rudderstack

The Data Maturity Guide

VERSION 2.0: HOW TO BUILD A PRACTICAL CUSTOMER DATA INFRASTRUCTURE IN 2025

Table of Contents

Phase 1 | Collection: Building the foundation

The journey toward data maturity	05
Why getting this right matters for your business	06
When to make the investment	07
Impact on the organization	08
Implementation strategy	08

Phase 2 | Centralization: A single source of truth

The Centralization phase solution	12
When to make the investment	13
Impact on the organization	13
Implementation strategy	14

Phase 3 | Machine learning: Getting predictive

Moving from historical to predictive	17
When to make the investment	18
Impact on the organization	19
Implementation strategy	20

Phase 4 | Real-time: The final frontier

The power of immediate action	24
When to make the investment	25
Impact on the organization	26
Implementation strategy	26

Data maturity assessment

Where are you in your data maturity journey	29
How to use this assessment	30

Key takeaways

After reading this guide, you'll have a solid understanding of the different phases of customer data maturity and clear direction for how to progress along your journey.

Here are a few key insights:

Oftentimes, customer data needs evolve through different stages, each requiring specific infrastructure capabilities that must meet business needs.

Rather than implementing the entire data stack all at once, focus on solving your current challenges while building the foundation for future capabilities—in other words, a good foundation is crucial.

Data governance and compliance must be foundational at every stage, not afterthoughts.

The right infrastructure empowers data teams to deliver value faster while reducing engineering overhead.

A staged approach maximizes ROI while minimizing disruption to existing operations.

Are you ready to take a more practical approach to customer data infrastructure that will solve your challenges today and scale with you as you grow? Let's explore where you are in your journey and how to advance to the next stage.



Introduction: Build a practical customer data infrastructure that scales with your business

The strategic use of customer data has become a critical competitive differentiator for organizations across every sector and scale, directly influencing market performance and growth potential. However, many organizations struggle with fragmented data infrastructure that can't scale to meet business demands for governance, compliance, and advanced use cases

The customer data ecosystem began as a simple solution but has grown into a sprawling landscape of specialized tools, leaving business leaders uncertain about which path to take to grow their business. What's needed is a pragmatic, enterprise-grade approach to building data infrastructure that evolves with your organization's needs.



This guide presents a four-phase framework for customer data maturity:

This framework provides a practical roadmap to help you identify your current state, optimize existing investments, and build robust data infrastructure to support your business objectives while maintaining strict governance and compliance requirements.

Wondering where you are in the data maturity journey?Take the assessment on page 29!

collection and integration

Phase 1

Getting started: Build your foundation with streamlined collection

The journey toward data maturity

Data maturity doesn't happen in a vacuum. The stages are interconnected, and while this concept is often positioned as a single major implementation of various tools, it's best approached as a practical *journey*. In other words, you don't have to do it all at once! You can take it one step at a time: Solve today's problems, do it methodically, and aim for future flexibility.

Customer data infrastructure should grow and evolve alongside your business needs. For data teams doing the work, the goal isn't adopting a particular architecture merely because it's considered "modern" at that moment–it's *progressively* evolving your infrastructure to impact the business's ability to make better decisions.

Modernization is a journey with many steps as your business grows. *The good news*? You don't have to make overwhelming changes. You can start enhancing your infrastructure wherever you are today, gradually building toward more advanced capabilities.

Beginning your data maturity journey



Collection is the first phase of your data maturity journey. This is where most companies establish their initial infrastructure or take their first step in modernizing existing systems, whether that's a collection of Google Tag Manager implementations, basic ETL jobs, manual CSV exports, point integrations, or brittle DIY pipelines.

This phase isn't about company size. Rather, it's about addressing fundamental challenges: data integration and consistency problems through a unified data layer.

Why getting this right matters for your business

After reading this guide, you'll have a solid understanding of the different phases of customer data maturity and clear direction for how to progress along your journey. Here are a few key insights:

- Data consistency: Unifying event data collection and routing standardized data downstream enables every team to align around the same data because there's consistency across tools and systems.
- **Integration complexity:** A centralized system for event collection and routing streamlines pipeline management and ensures reliability.
- Real-time data activation: Reliably connecting event data from various sources to downstream business tools enables teams to trigger activations on customer signals in real time.

