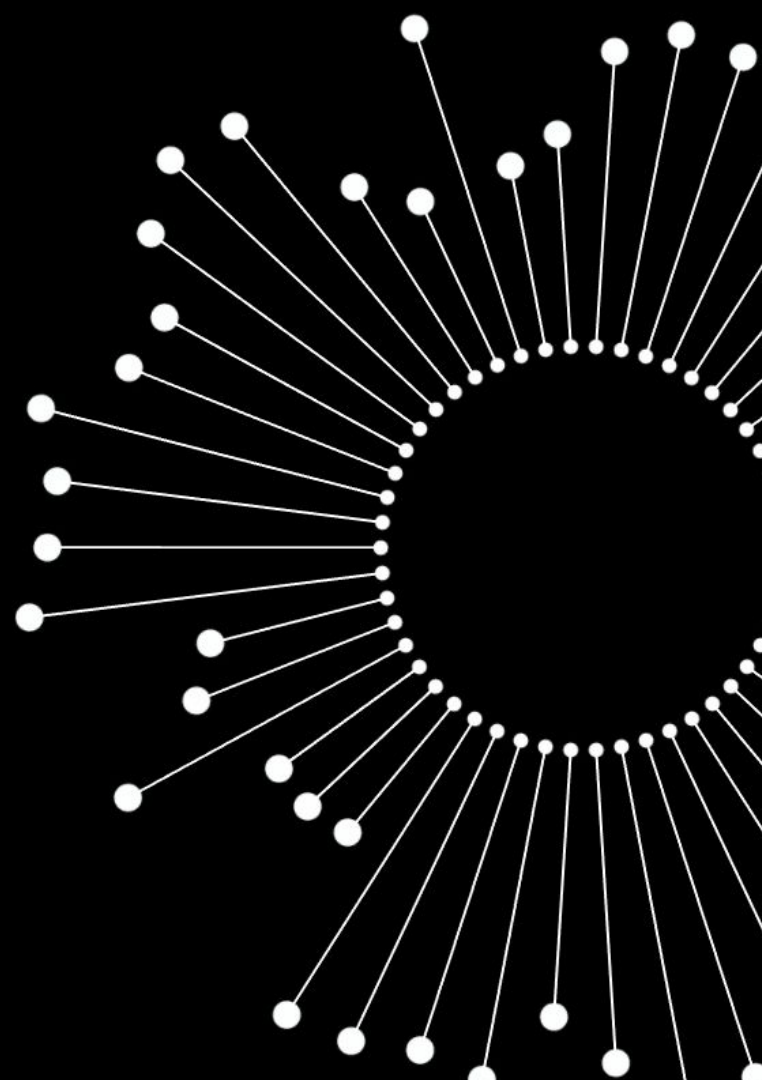# XES2.0

Process insights delivered.
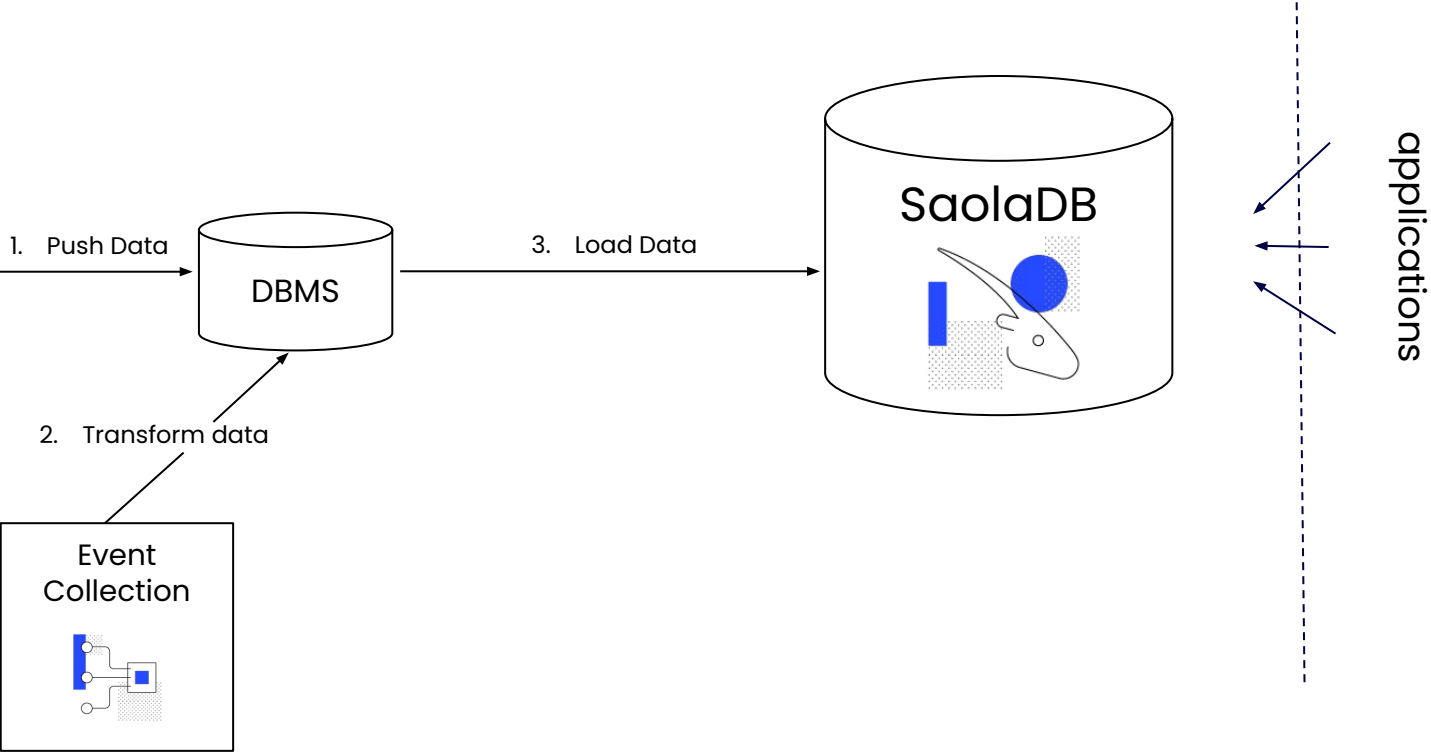
