



# DATA QUALITY: A PROBLEM IN THE REAL WORLD

ALI NOROOZIAN

BUSINESS DEVELOPMENT MANAGER, BEHFALAB

00.12.12





## EDUCATION:

Sharif University

Iran University of Science and Technology

## BEHFALAB EXPERIENCE:

Product Development

Project Management

International Affairs

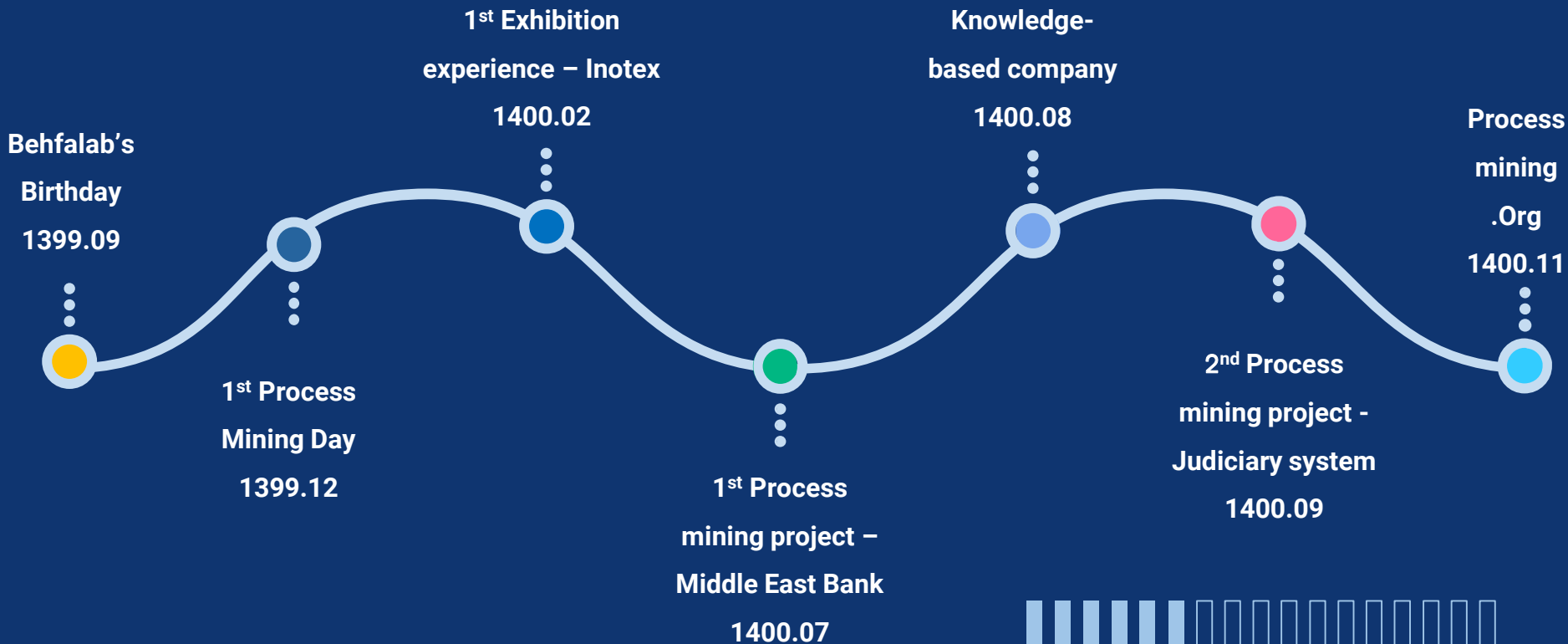




Spin-off of Behfa



Incubator of the Iran University of Science and Technology



# AGENDA



How to Experience a Good  
Process Mining Project?