Consider this scenario:

The user record in your email marketing system shows different information than your product analytics tool (e.g., it's **user_ID** in one system, but **userID** in another). This could be because of inconsistent event names, missing properties, or time discrepancies, and that's just scratching the surface. With inconsistencies like this, it becomes impossible to confidently answer even basic questions.

Without solving data consistency and integration at the root, it doesn't matter what kind of advanced technology you use in other parts of the stack. Other tools will have fundamental limitations because they are only as good as the data you feed them.

This foundation is crucial, especially for businesses that operate in regulated industries. Because inconsistent data not only hinders analytics. It also creates compliance risks. By solving data quality issues at the source, you create infrastructure that can scale with governance controls intact.

When to make the investment

The Collection phase is the foundation for all future data maturity efforts, but timing this investment correctly is crucial for maximizing impact. Consider implementing the Collection phase when your organization experiences:

- **Tool proliferation:** Your marketing, product, and engineering teams have adopted multiple SaaS tools, each requiring its own data pipeline.
- **Data inconsistencies:** The same metrics show different values across tools, creating confusion and eroding trust.
- Engineering bottlenecks: Your development team spends significant time maintaining custom integrations or implementing tracking requests.
- Compliance concerns: Privacy regulations require more control over what data is collected and how it's distributed.
 - **Activation limitations:** Teams can't respond to customer behaviors in real-time due to fragmented data flows.

For startups and smaller organizations, implementing the Collection phase early creates a scalable foundation that prevents technical debt. For larger enterprises, it's often the first step in modernizing legacy systems that have accumulated over time.

The key indicator:

When your teams spend more time debating whose numbers are correct than acting on the insights those numbers should provide, it's time to invest in a unified data layer.



Impact on the organization

A well-implemented collection layer transforms how your organization operates with data:

- Analytics alignment: Marketing, product, and executive teams work from consistent numbers, eliminating debates about whose data is correct.
- Engineering efficiency: Development teams reduce time spent on integration maintenance by 50-70%, freeing resources for product innovation.
- **Faster implementation:** New tools can be added to your stack in days rather than weeks, accelerating experimentation.
- Improved compliance: Centralized control over data collection creates consistent privacy standards and simplified audit trails.
- **Enhanced customer experiences:** Real-time event streaming enables immediate responses to customer behaviors across channels.

Beyond these operational benefits, the Collection phase creates a foundation of trust in your data. When teams know they're working with reliable information, they can make data-driven decisions confidently and rely less on intuition or opinion.

Perhaps most importantly, this phase sets the stage for future growth. By solving fundamental data consistency and integration challenges now, you create the infrastructure needed to support more advanced use cases as your business evolves—without requiring painful re-architecting later.

Implementation strategy

For data teams, the Collection phase implementation focuses on these principles:



This approach transforms enterprise architecture from point-to-point integrations to a controlled, centralized model that satisfies both technical and compliance requirements.

By establishing this foundation correctly, your teams can progressively build more sophisticated capabilities while maintaining data quality, governance, and compliance throughout their journey. In other words, **you can trust your data**.

Canada Drives: Building a strong data collection foundation

The challenge of online vehicle sales

Canada Drives, Canada's largest 100% online car shopping platform, faced unique data collection challenges in their business. Unlike traditional e-commerce with frequent repeat purchases, vehicle sales involve complex, high-value customer journeys where potential buyers often research anonymously for weeks before identifying themselves.

"Users don't identify themselves when they assess financing or calculate the value of a vehicle they're selling or trading in," explains Andrew Hall, VP of Data and Analytics at Canada Drives. "We needed better analytics tools to generate actionable insights into customer behavior at every level of our admittedly steep and unique funnel."

Implementing unified data collection

To address these challenges, Canada Drives implemented RudderStack, focusing first on establishing a strong event collection foundation:

- **Standardized event tracking:** Using RudderStack to capture anonymous and known customer interactions
- Sum Unified data layer: Creating a central collection point for all customer events
- Real-time data activation: Routing this data to both analytics platforms and operational tools
- Integration with warehouse: Establishing Snowflake as their source of truth for all customer data
- **Enhanced customer experiences:** Real-time event streaming enables immediate responses to customer behaviors across channels.

Measurable business impact

By building a strong Collection phase foundation, Canada Drives achieved significant results:



Enabling both immediate activation and long-term analytics

Setting the stage for more advanced data capabilities

By getting the fundamentals right with a unified data collection layer, Canada Drives created the foundation for their subsequent data initiatives, including centralization in Snowflake and sophisticated recommendation models that further improved their business outcomes.

