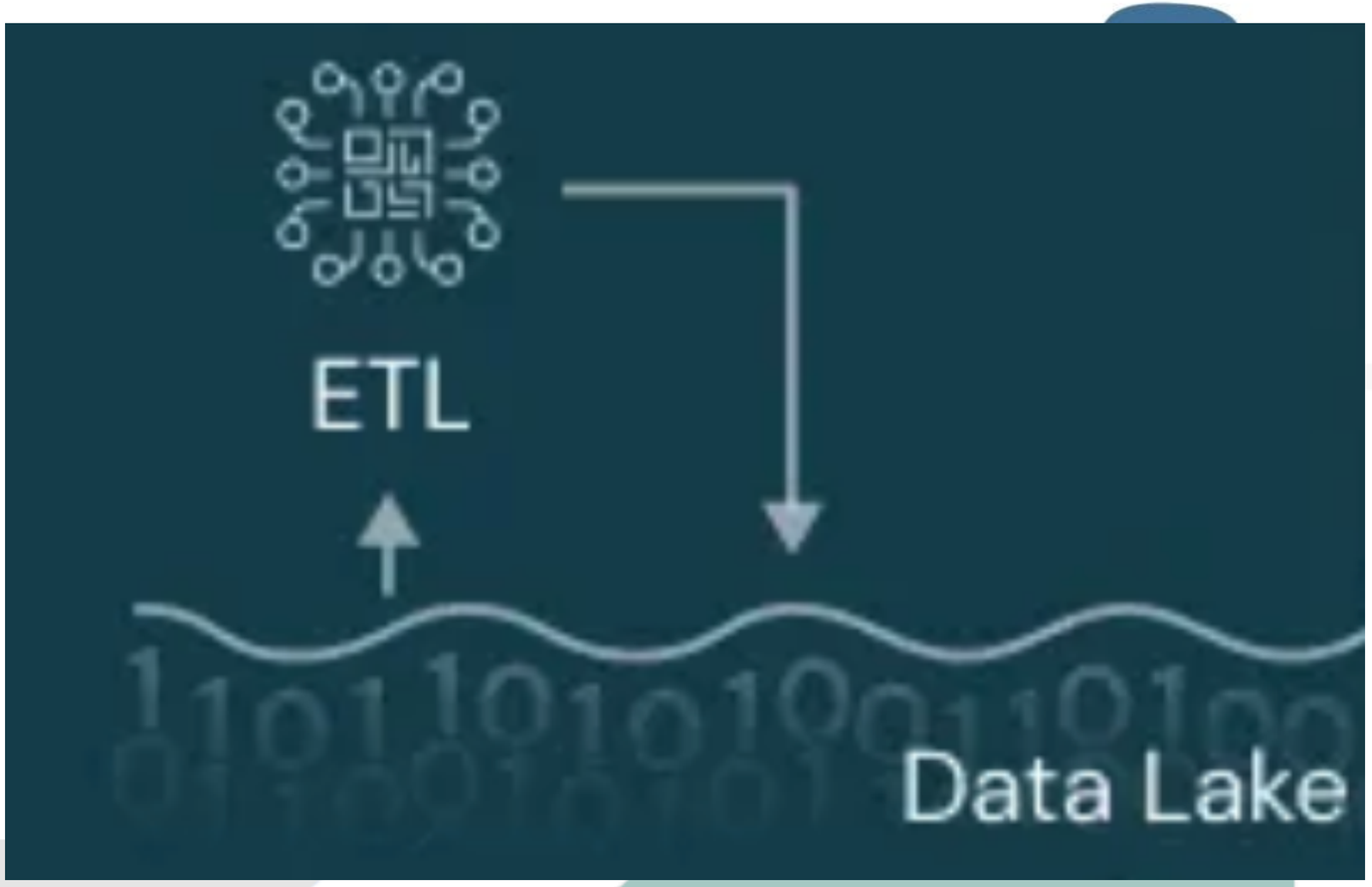
Additional Challenges

- Security
- Speed
- Development
- Cost
- Scalability
- Functionality
- …



Typical Data Scientist Approach

- code
  - in Python/R
  - using
    - public classes/libraries
    - data frames
- run
  - extract data from lake/database
  - execute on Spark cluster
  - materialise result in
    - data lake
    - data platform
    - database



