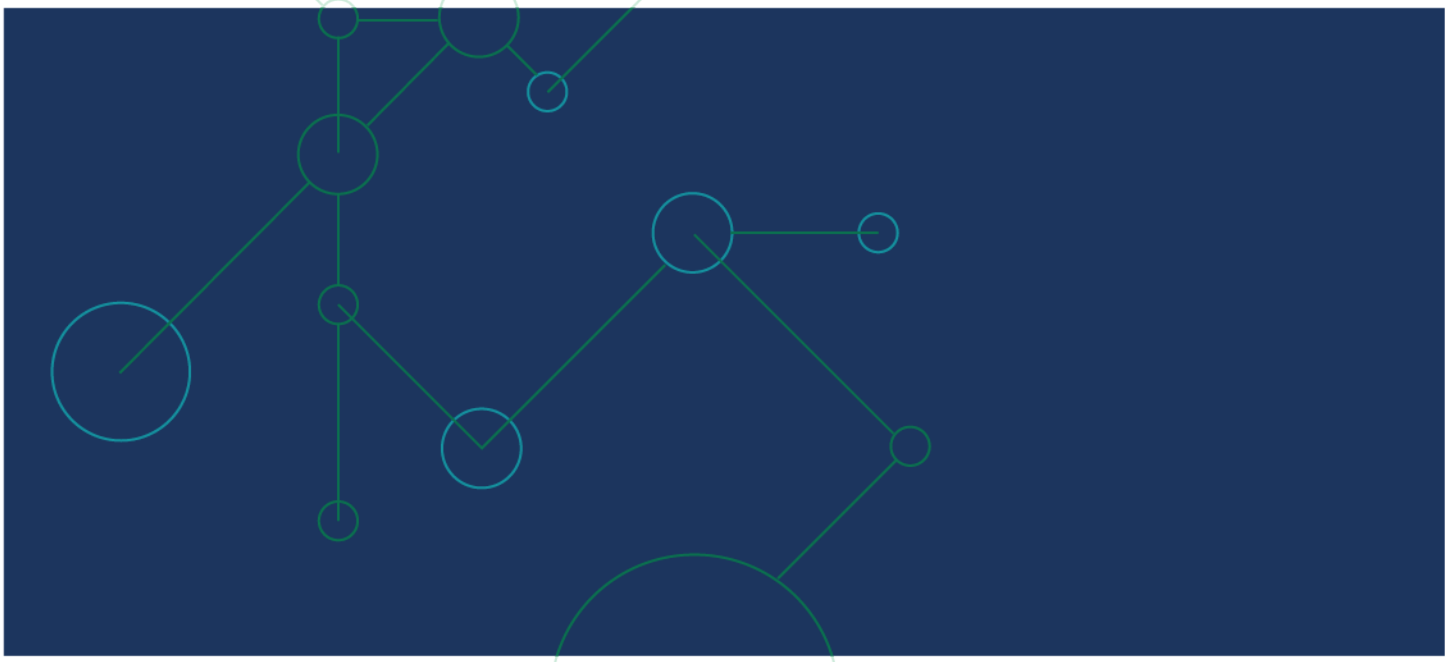
WHITE PAPER

Agile Data Warehouse Automation

The Time- and Cost-Saving Power of Qlik Compose



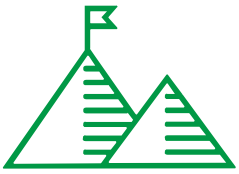
Qlik

SUMMARY

Your data architects and business intelligence (BI) teams want to rapidly integrate and transform data to meet fast changing, real-time business analytics requirements. But they're struggling with brittle, hand-coded, complex data warehousing processes causing project and decision delays. And that's led to management complexity, developer resource constraints, and cost overruns.

This white paper explains how our Qlik Compose™ solution overcomes these challenges, offering your data architects a fresh approach to data warehouse automation. It illustrates the way our solution automates the design, implementation, and change process for data warehouses and data marts. And how it replaces manual, error-prone coding with a metadata-based solution that automates data modeling and ETL commands and complex transformations while accelerating change propagation.

How We Got Here



Business Challenges

Traditional methods of designing, developing, and implementing data warehouses require large time and resource investments. The ETL stand-up development effort alone – multi-month and error-prone with prep times of up to 80 percent and expertise from specialized developers – often means your data model is out of date before your BI project even starts. Plus, the result of a traditional data warehouse design, development, and implementation process is often a system that can't adapt to continually changing business requirements. Yet modifying your data warehouse diverts skilled resources from your more innovation-related projects. Consequently, your business ends up with your data warehouse becoming a bottleneck as much as an enabler of analytics. What can and should you do differently?

Speeding time to analytics takes streamlining key aspects of your data warehouse creation and management lifecycle. It requires reimagining how you deliver on your BI team's goal – addressing strategic business questions with minimum tactical and software development work. And rethinking your processes for extracting data from disparate source systems and wrangling it into your data warehouse as well as how you update your data warehouse to meet changing requirements. The critical ingredient to greater speed at each step: automation.

Introducing Qlik Compose

Your business demands agility and responsiveness from your BI environment. Our Qlik Compose solution meets requirements with a metadata-based platform, simpler and purpose-built for your data architects, not your developers. It automates these traditionally manual, repetitive data warehouse tasks: design, development, testing, deployment, operations, impact analysis, and change management. It also accelerates tasks you can't completely automate. Here's how.

What Customers Are Saying

" We had initially planned on 45 days of ETL coding. With Qlik Compose we had it completed in two days."

