



WELCOME TO THE WORLD'S FIRST
SMART DATA WAREHOUSE



Technology Overview

A Smart Data Warehouse built for the cloud, Panoply automatically is tailored to fit data and scale towards the goals of Analysts and Data Scientists. Panoply's unique self-optimizing architecture utilizes machine learning and natural language processing (NLP) to streamline the data journey and reduce the time from data to value to as close as possible to none.

Introduction

A constant debate of utilization of MPP data warehouse technologies versus the massive scale and flexibility of Data Lake technologies is engulfing global developer communities. Today, the evolution of the Industry enables companies to use these building blocks to design extremely powerful Infrastructure solutions, by leveraging these technologies under a unified architecture. Panoply welcomes you to the world's first smart Data Warehouse.

Architecture

As with everything else in Panoply, simplicity and ease of use guide the user experience path. The Panoply stack is abstracted away behind a single JDBC end-point. So you can aim your SQL queries to run against this end point, guaranteeing a simple, always fast, infinitely scalable architecture.

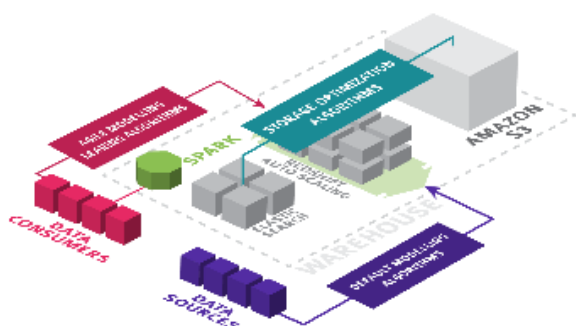
Panoply is constantly running periodic processes to mark the data and optimize the storage based on your usage. Data that you query frequently will be cached, aggregated and stored at a higher-level technology than aged raw data that's used less frequently.

The default architecture is based on Amazon's S3 and Redshift components although the multi tier stack can be adjusted based on specific organization needs.

Data Sources

Importing data into a traditional on-premise or cloud based data warehouse is a cumbersome, tedious and unscalable task. Traditionally, moving data into the analytical warehouse is done through a periodically running process, commonly known as ETL (Extract, Transfer & Load). This approach is dated, complex and unscalable. It's usually a hand-coded process that runs daily, making changes slow, manual and cumbersome. In the age of agile programming and continuous deployment, this approach is well overdue for a revolution.

In a world where a data source can be any technology or service where your raw data resides, from operational database like MongoDB or Postgres, to semi-structured data in CSV files, S3, Dropbox, to third-party services, like Facebook Insights and Salesforce Panoply offers a collection of predefined open-sourced data source integrations. **No schema or modeling is required. Panoply is capable of automatically discovering schemas, data formats, file compressions, nested objects, etc... Making it:**



“ The Only ETL-Less
Data Warehouse ”

Data Transformation

Instead of struggling in the quicksand of building complex ETL processes, Panoply solves the data transformation challenge with a two-step solution.

First, Panoply automates common transformations, for example:

- Compressed files are discovered and extracted.
- Common data formats are identified automatically, and parsed accordingly.
- Nested structures, are flattened into different tables.
- Data types are automatically discovered and a schema is generated based on the initial data structure.
- Enhancement modules are automatically applied to the data.
- Likely relationships between tables are automatically detected and used to model a relational schema.
- Slowly-changing tables are automatically generated.
- Aggregations are automatically generated.

Secondly, Panoply enables building of Transformation Views, which are essentially materialized views, with programmable user-defined functions.

Remodelling

Building a well-optimized analytic infrastructure may take months of development, while it's maintenance is an ongoing labor intensive task. Panoply offers several tools for automated maintenance of your analytic infrastructure.

- Reindexing happens automatically whenever the system detects changes in query patterns.
- Panoply automatically identifies columns used for joins, and re-distributes the data across nodes to improve data locality and join performance.
- Incremental backups are done immediately whenever changes are made.
- Full DB snapshots are done periodically.

Panoply provides full control over all processes, and enables you to apply changes manually if needed.

Data Utilization

Panoply exposes a standard JDBC end point with ANSI-SQL support. Utilization means plugging in your Chart.io, Tableau, Spark or R, or any other standard analysis tool including writing direct SQL, and beginning to analyze. Panoply:

- Automates data source connections.
- Utilizes NLP for data modeling and joining data.
- Automatically optimizes performance to up to 5x faster than traditional cloud data warehouses.
- Enables machine learning query optimization and views to ensure your data is always fast.
- Automates elastic storage and CPU scaling.
- Scales by anticipating your requirements.