Challenges With Common Approach

Security	hard to control
Speed	may be lower than DB with comparable HW
Development	efficient, skills available
Cost	potentially high, due to resource consumption
Scalability	depending on engine/ platform used
Functionality	rich
…	…

DB-centric Approach

- code
  - in SQL, PL/SQL
  - using
    - standard packages
    - database engine
    - custom code
- run
  - keep data in database
  - execute code inside database
  - materialise result in tables

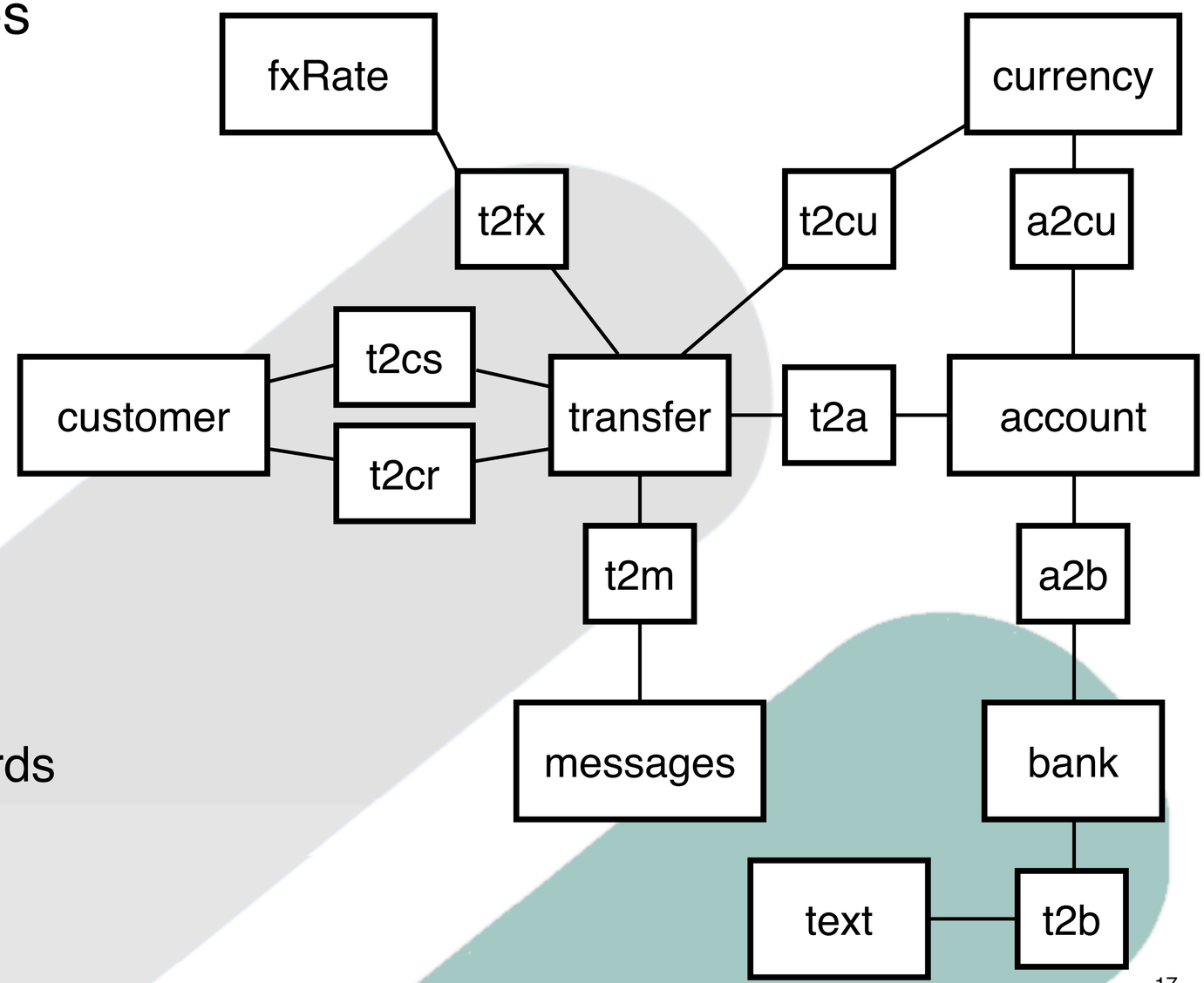
The Power of Oracle 26ai

Security	high, everything stays in DB context
Speed	high, due to efficiency of DB engine
Development	efficient, but available skills limited
Cost	bearable, due to platform efficiency
Scalability	high
Functionality	rich, PL/SQL, SQL, Javascript natively
…	…

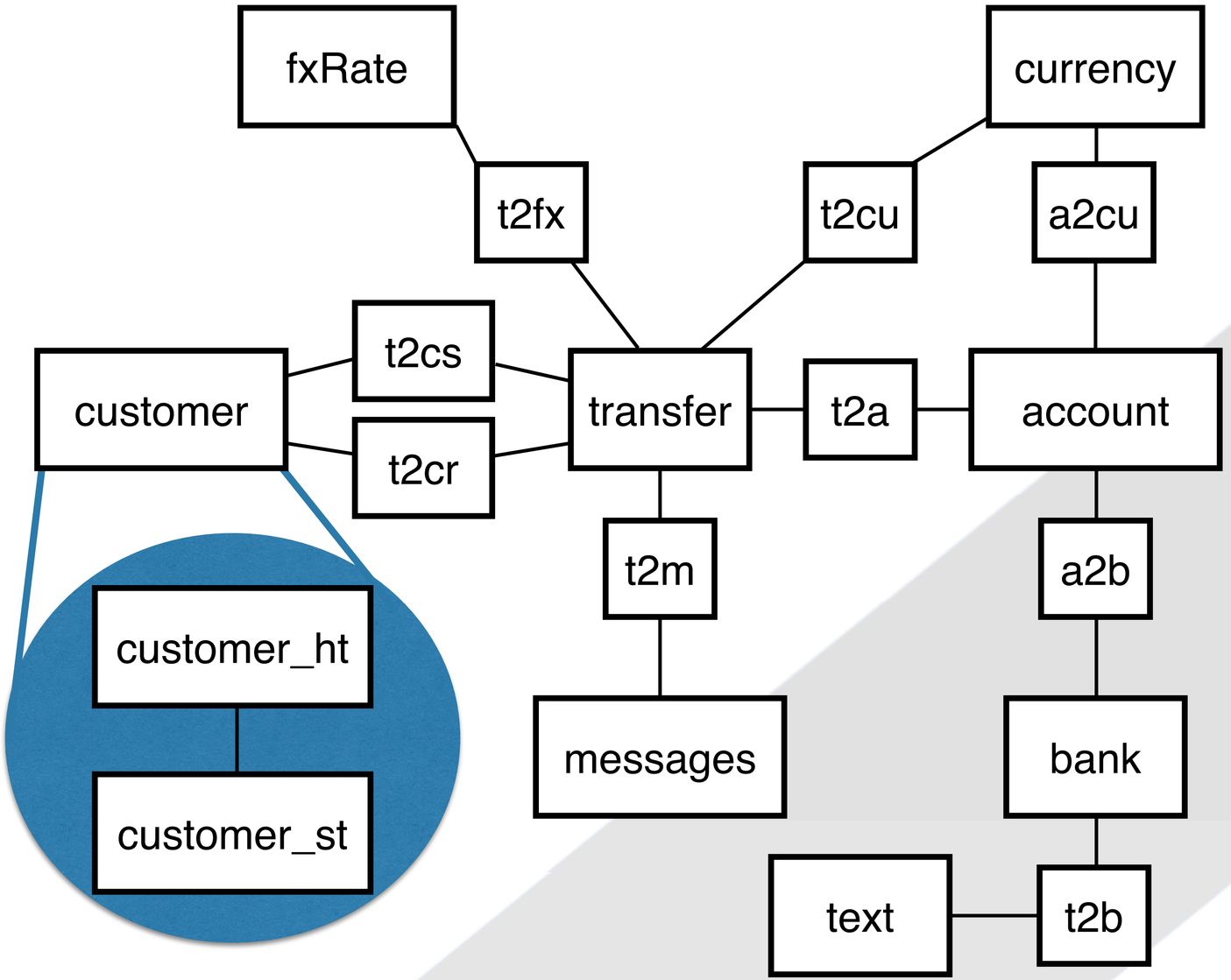
Efficient implementation is key!

