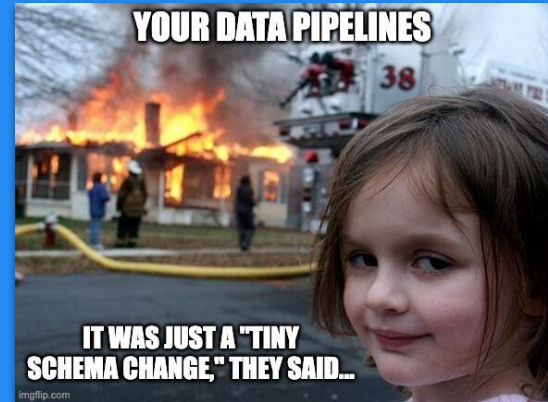


Optimizing ROI on Data Analytics with Data Observability

Ethan Post
Sales Engineer

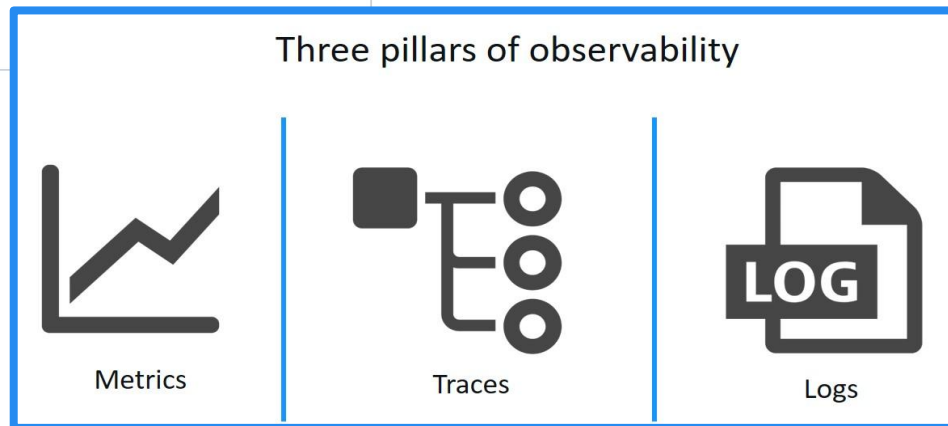


Agenda

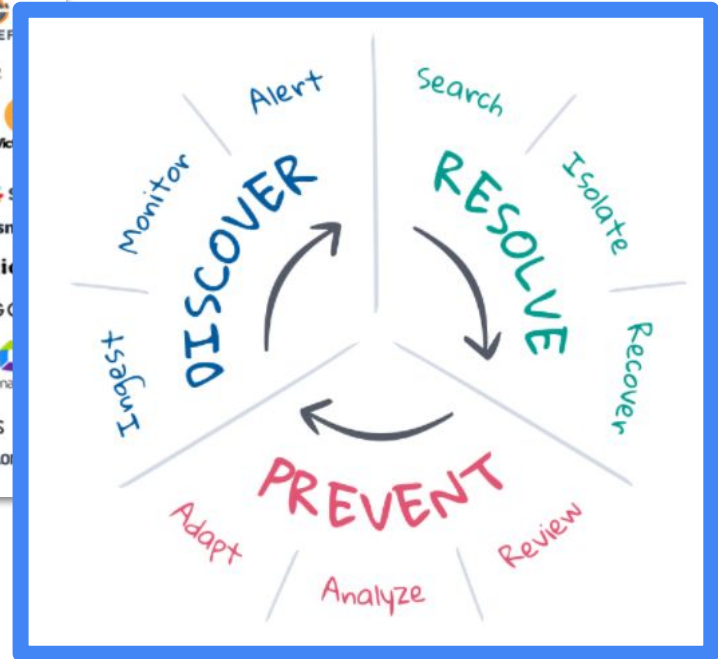
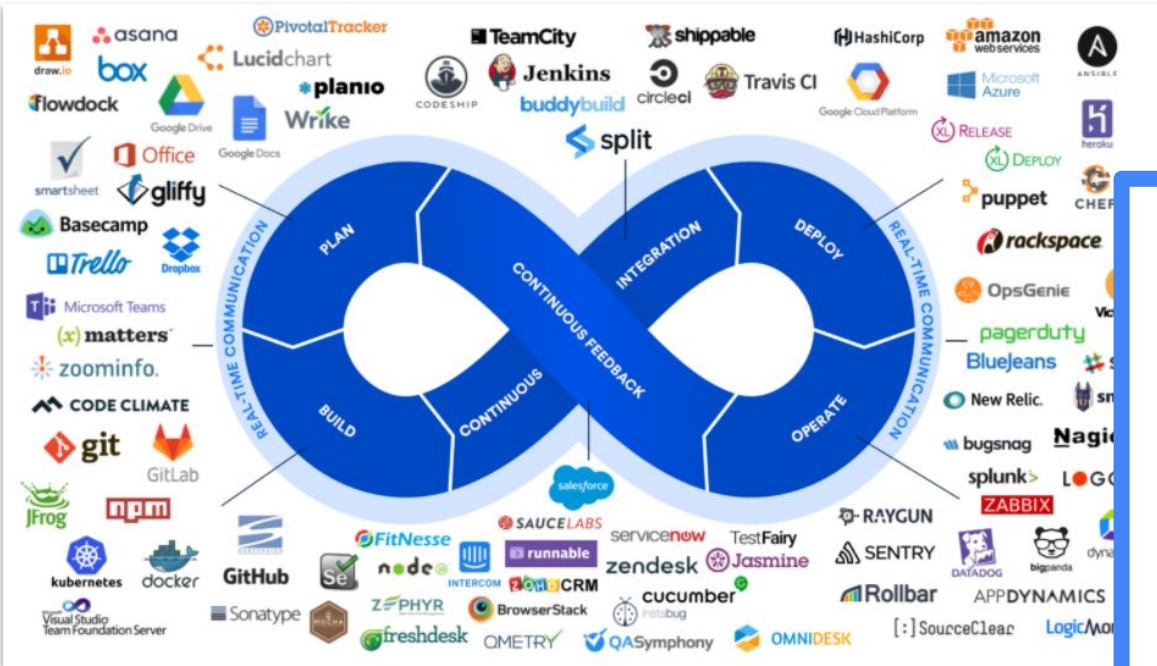
- What is Data Observability
- What is Monte Carlo
- Product Demonstration
- Customer Case Studies