Moving beyond basic integration

With accurate data visibility across tools established in the Collection phase, your teams can now understand what's working and what's not. As your organization grows, you'll face new challenges as teams begin asking more sophisticated questions that require additional data beyond basic user events.

At this point, multiple stakeholders-including marketing, product, and executive teams-will push for more sophisticated data insights:

Marketing teams will move beyond firsttouch attribution to explore metrics like CAC and ROAS by channel and campaign.

8

Product teams will want to analyze customer segments and behaviors in greater detail to guide feature development.

.....

Executive leaders will expect comprehensive financial reporting combining multiple data sources for strategic decision-making.

These advanced questions require additional data that your unified collection layer alone can't provide. This is when it's time to move from the Collection phase to the Centralization phase. (In fact, an alternate title of this section could be: *When to go beyond the Collection phase*)

Phase 2 Centralization: A single source of truth

			Analytics	
	Data Laver		 Product analytics 	
		\mathbf{r}	 Marketing analytics 	
Data Sources	Data Transformation	· ·	• Business Intelligence	
	Storage			
		· · · · · · · · · · · · · · · · · · ·	Activation	
Saas	Cloud data platform		Email marketing	
	Cloud data warehouse		• CRM	
			Customer support	
			Ad platforms	

The Centralization phase solution

The Centralization phase is where you'll build on your unified data collection to establish a single source of truth. This phase introduces bi-directional data flows and a centralized data warehouse or data cloud to address two critical problems:

Incomplete data: Valuable information created in downstream tools remains trapped in those systems.

Activation barriers: Teams struggle to use centralized insights for timely action.

By implementing a data cloud as your central storage/compute layer and adding ETL pipelines from your SaaS tools, you create a comprehensive source of truth combining event data, CRM records, payments data, and other customer information.

More importantly, the Centralization phase involves activating this data through reverse ETL, pushing insights from your warehouse back to business tools where teams can take action.

When to make the investment

You know you're ready for the Centralization phase when:

- Teams hit the limits of their SaaS analytics tools
- Reporting requires combining data from multiple disconnected systems
- Stakeholders need complex customer segments that aren't possible in existing tools
- Manual exports and spreadsheets become unsustainable

For organizations of all sizes, from mid-market to enterprise, the Centralization phase provides the governance and scale needed to manage complex customer data while meeting compliance requirements that are impossible to manage across fragmented stacks.

Impact on your organization

When properly implemented, the Centralization phase delivers significant benefits:

- No more data silos! Elimination of data silos across the entire business.
- Single source of truth: A single, governed source of truth for all customer data
- Full customer 360 views: Complete visibility into customer journeys
- Accessibility across the org: Self-service analytics for business teams-plus, the ability to activate customer segments across tools

This infrastructure enables your organization to break through the barriers of localized optimization, so every team can benefit from the full power of your customer data—while also maintaining consistent governance controls.

Implementation strategy

Building a successful Centralization phase infrastructure involves several key steps:

1 Select and configure your cloud data warehouse Choose a modern cloud warehouse like Snowflake, BigQuery, or Redshift • Set up appropriate data access controls and governance Establish consistent naming conventions for schemas and tables 2 Implement ETL pipelines from key data sources • Connect your SaaS tools (CRM, support, marketing platforms) to your warehouse Set up appropriate sync schedules based on data freshness requirements Document transformation logic for data normalization 3 Centralize your event stream data Configure your event collection system to load data into your warehouse Create consistent schemas for events and user properties • Establish identity resolution processes to connect users across systems 4 Build your customer data models Create unified customer profiles by joining data across sources • Develop and document business-relevant metrics and KPIs Document data lineage for transparency and trust 5 Implement reverse ETL capabilities Set up pipelines to sync enriched data back to operational tools

- Create processes for audience/segment management
- Establish monitoring to ensure data freshness

Throughout implementation, focus on creating documentation, building team capabilities, and establishing governance processes that will scale with your organization.

Centralizing customer data: How InfluxData built their single source of truth

When organizations hit the limits of fragmented analytics, centralizing customer data becomes essential. InfluxData's journey illustrates how implementing a proper data foundation can transform operations and enable advanced analytics capabilities.

