Analytics Applications

Lessons learned

BreezyDevConf
8.9.2022

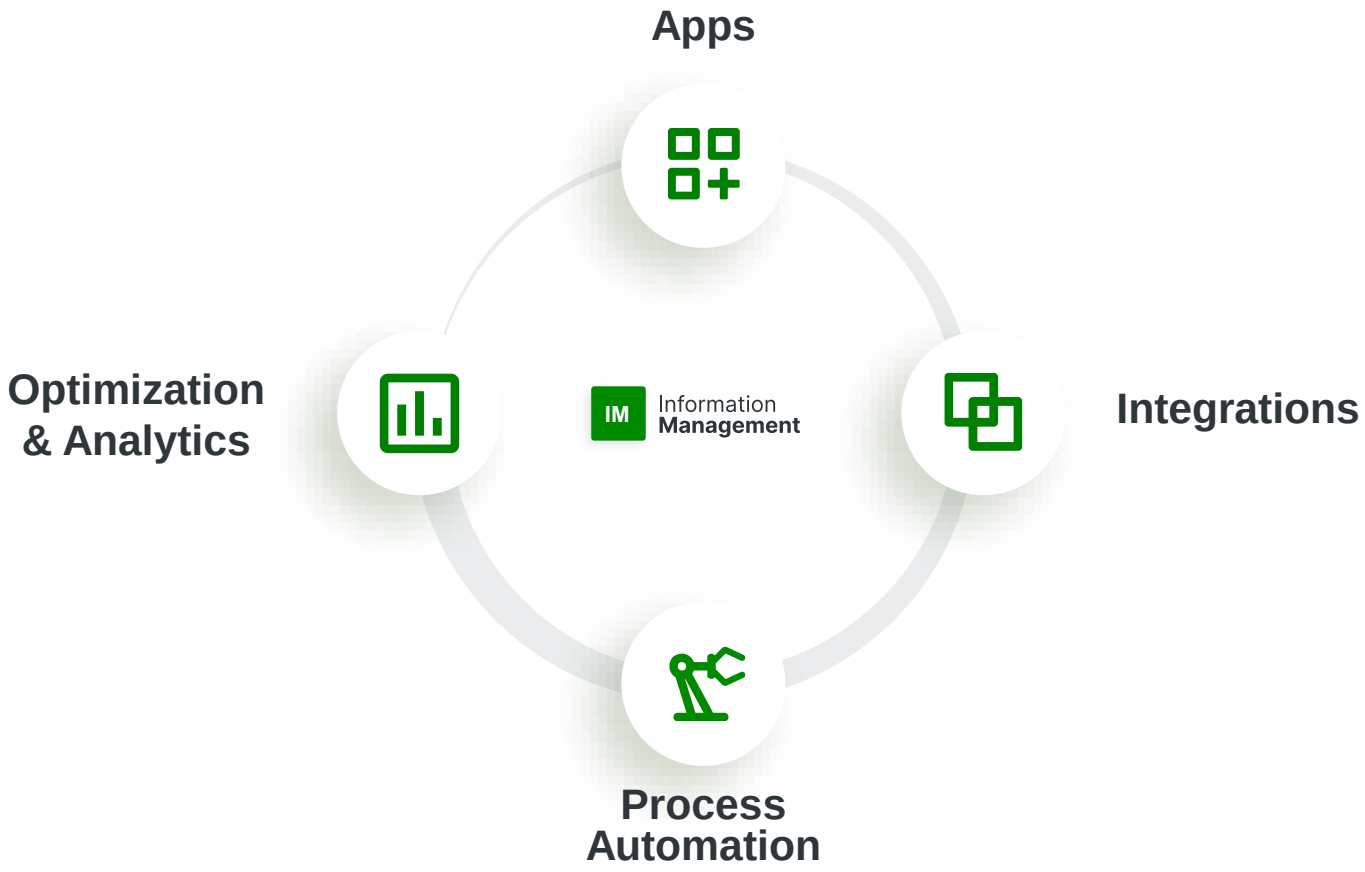
Daniel Djupsjöbacka, Head of Analytics & Optmization