Philipp Hoch, Product Lead, Celonis SE
02.11.2021

Celonis | ETL pipeline
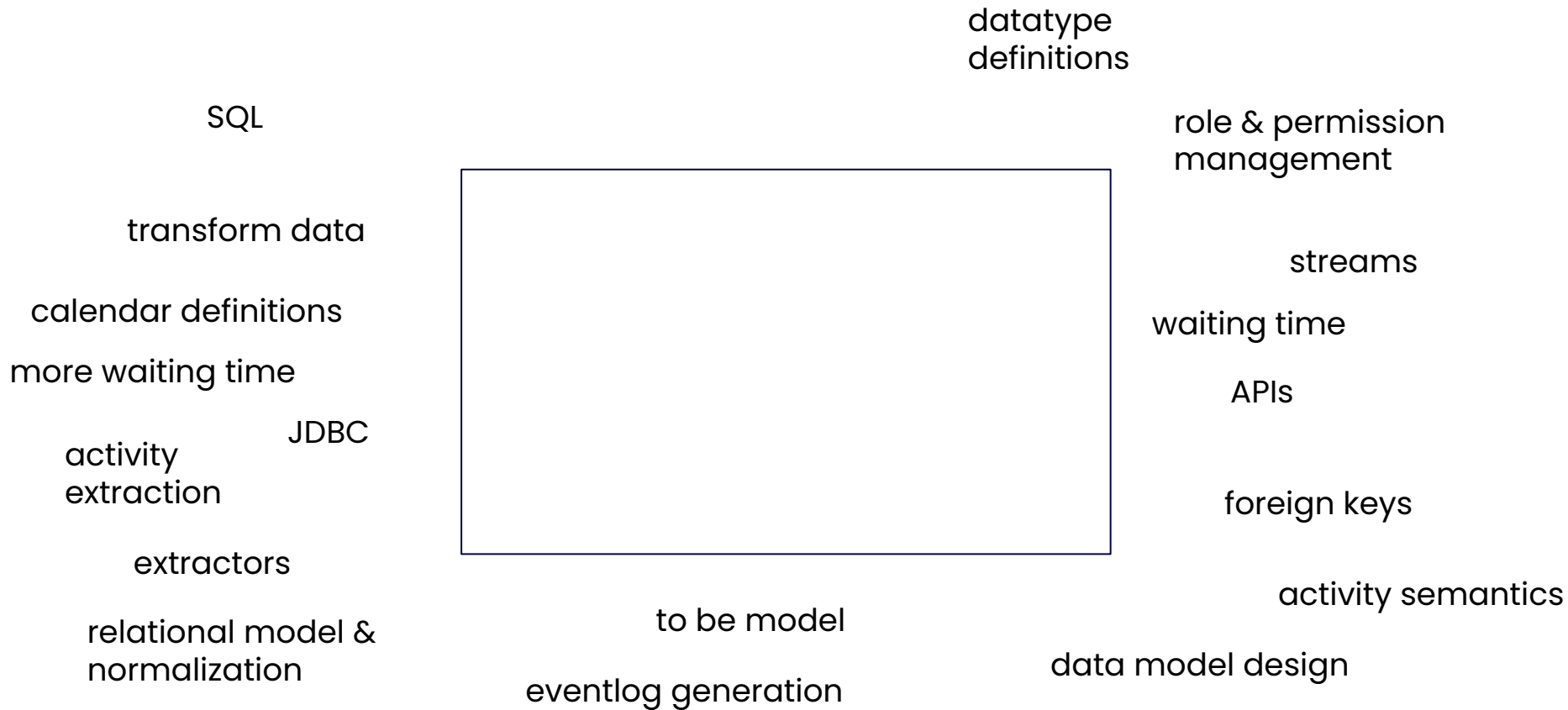
I want to generate insights

