

E-BOOK

The Looker Approach to Modern Data Modeling

Identify problems, automate solutions, and future-proof your data strategy.



Table of Contents

- 3** Introduction
- 4** Walk Away from The Old Way of Data Modeling
- 5** Build a Better Model with Looker
- 6** Three Looker Layers
- 9** ETL, ELT, and Looker: Change the Approach,
Not the Data
- 12** Borrowing from the Software Development Approach
- 14** Cultivating a Data Culture
- 16** Conclusion

Introduction

As an analyst, you've likely spent many hours writing SQL queries to provide data reports for your stakeholders.

Those long hours quickly turn into days, weeks and months. And, when things change—when new data sources are introduced or when once-irrelevant data fields suddenly become important to your queries—you have to start from scratch.

Your original queries are no longer usable. That's because you have to know what you want to extract from the data before you load it into the database and begin a query. And, you have to model your data every time you need to run a report. It's tedious and manual, which leaves you wishing for a better, more flexible way.

Looker's Approach to Data Modelling

We believe there is, with Looker's modular and adaptable approach to data modeling that separates the data model from the database. This enables you to create a model once and build flexible queries on top for reuse across your organization.

When things change in the natural course of business, or you want to look at the data in a new way, you can simply query the data differently rather than recreating the entire data model.

With Looker, you can expect greater efficiency, flexibility, consistency, and relevance. Having a unified truth about your data that everyone across the organization can access drives better business outcomes.



Walk Away from The Old Way of Data Modeling

Typically, when approaching data modeling in the past, you had to know what you wanted to ask your data before, collecting, transforming, and loading it into the target database.

With this approach, modifications to the data model fundamentally can alter the data itself, as data transformation and modeling take place before the data loads into the database. A data set that comes from many sources can further complicate matters.

Back in the day, databases and even raw data storage were very expensive, so limiting available data made sense. Unfortunately, it also forced you to narrow your data collection and limit querying possibilities.

This approach has some other downsides:

- **It's inflexible and brittle.** Changing the data model in the transformation layer is laborious and eats up a lot of time and resources. (See page 11 for the full story about a mobile applications company that used Looker to reduce task completion time from three weeks to less than 20 minutes.)
- **Data gets locked in and stagnant.** It's difficult and time-consuming to introduce fresh, new data sets or change the parameters you impose on your data—like doing data refreshes daily instead of weekly.



The Plummeting Costs of Data Storage

In 1981 it cost

\$500,00

to store 1 gigabyte on a hard drive.

By 2017, storing a single gigabyte cost less than

\$0.03

You can visualize the incredible downward curve of storage costs [here](#).

Build a Better Model with Looker

Approaches to data analysis have evolved. Data storage today is vastly less expensive than in the past, so you can load as much data into the data warehouse as your organization can collect. Today's data warehouses are also much faster and can process raw data with incredible compute speed.

Parallel to these evolutions that give you faster access to more data, Looker has introduced a revolutionary approach to data modeling. Instead of determining the data model and queries before the data load-and-transformation phases, Looker provides a user-friendly, SQL-based layer on top of the data. This lets you easily build a unified data model for your organization and query as flexibly and often as needed. Looker's modular architecture enables you to reuse business logic and write many different types of database queries.

You can easily make minor modifications to your data model that percolate up to your Explore layer. This added functionality makes it easy to adjust complex queries without making any modifications to your data transformation layer. You don't have to learn anything new since you and your team are already familiar with SQL or change how you've been doing things to get better at what you do. You can reduce the time it takes to glean valuable insights from your data from hours or days to just minutes—and it's easy.

To summarize, Looker data modeling technology provides:

- A model once and query many times approach
- Improved efficiency and fewer do-overs
- Power to gather more intelligence from data
- Evolution of a self-service data culture
- Better, smarter business decisions for your stakeholders

Use, reuse, and recycle. With Looker:

- There's a layer on top of the data specifically for modeling and querying.
- You can create components that anyone can use.
- There's no need to reinvent your data model for every use case.

Three Looker Layers

There are three primary layers in Looker, each with a unique function that handles a specific set of workflows.

Transformation Layer

The transformation layer is a data engineering layer. It enables users to perform data transformation functions like setting up sort keys, indexing, and performance-related transformations. As the name implies, this layer handles the transformation of the data that occurs in both traditional Extract, Transform, Load (ETL), and in more modern Extract, Load, Transform (ELT) processes (read more about the evolution of ETL in the next section).