ETL Challenges



Connectors



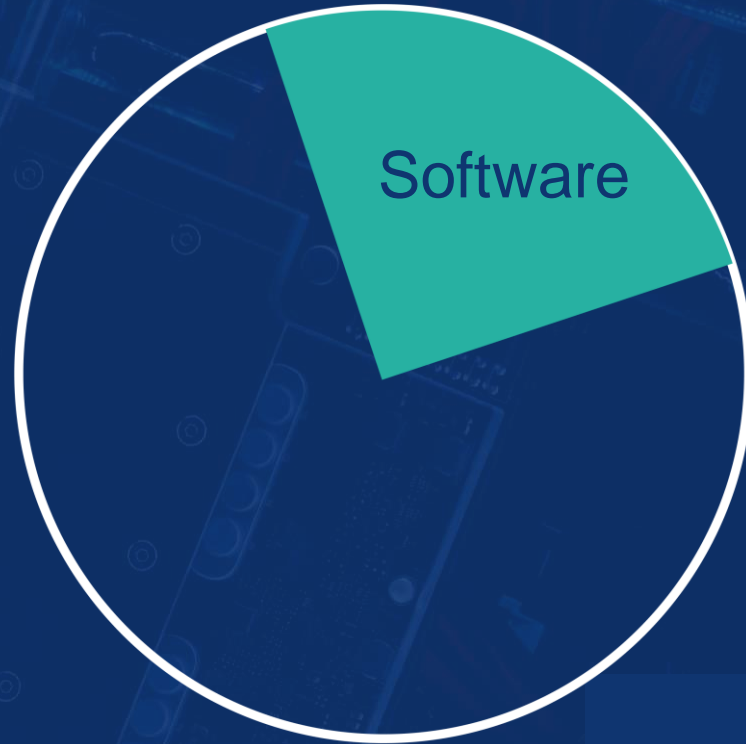
Data Quality Issues



Q&A



# HOW TO EXPERIENCE A GOOD PROCESS MINING PROJECT?

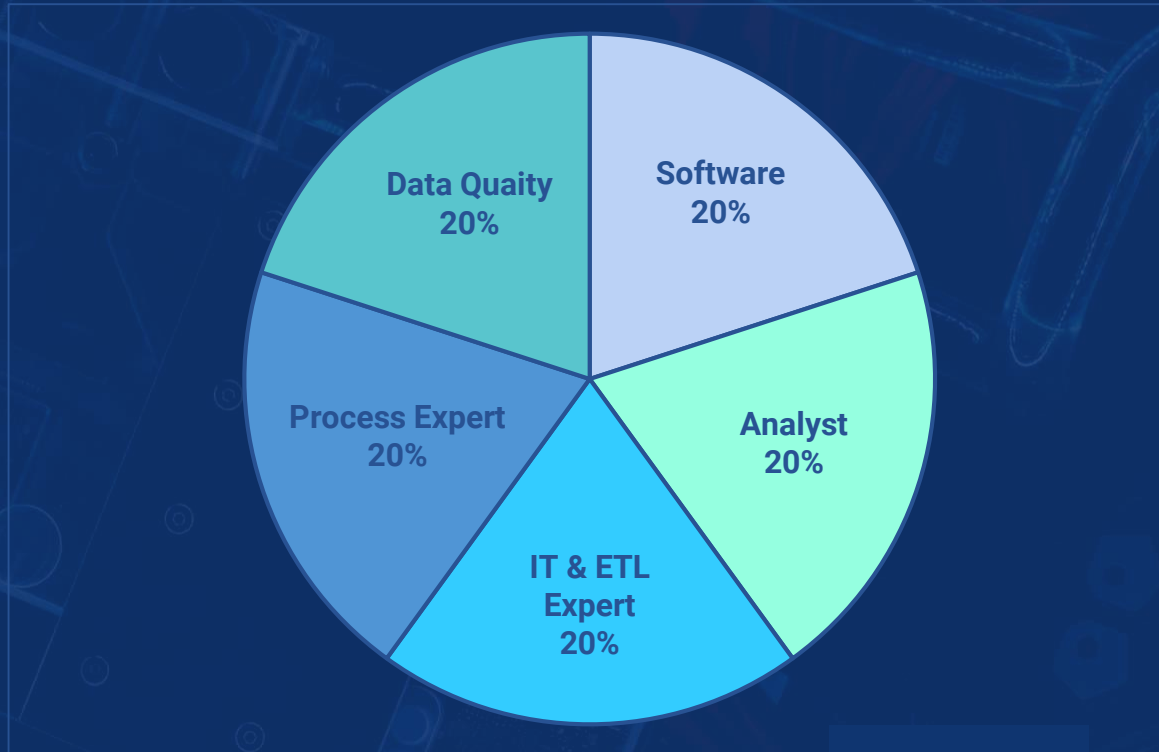




# Software



# HOW TO EXPERIENCE A GOOD PROCESS MINING PROJECT?