The challenge: Fragmented data creating business friction

InfluxData faced a common data maturity challenge: siloed information spread across PostgreSQL databases, SaaS applications, and analytics tools. This fragmentation created significant business problems:

- Teams wasted 5-10 hours weekly reconciling conflicting data
- Meetings began with debates about whose numbers were correct
- Engineers spent valuable time building and maintaining custom integrations
- Analysts couldn't deploy proper ETL processes for downstream applications

"We were wasting 10 minutes at the start of every meeting debating whose numbers tracked instead of talking about things that mattered".

"We needed consistent metrics, guardrails around data definitions, and better data hygiene".

Mona Sami Director of Data Analytics

🏈 influxdata°

The solution: Building a centralized data foundation

InfluxData implemented a modern data stack with RudderStack's customer data platform and Snowflake as their central warehouse. This combination allowed them to:

- Route data from PostgreSQL to Snowflake using a custom connector
- Establish bidirectional data flows between all systems
- Implement proper data governance and version control
- Enable analysts to manage ETL processes without engineering support

This approach allowed InfluxData to solve the fundamental data collection challenges that previously hindered their understanding of the complete customer journey.

The new infrastructure created a single source of truth while giving the analytics team complete control over data pipelines. "Instead of hiring and training a full-time senior engineer to build out reliable pipelines, we use RudderStack ETL tools out of the box to get the same results," notes Sami.

The results: From reconciliation to innovation

With a solid data foundation in place, InfluxData evolved from spending time reconciling numbers to leveraging data for business innovation:

- Expanded from basic ETL to implementing sophisticated event tracking
- Gained visibility into complete customer journeys across product and marketing touchpoints
- Implemented A/B testing to optimize user experiences
- Freed engineers to focus on product development rather than data integration

The transformation illustrates a key principle in the data maturity journey: establishing a centralized data foundation doesn't just solve immediate problems—it creates the capacity for teams to advance to more sophisticated analytics capabilities.

"Our primary motivation in choosing RudderStack was building ETL pipelines," Sami explains. "But additional use cases quickly presented themselves."



Phase 3 Machine Learning: Getting predictive

Moving from historical to predictive

After mastering deterministic analysis in the Centralization phase, the next phase requires anticipating future events and acting *proactively*. When your teams begin thinking about predicting customer churn, making personalized recommendations, or customizing experiences based on likely behaviors, you're ready for the Machine Learning (ML) phase.

This progression follows a natural evolution in data sophistication—from understanding **"What happened?"** to exploring and acting on, **"What will likely happen next?"** This shift transforms how teams operate, moving from reactive responses to proactive interventions that can impact business outcomes before problems even materialize.



The next level of optimization

The ML phase introduces a data flow enabling predictive analytics by addressing two fundamental challenges:

- Warehouse limitations: Data teams hit ceilings when they can't perform desired analysis in SQL or need to work with unstructured data
- **Operationalizing insights:** Making model outputs actionable in business tools requires specialized infrastructure

This phase uses the lakehouse architecture of a modern data cloud to store and process both structured and unstructured data (e.g., support transcripts, call recordings), enabling modeling tools to generate predictive outputs directly within a unified pipeline.

The introduction of a data lake is particularly transformative because it opens up entirely new categories of data for analysis. While structured transactional and behavioral data provides valuable signals, unstructured data often contains the richest insights about customer sentiment, pain points, and future intentions. Customer support interactions, product reviews, social media comments, and other text-heavy data sources become accessible for analysis when properly incorporated into your ML infrastructure.

When to make this investment

Not every predictive question requires formal machine learning. Simple models or regression analysis can often provide significant value. But you may be ready for the ML phase when:

- Teams want to proactively influence events before they occur
- You've hit the limits of SQL-based analysis
- Your organization needs to leverage unstructured data
- Specific use cases demonstrate clear ROI potential

The ML phase typically becomes relevant for larger mid-market and enterprise companies with significant data volume and dedicated technical resources to work on predictive initiatives.

Organizations with customer-centric business models often find the greatest initial value in ML implementation. Subscription businesses can predict and prevent churn, e-commerce companies can optimize product recommendations, and service businesses can forecast demand patterns to improve resource allocation.

Building a practical ML solution

Instead of implementing complex ML operations infrastructure, most organizations start with a simpler approach that leverages existing components:

- Implement a data lake for unstructured data or take advantage of this in your existing data cloud
- Add modeling capabilities that connect to your data sources
- Develop features and train models
- Push outputs to your warehouse as tables
- Activate insights through your existing reverse ETL pipelines

This approach enables you to deliver value from predictive analytics without extensive new infrastructure. The warehouse and reverse ETL components from your Centralization phase become the delivery mechanism for model outputs.