Example Money Transfers — Scenario

- data normalised in multiple tables
- contains sensitive information
  - names
  - web addresses
  - phone numbers
  - …
- has to be prepared as input for classification/prediction model training
  - without sensitive information
  - filtered against list of permitted words
  - with securely encoded identifiers
  - …



Example Money Transfers — Tasks

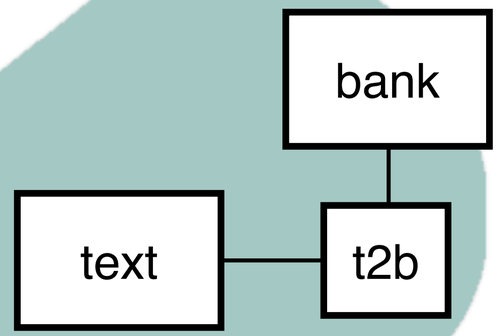


- Modularisation
- Denormalisation
- Filtering
- Standardisation
- Tokenisation
- White-Listing
- Anonymisation

Set-Based Modularisation & Denormalisation

- CTEs
- Joins
- (Un)Pivots
- …

```
with bank_filtered as (
    select b.id, b.uuid
         , cleanse_name_wrt_spec(
             coalesce(
               txt.namedef
             , txt.namede
             , txt.namefr
             , txt.textit
             , txt.texten)) name_cleansed
    from bank_st b
    left outer join b2t
      on (b.id = b2t.bank_id)
    left outer join text_st txt
      on (b2t.text_id = txt.id)
), prep_send_message as (
    select …
```



Row-Based Filtering & Standardisation

Convert all characters to lower case	`s = s.lower()`

Separate all emojis by whitespace	`s = re.sub(EMOJI_PATTERN, r' \1 ', s)`
	`EMOJI_PATTERN = re.compile(`
	`  "(["`
	`  "\U0001F1E0-\U0001F1FF"  # flags (iOS)`
	`  "\U0001F300-\U0001F5FF"  # symbols & pictographs`
	`  "\U0001F600-\U0001F64F"  # emoticons`
	`  "\U0001F680-\U0001F6FF"  # transport & map symbols`
	`  "\U0001F700-\U0001F77F"  # alchemical symbols`
	`  "\U0001F780-\U0001F7FF"  # Geometric Shapes Extended`
	`  "\U0001F800-\U0001F8FF"  # Supplemental Arrows-C`
	`  "\U0001F900-\U0001F9FF"  # Supplemental Symbols and`
	`  #  Pictographs`
	`  "\U0001FA00-\U0001FA6F"  # Chess Symbols`
	`  "\U0001FA70-\U0001FAFF"  # Symbols and Pictographs`
	`  #  Extended-A`
	`  "\U00002702-\U000027B0"  # Dingbats`
	`  "])"`
	`)`

Row-Based Filtering & Standardisation

Delete all white space separated strings that contain a dot or an @ sign	`s = re.sub(r'\S*[.@]\S*\s?', '', s)`
Delete all numbers	`s = re.sub(r'[0-9]', '', s)`