# PROCESS MINING PROJECT



Interview with  
process owners

Interview with IT  
experts

Product  
customization

Installing on the  
organization's  
server

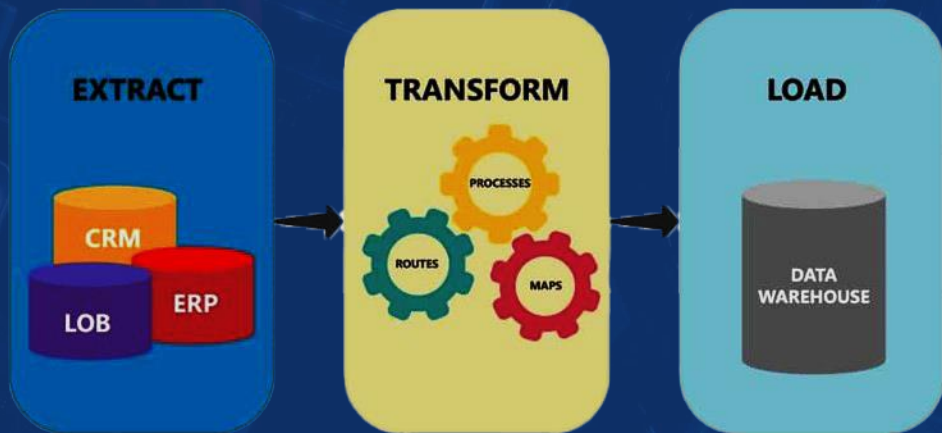
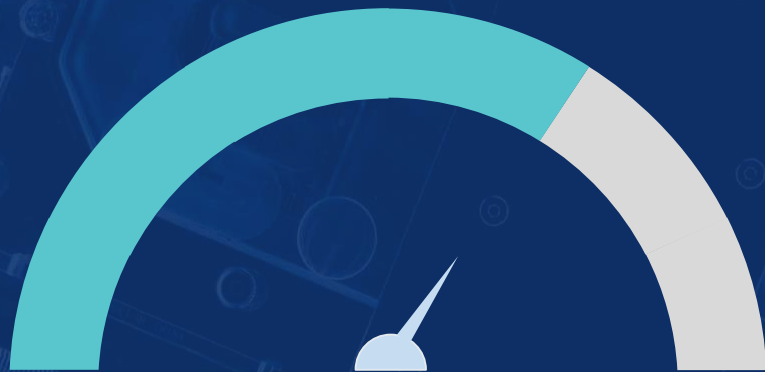