What makes this approach particularly effective is its pragmatic focus on outcomes rather than infrastructure complexity. By leveraging existing infrastructure and activation pathways, organizations can quickly demonstrate value from ML investments without lengthy implementation cycles or specialized ML operations expertise.

Impact on the organization

When implemented effectively, the ML phase enables your organization to:

- Mitigate problems proactively instead of reactively
- Leverage data that was previously unusable (e.g. if it was unstructured)
- Deliver insights that drive significant business outcomes

With predictive capabilities, you can identify at-risk customers before they churn, recommend products customers are likely to purchase, and optimize operations based on anticipated demand. These insights can revolutionize how your business operates. Beyond these specific applications, the ML phase creates a foundation for more sophisticated Al applications as your organization's needs evolve. The data infrastructure, governance processes, and technical capabilities established at this stage provide building blocks for increasingly advanced use cases.

ML capabilities also transform how teams think about their work. Rather than constantly reacting to past events, they begin to orient around shaping future outcomes—a fundamental shift that can create substantial competitive advantages in markets where customer experience and operational efficiency are key differentiators.

Implementation strategy

Implementing ML capabilities requires thoughtful planning and execution to ensure success. Follow these steps to build a foundation for predictive analytics:

1

2

3

4

Identify high-value use cases

- Begin with specific business problems that have significant ROI potential
- Prioritize use cases where proactive intervention delivers measurable value
- Consider starting with churn prediction, propensity modeling, or product recommendations

Establish your data lake infrastructure

- Enable unstructured data support in your data cloud or select a cloud data lake solution compatible with your existing infrastructure
- Configure appropriate access controls and governance
- Develop processes for ingesting and cataloging unstructured data

Implement data extraction and preparation pipelines

- Build processes to extract text from documents, transcripts, and other unstructured sources
- Develop natural language processing capabilities to derive insights from text
- Create feature stores to make model inputs easily accessible and reusable

Develop model building capabilities

- Start with simple modeling approaches before progressing to more complex techniques
- Establish clear metrics to evaluate model performance
- Build frameworks for model validation and testing

Create feedback loops for model improvement

- Implement logging to track model predictions and actual outcomes
- Establish processes for regular model retraining and improvement
- Develop monitoring to detect model drift or performance degradation

6 Build activation pathways

5

- Create data pipelines to output model predictions to your warehouse
- Configure reverse ETL workflows to send predictions to business tools
- Develop documentation to help business teams understand and act on predictions

Throughout implementation, focus on creating tight feedback loops between data science and business teams. The most successful ML implementations continually refine models based on real-world performance and evolving business needs.

Wyze: Leveraging the ML phase for personalized customer experiences

The challenge of predictive analytics

Smart home technology leader Wyze had successfully centralized their customer data in Snowflake, but faced a critical challenge that exemplifies the transition into the Machine Learning phase: how to transform their vast repository of customer data into predictive insights that could drive personalized experiences and increase customer lifetime value.

With tens of millions of customers generating data across multiple touchpoints-website, mobile apps, IoT devices, and Amazon storefront-Wyze needed to go beyond historical analysis to predict customer behaviors and preferences. However, they faced significant hurdles:

Inability to create unified customer identities across disparate platforms

Difficulty extracting meaningful signals from complex event data

Challenges in operationalizing insights for marketing activation

Limited ability to leverage ML models for personalized recommendations

Building an ML-powered infrastructure

Wyze's implementation of the Machine Learning phase demonstrates the key components needed for success:

- Identity resolution: Using RudderStack Profiles to automatically generate identity graphs that connect users across all touchpoints
- Feature engineering: Streamlining the creation of ML-ready user features from raw event data
- **Model development:** Enabling their AI/ML team to quickly build and deploy predictive models for churn, lifetime value, and product recommendations.
- **Activation workflow:** Creating a seamless pipeline to deliver model outputs to Braze for personalized marketing campaigns.

This infrastructure represents the essence of the Machine Learning phase– transforming centralized customer data into predictive insights that can be operationalized across the organization.