One of the most powerful transformation features is the ability to write a query to transform your data and store the resulting data table outside Looker. Looker stores the information directly in your database so it can be accessed going forward. That materialized table, also known to Looker as a “persistent derived table (PDT)” can be refreshed as new data enters the database on a timetable defined by the user. The next time you query that data, you don’t have to re-transform or re-compute the entire dataset because the PDT can then be queried by other applications.

LookML Layer

The LookML data modeling layer is the heart of Looker, and houses business logic that is consistently reused across the organization. SQL statements written against data via the LookML proprietary modeling language exist at this level. LookML dynamically joins different tables, enabling more complex analysis than traditional tools.

The biggest value driver for Looker’s modeling layer is that it aids collaborative development in real-time for teams. Analysts writing SQL against their database or building workbooks aren’t able to collaborate with each other on projects, which slows development or risks duplicating work. Borrowing from modern software development practices, Looker’s modeling layer contains an IDE (integrated development environment) that makes it easy to make business logic changes. Looker’s IDE seamlessly integrates with version control software, and supports shared branches of code, allowing multiple analysts to work together on one single shared version of LookML (read more about the evolution of ETL on page 9).

The consistency and reusability of the business logic that results from the data modeling layer also are key. For example, some business units in your organization may use the term “profit” in a calculation, while others use the term “net profit.” If the data model is in LookML, calculations are the same for everyone working from the same data set (even with varying terminology). The data represents a single source of truth.



Example: Integrating Business Logic into LookML

Groupon used Looker to determine the inflection point for customer acquisition as analysts experimented in Explore and determined that **after customers attended three events they were highly likely to become long-term customers.**

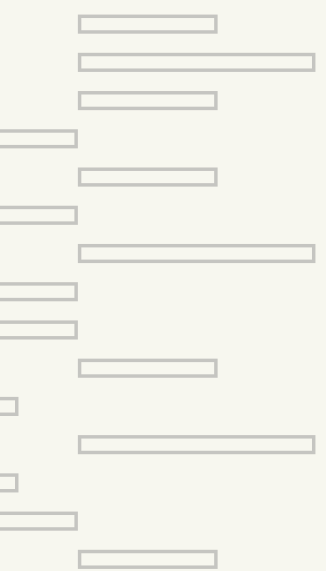
Groupon then moved the business logic into the LookML layer. They were comfortable that this was a static and proven data point for moving forward.

Explore Layer

The explore layer allows you to experiment with ad-hoc queries or perform one-off calculations. The Explore layer looks and feels like Microsoft Excel that includes analytic functions and output tables. If you aren't ready to permanently add something to your data model, this is the layer for iteration.

For example, you may want to assign rankings to each of your customers based on certain attributes, such as loyalty score, share of revenue, most recent interaction, etc. In the Explore layer, you can experiment with metrics and thresholds to determine the most accurate and relevant queries.

Once you've settled on the key customer metrics and a way to define rankings consistently so customer success, sales, and demand-generation teams are all ranking customers the same, you are ready to add this information to your data model by moving it into the next LookML data modeling layer.



USE CASE EXAMPLE

Defining a Customer “Lead”

Your organization may not have a standardized definition of a “lead,” because it may differ within sales, marketing, and finance. The way these groups define “lead” affects how they approach, categorize and query the data. It also affects how they report the efficacy and ROI of campaigns, customer spend, and other metrics.

The traditional data modeling and analysis approach imbedded the definition of a lead into the lowest level of data collection and modeling or defined it at query runtime, every time SQL statements ran. It impacted which data was collected as well as the method of query. If the definition was changed even slightly, it could fundamentally alter the underlying data.

With Looker, you can model the definition of a lead in a variety of ways and experiment with queries in the Explore layer.

Different groups can separately and collaboratively experiment with slightly different ways to slice and dice the definition of a lead within the data set.

When you land on the optimal, universally accepted definition, you have not lost time or resources by including it in an expensive, inflexible data model. Downstream reports can shift easily and flexibly as you gain cohesion around definitions of a lead, and the resulting analysis will be more useful to the entire organization.

ETL, ELT, and Looker: Change the Approach, Not the Data

The most important differences between the ETL and ELT approaches within Looker are where and when data transformation occurs and how to create a data model that enables flexible analysis.

ETL

In the past, systems were not scalable and storage was expensive. Transformation occurred early in the data lifecycle and low in the stack for processing efficiency based on predefined business logic and parameters, tightly coupled with specific queries. It was difficult to change or adapt the business logic later in the process or be flexible and nimble in changing query sets.