I want to generate insights

datatype
definitions

SQL

role & permission
management

transform data

streams

calendar definitions

waiting time

more waiting time

JDBC

APIs

activity
extraction

foreign keys

extractors

activity semantics

relational model &
normalization

to be model

data model design

eventlog generation

I want to generate insights

Empty View

Toggle Edit Mode on to start creating

And if you made it till here.

Congrats!

You have an empty dashboard.

I want to generate insights
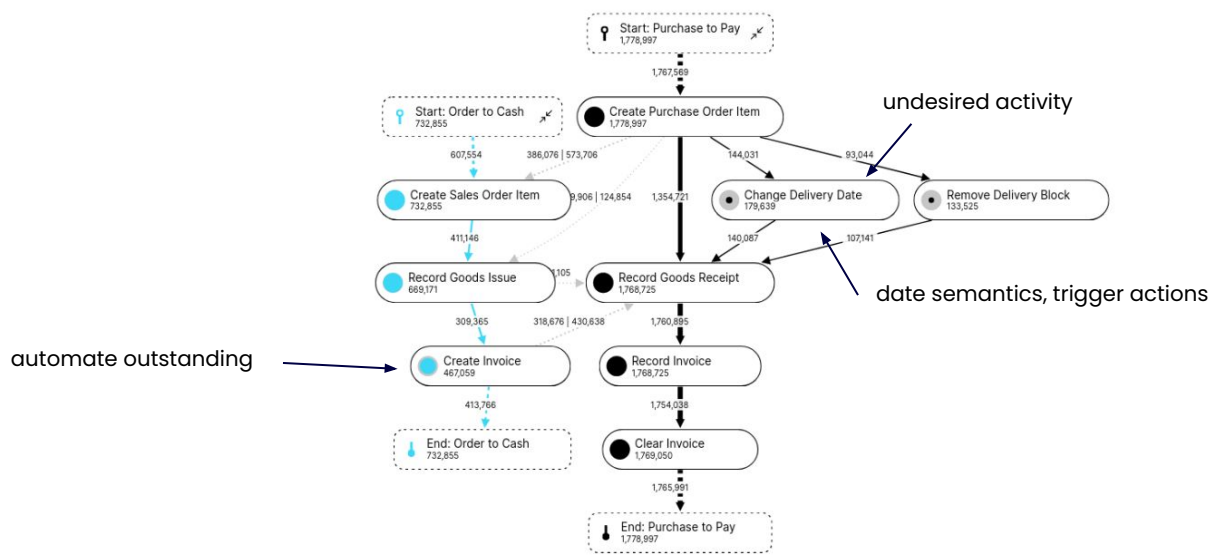
XES document