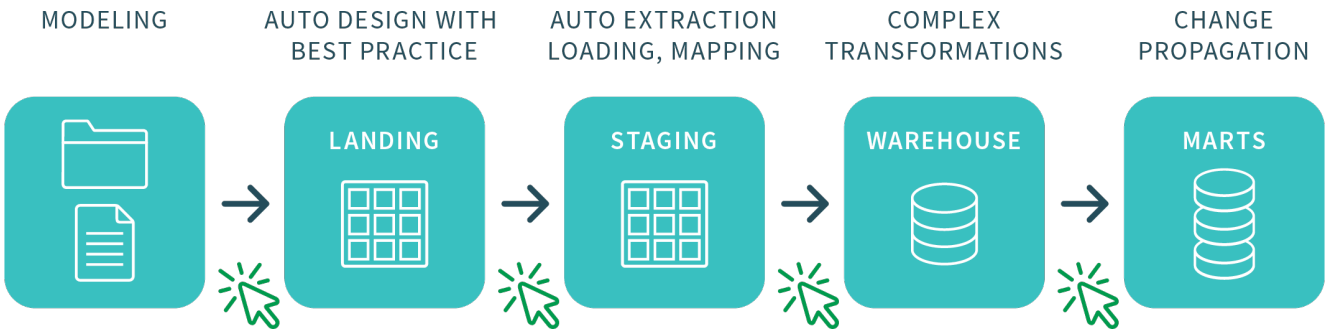
— Senior Information Management Analyst at a Global Insurance Company

Qlik Compose automatically generates the ETL commands, data warehouse structures, and documentation your team needs to efficiently execute projects while tracking data lineage and ensuring integrity. Using it, your IT teams can respond fast – in days – to new business requests, providing accurate time, cost, and resource estimates. Then once projects are approved, your IT staff can finally deliver completed data warehouses, data marts, and BI environments in far less time. And best of all for your enterprise – efficiencies, lower risk, and more effective analytics.

With our Qlik Compose solution, your business can also dramatically cut the time, cost, and risk of your cloud analytics projects. Even if you're using Amazon Redshift, which may be easily deployed and scaled but involves the same lengthy and manual processes to manage in the cloud as an on-premises data warehouse. For Redshift users, our solution is the answer to rapidly spinning up, scaling up, and iterating data warehouses, then dynamically adjusting data sources and models based on changing business requirements.

Agile Data Warehouse Automation with Qlik Compose

Workflow Management and Monitoring



Key Qlik Compose Features

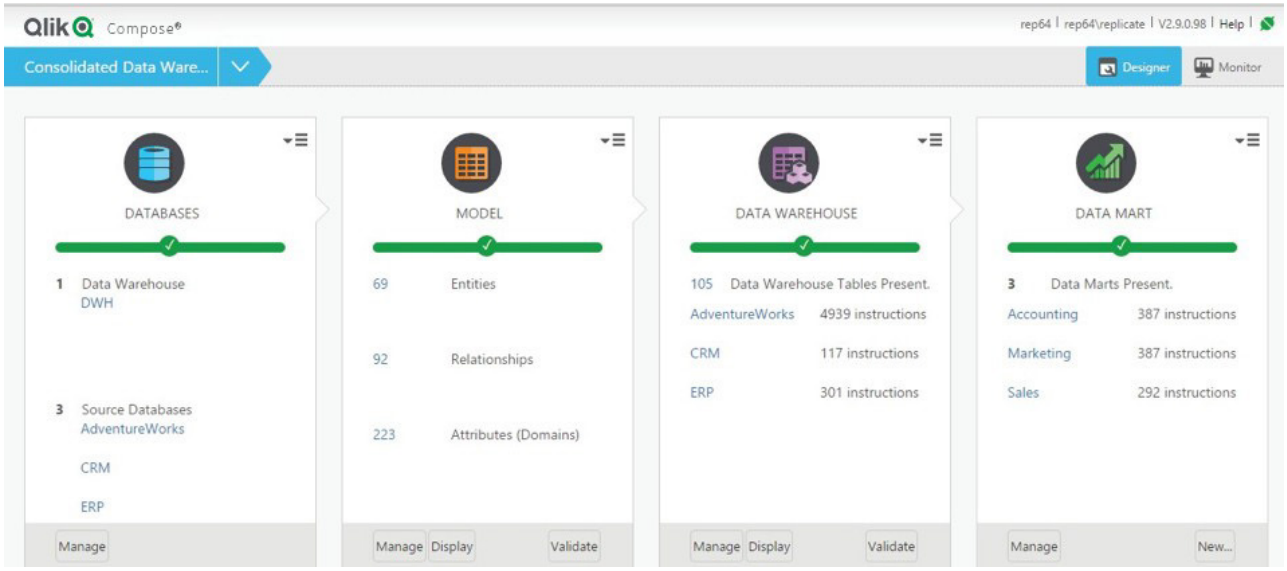
The comprehensive set of automation features in our Qlik Compose solution simplifies data warehousing projects. It eliminates the cumbersome and error-prone manual coding required by legacy data warehouse design and implementations' many repetitive steps. In addition, our solution includes the operational features your business needs for ongoing data warehouse and data mart maintenance.

Automation Features

- Optimized for either model-driven or data-driven data warehousing approaches
- Real-time source data integration
- Automated ETL generation
- Physical data warehouse management
- Data mart generation

Operational Features

- Monitoring
- Workflow designer and scheduler
- Notifications
- Data profiling and quality enforcement
- Lineage and impact analysis
- Project documentation generation
- Migration between environments



Agile Data Warehouse Automation

Create, Import, and Reverse Engineer Your Data Models

Traditional, conceptual data modeling is complicated by source-to-target data mappings. Our Qlik Compose solution eliminates this complexity by first, defining the physical model in several ways, then second, automatically mapping the source databases to the model using best practices for data warehouse generation.