What is Data Observability?

The three pillars of application observability



The reliability lifecycle



What is Data Observability?



Detect

- ML-powered anomaly detection
- Rule-based detection
- Targeted alerts to impacted owners & downstream users

Resolve

- Automated field-level lineage
- Impact radius assessment
- Code, data, and operational diagnostics

Prevent

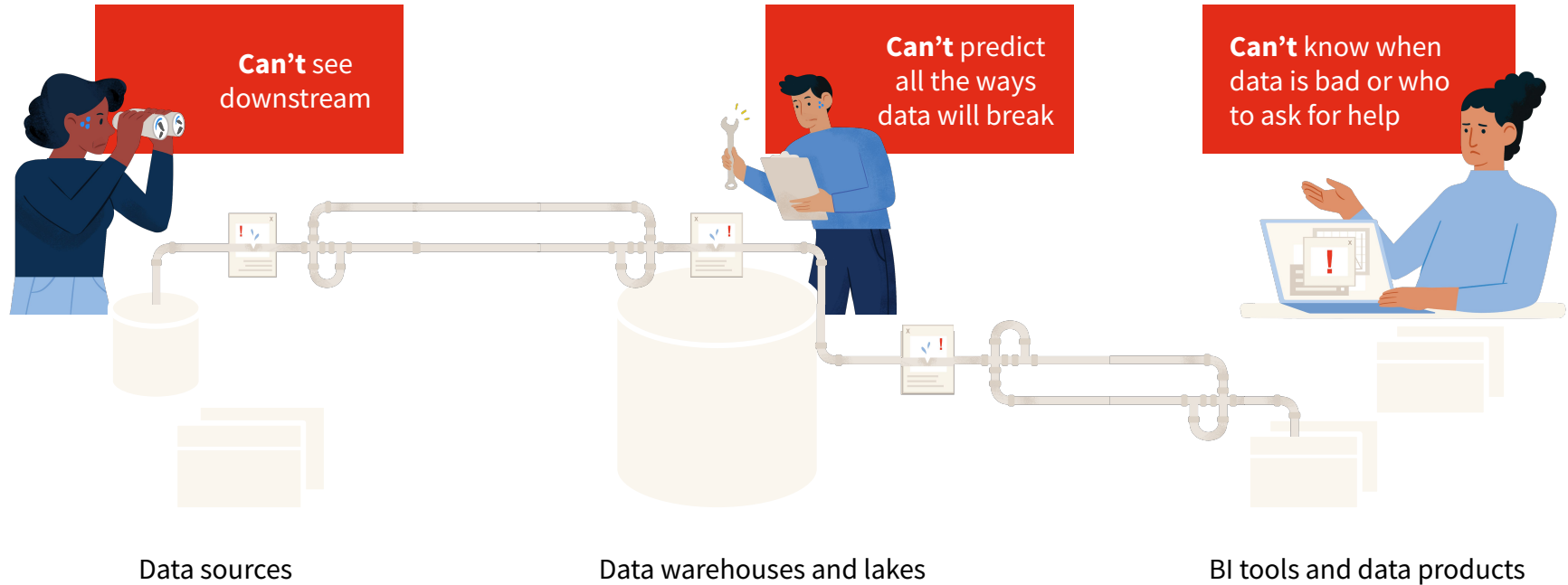
- Auto-generated and on-demand insights
- Schema change notifications
- Automated circuit breakers



