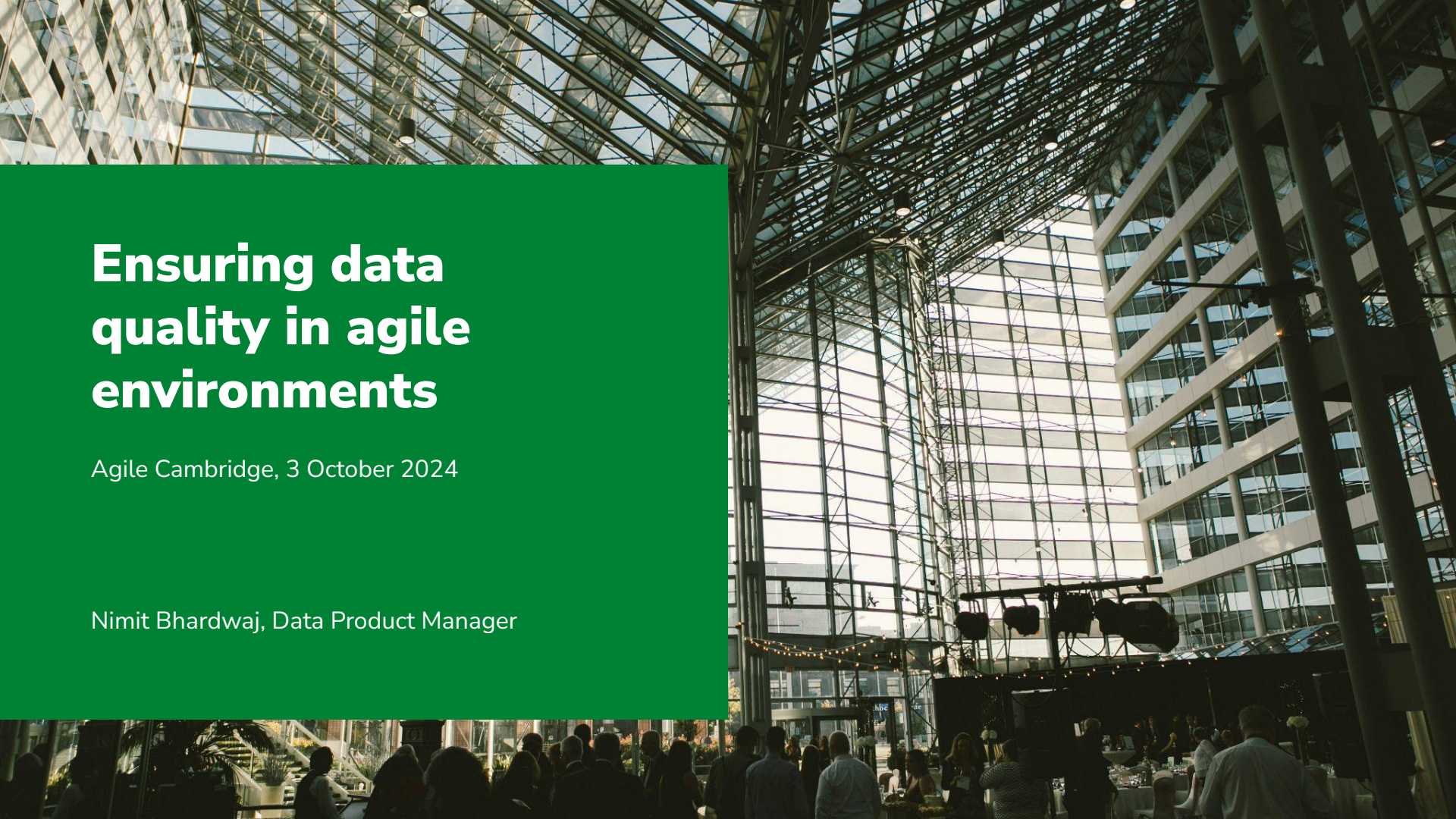




Ensuring data quality in agile environments

Agile Cambridge, 3 October 2024

Nimit Bhardwaj, Data Product Manager

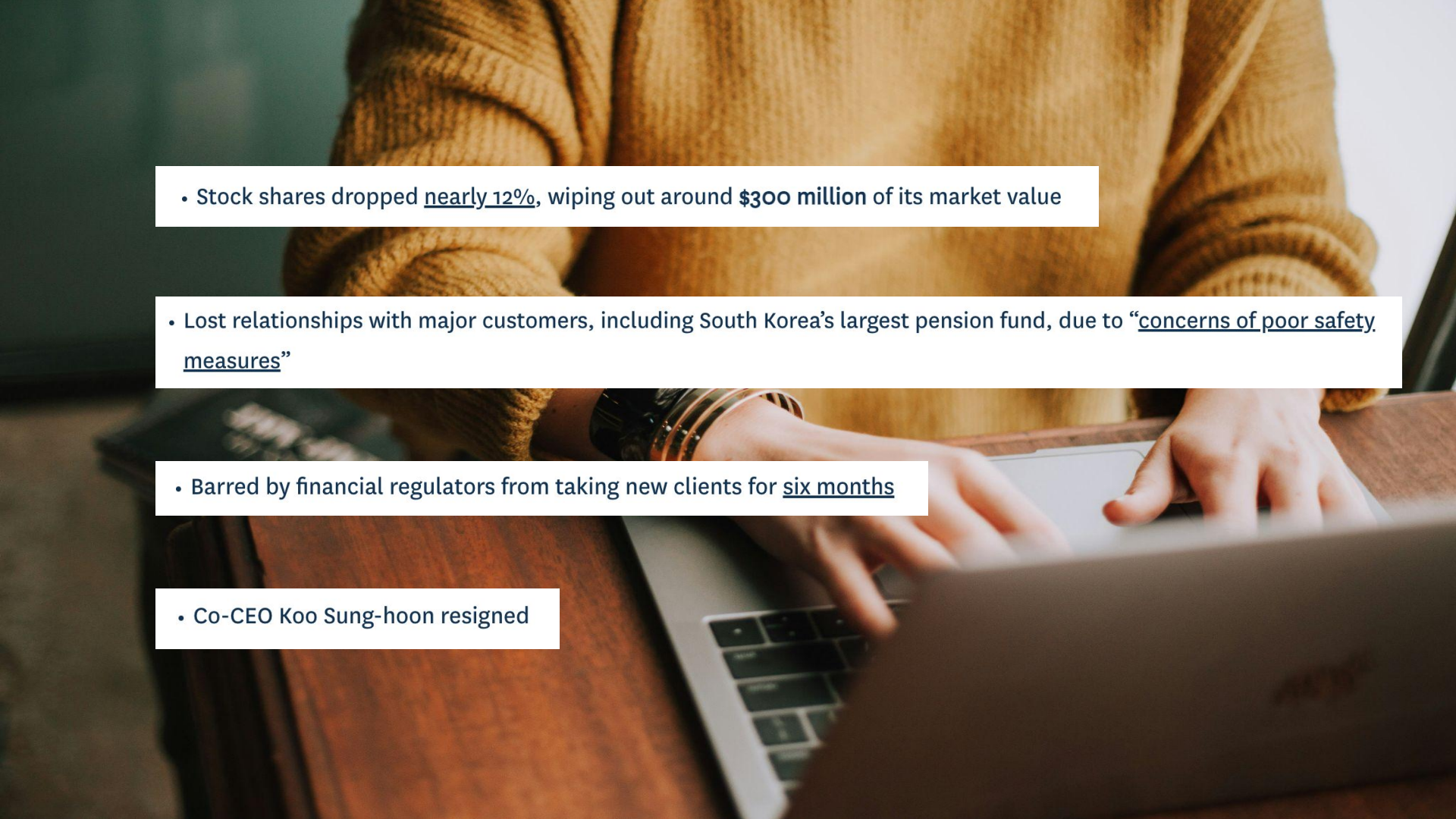


Space

The imperial units error that downed a multi-million-dollar Mars probe

20 years ago, NASA's Mars Climate Orbiter dived too deep and was crushed in Mars's atmosphere. The reason? A mix-up between metric and imperial measurements



A person wearing a yellow sweater is sitting at a wooden desk, working on a laptop. Their hands are visible on the keyboard and trackpad. The background is slightly blurred, showing a green wall and a window.