Essentially, the “Transform” stage was where business logic calculations were applied, with metrics and data definitions included in the bottom of the data stack.

Moving to ELT

You can load first and model your data afterward. With ELT, the bulk of data transformation occurs in the final target environment (databases and business intelligence tools) at query time. ELT addresses some of the limitations of ETL, as transformation moves to the query layer.

The limitation of ELT is that your organization doesn't have a unified, consistent understanding of which metrics are the “actual” metrics -- and data are modeled at query time, every single time.



Bonobos: A Single Source of Truth

This men's clothing retailer was launched online in 2007. With its digital focus, Bonobos could track every touchpoint of the customer journey across all devices and channels. However, Bonobos lacked a center of data truth, because everyone in the organization did their own analysis with different results.

Siloed data sources left different departments using different data sources and basing results and interpretations accordingly. Insights derived from these divergent data sources did not present a reliable or coherent picture of what was really happening with the business.

After implementing Looker, all employees can agree on and access a centralized and reliable data source. They are also aligned on key metrics and have a 360-degree view of the customer journey—revenue, conversion rates, and channel sources—from a single source of truth. With Looker, Bonobos has a clearer vision of its business strategy going forward.

The Looker Approach

Drawbacks exist whether you apply transformations at the lowest layer with ETL or at the query layer with ELT. The Looker alternative provides a layer in between, with the governance and consistency of the Transformation layer and the flexibility of the Explore layer. It's a "model once, query many times" approach.

The Explore layer enables all stakeholders to look at the same, unified data, ensuring a single source of truth. Decisions about business logic can be made when the data reaches its target environment, with experimentation available in both the LookML and Explore layers.

Data integrity is maintained regardless of how calculations are done, or how non-technical the stakeholder exploring the data is. With data unfrozen at the database level, you have the freedom and flexibility to make queries on the fly in the Explore and LookML layers. And, you can move some of those decisions into the LookML or Transformation layers when you choose.

Having more flexibility as an organization to experiment and pivot around your analytics lets you land on the most useful and revealing queries with the highest ROI.

With a single, centralized source of truth for all data, the amount of time spent on data analysis is reduced from hours to minutes. Everyone can access the same data and the right data, which improves the quality of insights and allows more time to work on more strategic data analysis projects.

USE CASE EXAMPLE

Rapid Acceleration with Looker: From Three Weeks to 20 Minutes

Your organization may not have a standardized definition of a “lead,” because it may differ within sales, marketing, and finance. The way these groups define “lead” affects how they approach, categorize and query the data. It also affects how they report the efficacy and ROI of campaigns, customer spend, and other metrics.

The executive board of a mobile business applications company used to ask its in-house data analyst to calculate a huge number of metrics from a spreadsheet with 6,000 data points.

These metrics ranged from the number of website visitors by country, day, and sales funnel stage to revenue by day, month and customer type. The analyst locked himself in a conference room for three straight weeks and wrote thousands of custom SQL queries to extract answers from the spreadsheet.

The organization now uses Looker, and that same request takes the data analyst less than 20 minutes—a 1500-times acceleration.

Looker enabled the customer to completely model 100% of their data only once instead of modeling it repeatedly, as they would with other tools.

Additionally, LookML dynamically joins different tables into a data model. For example, by joining company and revenue tables, the organization now makes flexible queries and has evolved to a more relevant data model, with the LookML generator writing much of the SQL code itself, freeing the analyst from writing many queries.

Today, that same data analyst runs reports on the dimension fields the board requests and has answers in minutes, giving him time to focus on strategic data initiatives.

Borrowing from the Software Development Approach

For decades, software engineers have coalesced around a set of best practices for collaboration that is both cultural and practical. Looker has taken the processes and best practices from software and applied them to solving the challenges of data modeling. Embracing this model in terms of how to agree on a data model as an organization and how to approach and implement Looker as a data analysis tool ensures best practices for data modeling.