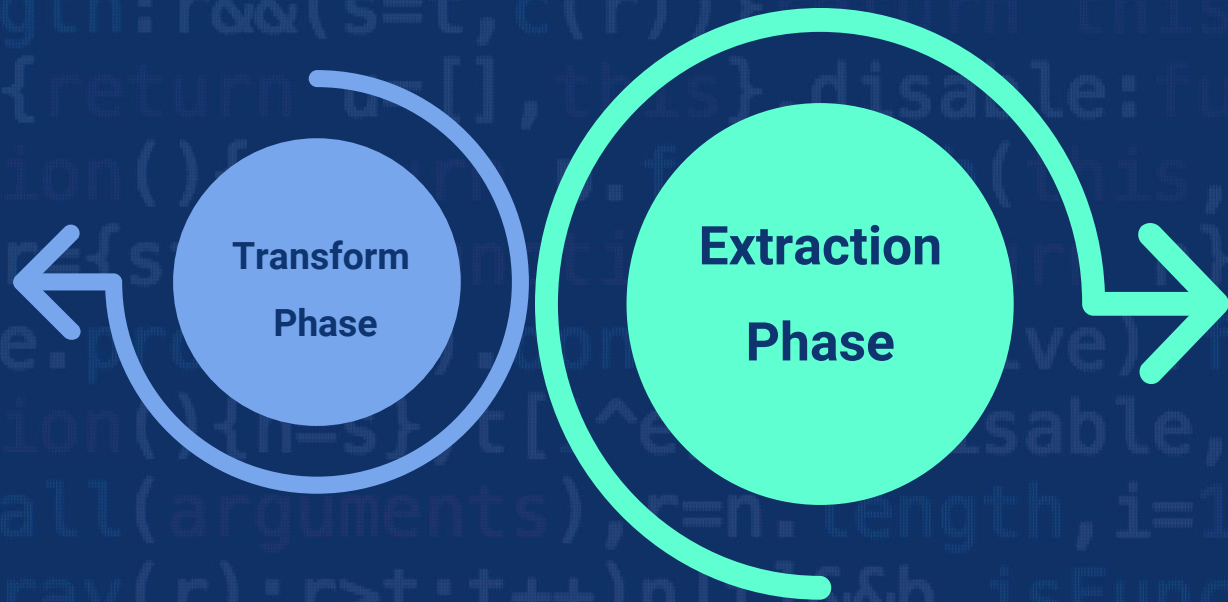
# ETL



60%



# ETL PROBLEMS



# CONNECTORS

Utilize Data That's  
Spread Across  
Disconnected Systems  
And Applications



Microsoft  
Dynamics

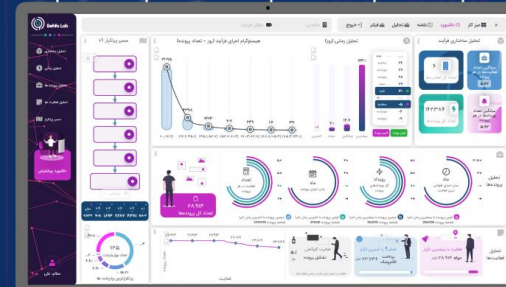


ORACLE

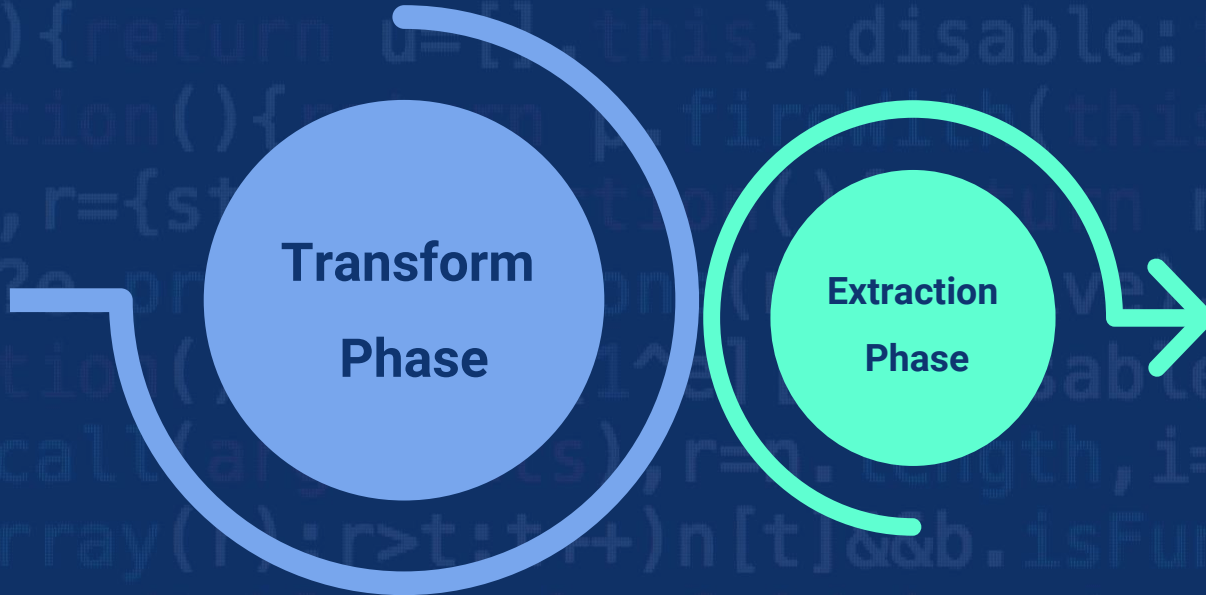
READSOFT

salesforce

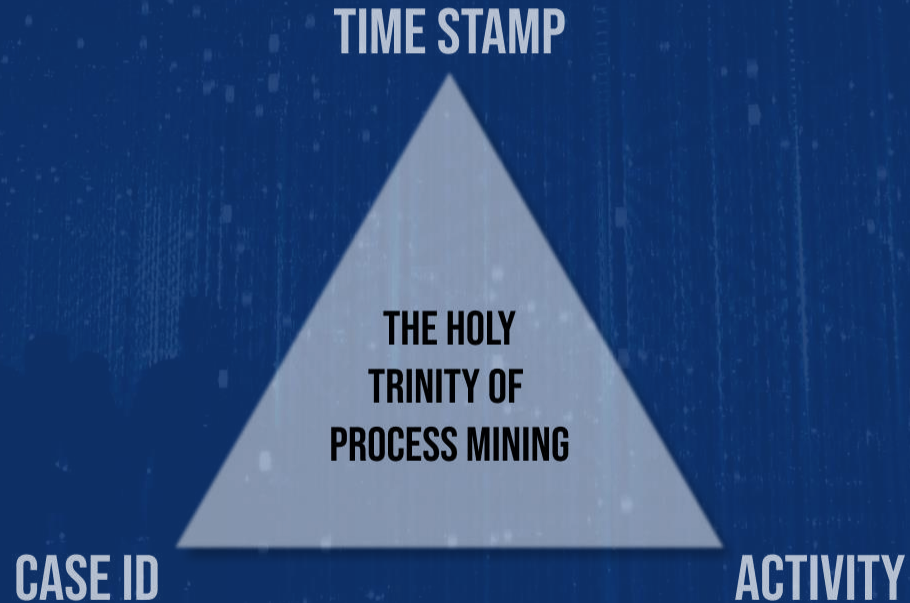
SAP



# ETL PROBLEMS



# KEY FACTORS IN PROCESS MINING



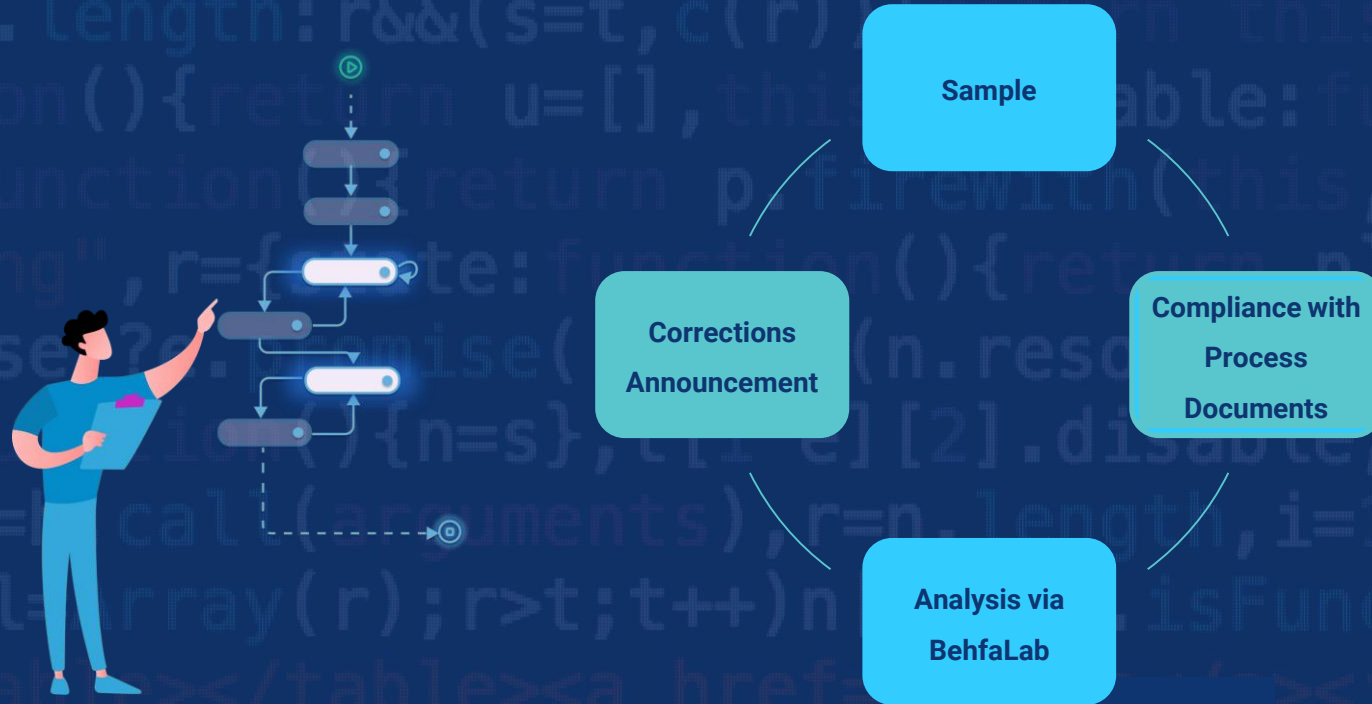




Case ID	Timestamp	Activity	Resource	Loan goal	Requested amt	Offered amt
C001	18-10-2016	Check completeness	Sue	Mortgage	100 000	-