Delete all but emojis and letters	`s = re.sub(CLEAN_PATTERN, ' ', s)`
	`CLEAN_PATTERN = re.compile(`
	`"[^\U0001F600-\U0001F64F"`
	`"\U0001F1E0-\U0001F1FF" # flags (iOS)`
	`"\U0001F300-\U0001F5FF" # symbols & pictographs`
	`"\U0001F600-\U0001F64F" # emoticons`
	`"\U0001F680-\U0001F6FF" # transport & map symbols`
	`"\U0001F700-\U0001F77F" # alchemical symbols`
	`"\U0001F780-\U0001F7FF" # Geometric Shapes Extended`
	`"\U0001F800-\U0001F8FF" # Supplemental Arrows-C`
	`"\U0001F900-\U0001F9FF" # Supplemental Symbols and`
	`# Pictographs`
	`"\U0001FA00-\U0001FA6F" # Chess Symbols`
	`"\U0001FA70-\U0001FAFF" # Symbols and Pictographs`
	`# Extended-A`
	`"\U00002702-\U000027B0" # Dingbats`
	`"\U0001F600-\U0001F64F"`
	`" \w]"`

PL/SQL Based Filtering & Standardisation

```
create or replace function prep_str_pl (P_STRING IN VARCHAR2)
  return VARCHAR2
  is
  l_result VARCHAR2(4000);
begin
  l_result := regexp_replace(p_string, '[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,4}$', '');
  l_result := regexp_replace(l_result, '[0-9]', '');
  l_result := lower(l_result);
  l_result := regexp_replace(l_result
                             , '[^' || 'a-zäöüàáçèéêùû ' || unistr('\D83C\DC00') || '-' || unistr('\D83D\DEFF')
                                    || unistr('\2600')      || '-' || unistr('\27FF') ||']'
                             , '');
  l_result := regexp_replace(l_result
                             , '([' || unistr('\D83C\DC00') || '-' || unistr('\D83D\DEFF')
                                    || unistr('\2600')      || '-' || unistr('\27FF')
                                    ||  ']{1})'
                             , ' \1 ');

  return(l_result);
end;
/
```

Delete all white space separated strings that contain a dot or an @ sign

Delete all numbers

Convert all characters to lower case

Delete all but emojis and letters

Separate all emojis by whitespace

SQL Set Based Filtering & Standardisation

```
), prep_send_message as (
  select id, regexp_replace(
                  regexp_replace(
                     lower(
                         regexp_replace(
                             regexp_replace(m.snd_message
                                          , '[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,4}$'
                                          , '')
                           , '[0-9]'
                           , '')
                          )
                    , '[^' || 'a-zäöüàáçèéêùû ' || unistr('\D83C\DC00') || '-' || unistr('\D83D\DEFF')
                                                || unistr('\2600')      || '-' || unistr('\27FF') ||']'
                    , '')
                , '([' || unistr('\D83C\DC00') || '-' || unistr('\D83D\DEFF')
                       || unistr('\2600')      || '-' || unistr('\27FF')
                       ||  ']{1})'
                , ' \1 ') as snd_message
   from messages m
), wl_send_message as (
```

