

# Pharmaceutical Manufacturing Process Anomalies Detection And Diagnosis Using Machine Learning

**Marco Vicentini**

**Analyst, IDS Digital Solutions  
And Data Engineering**

**Eli Lilly Italia**

**Giuseppe Salerno**

**Sr Associate, IDS Digital Solutions  
And Data Engineering**

**Eli Lilly Italia**



**connect**

# Lilly Fast Facts

A heritage 148 years strong, founded on May 10th, 1876



Headquarters located in **Indianapolis, Indiana, U.S.A.**



More than **42,000 employees** worldwide



Approximately **10,000 employees** engaged in research and development



Clinical research conducted in more than **55 countries**



Research and development facilities located in **7 countries**



Manufacturing plants located in **9 countries**



Products marketed in approximately **110 countries**



# Manufacturing Processes Anomalies Detection & Diagnosis In a Nutshell

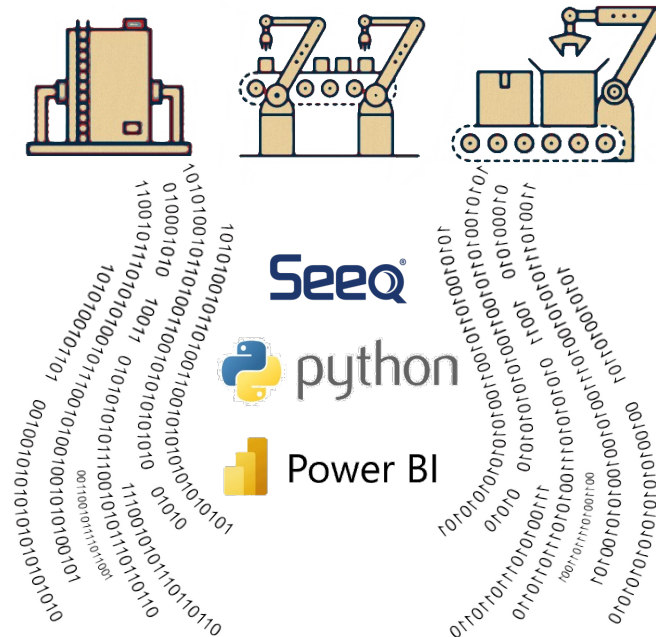
## WHAT

*Progress from Descriptive to Predictive Machines Performance Monitoring*



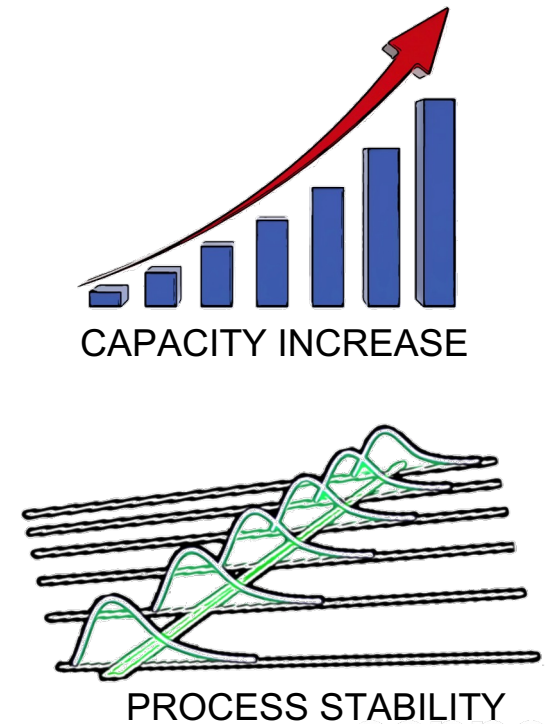
## HOW

*Embedding Advanced Analytics in Machine Performance Data Analysis*



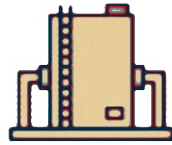
## WHY

*To Support Reliable Supply*

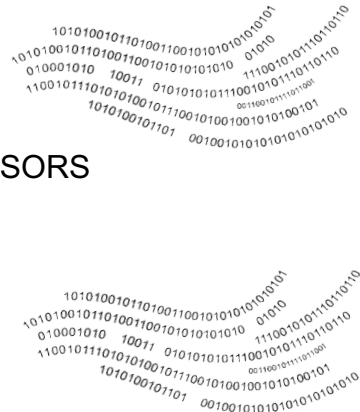


# Manufacturing Processes: Phases and Variables

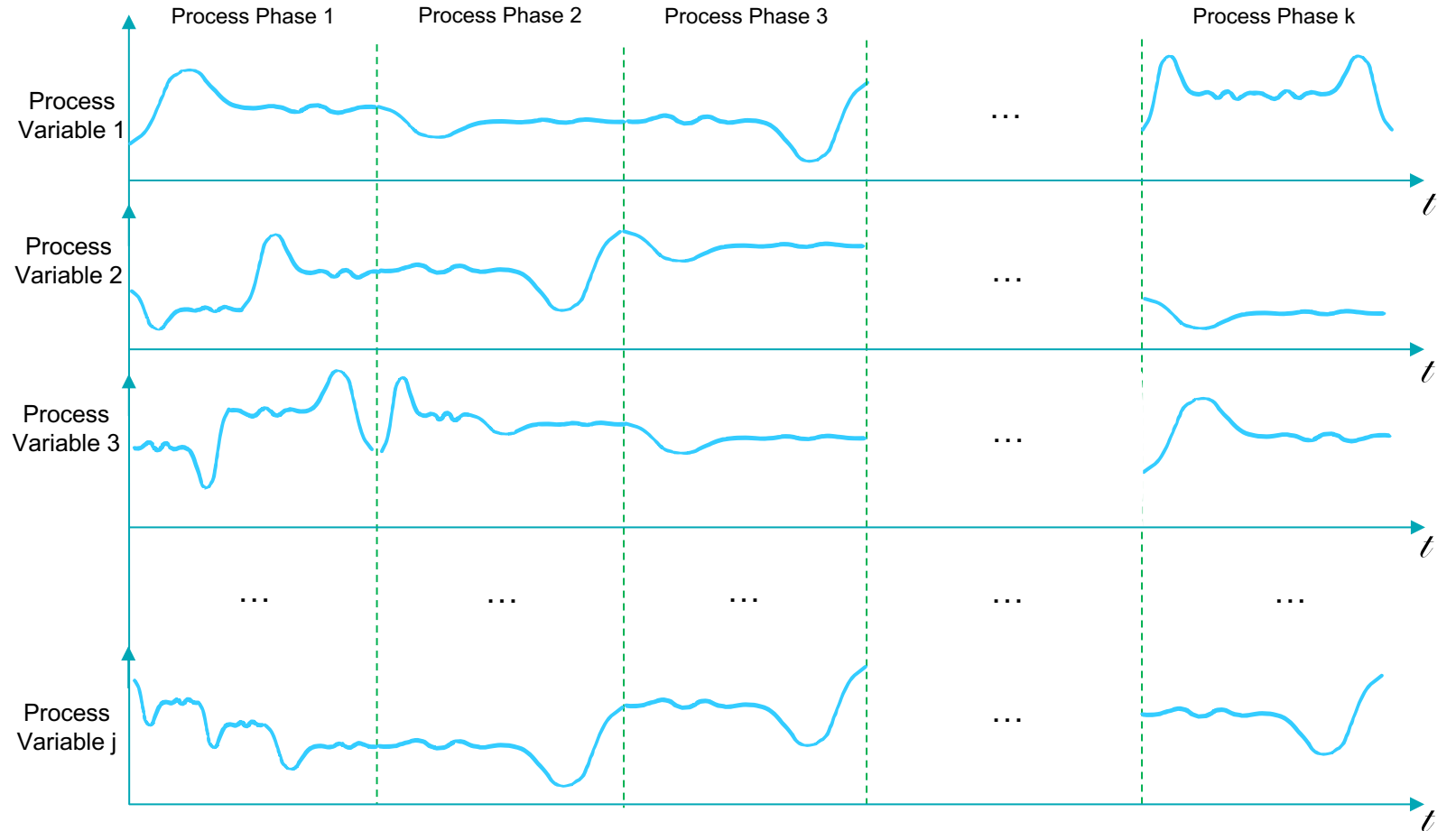
PROCESS MANUFACTURING



- CIP
- SIP
- AUTOCLAVES
- COMPONENT PROCESSORS
- PARTS WASHERS
- VPHP
- FORMULATION
- FILTRATION
- ...