C001	19-10-2016	Check credit history	Sue	Mortgage	100 000	-
C001	19-10-2016	Calculate risk score	Bob	Mortgage	100 000	-
C001	20-10-2016	Make offer	Mike	Mortgage	100 000	70 000
C001	25-10-2016	Make offer	Mike	Mortgage	100 000	80 000
C002	20-10-2016	Check completeness	Sue	Car	15 000	-
C002	20-10-2016	Check credit history	Sue	Car	15 000	-
C002	22-10-2016	Calculate risk score	Elsa	Car	15 000	-
C002	24-10-2016	Reject application	Elsa	Car	15 000	-
C003	02-11-2016	Check completeness	Maria	Mortgage	30 000	-
C003	04-11-2016	Ask for additional data	Maria	Mortgage	30 000	-
C003	10-11-2016	Check credit history	Maria	Mortgage	30 000	-
...	...	...	...	...	...	...



# Data Quality Improvement Cycle





# DATA QUALITY PROBLEMS IN THE REAL PROJECTS



01



Activities Recording

Naming Issues



02

03



Null Data



# DATA QUALITY PROBLEMS IN THE REAL PROJECTS



04



Timestamp Scope

Same Timestamp



05

06



Recording Activity as an Attribute



# CORRECTIONS ANNOUNCEMENT



Activities Naming

Process Transparency

Find New Activities

01

02

03

04

05

06

Distinguishing Between  
Resources

Database Development

Attribute Selection





# Thanks For Attention



Ali.Noroozian@Behfalab.com