Delete all white space separated strings that contain a dot or an @ sign

Convert all characters to lower case

Delete all numbers

Delete all but emojis and letters

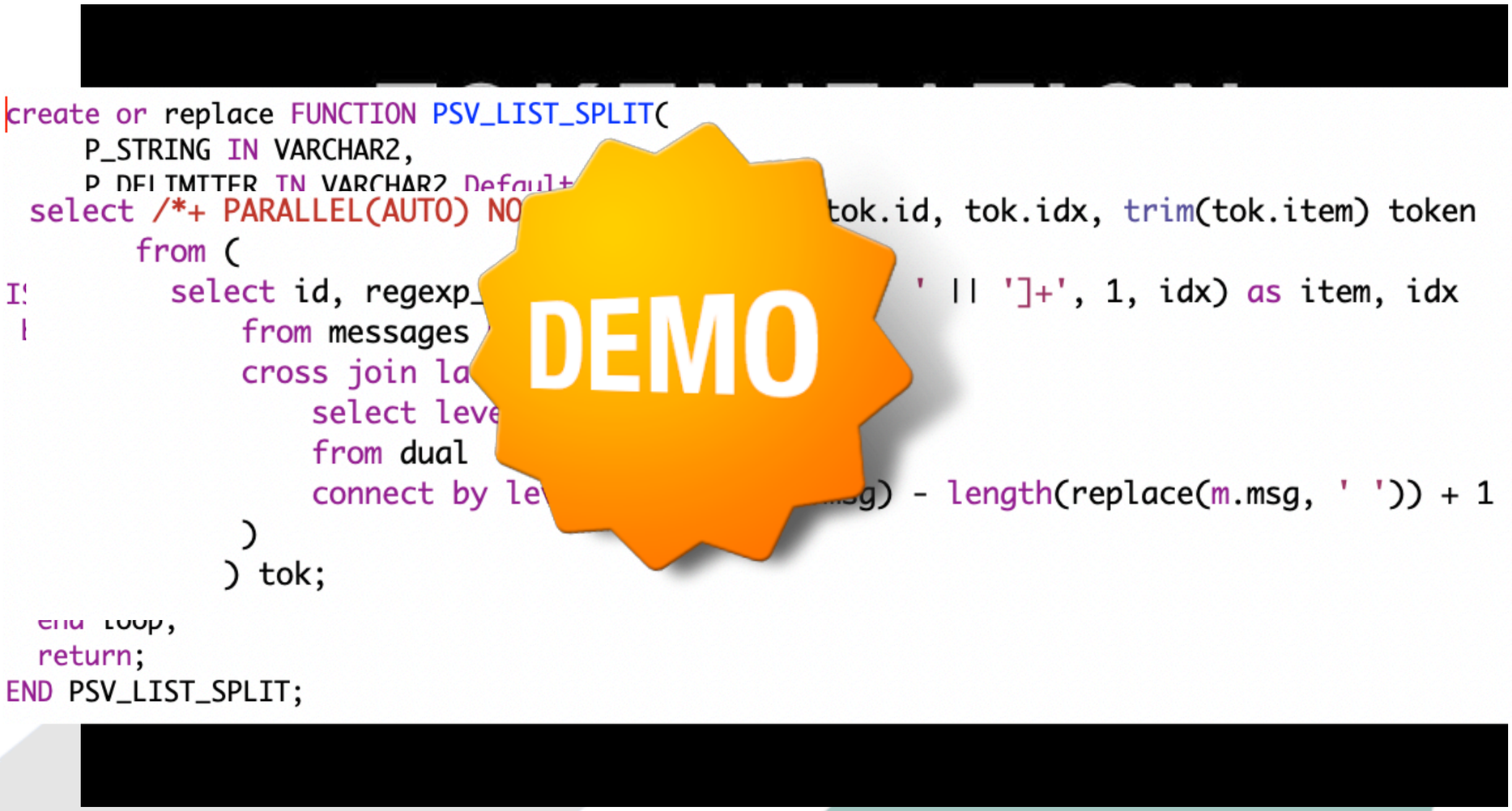
Separate all emojis by whitespace

PL/SQL vs. SQL vs. Macros For Filtering & Standardisation

- cascade of regexp & text actions
- single row function calls in PL/SQL or Python
- set-based processing in SQL
- single definition & maintenance with SQL Macros

PL/SQL & SQL Tokenisation

- splitting a series of values in a set of single values w.r.t. a given delimiter
- In Python it's a transparent function call
- table function call in PL/SQL
- pure SQL using cross lateral join is more efficient



PL/SQL & SQL White Listing

- reduce a stream or set of tokens to instances of approved tokens
- keep token order
- deliver at scale
- function calling tokenisation table function in PL/SQL
- faster is pure SQL using SQL-tokenisation and inner join against white list
- SQL Macros allow reuse, improve readability & code quality