GAMBIT ATEA

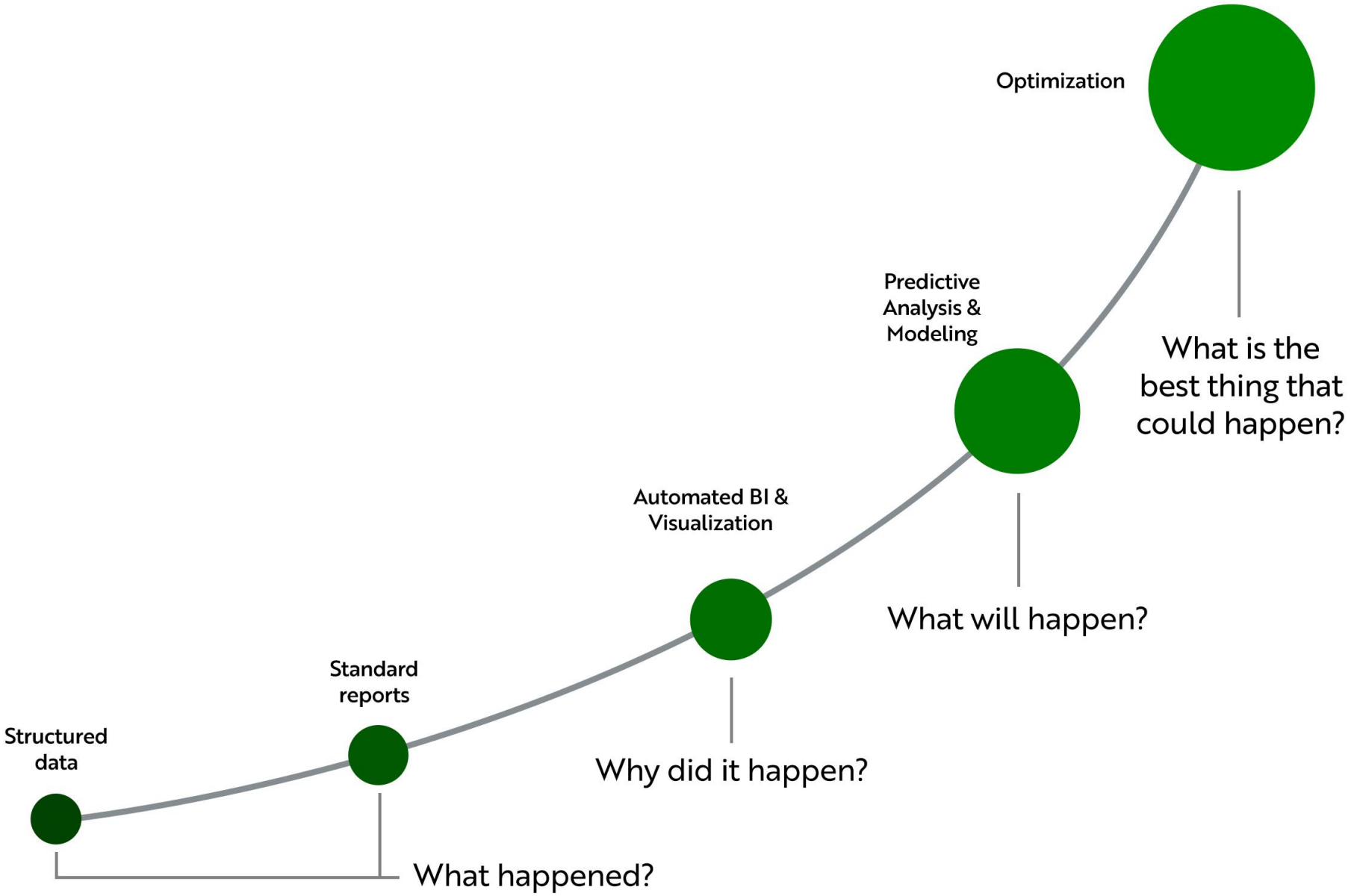
Our journey

What we do

Helping our customers on their digital journey

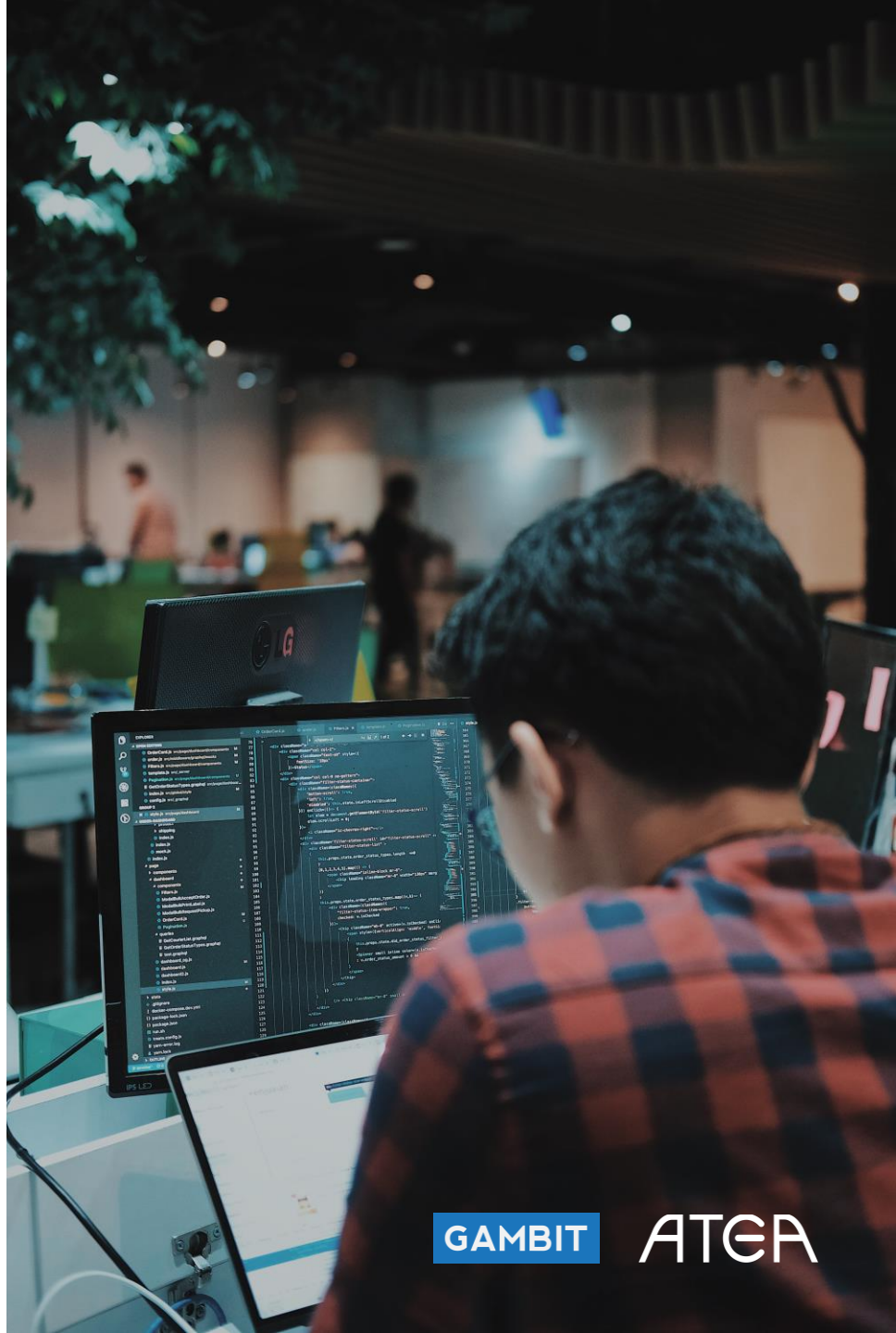


GAMBIT ATEA


Optimization
Predictive Analysis & Modeling
Automated BI & Visualization
Standard reports
Structured data
What is the best thing that could happen?
What will happen?
Why did it happen?
What happened?

How can developers support analytics with good practices?

- How is this related to what I do?
- How can I develop my way-of-working?



Case: Modelling relationships

- Data: 10 years of hourly samples of process input, states and output.

- ”Does the temperature during the production process have any impact on the output concentration?”