- Stock shares dropped nearly 12%, wiping out around \$300 million of its market value

- Lost relationships with major customers, including South Korea's largest pension fund, due to "concerns of poor safety measures"

- Barred by financial regulators from taking new clients for six months

- Co-CEO Koo Sung-hoon resigned



About me

Data Product Manager @ Booking.com

- Passionate about Data Quality
- Formerly at BCG & SWIFT
- Certified Product Owner
- Love traveling, reading, flying
- New puppy dad

Who is this talk for?

Data producers, data consumers, and those who do both

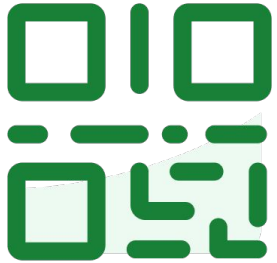
- **Data producers**
People who create, collect, or store data—**engineers, sales associates, customers.**
- **Data consumers**
People who use data to make decisions—**analysts, business leaders, marketers.**
- **Both roles**
Those who both create and use data—**product managers, data scientists, cross-functional teams.**

Agenda

1. What is data quality?
2. Why is DQ important?
3. Common DQ issues
4. Typical agile teams
5. Challenges
6. Strategies & best practices
7. Case study
8. Conclusion



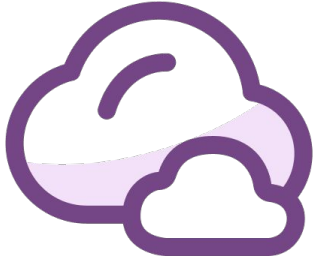
slido



Join at slido.com
#1736927

① Click **Present with Slido** or install our [Chrome extension](#) to display joining instructions for participants while presenting.

slido



What is your primary role?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

slido



How would you categorise yourself/your team?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

slido



What does data quality mean to you? (in a few words)

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

slido



What emoji best represents
your relationship with data
quality?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

What is data quality?

Good data quality means the data is useful and relevant for its intended purpose.

01

Accuracy and validity

The data must be correct and error-free. It should conform to the defined format or standards.

02

Completeness

All required data is available.

03

Consistency

Data should be the same across all sources.

04

Timeliness and availability

Data should be up-to-date and available when needed.

Why is data quality important?

01

Informed decision-making

Quality data leads to better, more confident decisions.

02

More efficient operations

Reduces time spent on fixing data issues and rework.

03

Increased customer trust

Accurate data enhances credibility with customers.

04

Regulatory compliance

Ensures adherence to laws and standards.

05

Improved business outcomes

Drives better analytics, predictions, and overall business performance.

slido



What percentage of
business data is estimated
to be inaccurate?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

slido



What percentage of companies characterise their data as 'very good'?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

slido



Which of these issues have caused a massive crisis for companies?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

Common data quality issues

Data quality problems can significantly impact operations and decision-making.

Incomplete data

JP Morgan Chase's risk model failure

Duplicate records

HSBC customer record duplications

Inconsistent data

Tesla's production data discrepancies

Inaccurate data

TSB Bank IT failure

Human error

NASA's Mars orbiter crash

Samsung Securities' "Fat Finger" error

Data quality challenges in agile teams

Why maintaining high data quality can be difficult in fast-paced agile teams.



Limited time for testing

Prioritizing speed over quality may result in shortcuts in data validation.



Frequent changes in requirements

Evolving requirements can introduce inconsistencies in data models or schema.