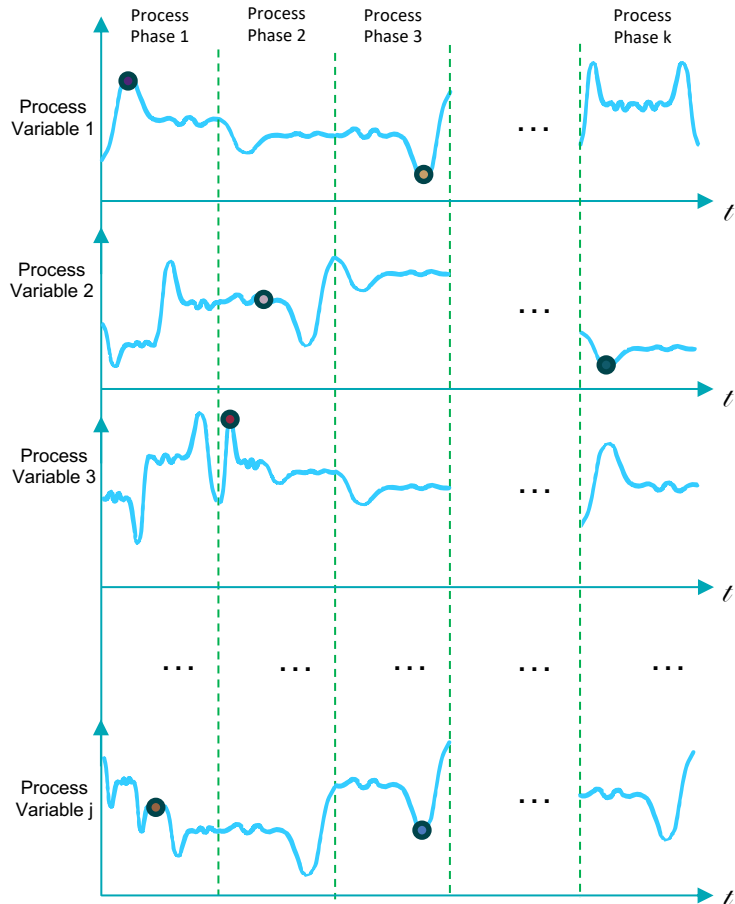
## PROCESS MANUFACTURING CYCLE



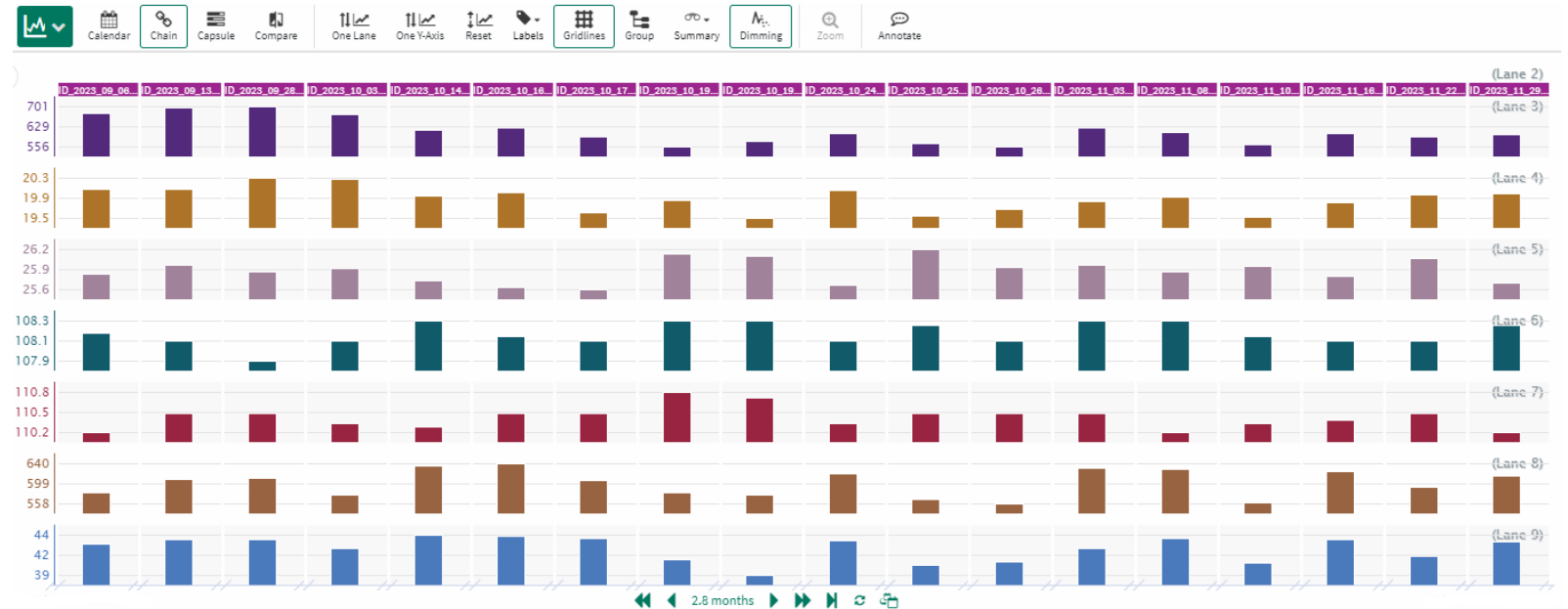


# Anomaly Detection: Technical Knowledge Approach

## Technical Knowledge Based Leading Indicators Definition



## Process Leading Indicators Monitoring



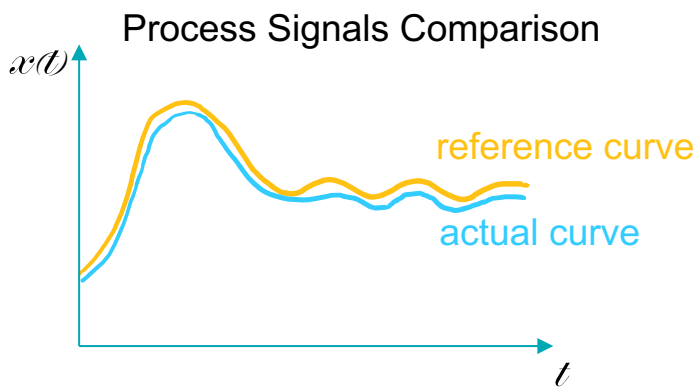
### PROS:

- Easy to be interpreted

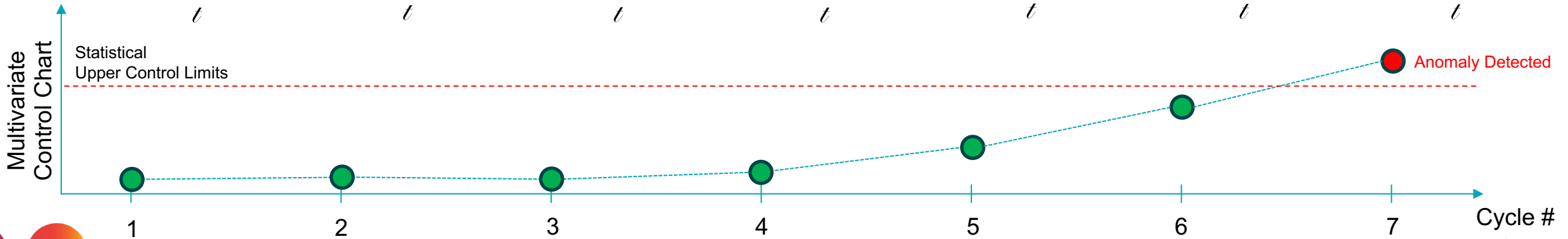
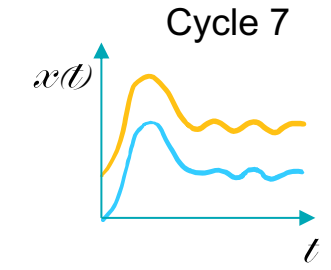
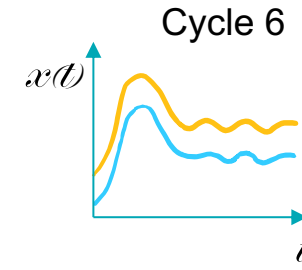
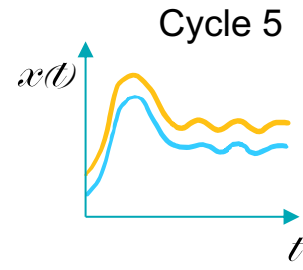
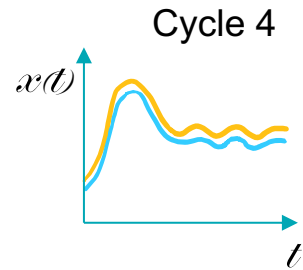
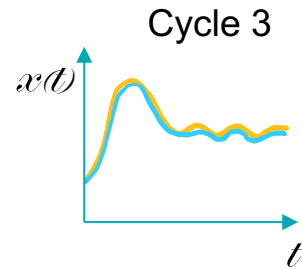
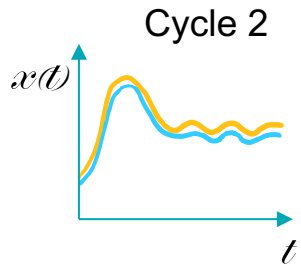
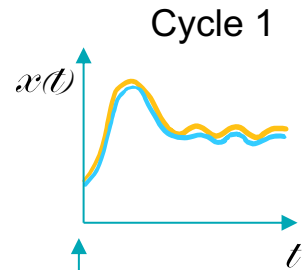
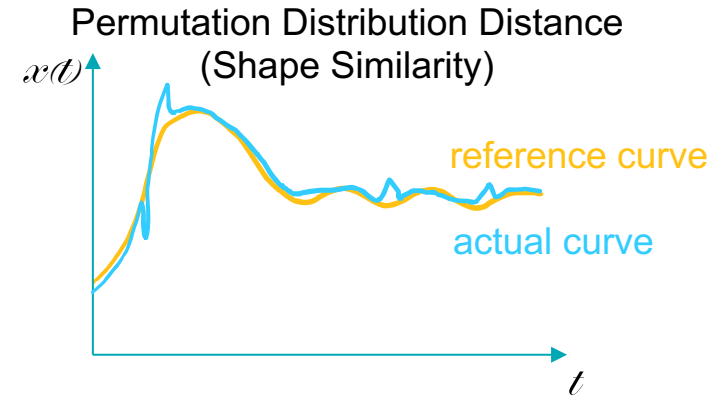
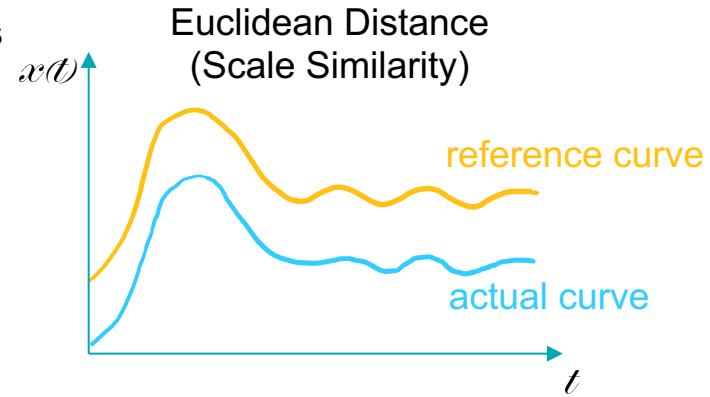
### CONS:

- Potential technical knowledge bias
- Process dependent approach, difficult to scale up

# Anomaly Detection: Statistical Learning Approach



Time Series Distances

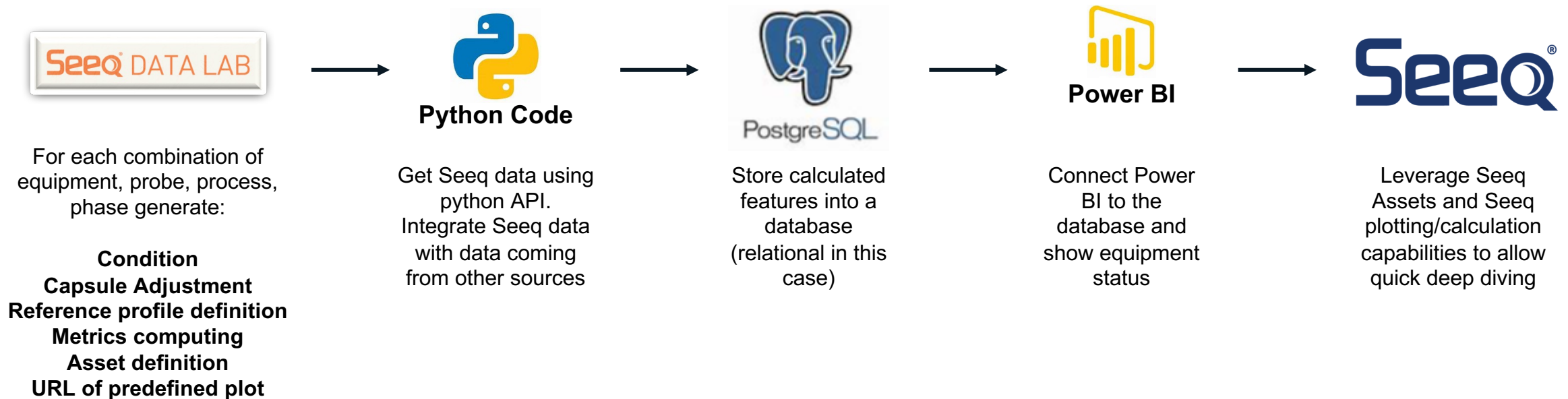


# How to scale up with Seeq

The applicability of the timeseries distance approach encounters a problem of scalability: for example, given **2** equipment with **10** probes which implements **5** different processes, and each process is made of **7** phases, the final number of timeseries distances is:

$$n \text{ ts distances to monitor} = 2 \text{ equipment} \cdot 10 \text{ probes} \cdot 5 \text{ processes} \cdot 7 \text{ phases} = 700 \text{ distances}$$