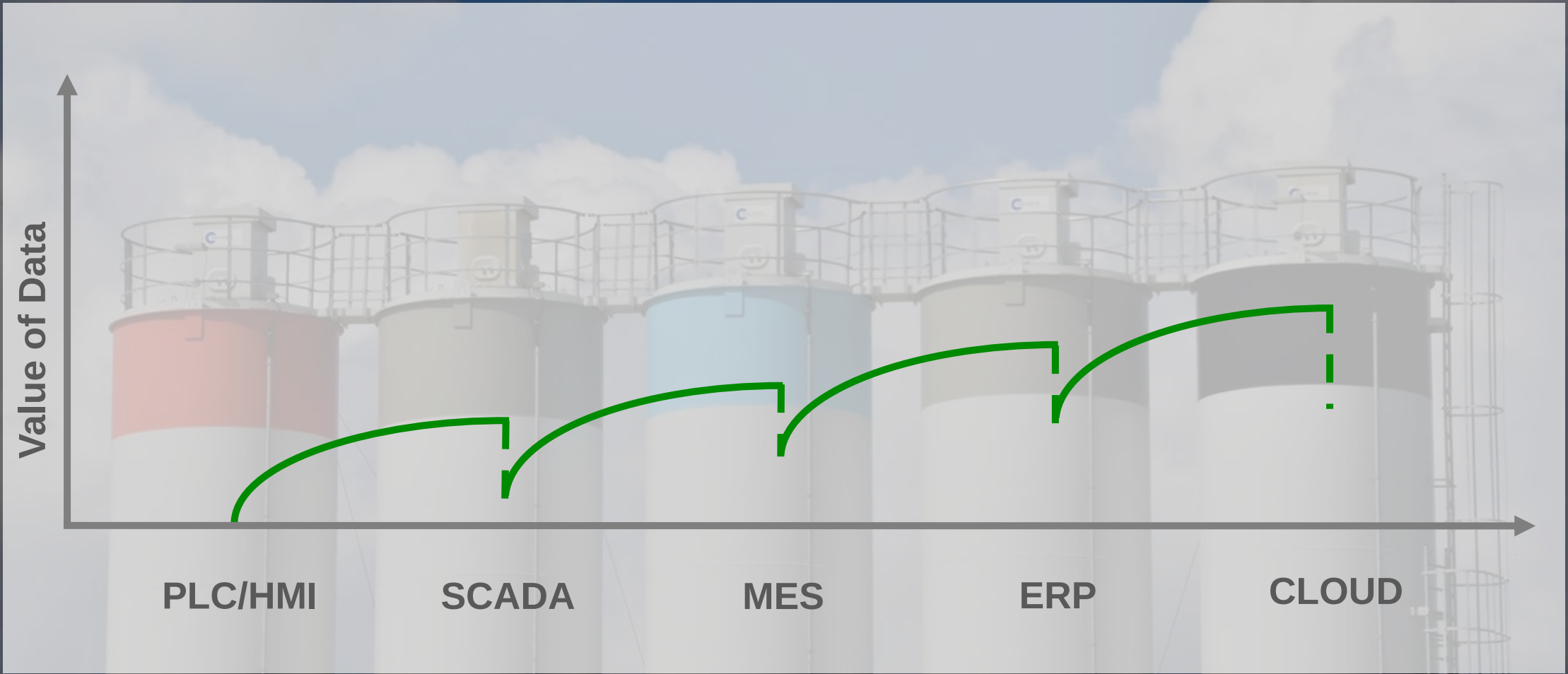
GAMBIT ATEA

Case: Identifying related data in different parts of enterprise

- IoT devices stream data to the cloud
- Can the data context be found for these devices in other systems?