Lack of dedicated DQ support

Teams may lack data specialists, leading to oversight in data governance.



No data contracts in place

Changes lead to data downtime without clear agreements on data format.



Lack of DQ measurement

Teams rarely measure data quality as a KPI, making it harder to improve.



Strategies & best practices

Measure your data quality

Data quality measurement, should be considered table stakes in any modern business environment, not just in situations where data-driven decisions need to be accurate and effective.

01

Define your goals

Start by outlining clear goals based on business needs. Based on that, categorise which aspects of data quality are most relevant to you.

02

Determine KPIs

Determine which KPIs can best be used to measure each aspect of data quality you've outlined as part of your goals.

03

Implement continuous monitoring

Establishing monitoring as early as possible will help you constantly check and report on how the quality of your data.

04

Audit and review regularly

Review your measurement thesis to account for how data sources, tools, stakeholder expectations, and the overall business environment evolve.

Treat data as a product

Viewing data as a product naturally results in data quality becoming critical and justifies the resources required.

01

Assigning data quality ownership

Make data quality a shared responsibility across the Agile team.

02

Embed data quality in definition of done

Ensure data quality requirements are part of the sprint goal.

03

Implement quality assurance

Allot resources to ensure data is high quality, with data integrity checks and process validations.

04

Documentation & training

To ensure usability, clear documentation should be established relating to the organization's data.

Incremental data quality improvements

Improve data quality progressively through Agile iterations.

01

Start with high-impact areas

Focus initial efforts on the most critical data elements or processes.

02

Iterative enhancements

Continuously improve data quality with each sprint, refining data validation and governance over time.

03

Leverage retrospectives

Use Agile retrospectives to identify data quality issues and plan improvements in the next sprint.

04

Gradual technical debt reduction

Address data-related technical debt incrementally, without disrupting the sprint flow.

slido



Which strategy would you like to try implementing in your team?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

Case study



Data quality challenges in finance team

Incorrect data at the source

- Some cost information was not recorded correctly.
- Prices were wrongly converted between currencies.
- Supplier information was incomplete.

Downstream issues

- Payment systems couldn't process the data.
- Some suppliers weren't getting paid on time.
- Finance team spent enormous manual effort fixing ~4% of transactions.

Negative impact

- Resources wasted on manual data fixes.
- Delays in payments to suppliers.
- Strain on finance department's productivity and efficiency.

Key objectives

Fix data issues at source & protect financial data consumers



Objective 1

- Support data producing teams to fix data issues at the root.
- Prioritize which issues to fix first (based on impact and complexity).



Objective 2

- Prevent bad data from reaching downstream financial systems.
- Create a buffer to stop bad data from causing further issues in payment systems.



Challenges

- Financial systems were immutable, so bad data could not be easily reversed.
- Fixing all the data issues would take huge time and effort.

Solution

Data quarantine & remediation



Implemented data quarantine mechanism

- Built a quarantine process in the pipeline to catch bad data.
- Developed business logic driven data quality rules based on quality standards defined by data consumers.



Created a data correction UI

- Lightweight UI for data producers to fix issues directly.
- UI writes back to the source and reprocesses the corrected data.



Collaboration and accountability

- Data producers were made accountable for fixing their data.
- Regular conversations with source teams helped prioritize fixes.

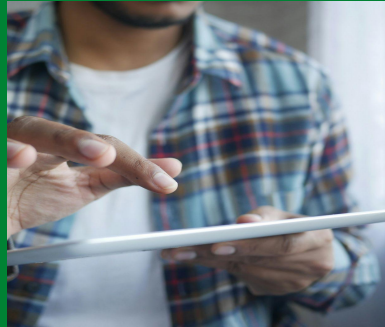
Outcomes

Reduced manual effort and improved accountability



Significant reduction in manual effort

Finance team now manually fixes only 0.2% of transactions (down from 4%).



Improved data ownership

Source teams began taking ownership of their data issues. The cost of fixing data became a higher priority for producers.