You use Qlik Compose Data Warehouse model designer to generate models in three ways:

- As a modeling tool, and then Qlik Compose also helps you define the entity relationship diagram.
- To import third-party models like CA Erwin or other industry-standard models.
- To discover the model by reverse engineering the source databases.

These are some key Qlik Compose features, delivering compelling benefits:

- Attribute Domain glossary for all the attributes in your data warehouse
- Definition of entities with attributes and relationships between the entities
- Definition of history type for each attribute (type 1 and type 2)
- Definition of one or many satellites per entity
- Built-in date model that can be linked to any date attribute

Together, these capabilities enable you to create a model based on your business requirements, then automatically generate and implement that logical model without additional manual work. Our solution automatically generates the mappings using data warehouse best practices for higher consistency and quality.

Enable Agile Development (Implement-Change-Adjust)

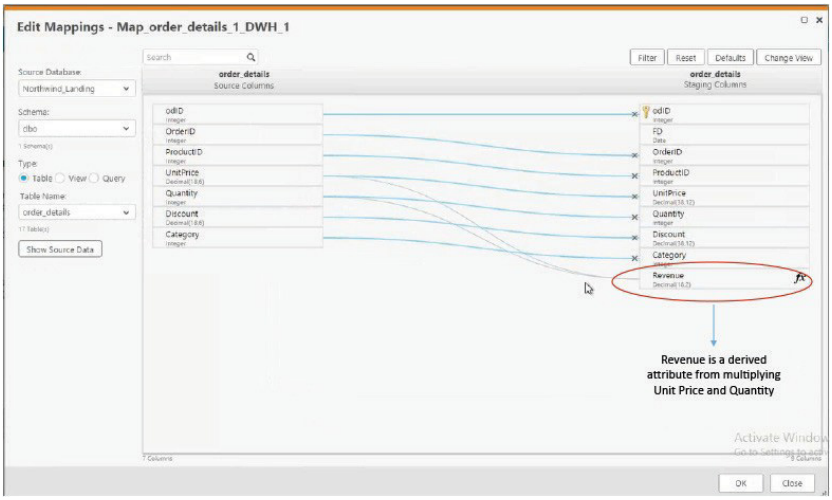
Starting a new data warehouse or BI initiative can be a lengthy, expensive, and risky proposition. Our solution reduces your risk and helps you deliver a product that satisfies your end users. Because a variety of factors often change data warehouse requirements throughout the project lifecycle, your business needs an incremental and agile development approach. That's what our solution delivers.

Qlik Compose is purpose-built to accommodate change. You can introduce and rapidly deploy any changes to your data sources, models, and business rules, speeding your project execution and learning processes while reducing risk.

Modify and Enhance Your Data Model

You use derived attributes in any data warehouse to enhance the data model. They're usually calculated from other attributes, such as multiplying an employee's monthly salary by 12 to calculate an annual salary or uncovering a person's full name from first name and last name attributes. To create derived attributes, your data warehouse administrator and ETL developers must typically manually modify the data warehouse tables and write ETL code that calculates values. With our solution, derived attributes are established as part of the model definition. And they're continuously evaluated and updated as part of ongoing, automated data warehouse management.

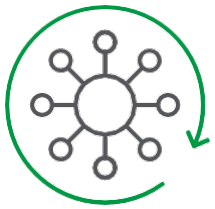
Plus, our Qlik Compose solution helps you automate the process of replacing values in the source data with the actual data you want to store in your data warehouse. How? By implementing lookup tables functionality. For example, using a lookup table to replace a zip code with a full address or, conversely, to replace a full address with a zip code.



Support Date/Time-Based Modeling

Every team wants more flexible BI and analytics. Our solution delivers it by automating the construction, management, and ongoing updating of time dimensions. Incorporating time into data warehouse dimensions is one of the most common BI and analytics project requirements. Time dimension attributes – representing periods such as years, semesters, quarters, months, or days – provide varying levels of granularity for analysis and reporting. These attributes are organized in hierarchies, and the scale of the time dimension is determined largely by your business and reporting requirements for historical data. For example, most financial and sales data in BI applications use months or quarters. Our Qlik Compose solution completely automates time dimensions processes.

Real-time Data Integration and Automated ETL



Easy-to-Use Graphical Interface

The clearly defined, easy to navigate graphical user interface (GUI) in our Qlik Compose solution hides data integration and transformation complexity while still offering users a rich feature set. This innovative GUI makes it possible for you to easily design data models and build data warehouses and data marts with just a few clicks. Once you've finished designing and building the environment you want, simply switch to a monitoring console to track your progress.