- establish data connection
- extract data
- create data model
- utilize activity semantic
- transfer BPMN model
- connect to existing processes/event logs
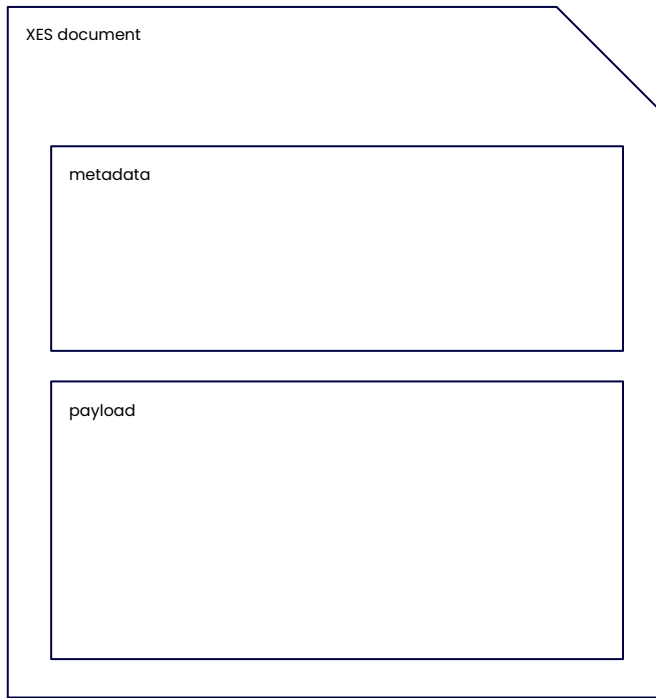
I want to generate insights

KPIs

Customer OTD	Singapore OTD	SO Items	PO Items	Create POI to Create SOI	Create SOI to Inv
47%	40%	733K	1.78M	1 d	7 d