Measurable business impact

Wyze's implementation of ML capabilities delivered dramatic results:



🖄 Significantly shortened time from idea inception to campaign testing and deployment

"When you have the power of RudderStack in hand, you can blast off right away."

"It's so much easier to build a machine learning model once your designs are driven by clean data, useful user features, and 360 customer views."

Wei Zhou Director of Data Engineering at Wyze

WYZE

The Machine Learning phase in action

Wyze's journey illustrates how organizations in the Machine Learning phase can transform their approach to customer data:

- Expanded from basic ETL to implementing sophisticated event tracking
- Gained visibility into complete customer journeys across product and marketing touchpoints
- Implemented A/B testing to optimize user experiences
- Freed engineers to focus on product development rather than data integration

By implementing this sophisticated ML infrastructure, Wyze hasn't just improved their analytics capabilities—they've created a foundation for AI-driven customer experiences that deliver measurable business results through personalized engagement and targeted offerings across all customer touchpoints.

Phase 4 The final frontier: Real-time phase

The power of immediate action

For many organizations, the ML phase represents the final stage of data maturity. However, companies with millions of customers may find significant value in taking this one step further: enabling real-time personalization based on immediate user behavior.

The Real-time phase delivers on the same fundamental outcome as previous phases– optimizing the customer journey–but makes it instantaneous, customizing and personalizing experiences as they happen.



Two flavors of real-time

Real-time personalization typically manifests in two forms:

Next behavior personalization:

Delivering optimized experiences when users take future actions (like showing relevant content when a user logs in) -

In-session personalization: Adapting the experience during the current session based on real-time behavior (such as personalized search results)

The first approach requires real-time delivery, but not real-time computation. The second demands both immediate analysis and activation.

Building for true real-time

The infrastructure you'll build in the Real-time phase addresses two critical challenges:

- Data freshness: Real-time personalization requires immediate access to user actions
- Latency requirements: Infrastructure must deliver insights while users are still engaged

To enable these capabilities, this phase introduces a few benefits:

- An in-memory data store for real-time access to customer profiles
- Enhanced ML infrastructure that can serve predictions instantly
- Integrations with your customer-facing applications

When to make this investment

The Real-time phase requires a significant undertaking. Typically, you should invest in progressing to this phase when:

Small conversion improvements drive substantial business impact

- Personalization has reached the limits of batch processing—and thus warrants a more real-time approach
- Vour business is particularly time-sensitive (e.g., media, e-commerce, healthcare)
- Vou have the technical resources to implement customer-facing integrations

Unlike previous phases, the Real-time phase applies primarily to larger enterprises with significant engineering resources and clear use cases with demonstrable ROI.

Impact on the organization

When implemented effectively, the Real-time phase transforms how businesses interact with customers, creating a fundamental shift from reactive to proactive engagement. Marketing teams gain the power to dynamically adjust promotions and offers based on immediate signals, dramatically increasing conversion rates. Product teams can personalize user experiences on-the-fly, reducing friction points and increasing engagement metrics in many cases. Customer service operations benefit from immediate context awareness, enabling representatives to address issues before customers even report them.

Perhaps most significantly, the Real-time phase helps businesses create truly differentiated experiences that competitors operating on batch processes simply cannot match. Organizations with time-sensitive offerings (like flash sales for retailers, breaking news for media companies, or fraud detection for financial institutions) see particularly transformative results, often reducing decision latency from hours to milliseconds. This capability becomes a genuine competitive advantage, as customers increasingly expect hyper-personalized experiences tailored to their immediate context and needs.

Implementation strategy

Implementing real-time capabilities requires collaboration across teams:

- 1 Data engineering establishes the in-memory data store
- 2 Data science develops models that can serve predictions in milliseconds
- **3** Product engineering builds integrations into customer-facing applications

Most organizations implement this phase incrementally, starting with next-behavior personalization before advancing to in-session capabilities.

loveholidays: Real-time personalization at scale

After successfully centralizing their customer data in BigQuery, travel booking site loveholidays faced a critical challenge: how to activate this rich data in real-time to create personalized customer experiences without sacrificing performance.

With 20 million monthly users expecting instant results when searching for vacation packages, loveholidays knew that even small delays in page load or personalization could significantly impact conversion rates. Third-party personalization solutions required API calls that took around 200 milliseconds—an eternity in the competitive online travel space. So, loveholidays took matters into their own hands: They utilized RudderStack and Redis to enable real-time capabilities on their existing infrastructure.