Real-Time Source Data Loading

Our Qlik Compose solution is integrated with Qlik Replicate™ which is our high-performance data replication and loading software application. Together, they support the entire lifecycle of data warehouses and data marts with features such as:

- Source mapping
- Model definition
- Data warehouse and data mart creation
- Data warehouse/data mart population
- Ongoing updates

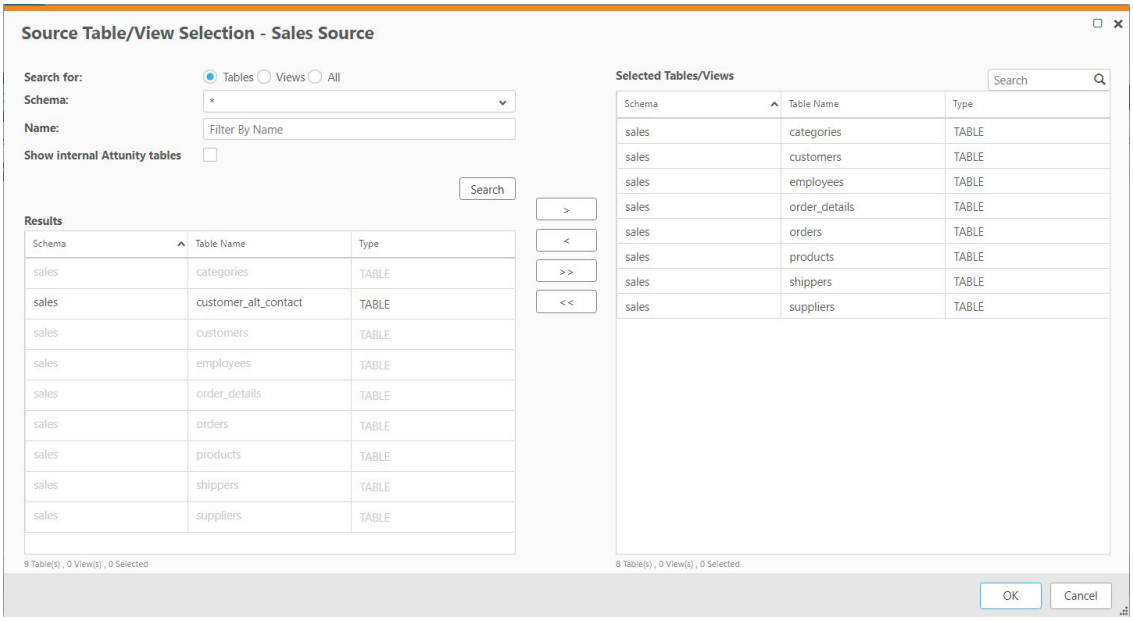
Your users can load data from many heterogeneous data sources simultaneously with real-time change data capture (CDC) technology.

Zero-Footprint Technology

Our solution is a zero-footprint technology. No agents need to be placed on your source databases. This eliminates performance overhead and administrative burdens on your mission-critical systems.

Support for Heterogeneous Source Systems

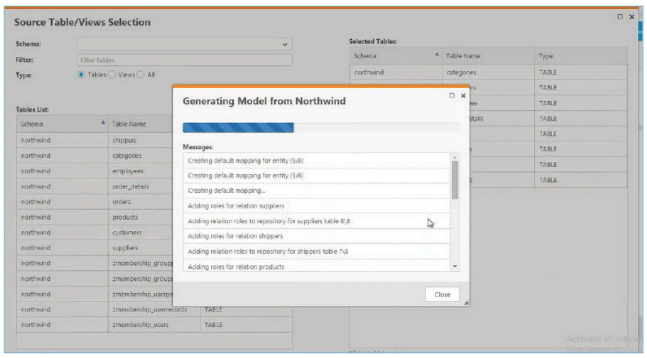
Through Qlik Replicate integration, our Qlik Compose solution supports the industry's widest ecosystem of heterogeneous database sources with real-time and query-based CDC. Our platform supports all major databases – from legacy to modern data stores. Whether sourcing from or to RDBMS, Hadoop, or enterprise data warehouses deployed in the cloud or on premises, our Qlik Replicate management environment works across endpoints similarly. It's purpose-built for heterogeneous data replication.



ETL Automation

Our Qlik Compose software provides automated ETL generation and execution for loading data to the staging[1] area from the landing[2] area in your data warehouse:

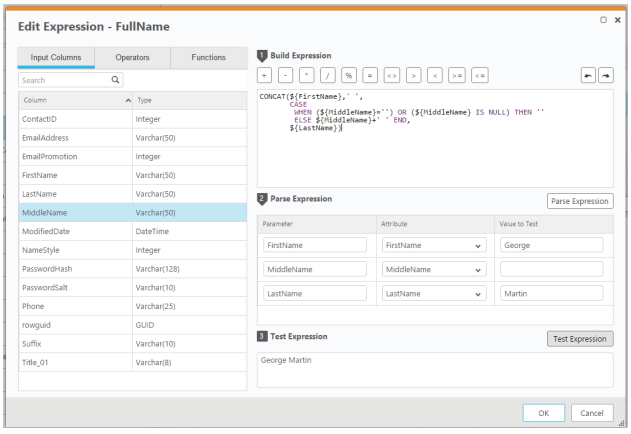
- Every combination of source table to staging table can include a mapping definition.
- Mapping definitions connect source columns to attributes in a data warehouse entity.
- Qlik Compose uses the same mapping and transformations for the initial load of the data warehouse from replica tables, and for updating the data warehouse from change tables.
- Supported transformation functions include:
  - Mapping an attribute in the staging area to an expression based on the source
  - Filtering records from the source according to a logical expression
  - Providing mappings based on a landing table, view, or SQL query



Your users may also include custom SQL-based ETL steps in their processing flow to leverage pre-defined procedures and enhance supported transformation features.

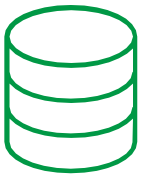
Advanced Transformations

The GUI-based Expression Builder lets you easily build an expression that calculates new derived attributes or transforms existing fields. It streamlines access to the required elements for an expression – without manually typing any information. After building an expression, your users can list and test drive it from within the Expression Builder editor. The benefits: More easily create and apply transformations to data elements this way, and ensure your users get results.