Purchase to Pay

Start: Purchase to Pay
1,778,997

1,767,569

Start: Order to Cash
732,855

Create Purchase Order Item
1,778,997

undesired activity

607,554   386,076 | 573,706   144,031   93,044

Create Sales Order Item
732,855

9,906 | 124,854   1,354,721

Change Delivery Date
179,639

Remove Delivery Block
133,525

411,146

140,087   107,141

Record Goods Issue
669,171

1,105

Record Goods Receipt
1,768,725

date semantics, trigger actions

309,365   318,676 | 430,638   1,760,895

automate outstanding

Create Invoice
467,059

Record Invoice
1,768,725

413,766

1,754,038

End: Order to Cash
732,855

Clear Invoice
1,769,050

1,765,991

End: Purchase to Pay
1,778,997
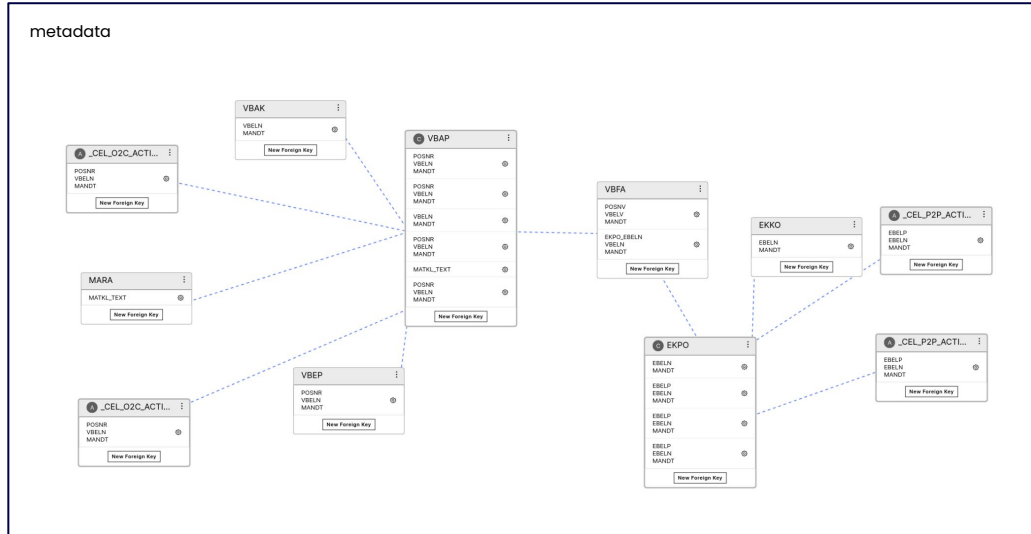
75%

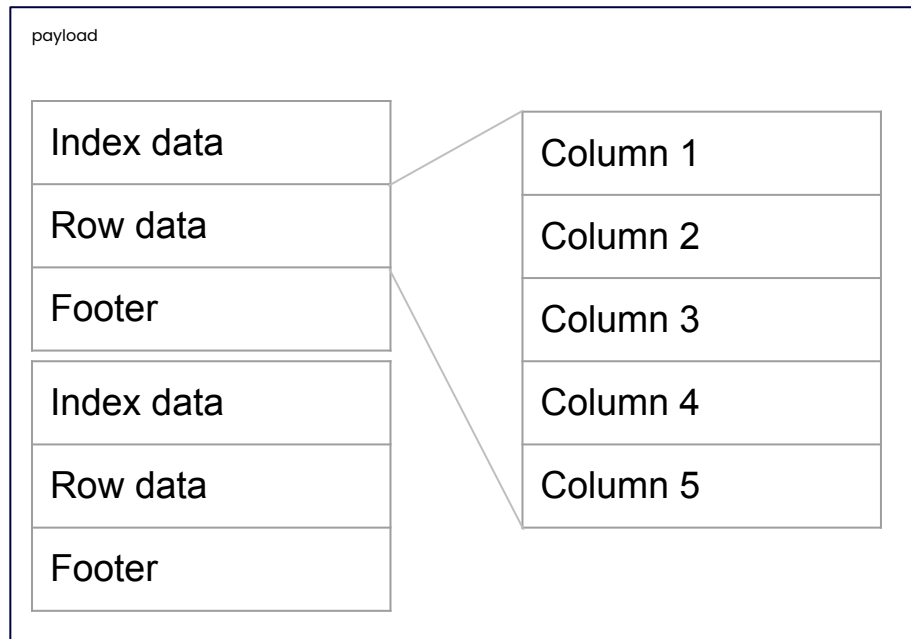# The document

XES document

metadata

payload

Metadata

metadata

- tables and relationships
- event log definitions
- activity type semantics
- BPMN model
- rules and actions
- multi event log connections
- domain specific elements

⟹ extensibility

Payload

(C)

payload	
Index data	Column 1
Row data	Column 2
Footer	Column 3
Index data	Column 4
Row data	Column 5
Footer	

e.g. ORC, parquet

- column oriented
- data compression
  - dictionary encoding
  - bit packing
  - run length encoding
- predicate push down
- limited IO operations
- self-describing
- built-in index
- built-in aggregates
- industry standard
- built-in encryption
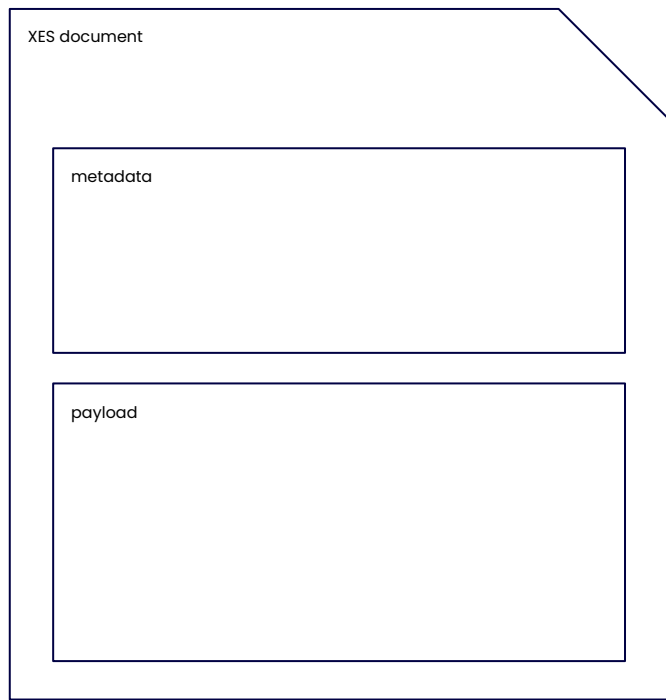- made for "big data"

Payload

C

payload	
JDBC Driver	oracle.jdbc.OracleDriver
JDBC URL	jdbc:oracle:thin:@32.12.31.76:1521:orclpdb1

exemplary

- APIs
- streams
- databases
- web services

⟹ any kind of data sources

The document

c

XES document

metadata

payload

- encryption
- profiles/roles
- confidentiality
- extensibility
- scalability
- speed

Thank you.

File Size Comparison Across Encoding Methods
Dataset: TPC-DS Scale 500 Dataset

585 GB
(Original Size)

505 GB
(14% Smaller)

Impala
221 GB
(62% Smaller)

Hive 12
131 GB
(78% Smaller)

- Larger Block Sizes
- Columnar format arranges columns adjacent within the file for compression & fast access

Encoded with
Text

Encoded with
RCFile

Encoded with
Parquet

Encoded with
ORCFile

source: community.cloudera.com