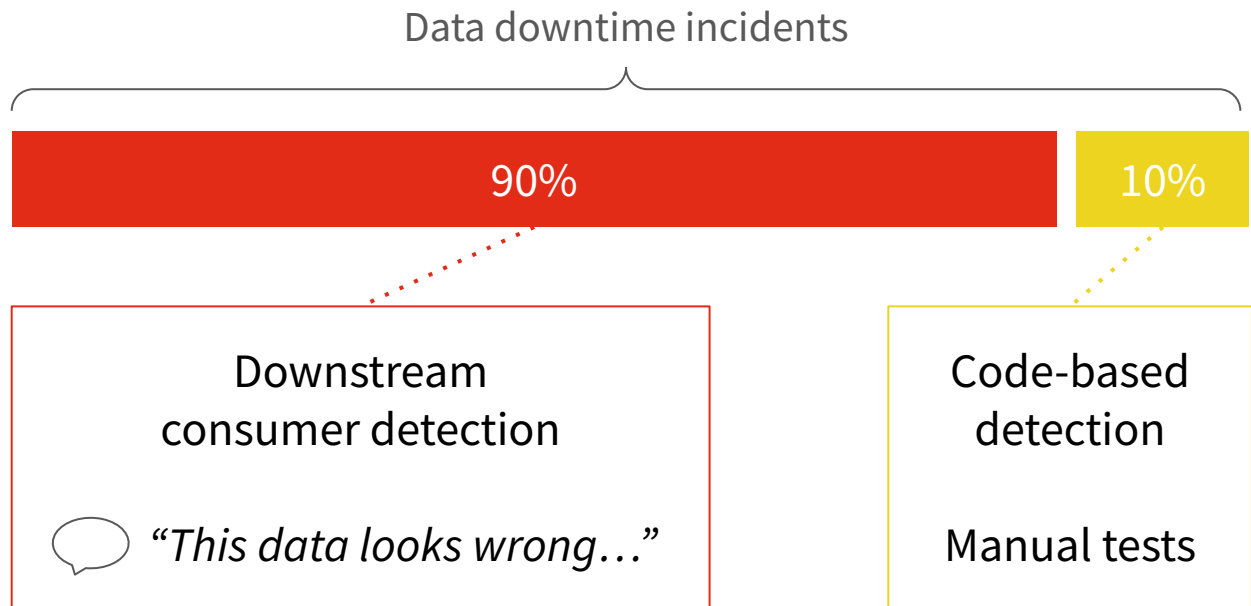
DATA OBSERVABILITY PILLARS

Freshness | Volume | Quality | Schema | Lineage

The problem: data downtime



Data quality incidents are detected **reactively** today



The result?

Days to weeks pass before incidents are detected and resolved

Business impact from poor data quality

~70

high severity events each year
per every 1k tables ¹

30-50%

data engineering time
spent on fire drills ²

12-27%

avg. annual revenue lost for
companies resulting from
poor data quality ³

1. Benchmark data based on Monte Carlo customer production deployments
2. Monte Carlo market research and customer-reported benchmarks
3. Experian Data Quality; MC research via Wakefield Research Survey

What is Monte Carlo?

Good news: data downtime looks similar across companies

- Is this data up-to-date?
- Why does this data size look off?
- Isn't this value suspiciously high?
- Why are there so many nulls?
- Why do we have duplicate IDs?
- What reports will I break with this schema update?
- Why are there 0s on tiles that usually show 100s?
- ...

The five pillars of data observability

DATA OBSERVABILITY PILLARS



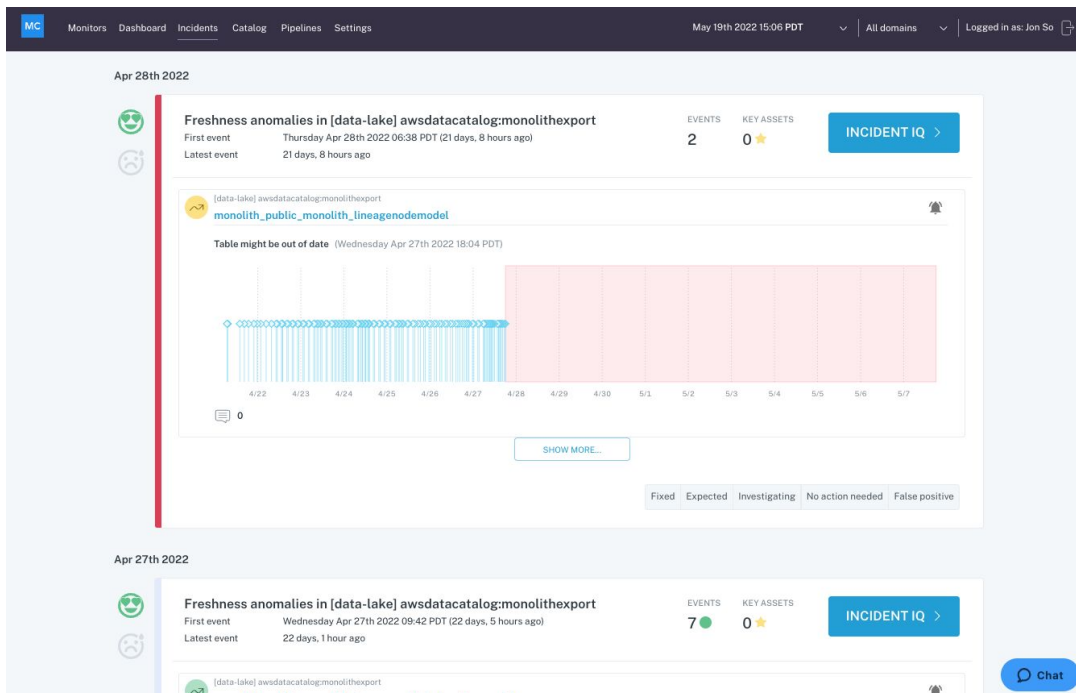
Freshness

Volume

Quality

Schema

Lineage



The 5 pillars of data observability

DATA OBSERVABILITY PILLARS

Freshness

Volume

Quality

Schema

Lineage

Volume anomalies in [snowflake] analytics:prod_detectors

First event Sunday May 1st 2022 13:58 PDT (10 days, 22 hours ago)