[1] The data warehouse staging area includes tables in the structure of the data warehouse model – the source for the data to load or update the data warehouse.
[2] The data warehouse landing area (per operational data source) includes tables replicated from the source, as well as change tables.

Physical Data Warehouse Management



Automatically Generate Code Creating Data Staging, a Data Warehouse, and Data Marts

Data Staging

Staging tables are mirror images of the source table inside your data warehouse. You use them to collate and prepare data for loading into your data warehouse tables. Our Qlik Compose solution automatically generates staging tables when discovering the source databases.

Data Warehouse

Once our solution discovers and maps source metadata, then generates the model, it builds your data warehouse with the click of a button. It also generates data definition language (DDL) statements based on normalized enterprise data warehouse principles. But our software minimizes issues teams using the traditional approach to normalized data warehouse implementation encounter: long development time, difficult maintenance, and change propagation.

Once our technology generates the DDL, it executes the statements. Over the life of your data warehouse, your existing model is constantly compared to the sources. As your data warehouse model evolves, Qlik Compose automatically introduces physical adjustments to your data warehouse structure without generating ETL code.

As a result, data warehouses generated with our solution are compliant with data vault modeling techniques, including a(n):

- Completely auditable architecture
- Data warehouse model aligned with the business model
- Approach extremely adaptable to (business) changes
- Design purpose-built and optimized for the data warehouse
- Approach that lends itself to real-time processing
- Way to easily load to a dimensional model
- Model with layered isolation from change
- Approach that can be incrementally built, easily extended

Our Qlik Compose data warehouse automation features include:

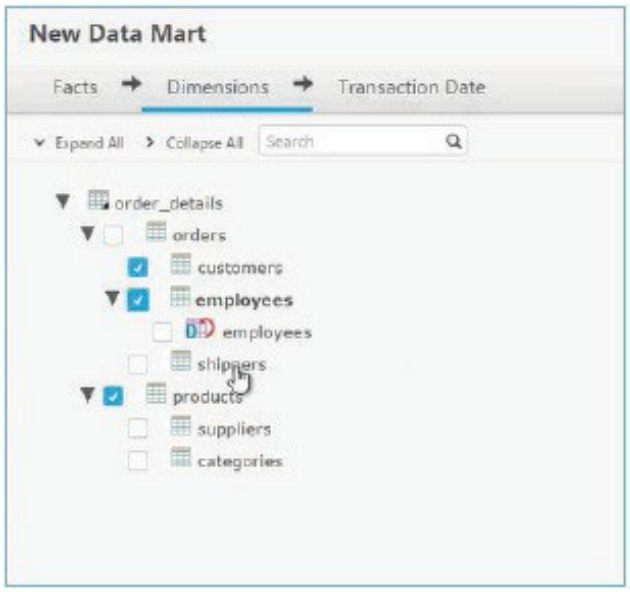
- Transparently managing surrogate keys and natural- to-surrogate key[3] mappings
- Handling of history types and slowly changing dimensions
- Supporting and integrating history information coming from operational sources with history
- Not dealing with the order in which the data warehouse is loaded via handling of premature facts[4]
- Handling updates to the data warehouse based on partial records (updating only changed fields)
- Asynchronously loading the entity from several sources (In other words, sources loading a single record in your data warehouse don't have to use the same loading schedules as standard ETL processes.)

Data Mart

Our Qlik Compose solution also automates dimensional modeling (star schema) of your data marts. Generated data marts are suitable for multi-dimensional analysis and organized per subject area. That makes them easy to understand for your business users.

When you build a data mart, our solution wizard walks you through all the steps required to build and populate it:

1. Selecting fact tables
2. Selecting dimension tables
3. Defining transaction date(s)
4. Creating deformalized tables based on facts and dimensions defined above
5. Generating ETL code to incrementally populate the data mart



[3] A key is one or more data attributes that uniquely identify an entity. A natural key is formed of attributes that already exist in the real world. For example, U.S. citizens are issued a social security number (SSN). A surrogate key has no business meaning and is internally generated by the system. Surrogate keys are used to simplify the management of complex natural keys (more than one attribute in the key) or when integrated data sources have a different representation of the natural key for the same values.

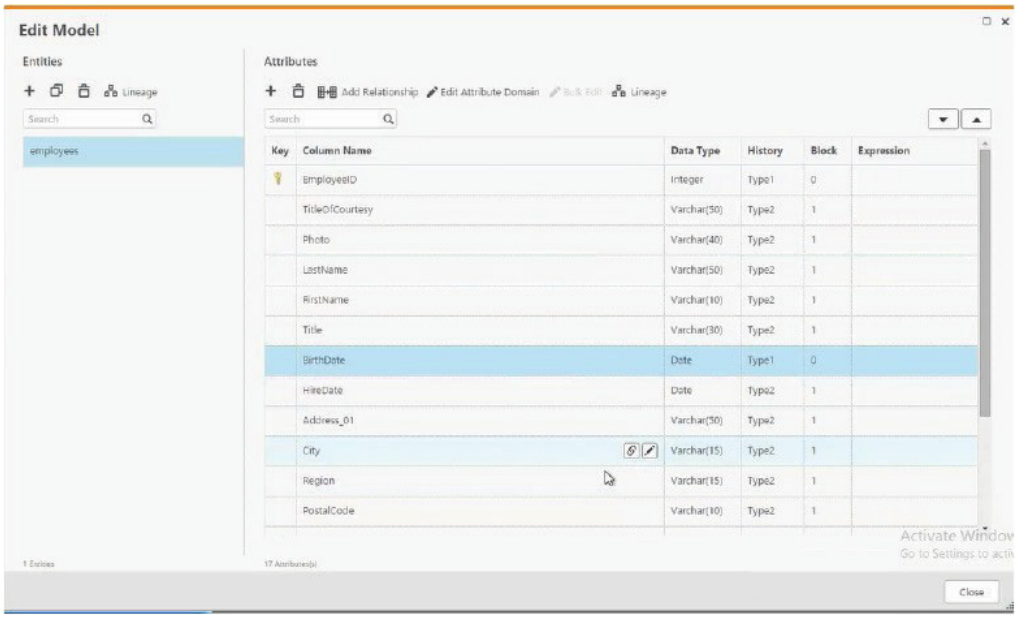
[4] Premature facts are facts that are loaded to the data warehouse before the dimensions that describe them are loaded. For example, an order record loaded before the customer record that made that order.