Building a real-time infrastructure

Loveholidays exemplifies the Real-time phase of data maturity by implementing a sophisticated architecture that enables millisecond-level personalization:

- **Event stream data collection:** Using RudderStack's JavaScript SDK to capture user behavior across their sitec
- Centralized data warehouse: Storing and modeling comprehensive customer data in BigQuery
- In-memory data store: Leveraging Redis to make personalization data instantly accessible
- **Reverse ETL pipeline:** Moving enriched customer profiles and models from BigQuery to Redis
- Real-time personalization engine: Building custom logic to serve hotel recommendations in just 20 milliseconds

This infrastructure represents the pinnacle of data maturity–going beyond historical analytics and batch processing to deliver truly real-time experiences based on all available customer data.

Measurable business impact

The results of loveholidays' real-time data infrastructure speak for themselves:



By implementing this sophisticated data infrastructure, loveholidays hasn't just improved their analytics-they've fundamentally transformed how customers experience their service, driving measurable business results through the power of real-time data activation.

The continuous journey

Even for companies that implement all four phases, data maturity remains an ongoing process. New use cases will emerge, technologies will evolve, and business needs will change.

The most successful organizations view data maturity not as a destination, but as a continuous journey of improvement-building capabilities that solve real business problems while establishing the foundation for future innovation.

Data maturity assessment

Where are you in your data maturity journey?

Answer these questions to identify your current phase and determine your next steps toward a more mature data infrastructure.

Collection phase indicators

Do multiple teams have different versions of the same customer data?

Are you struggling to get clean, consistent data into your analytics tools?

Is your engineering team spending significant time on custom integrations?

Do you have difficulty tracking users across different platforms?

Are you unable to answer basic questions about customer journeys?

If you answered **yes** to most of these questions: You likely need to establish a unified data layer to collect consistent customer data across touchpoints.



Centralization phase indicators

Do you have clean data collection, but information remains trapped in separate tools?

Are teams requesting data from other systems they can't access?

Do reports require manual exports and spreadsheet manipulation?

Is there a growing need for custom customer segments?

Are downstream teams unable to act on insights without engineering help?

If you answered **yes** to most of these questions:

You're ready to implement a data cloud with bidirectional data flows.

Ę

Machine Learning phase indicators

Are you looking to predict customer behavior rather than just analyze it?

Do you need to incorporate unstructured data (support transcripts, etc.)?

Are you hitting the limits of what SQL can accomplish?

Do you want to proactively address issues like churn before they happen?

Do you have resources dedicated to data science initiatives?

If you answered **yes** to most of these questions: It's time to explore adding data lake and modeling capabilities.

Real-time phase indicators

Does your business model require immediate personalization?

Would small improvements in conversion rates significantly impact revenue?

Do you need to modify user experiences during the same session?

Can you justify dedicated engineering resources for real-time capabilities?

Have you already implemented predictive modeling successfully?

If you answered **yes** to most of these questions:

You're ready to implement a data cloud with bidirectional data flows.

How to use this assessment

Go through each section honestly assessing your current capabilities

Identify the highest section where you answered "yes" to multiple questions

Focus on completely solving challenges in your current phase before moving to the next

Remember that not every organization needs to progress through all phases

Ready to discuss your results? Contact RudderStack's data infrastructure experts for a personalized consultation on your next steps.



Learn more at https://www.rudderstack.com/contact

Next steps: Your practical path forward

The power of immediate action

The journey to data maturity isn't about implementing the perfect architecture overnight. It's about making incremental improvements that deliver business value at each stage while building toward more advanced capabilities.

By understanding where you are in this journey–whether establishing your first unified data layer, centralizing in a warehouse, implementing predictive models, or enabling real-time personalization–you can focus your investments where they'll make the most impact.

Remember that each organization's path is unique. Some may progress through all four phases, while others may find their optimal solution at an earlier stage. The key is matching your data infrastructure to your specific business needs, technical resources, and growth objectives.

As you advance through your data maturity journey, focus on building strong foundations, maintaining governance throughout, and continuously aligning your infrastructure with measurable business outcomes. This practical approach will help you navigate the complex data landscape while delivering sustainable value for your organization.

Ready to take the next step in your data maturity journey? Contact our team to discuss where you are today and how RudderStack's enterprise-grade infrastructure can help you reach your data goals.

Learn more at https://www.rudderstack.com/contact