Latest event 10 days, 22 hours ago

EVENTS

2

KEY ASSETS

2★

INCIDENT IQ >



[snowflake] analytics:prod_detectors



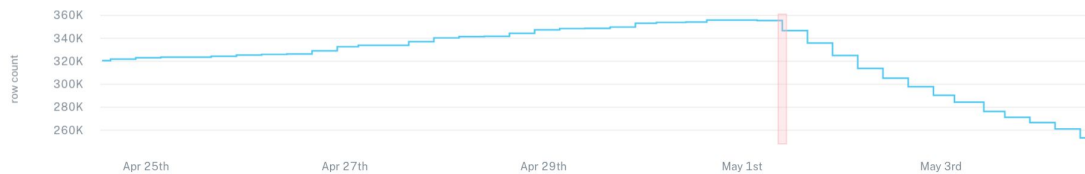
numeric_status_indicator



Reports affected by this event.



Exceptionally large number of rows (9K) deleted, typically ≤ 442 (Sunday May 1st 2022 09:40 PDT)



0

SHOW MORE...

Fixed Expected **Investigating** No action needed False positive

Status updated: May 9th 2022 10:57 PDT

Last updated by epost@montecarlodata.com

The 5 pillars of data observability

DATA OBSERVABILITY PILLARS

Freshness

Volume

Quality

Schema

Lineage

Dimension anomalies in [snowflake] analytics:prod_lineage

First event Tuesday May 10th 2022 22:38 PDT (1 day, 13 hours ago)
Latest event 1 day, 12 hours ago

EVENTS

2

KEY ASSETS

2★

INCIDENT IQ >



[snowflake] analytics:prod_lineage

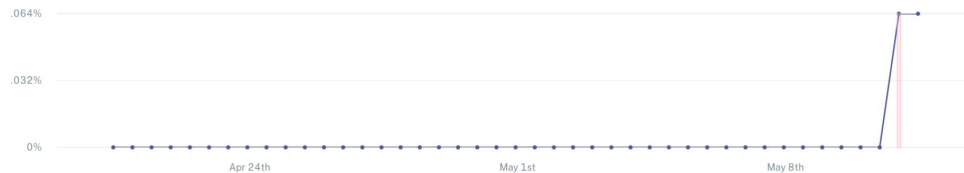


looker_dashboard_nodes



Distribution anomaly in field "account_id" (Tuesday May 10th 2022 23:08 PDT)

New value ("59abclcd-f083-447e-8c7f-53404f219e75") detected



Total rows

Log scale



0

SHOW MORE...

Fixed Expected Investigating **No action needed** False positive

Status updated: May 11th 2022 01:24 PDT
Last updated by epardes@montecarlodata.com

MC

The 5 pillars of data observability

DATA OBSERVABILITY PILLARS



Freshness

Volume

Quality

Schema

Lineage

Schema changes in [snowflake] analytics:prod_detectors

First event Monday May 9th 2022 08:06 PDT (3 days, 4 hours ago)
Latest event 3 days, 2 hours ago

EVENTS

4

KEY ASSETS

3★

INCIDENT IQ >



[snowflake] analytics:prod_detectors



freshness_training



Reports affected by this event.



Schema changes detected (Monday May 9th 2022 10:07 PDT)

new_n days_existing update_pattern were added

reason changed from "varchar(99)" to "varchar(126)"

max_delay_sec changed from "number(18,0)" to "number(24,0)"

0

SHOW MORE...