Rapidly Propagate Data Model Changes to Physical Structures – No Manual ETL Coding

With our Qlik Compose solution, it's easy to introduce a new data source, modify an existing source, or simply add new attributes to an existing model without any ETL code modifications. Your users simply follow the same steps and wizards they used during the original design stage. Our technology automatically generates and reflects Table/ETL code changes in the new version of your data warehouse. This way your users can implement business changes with a few clicks. Our solution also automatically generates all ETL code to implement a change, so you don't have to do manual coding anymore.

Preserve History for Slow Changing Dimensions (Type 2) – No Manual Coding

Teams use this method to track historical data by creating multiple records for a given entity in the dimensional/hub tables. For example, you might need your data warehouse application to track and store customer addresses every time you register a new customer location. Qlik Compose automatically identifies Type 2 attributes and recommends a table structure accordingly. But before deploying the model, your user can manually overwrite the recommendation and change the attribute to Type 1 – if history tracking isn't required. Because our software automatically tracks and handles changes in the history, it significantly reduces implementation complexity and eliminates the need for manual coding.



Generate an Optimized Data Warehouse Architecture Using Hubs and Satellite Table Layout

Our Qlik Compose software features a data warehousing design pattern that simplifies designing a robust, agile, and historical data warehouse. It uses Hub and Satellite tables to model an optimized data warehouse. [Note: These are the names given to a type of table, in the same way a star schema is made up of dimensions and facts in the data mart.]

As the first modeling step, our software helps you break the model into core entities, such as Customer, Sale, Employee, etc. Each core subject area is then turned into a Hub table. The Hub table contains just Type1 attributes, like the business key from the source system(s), such as Customer_Id.

Satellite tables contain all the descriptive information about a Hub. Each Hub can have one-to-many Satellite tables that contain information about that business key (Type2 attributes). In the context of the Customer_Hub, there could be a Customer_Name_Satellite and a Customer_Address_Satellite.

The Hub and Satellite-based design makes it easy to add sources to your data warehouse in the future. You can create new Hubs and Satellites from new source systems and join them to already existing Hubs. Similarly, you can create new Satellites for pre-existing Hubs. This prevents modifications to already existing tables.

The benefit of our approach is this: No more feeling pressure to create the perfect data warehouse design from the start. With our Qlik Compose solution, you and your team can continually edit and add to your data warehouse as business needs change. An agile approach lets you build segments of your data warehouse quickly, which pleases business owners who want immediate results faster. Traditionally, your developers may have spent months or even years building and testing data warehouse logic, while business users waited to see any benefit. Not anymore.

Data Mart Generation

Automate the Data Mart Definition Generation Process

Since a data mart is essentially a subset of your data warehouse, anyone using Qlik Compose can create any number of data marts according to their business needs.

They also can create multiple star schemas for a single data mart using pre-coded design patterns. Star schemas allow you to reuse existing dimension tables within the same data mart, saving space in the data warehouse platform while simultaneously improving query performance. For example, you could create one star schema with an Order Details fact table and Customers and Products dimensions and another star schema with the same dimensions, but a different fact type (or the same fact type, but different dimensions). This allows you to generate BI reports using different facts that share the same dimensions.

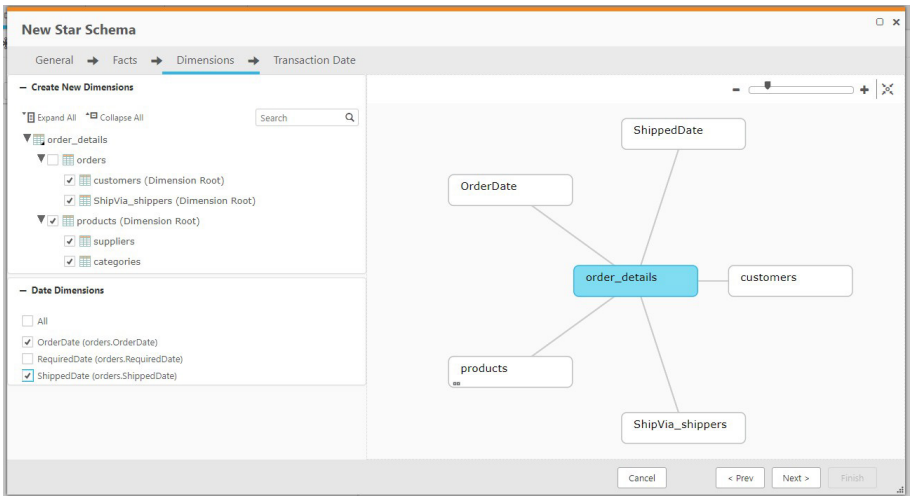
And in a star schema, your dimensions are linked with each other through one join path intersecting the fact table, facilitating accurate and consistent query results.