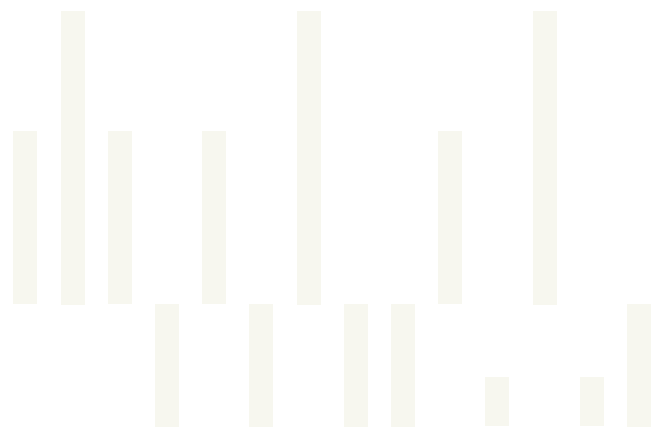
Success with analytics requires everyone in your organization to agree on what data represents and what you are asking of it — but getting to that agreement can be tricky. In open source projects, disparate teams and individual software developers committing code to the same repository must agree on best practices, otherwise the software will break. Your team needs to agree on a data model to get the best results. (See “Defining a Customer Lead,” page 8.)

LookML is a code-based environment, as opposed to a WYSIWYG/GUI-style tool. Unlike GUI tools that lock you into predefined workflows, LookML enables customers to improve and customize the Looker experience for each user and is intentionally designed to directly represent the LookML experience in the Explore layer.

A field called “username” in LookML is accessible from a link called “username” in Explore. Clicking that link shows the username line of code, the data type, the string and the underlying SQL statements. LookML generates the code for you, and reflects what is represented in a declarative way. The Git (version control) workflow and the LookerIDE (integrated development environment) are simple and straightforward, making code-based modeling very approachable and familiar to someone comfortable with a software environment.

Many software development teams adopt the approach of creating a minimum viable product and then iterate collaboratively on that product based on feedback until they have a product “ready” for the market.

Implementing Looker is very similar. Customers create an initial data model, either using pre-built “Blocks,” or packages for commodity databases, and then work together to improve the model over time as stakeholders experiment in the LookML and Explore layers.



Strategies for Success: Rapid Results for Different Data Models

The Looker team and customer community provide a variety of tools and strategies to help kickstart your data model and fine-tune it for rapid success. Looker implements on top of your existing database with a LookML generator built on an initial codebase that those familiar with SQL can easily add to.

Preparing for implementation and the implementation itself are straightforward and rapid. Customers with commoditized data sets or platforms can be up and running in minutes with drag-and-drop dashboards. Implementation for customized, bespoke data sets can be automated and resemble a land-and-expand model, as customers select a target data set to begin modeling and build out from there.

When using commoditized data sets, platforms or pre-built databases, Looker has pre-built templates to immediately build models and dashboards. Data vendors and Looker have worked together to optimize compatibility, workflow, and data analysis.

Today, pre-built templates are available, and, in the future, full applications will be available for rapid implementation. These packages can be downloaded and implemented on top of your database schema as a fully-fledged data model, ready for customization. Looker enables you to extract the best results and maximum return from these commoditized platforms.

For example, Google BigQuery costs vary based on the amount of data utilized, and Looker enables you to measure usage and optimize costs. Other providers Looker optimizes include: Fastly, Pendo, Snowflake, Cloudflare, BigSquid, AWS, Kustomer, Stitch, Tenjin, Qubit, Fivetran, Heap, Funnel, Snowplow, Salesforce, Permutive, Elasticity, and Adobe Analytics. Leveraging pre-built packages and models that other analysts already have worked on saves time and accelerates time to results.

For managing complex or bespoke data sets, you can leverage Looker's Model Generator. After connecting Looker to the target database, the LookML generator automatically analyzes database column and row structures to infer schema relationships. Based on the inference, LookML generates an initial data model and creates a code set for adding additional SQL queries.

In the future, existing SQL queries will import into the model, with the model updated automatically with machine learning. For this, Looker recommends implementing with smaller data sets as part of a minimum viable product and adding additional data inputs as the model proves out.

Another useful strategy is to leverage Looker Blocks pre-built portions of LookML code that help accelerate analytics. From optimized SQL patterns to fully built-out data models, Looker Blocks can serve as a starting point for quick and flexible data modeling. Blocks can be downloaded and are available from a directory on the Looker website.

Self-Service Takes the Load Off Analysts

OnPeak uses Looker for all internal marketing, sales, and account management teams, giving them self-service access to the specific data they each need to do their jobs better.

Now, 70% of analytics requests take minutes instead of hours.

Cultivating a Data Culture

Data Analytics as a Product, Not a Process

Data analysts are the core of your company's data culture. They understand organizational data, workflows, queries, and output requirements better than anyone else. This uniquely positions them to help empower the shift from a manual, operational process to an internal company culture of perceiving—and participating in—data analytics as a critical internal function.