Fixed Expected Investigating No action needed False positive

Status updated: Apr 25th 2022 01:21 PDT

Last updated by epardes@montecarlodata.com

The 5 pillars of data observability

DATA OBSERVABILITY PILLARS

Freshness

Volume

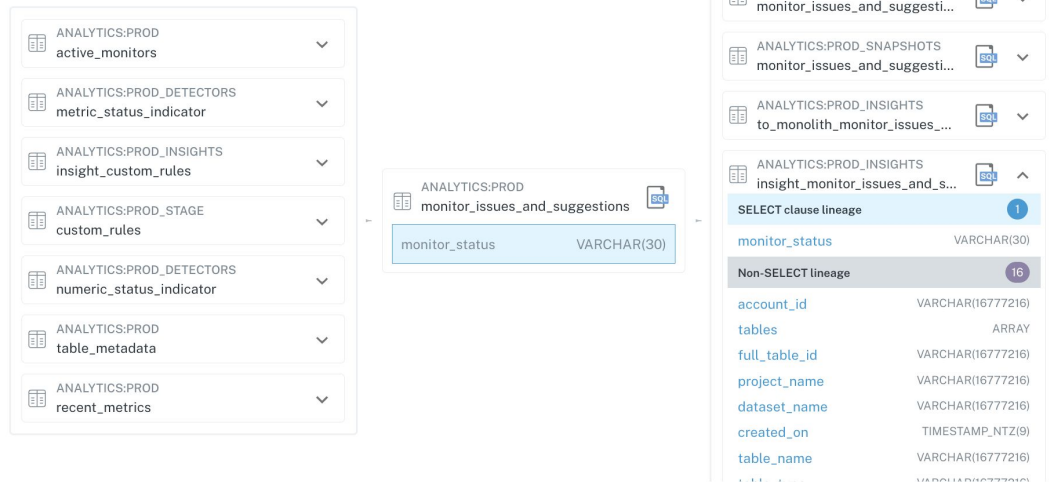
Quality

Schema

Lineage

Field Lineage

Column-level lineage across data assets



Field	Type	Description	Tags
agg_type	VARCHAR(16777216)		--

So... How does Monte Carlo help?

Make Data More
Reliable and
Trustworthy

Reduce Time and
Resources Spent on
Data Quality

Mitigate Risk and
Impact of Bad Data

Platform Demo

Why do teams choose Monte Carlo?



Optimize proactive incident detection



Robust alert routing and noise reduction



Democratize data quality



Enhance data trust



Prevent data downtime



Reduce time to resolution



Automate monitoring for scalability



Gain visibility on data health

Tactical

Strategic

Make Data Reliable and Trustworthy



A book-summarizing subscription service with over 16 million users worldwide

“The self-service capabilities of data observability helped build back trust in data, as users were seeing us in action: going from a red alert to a blue “work-in-progress” to “resolved” in green.”
-Gopi Krishnamurthy, Director of Engineering

Outcomes

- 120 engineering hours saved/week
- Faster data incident resolution
- Increased revenue: Highly reliable data, including new data SLAs, ensured marketing spend was allocated appropriately

Challenge

When COVID-19 changed user behavior overnight, data quality issues eroded trust in the company’s data.

Solution

Implemented Data Observability to deliver visibility into the health of their Redshift warehouse, ETL solutions, and Periscope dashboards.



An online pet food retailer delivering millions of meals monthly

"I now have confidence if I see a change in my numbers, it's a business change and not a data issue." – Head of Data Strategy & Insights

Outcomes

- 200+ engineering hours saved by using Monte Carlo (vs. building in-house)
- Simple validation for business changes and proactive monitoring for data issues
- 20-minute implementation

Challenge

The Farmer's Dog lacked clarity around the root cause of their data incidents and needed a way to detect changes.

Solution

Implemented Monte Carlo to monitor freshness, volume, data, schema and lineage.

"This was one of the easiest setup process that I've ever experienced."

A news, media and sports powerhouse

“Data observability has become a necessity, not a luxury, for us. As the business has become more and more data-driven, nothing is worse than allowing top executives to make a decision based upon data that you don’t have trust in.”

Outcomes

- Increased cost savings
- Better collaboration between data engineers and analysts

Challenge

Fox was exploring a decentralized, self-service approach to their data architecture, they faced the challenge of ensuring all data available to stakeholders reliable and trustworthy.

Solution

They implemented Monte Carlo's data observability platform to proactively detect and resolve incidents when they occur, before stakeholders lose trust in the data.

A new, smarter take on auto insurance that promises fast, hassle-free and easy to understand plans

“We no longer had to tailor specific tests to every particular data asset. All we really had to do was sign up, add the security implementation to give Monte Carlo the access that it needed, and we were able to start getting alerted on issues. Monte Carlo gave us that right out of the box.”

Outcomes

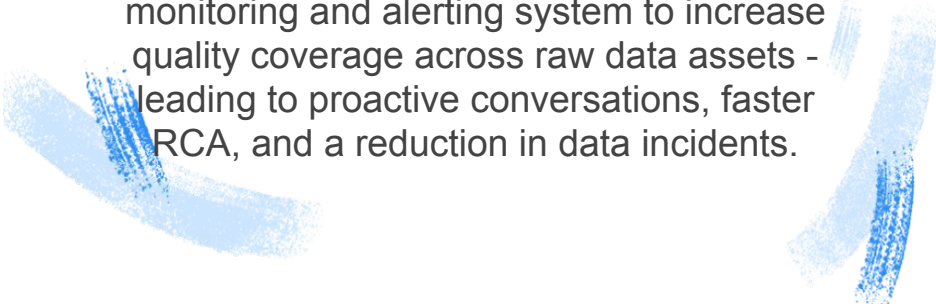
- 70% increase in data quality coverage
- Solve data issues faster
- 50% faster resolution times stemming from common changes such as schema drift or additional sources

Challenge

With the proliferation of data sources, it became harder for Clearcover's data engineers to scale data quality testing across pipelines manually.

Solution

They implemented Monte Carlo's automated monitoring and alerting system to increase quality coverage across raw data assets - leading to proactive conversations, faster RCA, and a reduction in data incidents.