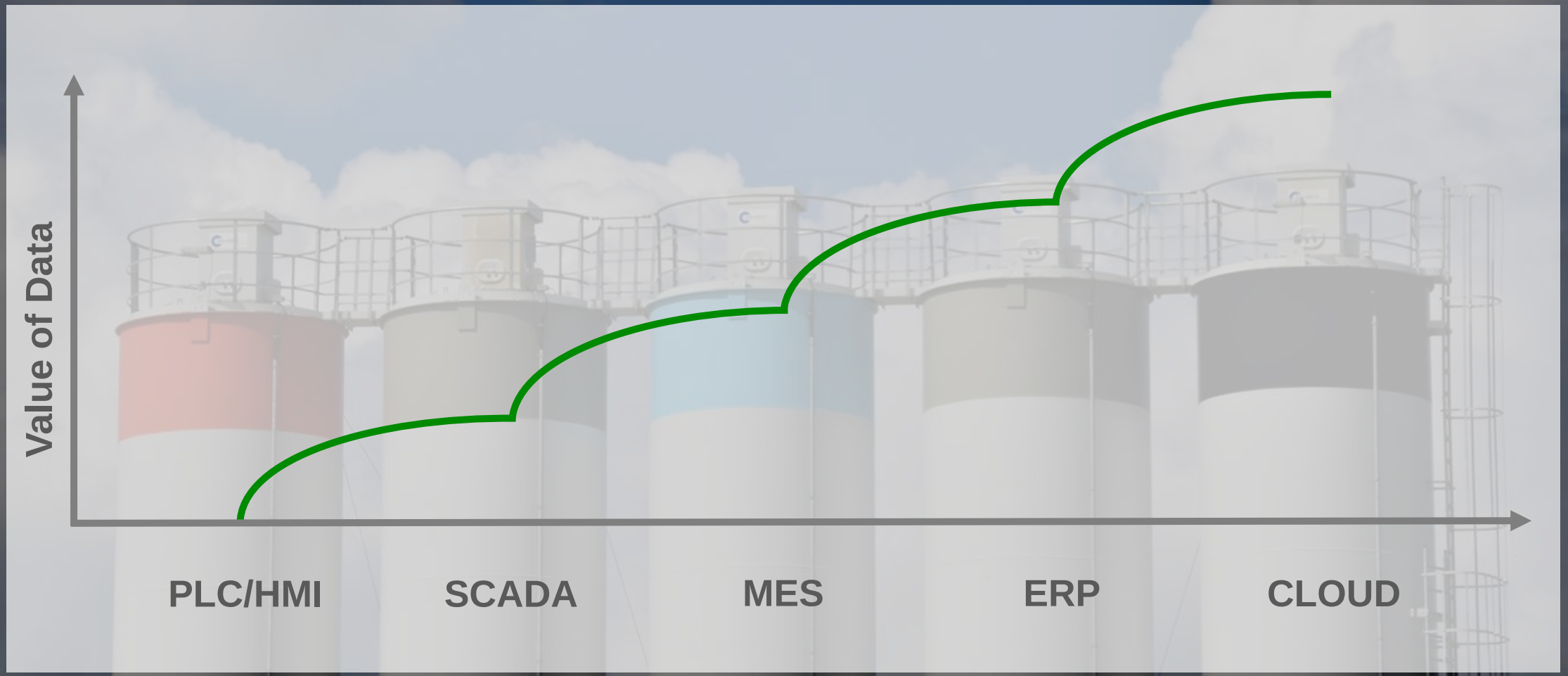
GAMBIT ATEA

Value of enterprise data with context loss



GAMBIT ATEA

Value of data with no context loss