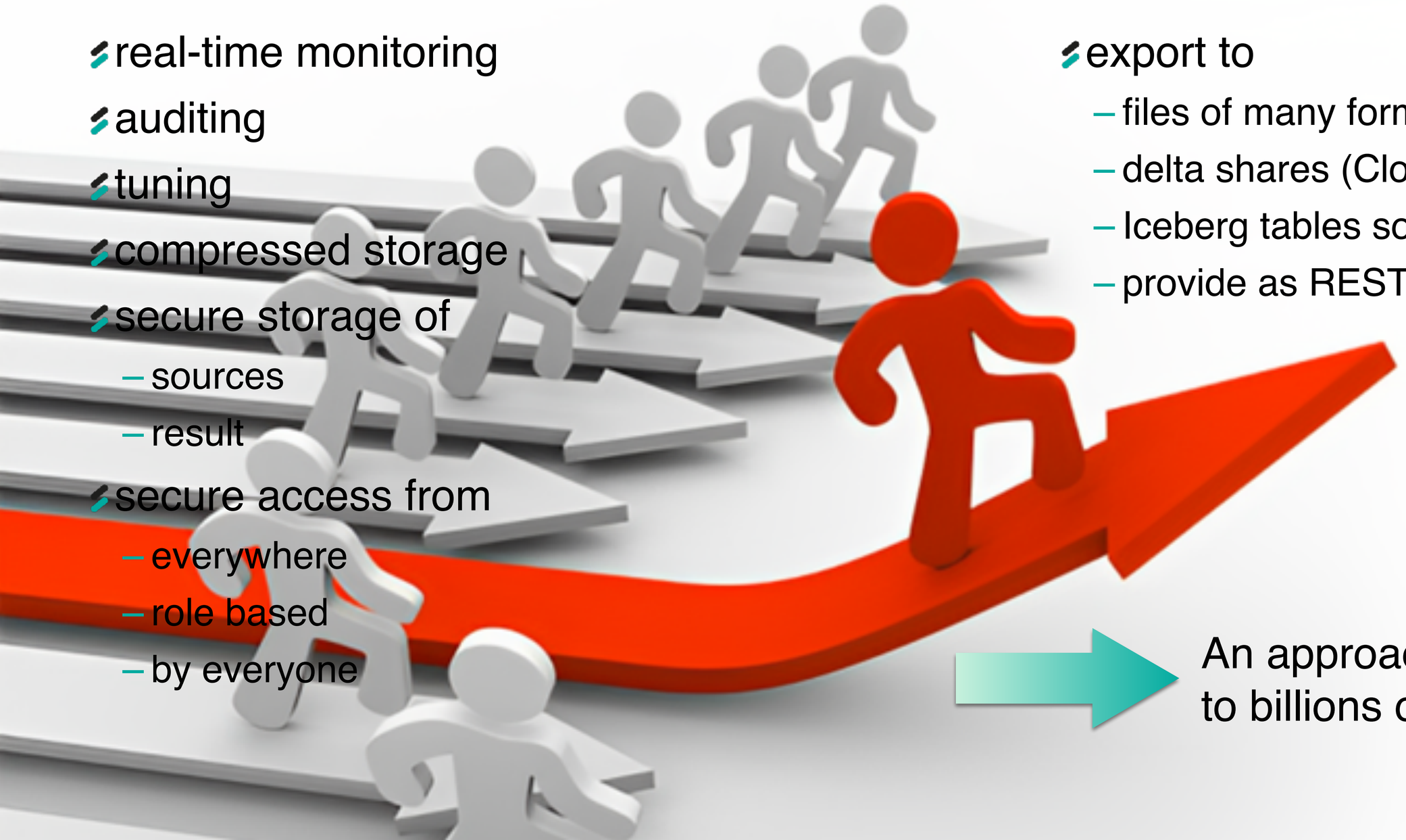
```
-- function that
-- + tokenizes each message
with messages as (
    select id, msg
    from (VALUES
           (1, 'a b c d')
         , (2, '1 2 3 900')
         ) a (id, msg)
)
    select /*+ PARALLEL(AUT
      m.id, substr(listag                                    te) within group (order by m.idx),0, 4000) message
    from (
      select
         tok.id, tok.idx
      from (
        select id, regexp_                                   ]+', 1, idx) as item, idx
            from messages
            cross join latera
                select level i
                from dual
                connect by level <= le    h(m.msg) - length(replace(m.msg, ' ')) + 1
            )
           ) tok
      inner join export_whitelist_t wl
        on (tok.item = wl.wl_item)
       ) m
    group by m.id;
    id, filter_message_wrt_whitelist(msg)
from messages;
```