Our Qlik Compose solution supports three types of data marts out of the box:

- **Transactional** – A star schema with a transactional fact table allows you to retrieve the desired data, even if a dimension table contains multiple versions of the same record. To use an example from the automotive industry, selecting "OrderDate" as the Transaction Date would allow you to generate a report for the number of customers who bought cars in New York between 2013 and 2016, even if a customer moved to a different city (which would also result in a new record being added to the Customers dimension).

- **Aggregated** – A star schema with an aggregated fact table allows you to make aggregate calculations based on the fact table attributes. For instance, you could create an aggregated fact that shows the total freight costs per shipping region and product category. Additionally, the presence of a transaction date in the fact table makes it possible to retrieve the desired data, even if a dimension contains multiple versions of the same record. In a transportation industry example, a shipper could use an aggregated fact to generate a report for the total cost of shipping rice to Australia from 2015-2016.

- **State Oriented** – A star schema with a state-oriented fact supports Type 2 columns in the fact table. This is useful in cases where the fact is not a singular event in time, but rather, consists of multiple “states” or events that occur over time. A typical example of facts with multiple states are insurance claims or flight reservations. There are also cases when the same entity is treated as both a fact and a dimension – for example, Customers. In such cases, you can generate a report that relates to the state of the fact, such as the time a claim was submitted to the time it was approved.



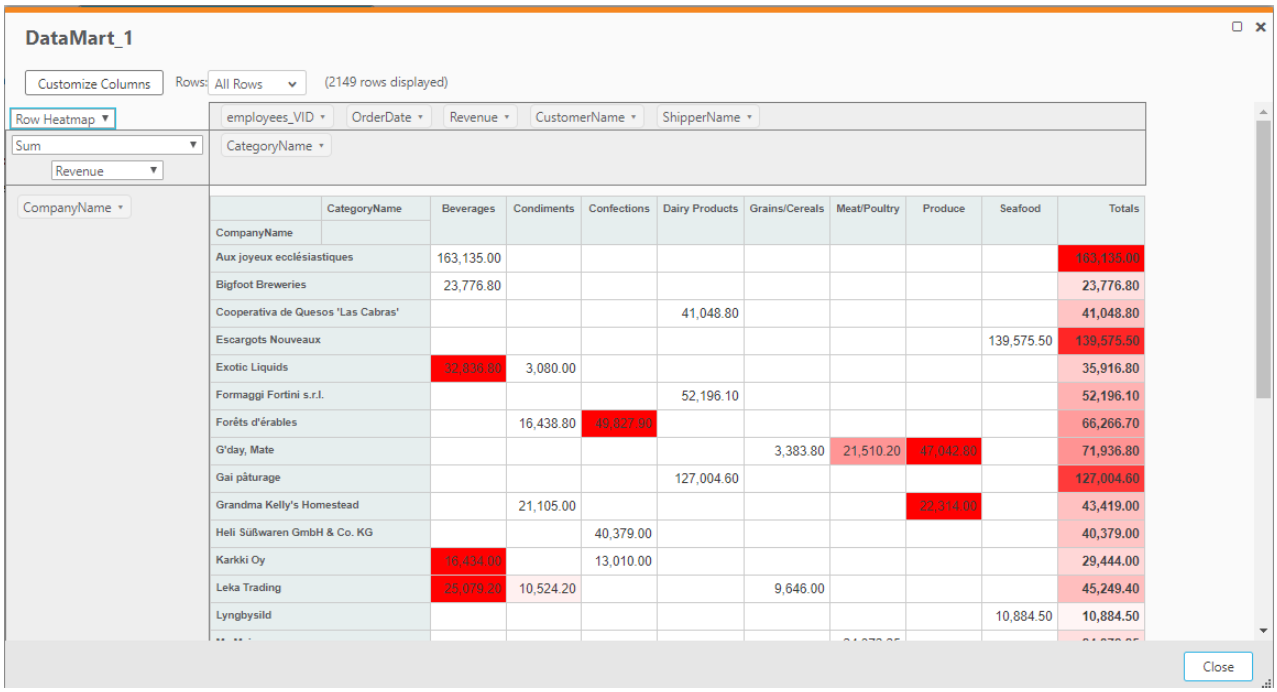
Deploy One Solution that Physically Creates and Manages Data Marts

Once defined, our Qlik Compose software streamlines and simplifies your data mart management:

- Automates the building of your physical data mart tables
- Generates the ETL set to load your data mart from your data warehouse
- Makes incremental updates to your data mart as your data warehouse is updated
- Associates facts with the matching version of their dimensions according to a user-selected date attribute

Enjoy Out-of-the-Box Enablement for BI Visualization Tools

Saving you time and money, our Qlik Compose software designs data marts to work with BI visualization tools out of the box. It does this by exposing the applications to all the primary and foreign keys, and by using an application-friendly attribute name that matches with the tables in the data marts. This enables visualization applications such as Microsoft PowerPivot, Tableau, and Qlik to work on top as our solution generates data marts out of the box.