A decorative graphic consisting of several overlapping, horizontal brushstrokes in various shades of blue, located at the bottom of the slide.

Reduce Time and Resources Spent on Data Quality

Software that provides an intuitive, visual way to discover, consolidate, and communicate user needs

“Monte Carlo gives us the power to know what’s going on with our data at any given point in time so we can ask the right questions when data downtime strikes, for instance ‘we think something’s wrong here, did you change anything, or is this expected?’”

Outcomes

- 3x reduction in infrastructure costs through isolating an unexpected spike in marketing events
- Faster time-to-detection (TTD)
- 99% faster incident det

Challenge

Hotjar’s in-house built solution through dbt frequently had alerting delays, and upstream tool notification settings were too infrequent

Solution

evaluated data observability solutions requiring:
Timely incident alerting
Root cause analysis with field-level lineage for speedy resolution
Field Health



Restaurant point of sale and management system that helps restaurants improve operations, increase sales and create a better guest experience.

“With Monte Carlo, we got the thing up and running within a few hours and then let it go...It just cuts down on time. You’re directed exactly toward what the problem could be, and from there, you can expand.”

Outcomes

-

Challenge

Toast experienced hypergrowth with larger data volumes and less visibility into data health. Their homegrown data quality tool could not keep up with business needs.

Solution

Monte Carlo provided out-of-the-box support and fast deployment for comprehensive data quality monitoring and lineage, resulting in saved time and money building and manually deploying data quality tests and more time spent on ML initiatives.

Mitigate Risk and Impact of Bad Data

A leading digital advertising software company that gives small and medium-sized businesses access to enterprise grade advertising technology

"I can't imagine a situation where I would fire up Snowflake and not put Monte Carlo on top of it. It's become part of my go-to stack along with Looker and machine learning in GCP...I'd recommend it as a first step prior to data catalogs and other investments."

- Adam Woods, Chief Technology Officer

Outcomes

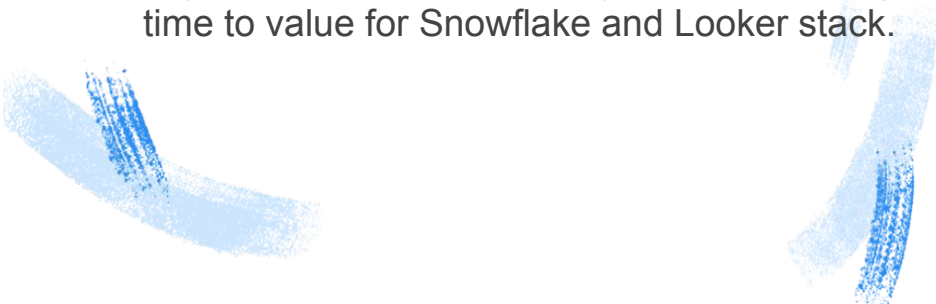
- 88% reduction in data downtime
- Time-to-detection (TTD):
"Accelerated from days to minutes."
- Time-to-resolution (TTR): 96% reduction

Challenge

When Choozle rolled out a unified reporting capability, they faced the problem of table sprawl and data fragmentation, which caused customers to lose trust in their platform.

Solution

They overcame this problem by integrating an easy-to-use data observability solution with high time to value for Snowflake and Looker stack.

A decorative graphic consisting of several overlapping, light blue brushstroke-like shapes, primarily located in the bottom right quadrant of the slide.

Solution that provides retailers with an all-in-one returns platform to delight customers, drive revenue, and preserve the planet

“Being able to quickly identify client-facing issues and be proactive is really the key to building trust in our data. And this feature makes the data engineers’ jobs much much easier—I can tell you definitely from experience here.”

Outcomes

- 50% of engineering capacity recovered from faster incident resolution

Challenge

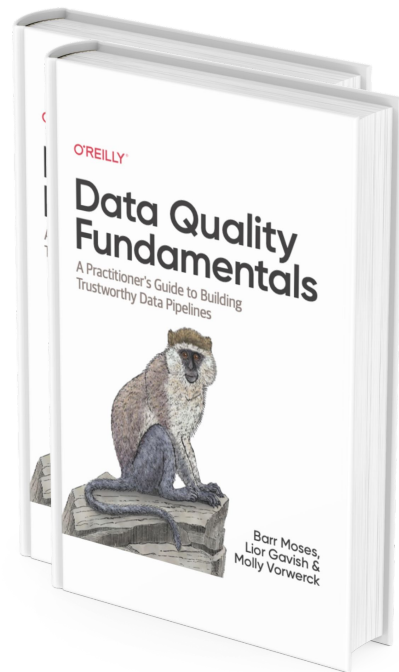
The team at Optoro didn’t have a method for understanding when data might be missing, when it might go stale, or if the data wasn’t what they expected. When data issues did occur, customers were often the first to know.

Solution

With data observability in place, the data team is now able to automatically and proactively identify and resolve - saving their engineers 44 hours each week on support tickets.

O'Reilly's Data Quality Fundamentals

Learn what it takes to build more reliable data lakes at scale.



<https://www.montecarlodata.com/oreilly-data-quality-fundamentals-early-release/>

Customer Success Story



Overview

Restaurant point of sale and management system that helps restaurants improve operations, increase sales and create a better guest experience.