GAMBIT ATEA

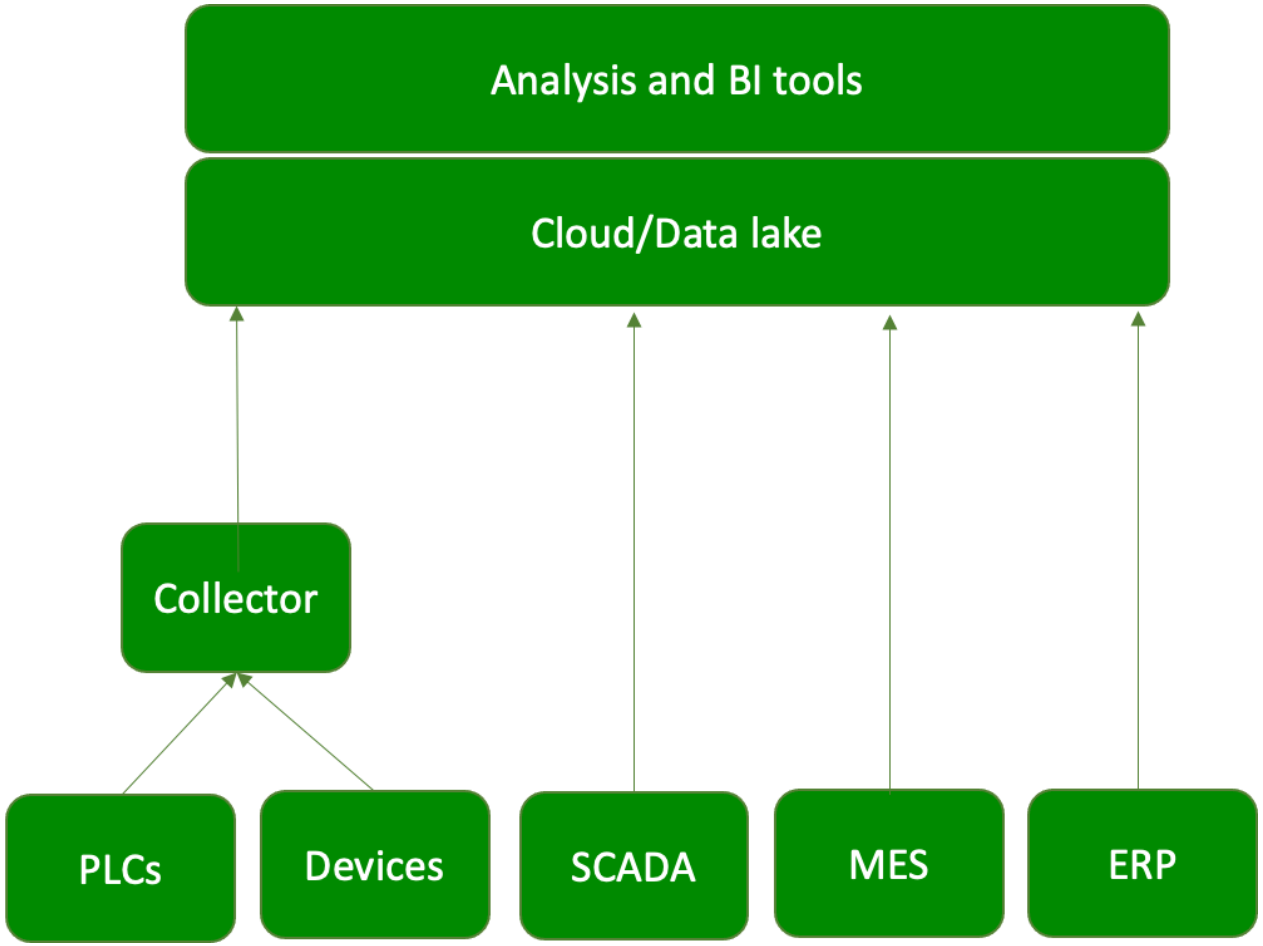
Solution driven approach

Streaming raw data directly into the cloud with no or little context

Problems:

- No abstraction
- No normalization
- Data gaps
- Same data silos

Only selected cases are solved



GAMBIT ATEA

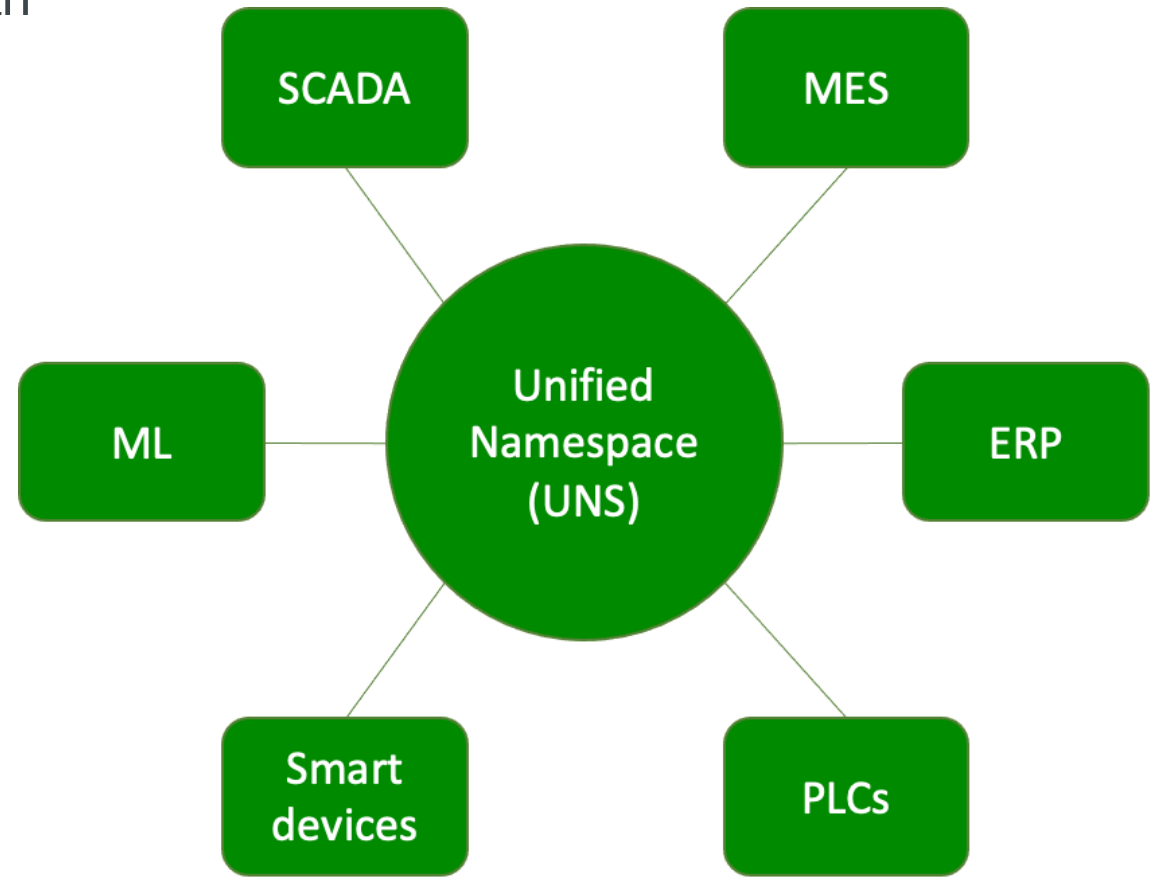
Technology driven approach

Nodes takes part in an eco-system and can both **produce** and **consume** data

Requirements:

- Report by exception
- Lightweight
- Open architecture

Current state of business



GAMBIT ATEA

- Corporation
  - Site 1
    - Line 1
      - Cell 1
        - PLC
          - Temperature
          - EnergyConsumption
        - MES
          - OEE
          - ProductType
          - Parameters
        - ERP
          - OrderNumber
        - AI
          - Suggestions
            - NextMaintenance
      - Cell 2
        - ...
    - Line 2
      - Cell 1
        - ...
  - Site 2
    - ...

```
{
  "energyMeter": {
    "timestamp": 1662633420,
    "value": 1337,
    "unit:" "kWh"
  }
}
```

Topic examples:

PartsUnlimited/Site1/Line1/Cell1/EnergyConsumtion

PartsUnlimited/Site1/Summary/OEE

GAMBIT ATEA

Case: Data aggregation

- Client developed a Master Data Management (MDM) system
- What is MDM?
  - Master data for all things, persons or places that are relevant for different parts of business.
  - Important data – not all data!
- Why MDM?
  - ensure that organization's shared data is consistent and accurate

https://www.informatica.com/resources/articles/what-is-master-data-management.html
https://www.tibco.com/reference-center/what-is-master-data-management

Case: Data aggregation

- Business: B2C distribution
  - Big Orderdata for large amount of clients over long time

- What to keep in MDM?

- What information is lost?

Case: Non-consistent data

- Client had user user interface for storing numeric project data. Data treated as strings both in interface and database.
- Result:
  - 16,5 and 16.5 both ok
  - Empty values are represented by null, ” ” (1 space), ”-” (minus), ”” (empty string)
  - Unknown values are represented by ”???”, ”?”, ”16?” etc
  - On limits
    - numbers mixed with: ”<=4”, ”>6”
    - In some cases, ”X” or ”x” for ”yes, these is a limit”
  - Ranges indicated by ”0-400”, ”400-x”, combinations like ”A > 2, B < 4” and ”600-x €/month”

text
0-100
0-x00
0-99999999
0-10000
5-50
0-40
?
?
0-160

x
1.5
4,5
5
1.5
55

GAMBIT ATEA

Case: Non-consistent data

- Locally, this mix of numbers, text and operators can possibly be understood
  - The data is in many cases useless, or even harmful if used outside the specific function
- This happens because the user is allowed to input it
- ”Users gonna use”

text
0-100
0-x00
0-99999999
0-10000
5-50
0-40
?
?
0-160

x
1.5
4,5
5
1.5
55

User gonna use – part 2

- Users input production plans into system

- Users have different naming conventions resulting in difficulty in analyzing the data
  - Is a production run for one or two batches?
  - Different usage of labeling of production lines across users.

Summary

- The value of data is directly related to how well it can be used inside and across data silos in an organization
- Useless data is more costly than no data
- Developers can do a lot to improve the usability of data