CompanyName	Beverages	Condiments	Confections	Dairy Products	Grains/Cereals	Meat/Poultry	Produce	Seafood	Totals
Aux joyeux ecclésiastiques	163,135.00								163,135.00
Bigfoot Breweries	23,776.80								23,776.80
Cooperativa de Quesos 'Las Cabras'				41,048.80					41,048.80
Escargots Nouveaux								139,575.50	139,575.50
Exotic Liquids	32,836.80	3,080.00							35,916.80
Formaggi Fortini s.r.l.				52,196.10					52,196.10
Forêts d'érables		16,438.80	49,827.90						66,266.70
G'day, Mate					3,383.80	21,510.20	47,042.80		71,936.80
Gai pâturage				127,004.60					127,004.60
Grandma Kelly's Homestead		21,105.00					22,314.00		43,419.00
Heli Süßwaren GmbH & Co. KG			40,379.00						40,379.00
Karkki Oy	16,434.00		13,010.00						29,444.00
Leka Trading	25,079.20	10,524.20			9,646.00				45,249.40
Lyngbysild								10,884.50	10,884.50

Operational Features and Lifecycle Management

Workflow Designer, Monitor, and Scheduler

Keeping your data warehouse and data marts up to date is a complex process with multiple tasks. These include collecting data from multiple sources, staging it in the landing area, transforming and integrating the data before loading it into your data warehouse, and then loading the relevant information into your data mart(s).

Our Qlik Compose software has a built-in workflow designer so you can run all your data warehouse and data mart ETL tasks as a single, end-to-end process. If you have one data mart project, all tasks run sequentially, starting with the data warehouse ETL tasks and ending with the data mart ETL task. If you have many data marts, the data mart ETL tasks can be designed to run in parallel, following the completion of the data warehouse ETL tasks.

Our solution monitor shows you the current status of the workflows and independent tasks, enabling your users to drill down for additional information. You can enable workflows to run immediately or schedule them to run in the future (either once or at set intervals).

Notifications

Our solution can be configured to send notification messages about events that occur when Qlik Compose tasks are running. You set these via our Qlik Compose monitor. Select events, and when they happen, specified recipients will be alerted.

Your users can manage created notifications from the Notifications list. It includes details about each defined alert and lets users activate/ deactivate a notification. In addition, your users can change the definitions of existing notifications or delete them.

Lineage and Impact Analysis

Automatically collecting and storing metadata during your design and implementation stages is core to our solution. Qlik Compose connects directly to your different data sources and collects information about the source schemas, tables, attributes, and primary and secondary keys. Our software stores this collected information in its central repository, giving you a visual map of your data flow from source data to your data warehouse and data marts, then regenerates data lineage when you implement changes.

Before editing an entity or attribute, your users should run the Impact Analysis report to check which other entities/ attributes in the entity's/attribute's lineage will be impacted by the change. For example, removing the "Discount" attribute from a table will affect the "Total Price."

Data Profiling and Quality Management

You can use our Qlik Compose software to profile data and improve its quality before loading it into your data warehouse for accurate reporting.

- Data profiling features enable your BI teams to validate data in the landing area tables to eliminate surprises as they execute on their data model. Teams can identify format discrepancies or issues such as null or blank fields, then repair them, for example, adjusting source tables or creating new rules, before loading the data.

- Your BI teams also can improve data quality by configuring and enforcing rules before loading data to automatically discover issues with values, formats, data ranges, duplicates, etc., and define exception policies (i.e., diverting exceptions to an error mart) and remediation tasks.

Migration Between Environments

An Export/Import Project Configuration facility in Qlik Compose lets your users capture the configuration settings of an existing data warehouse project including databases, sources, data warehouse, data marts, transformations, filters, scheduling, and notifications. This is helpful, for example, when your users need to move configuration settings from a test to a production environment.

Automated Development, Testing, Acceptance, and Production

Our Qlik Compose solution includes automated Development, Testing, Acceptance and Production (DTAP) capabilities that accelerate your data warehouse projects and improve your agility. Your users can streamline the creation and deployment of data warehouse models, ETL code, and other components. And they can easily generate deployment packages, adjust global variables for specific environments, and move between environments using either the GUI or command line.

What Customers Are Saying

" Thanks to Qlik Compose, our ability to make decisions faster and more effectively has increased at least tenfold."

— Sean Rassi,
Vice President of Design and Technology, Poly-Wood, LLC

Version Control

Further optimizing your data warehousing process, our Qlik Compose software centralizes and automates project version control. Your users can roll back, compare, and merge versions; lock projects; and track all changes. By reconciling multiple changes from multiple users at any time, this capability improves your team collaboration, productivity, and agility. And importantly, our solution integrates with Git and other leading version control tools.

Agile Data Warehouse Automation

Advance your organization's BI and analytics goals with Qlik Compose, our comprehensive, agile data warehouse automation platform for data architects. It uses key technologies and introduces innovations that enable efficient, effective data warehouse lifecycle management.

Your BI staff won't have to worry about manual coding anymore when it comes to designing, creating, populating, and managing your data warehouses and data marts. They can implement our solution their way – using either a model-driven approach based on business process or a data-driven approach based on source structures.

Improve your analytics productivity, time to value, quality, and consistency with our solution's controls and process optimization. Realize better insights using fewer resources and accelerate your BI project ROI with our Qlik Compose solution.

Learn more at **qlik.com/products/qlik-compose-data-warehouses**

Qlik

About Qlik

Qlik's vision is a data-literate world, where everyone can use data and analytics to improve decision-making and solve their most challenging problems. Our cloud-based Qlik Active Intelligence Platform® delivers end-to-end, real-time data integration and analytics cloud solutions to close the gaps between data, insights and action. By transforming data into Active Intelligence, businesses can drive better decisions, improve revenue and profitability, and optimize customer relationships. Qlik does business in more than 100 countries and serves over 38,000 active customers around the world.

qlik.com