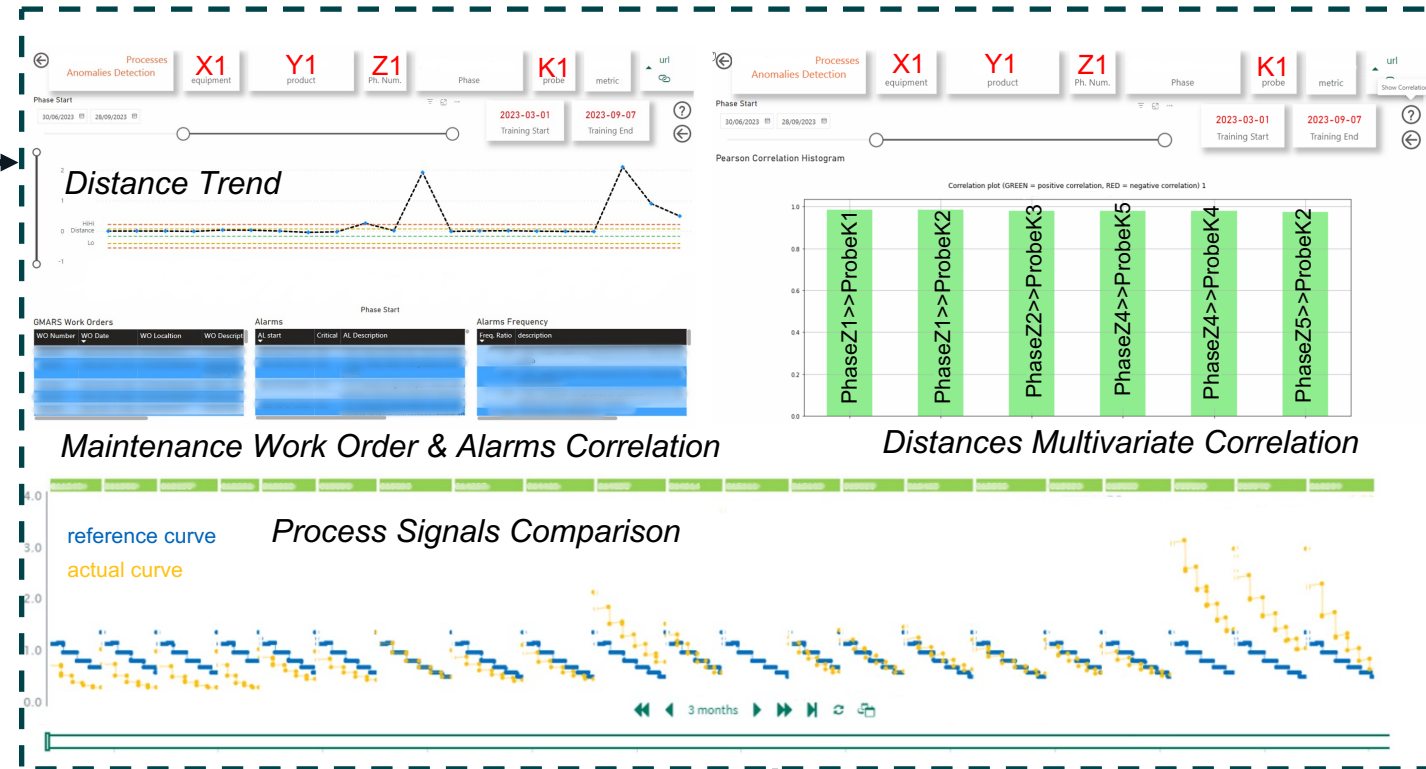
This high number of distances to be monitored poses a scalability problem that has been solved with the following **architecture**:



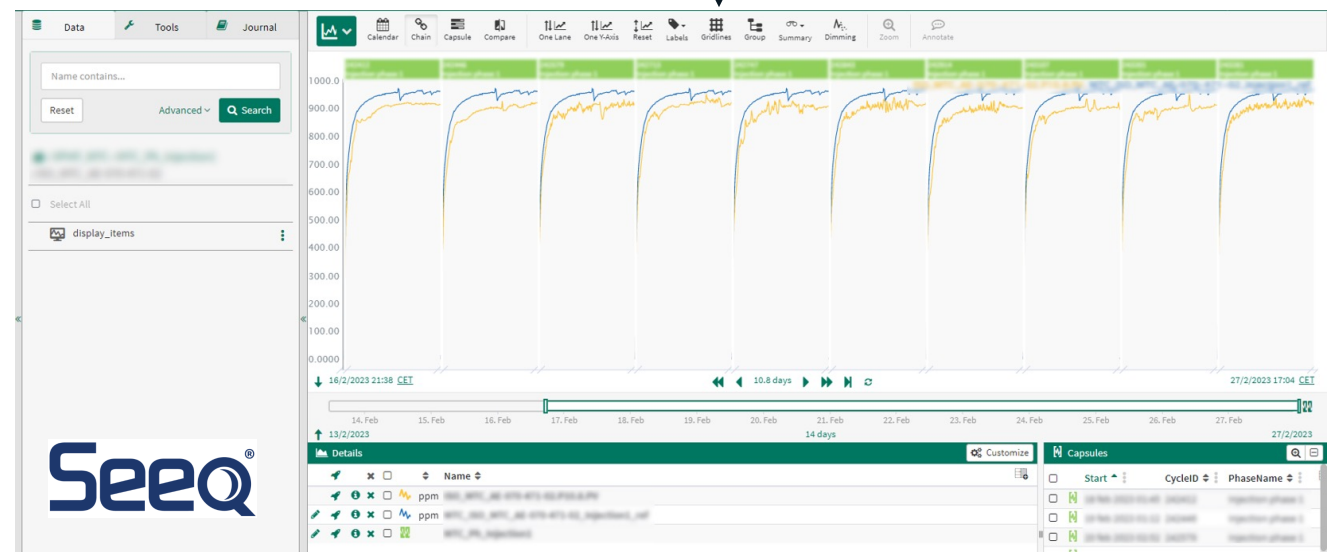
# Architecture Example

## Power BI control tower

eqp >> product >> phase >> probe	1	2	3	4	5	6	7	8	9
Equ X1>> ProductY1 >> PhaseZ1 >> ProbeK1	0%	0%	0%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ2 >> ProbeK1	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X2>> ProductY2 >> PhaseZ3 >> ProbeK3	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ1 >> ProbeK1	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ1 >> ProbeK1	100%	100%	100%	100%	50%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ2 >> ProbeK1	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X2>> ProductY2 >> PhaseZ5 >> ProbeK3	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X2>> ProductY3 >> PhaseZ1 >> ProbeK5	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ1 >> ProbeK1	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ2 >> ProbeK1	100%	100%	100%	100%	50%	100%	100%	100%	100%
Equ X2>> ProductY2 >> PhaseZ3 >> ProbeK8	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ1 >> ProbeK1	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ1 >> ProbeK7	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ2 >> ProbeK1	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X2>> ProductY2 >> PhaseZ3 >> ProbeK3	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X5>> ProductY3 >> PhaseZ1 >> ProbeK5	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ8 >> ProbeK1	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ2 >> ProbeK3	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X2>> ProductY8 >> PhaseZ3 >> ProbeK3	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ1 >> ProbeK1	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY5 >> PhaseZ1 >> ProbeK9	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X1>> ProductY1 >> PhaseZ2 >> ProbeK5	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X2>> ProductY2 >> PhaseZ4 >> ProbeK3	100%	100%	100%	100%	100%	100%	100%	100%	100%
Equ X2>> ProductY3 >> PhaseZ1 >> ProbeK5	100%	100%	100%	100%	100%	100%	100%	100%	100%



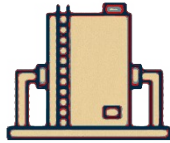
Equ X1	Equ X2	Equ X3	Phase Z1	Phase Z2	Phase Z3
Product Y1	Product Y2	Product Y3	Probe K1	Probe K2	Probe K3



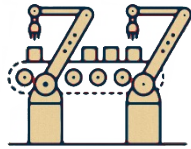
# Actual Implementation Status

## The Processes

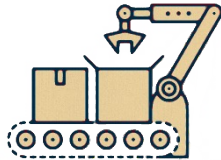
PROCESS MANUFACTURING



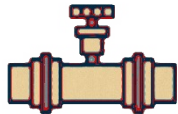
DISCRETE MANUFACTURING



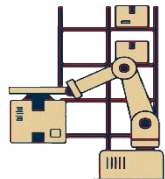
PACKAGING



UTILITIES GENERATION  
& DISTRIBUTION



AUTOMATED WAREHOUSE



## The Digital Capabilities

Fault Diagnosis

Components Anomalies Detection

Alarm Anomalies Detection

Variability Analysis

## The Business Values

**+15% Capacity Increase**  
on Early Stages Processes

**+20% Availability Uptime**  
on Mature Processes

**-15% Workload Reduction**  
of Industrial Engineers



# Key Take Aways

Leverage Seeq to:

- Historian data ingestion
- Process data segmentation  
Define conditions to segment the data into phases. Data Lab gives the possibility to automate this step.
- Process features computing  
Timeseries Euclidean distance can be calculated directly in Seeq
- Time series data visualization  
Seeq Asset with predefined plot help the user to visualize the raw data and integrate with more variables and information
- Easy scalability  
Seeq Data Lab and Seeq API for Python allow easy replication of the same approach on multiple assets