Industry

SaaS / Fintech

Data stack & comms channel



Situation

- **Context:** The data team was deploying a new data platform that with a decentralized model that emphasized self-service for analysts and positioned the data team in a more consultative role; the change arose due to:
 - Rapid growth of the business
 - Growing volume of data
 - Increased number of people relying on that data
 - Limited data engineering resources
- **Challenge:** ensure data quality for ever-growing volumes of mission-critical data
 - Attempted to build monitors in house, but couldn't keep up with new data sources and integrations
- **Solution:** data observability implementation that provided:
 - Fast implementation
 - Scalability as the business and data dependencies continued to grow

Results

Speed of implementation + ease of use

After spending months building and maintaining freshness monitors, the team implemented Monte Carlo within a few hours, including automated:

- Freshness monitors
- Volume monitors
- Quality monitors
- Schema change monitors
- Custom rules & tests
- Root cause analysis tools, including lineage

Data engineering hours saved

Data engineering team now has the time to focus on enriching data and enabling Toast teams

“With Monte Carlo, we got the thing **up and running within a few hours** and then let it go...It just cuts down on time. You're **directed exactly toward what the problem could be**, and from there, you can expand.”

Angie Delatorre
Data Engineer

MC

Customer Success Story



Overview

Solution that provides retailers with an all-in-one returns platform to delight customers, drive revenue, and preserve the planet

Industry

SaaS (Tech), E-commerce

Data stack & comms channel



Situation

- **Context:** high data volumes, with regular updates generated by their customers inventory tracking systems feeding into their application to identify areas for improvement
- **Challenge:** prevent data quality issues from negatively impacting the customer experience
 - The need to alert on stale data and pipeline changes such as schema drift
 - Ongoing monitoring, with new rules automatically being generated to handle new datasets
- **Solution:** data observability implementation that provided:
 - Automated lineage graphs to help quickly identify how data flows and where resolution is needed
 - Out of the box monitors that create checks freshness, volume, and schema changes based on historical behavior

Results

Save engineering time spent on issue discovery

50% of engineering capacity recovered from faster incident resolution

End-to-end visibility

Ability to quickly deploy Monte Carlo across all data marts and tables in the warehouses and BI tools to understand both downstream and upstream impact

Increased customer trust in results

Proactively discovering when data quality has been affected to quickly resolve customer-facing incidents

“ **Being able to quickly identify client-facing issues and be proactive is really the key to building trust in our data.** And this feature makes the data engineers’ jobs much much easier—I can tell you definitely from experience here.

Patrick Campbell
Lead Data Engineer

Customer Success Story



Overview

A new, smarter take on auto insurance that promises fast, hassle-free and easy to understand plans

Industry

Fintech (insurance)

Data stack & comms channel



Situation

- **Context:** 50+ data sources being brought into their data warehouse along with multiple teams working on modeling and curating data
- **Challenge:** keeping up with a proliferation of data generated by multiple business departments without requiring data source domain expertise
 - Data stack also continued to evolve with new orchestration, transformation, and warehouse tooling
- **Solution:** deploying Monte Carlo across every table to identify trends in data resulting in:
 - Reduced implementation times with out of the box coverage for new datasets
 - Alerting with customized tags to a dedicated Slack channel
 - Automated data lineage creation to isolate and remediate incidents
 - Custom field health thresholds to meet internal SLAs

Results

Holistic dataset coverage

70% increase in data quality coverage

Solve data issues faster

50% faster resolution times stemming from common changes such as schema drift or additional sources

Futureproof their data strategy

Out of the box ML monitors deployed across their dataset in days ensures that all future initiatives are covered

“ We **no longer had to tailor specific tests to every particular data asset.** All we really had to do was sign up, add the security implementation to give Monte Carlo the access that it needed, and we were able to start getting alerted on issues. **Monte Carlo gave us that right out of the box.** ”

Braun Reyes

Senior Manager of Data Engineering

Customer Success Story



Overview

Software that provides an intuitive, visual way to discover, consolidate, and communicate user needs.

Industry

Tech (SaaS)

Data stack & comms channel



Situation

- **Context:** a need to monitor end-to-end pipelines, including marketing data, being hosted and analyzed in Segment, their customer data platform
 - Other data sources include CRM, digital ads, and payments software
- **Challenge:** in-house built solution through dbt frequently had alerting delays, and upstream tool notification settings were too infrequent
 - Another recent example includes a broken link in a live customer survey, requiring immediate attention to ensure a successful campaign
- **Solution:** evaluated data observability solutions requiring:
 - Timely incident alerting
 - Root cause analysis with field-level lineage for speedy resolution
 - Field Health

Results

Decrease inefficient spend on data tools

3x reduction in infrastructure costs through isolating an unexpected spike in marketing events

Faster time-to-detection (TTD)

99% faster incident detection(**2 hours** vs default notification time of **8 days**)

In-app resolution tools

End to end lineage graphs used for root cause analysis

“**Monte Carlo gives us the power to know what’s going on with our data at any given point in time so we can ask the right questions when data downtime strikes**, for instance ‘we think something’s wrong here, did you change anything, or is this expected?’”