DEMO

Additional Advantages of 26ai

- real-time monitoring
- auditing
- tuning
- compressed storage
- secure storage of
  - sources
  - result
- secure access from
  - everywhere
  - role based
  - by everyone

- export to
  - files of many formats
  - delta shares (Cloud)
  - Iceberg tables soon (Cloud)
  - provide as REST API



An approach that scales to billions of records!

Conclusions

- PL/SQL and SQL may be a bit out of fashion, but they are still great or even superior
- Database 26ai provides extremely powerful performance & engineering features
  - lateral Joins
  - table functions
  - SQL macros
  - and many more
- Autonomous AI Database provides additional performance advantages through
  - flexible resource scaling
  - delta share and Iceberg integration
  - underlying Exadata architecture

Give it a try!

Past & Future TechCasts:

July 23rd

Temporal Data Lineage and Impact Analysis in Oracle Database: Evolution to Oracle Graph Architecture with Time-Travel Capabilities

Presented by **Karin Patenge & Tomislav Zarkovic**

August 20th

How and Where Oracle 26ai Beats Python for Data Preparation

Presented by **Dr. Holger Freidrich**

September 17th

Graph Analytics with Oracle AI Database 26ai and Linkurious Ogma

Presented by **Leo Nicolle & Karin Patenge**

TechCast Archive

2026 | 2025 | 2024 | 2023 | 2022 | 2021

Date	Title	Presenter(s)	Replay	Download(s)
Jun 25	Our Favorite Features of OAC: March & May 2026 Releases	Gautam Pisharam & Carter Beaton	Video	Slides
Apr 30	Pythian's AI Journey with UMass: From AI Assistant to AI Agents!	Jon Mead & Ambili Thottathil, Pythian	Video	Slides
Apr 2	Build Practical AI Prototypes with Oracle Analytics (Even with a Small Team)	Dennis Gray & Adam Hume, Definian	Video	Slides
Mar 5	Fraud Stops with Oracle Database 26ai Vector Search + OML	Dr. Abigail Giles-Haigh	Video	Slides
Feb 5	Our Favorite Features of OAC: November 2025 & January 2026 Releases	Branden Pavol, Gautam Pisharam, & Carter Beaton	Video	Slides

Submit a topic to share at **https://andouc.org/techcasts/**

Analytics and Data SUMMIT 2027

APRIL 13-15, 2027

SAVE THE DATE!
April 13-15, 2027

Call for Speakers is OPEN!!

https://andouc.org/analytics-and-data-summit-2027/

ANALYTICS AND DATA TechCasts

Helpful Links –

ORACLE ANALYTICS VIDEOS:
https://www.youtube.com/@OracleAnalytics/videos

OAC MARCH AND MAY 2026 NEW FEATURES VIDEOS BY ORACLE:

https://bit.ly/OACMar26Features
https://bit.ly/OACMay26Features

OAC NEW FEATURES DOCUMENTATION BY ORACLE:
https://docs.oracle.com/en/cloud/paas/analytics-cloud/acswn/index.html#GUID-CFF90F44-BCEB-49EE-B40B-8D040F02D476

ORACLE ANALYTICS COMMUNITY:
https://community.oracle.com/products/oracleanalytics

ORACLE ANALYTICS LIBRARY/EXAMPLES:
https://www.oracle.com/business-analytics/data-visualization/examples/

ORACLE ANALYTICS LIVE DEMOS:
https://www.oracle.com/business-analytics/data-visualization/demos/

ANALYTICS AND DATA TechCasts