As data modeling moves out of a manual IT ticketing workflow, data analysts become critical to enabling and managing the internal user experience and for accessing company data. Looker offers data project management training and resources to assist in this transformation.

True Self-Service Data

Traditionally, analysts spent the bulk of their time responding to ad-hoc query and report requests and had to pull data manually to fulfill workflows. Additionally, metrics were often high level and static.

Conversely, Looker is built on a foundation of data self-service, which is integral to the success of any forward-thinking organization. Looker layers the data model over the database, creating more consistency across the organization and less pressure on analysts to deliver one-off queries and reports.

When definitions change, revisions only need to be made in one place, which saves time, makes data accessible to everyone, reduces the burden on analysts and creates a truly data-driven culture. Everyone in your organization—even non-technical stakeholders— can query data in the same way, using the same definitions and metrics.

One customer in the digital marketing space transformed the data analyst workflow with Looker. Before implementing Looker, analysts were spending more than 20 hours a week assisting stakeholders with data requests and issues. Leveraging Looker's self-service capabilities, that time has been reduced by more than 90% to an average of only two hours per week.

Glossier. Looker Drives Self-sufficiency and Deeper Insights

Glossier is a global, fast-growing direct-to-consumer skincare and makeup brand. Before deploying Looker, Glossier's data analysts spent most of their time fielding ad hoc queries from business teams so they could get insights into the company's performance. At best, analysts were able to provide the business team with answers to questions based on relatively simple, high-level sales metrics.

With Looker, Glossier can now get answers to more complex, strategic questions, such as:

- “What is the lifetime value (LTV) of a customer?”
- “What is the order in which customers purchase products?”
- “Where are customers located?”

Looker's self-service capabilities enabled Glossier to adopt a truly data-driven culture.

Users can access dashboards to answer questions that help them identify potential growth areas, and they can make better and more informed decisions about providing the best possible customer experience.

Agreeing on Metrics

One of the biggest challenges organizations—especially large enterprises—face is consistency in reporting. For example, most analysts find it challenging to ensure that the marketing department agrees with the sales department on their definitions of a particular metric, since teams may have distinct ways to define and characterize how the business works.

In reality, there is only one business, so there should be only one way of measuring how the business performs in any given context. Business intelligence practitioners need one way to describe key metrics so that everyone can clearly align toward the same goals.

Some business intelligence tools allow users to create visualizations based on self-contained workbooks. In principle, this type of software is meant to empower business users to see and analyze data quickly.

Unfortunately, this approach too often results in a proliferation of reports generated from thousands of workbooks with metrics based on disparate calculations. This becomes a management nightmare for analysts, and because business users aren't making strategic plans and taking actions based on universally standardized data, companies do not benefit from empirically factual insights that can improve the bottom line.

With LookML, you gain a single consistent agreed-upon set of metrics that can be customized for each department. Regardless of how different departments may name their metrics, underlying calculations are the same. Finance, for example, may call it “net profit,” while marketing may call it “profit.”

To get varying functional areas at your company to coalesce around consistent definitions of metrics, there must be agreement about what the data means. This agreement usually happens as a result of frequent meetings at the beginning of an implementation to define the metrics that make up the data model.

Collaboratively coming together needs to be a best practice at your organization and can take many forms. Some companies have regularly scheduled meetings focused just on defining metrics. Larger enterprises often establish a formal Center of Excellence (CoE), complete with an articulated mission statement, structured processes and designated subject matter experts.

Whichever path to collaboration your organization chooses, gaining agreement on metrics is essential to building a model that will accurately reflect your company's performance.

Conclusion

Looker enables data analysts to spend more time building a business model that accurately reflects how their organization functions and is performing--and less time addressing one-off requests from stakeholders.

Data trust is established and can be accessed more easily by more people and put to immediate use. Looker's platform drives better business insights that lead to revenue wins and improved customer engagement, greater efficiency, and a data-driven culture of collaboration and consensus.

About Looker

Looker is a unified data platform that delivers actionable business insights to employees at the point of decision. Looker integrates data into the daily workflows of users to allow organizations to extract value from data at web scale. Over 2000 industry-leading and innovative companies such as Sony, Amazon, The Economist, IBM, Etsy, Lyft and Kickstarter have trusted Looker to power their data-driven cultures. Looker joined Google Cloud in February 2020. For more information, connect with us on [LinkedIn](#), [Twitter](#), [Facebook](#) and [YouTube](#) or visit [looker.com](#).