Pablo Recio

Senior Data Engineer

Customer Success Story



Overview

With over 16 million users worldwide, Blinkist helps time-strapped readers fit learning into their lives through their ebook subscription service

Industry

Media

Data stack & comms channel



Challenge

- **Context:** 40% YoY revenue growth target to be fueled by:
 - Increased marketing spend
 - User experience campaigns
 - Faster innovation cycles
 - A/B testing and experimentation
- **Challenge:** CXOs and campaign managers were reliant on data to make decisions, but data engineering team **spent 50% of their time firefighting data drills**
- **Solution:** selected Monte Carlo to adopt and meet data reliability engineering principles:
 - Data governance
 - Data quality
 - Refactoring systems

Results

Higher engineering productivity

120 engineering hours saved per week

Increased revenue

Highly reliable data, including **new data SLAs**, ensured marketing spend was allocated appropriately

Increased data product adoption

“ The self-service capabilities of data observability helped **build back trust in data**, as users were seeing us in action: going from a red alert to a blue “work-in-progress” to “resolved” in green. They knew who was accountable, they knew the teams were working on it, and **everything became crystal clear**.

Gopi Krishnamurthy
Director of Engineering

Customer Success Story



Overview

Software that provides an intuitive, visual way to discover, consolidate, and communicate user needs.

Industry

Tech (SaaS)

Data stack & comms channel



Situation

- **Context:** a need to monitor end-to-end pipelines, including marketing data, being hosted and analyzed in Segment, their customer data platform
 - Other data sources include CRM, digital ads, and payments software
- **Challenge:** in-house built solution through dbt frequently had alerting delays, and upstream tool notification settings were too infrequent
 - Another recent example includes a broken link in a live customer survey, requiring immediate attention to ensure a successful campaign
- **Solution:** evaluated data observability solutions requiring:
 - Timely incident alerting
 - Root cause analysis with field-level lineage for speedy resolution
 - Field Health

Results

Decrease inefficient spend on data tools

3x reduction in infrastructure costs through isolating an unexpected spike in marketing events

Faster time-to-detection (TTD)

99% faster incident detection(**2 hours** vs default notification time of **8 days**)

In-app resolution tools

End to end lineage graphs used for root cause analysis

“**Monte Carlo gives us the power to know what’s going on with our data at any given point in time so we can ask the right questions when data downtime strikes**, for instance ‘we think something’s wrong here, did you change anything, or is this expected?’”

Pablo Recio

Senior Data Engineer

MC

Customer Success Story



Overview

A new, smarter take on auto insurance that promises fast, hassle-free and easy to understand plans

Industry

Fintech (insurance)

Data stack & comms channel



Situation

- **Context:** 50+ data sources being brought into their data warehouse along with multiple teams working on modeling and curating data
- **Challenge:** keeping up with a proliferation of data generated by multiple business departments without requiring data source domain expertise
 - Data stack also continued to evolve with new orchestration, transformation, and warehouse tooling
- **Solution:** deploying Monte Carlo across every table to identify trends in data resulting in:
 - Reduced implementation times with out of the box coverage for new datasets
 - Alerting with customized tags to a dedicated Slack channel
 - Automated data lineage creation to isolate and remediate incidents
 - Custom field health thresholds to meet internal SLAs

Results

Holistic dataset coverage

70% increase in data quality coverage

Solve data issues faster

50% faster resolution times stemming from common changes such as schema drift or additional sources

Futureproof their data strategy

Out of the box ML monitors deployed across their dataset in days ensures that all future initiatives are covered

“ We **no longer had to tailor specific tests to every particular data asset.** All we really had to do was sign up, add the security implementation to give Monte Carlo the access that it needed, and we were able to start getting alerted on issues. **Monte Carlo gave us that right out of the box.** ”

Braun Reyes

Senior Manager of Data Engineering

MC

40

Customer Success Story



Overview

Subscription-based, online fresh pet food service

Industry

E-commerce

Data stack & comms channel



Situation

- **Context:** as the data strategy has evolved, so has the number of tools involved in the entire data stack, and the number of downstream team impacted by data delivery
- **Challenge:** a desire to improve visibility into how to identify, receive notifications for, and resolve data issues that may occur across the entire pipeline
 - Future-proofing trust in the data team by providing dependent teams on the ability to troubleshoot their own data issues
- **Solution:** evaluated data observability solutions requiring:
 - End-to-end visibility into the entire pipeline, to discover issues that were previously unknown
 - Ease of deploying the solution, including notifications into their preferred communications channel

Results

Save engineering resources

6 months saved time versus building an in-house tool

Broader incident coverage

Monitoring data and movement between sources provided insights into issues that otherwise wouldn't have been discovered

Improved self-service troubleshooting

Lineage graphs are used by others outside of the data team to identify where the issues are occurring

“ **Monte Carlo has been exceptional at catching upstream errors.** And being able to tell you that something is amiss and guide your focus with such precision! **Pointing not only to the specific table but the lineage that quickly gets you to the root of the error.** It's just absolutely incredible.

Rick Sapporta

Head of Data Strategy and Insights