Long-term data quality improvement

Fewer issues reached the financial systems. Source teams prioritized fixing bugs over repeated corrections.



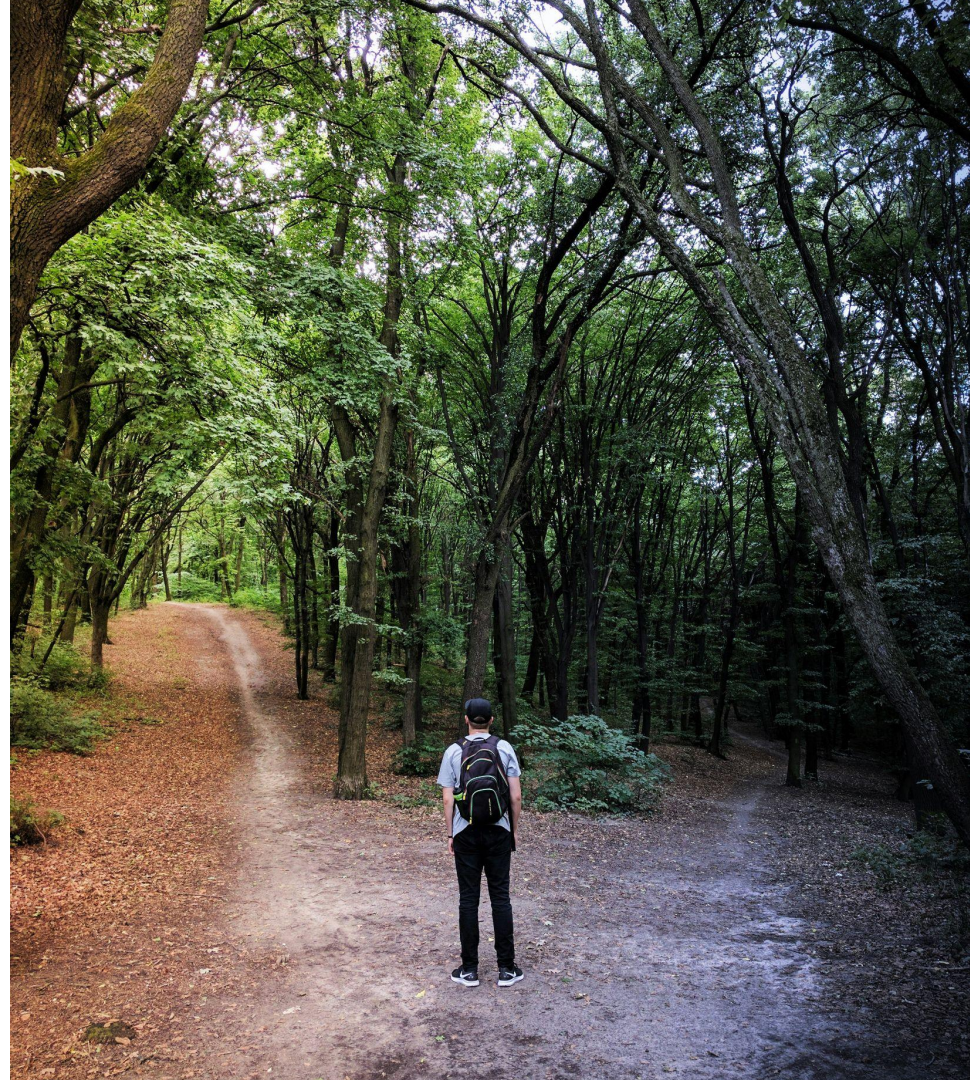
Conclusion

- Data Quality is everyone's responsibility
- Continuous improvement through agile
- Use automation and monitoring
- Collaboration and accountability

“

**Data is the life blood of
decision-making...**

**and quality data is what
separates confident
decisions from risky
guesses**



Thank you

Nimit Bhardwaj

[linkedin.com/in/nimitbh/](https://www.linkedin.com/in/nimitbh/)