Query and Schema Optimization

Panoply utilizes statistical algorithms to inspect query and dashboard runtime over the selected data to constantly looks for local optimizations. The platform automatically re indexes the schema and performs a series of optimizations on the queries and data structure to improve runtime. This process is performed in an automated, agile and continuous manner streamlining the entire process to the hands of the Data Scientist and Analyst.

Dashboards	Database	Vanilla Redshift Baseline	Panoply Baseline	Panoply - 2 day	Panoply - 7 day	Panoply 12 day
Big Dashboard	tpc-ds-1	152.3	310.4	185.1	171.4	48.8
Income and Profit	tpc-ds-1	30.2	91.3	47.6	54.6	32.7
Sales by Gender Dashboard	tpc-ds-1	21.8	135.9	67.1	72.6	5.8
Sales and Profit by State	tpc-ds-1	41.2	95.4	43.3	47.9	4.4
Sales Tax and Coupon by City	tpc-ds-1	16.9				
Education vs. Spend	tpc-ds-1	16.6	89.7	35.0	37.6	4.4
Sales by Time	tpc-ds-1	29.2	96.0	31.5	35.5	4.9
Sales Returns Dashboard	tpc-ds-1	17.1	65.7	21.0	23.5	7.8
Profit by Category and Brand	tpc-ds-1	15.4	42.0	26.1	25.7	5.3
		Avg. Total Load Time	Avg. Total Load Time	Avg. Total Load Time	Avg. Total Load Time	Avg. Total Load Time

Security

Enhanced Privacy

Panoply offers additional layers of security on top of your general Datacenter security, such as columnar encryption that enables you to encrypt your data thus providing you full control over your data. Keys are rotated on a daily basis.

Access Management

Panoply implements two-step verification. You can also require your colleagues to change your passwords at configurable intervals.

Panoply's permissions model enables you to restrict access to specific tables, views or columns for hierarchical security protocol. Panoply's platform enables you to set read-only permissions and lock down your data so it can't be exported. To prevent former employees from retaining access to data, user management options include the ability to automatically expire users after a configurable amount of time, requiring manual renewal of their accounts before they can resume access to the data.

Panoply enables you to audit your company's data access patterns and identify any suspicious activity. Panoply does not have access to any of its clients' raw data or encryption keys.

Test and Risk Assessment

Security is a top priority in the implementation of all R&D processes, across all system layers, from the physical layers up to the application layer. Strict development processes, coding standards, and a rigorous testing platform ensure adherence to industry-standard, best practices for security. In addition, Panoply's testing protocol includes code reviews and tests for quality assurance.

As part of Panoply's ongoing risk assessment process, our security professionals meet with the R&D leaders on a monthly basis to review risk assessments, system security configurations, and policies. In addition, the teams discuss and review new projects and their potential impact on security.

Network Security

Panoply takes every precaution to ensure that every layer involved in data transfer is secured by best-of-breed technologies. The company's network is segmented using AWS security groups, VPCs, ACLs, and additional custom measures. In addition, Panoply's security operations center (SOC) is kept up to date with security alerts that are analysed and addressed in real-time.

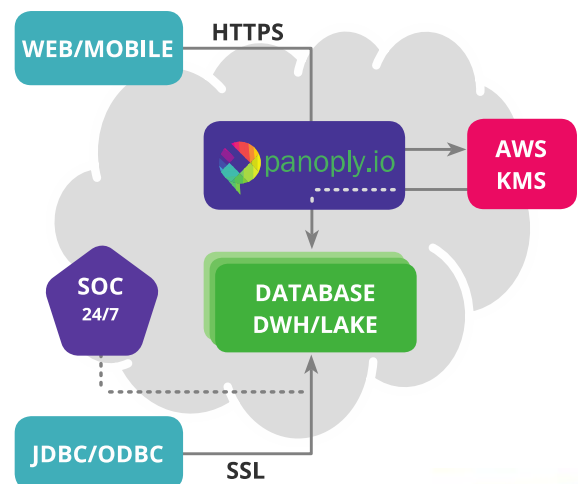
Access to the Panoply systems are monitored, logged, and analysed 24x7. Anomaly detection is used to identify strange connections to data.

Audits

As part of our long-term commitment to ensuring the security and privacy of the data, Panoply conducts annual gray-box penetration tests to all of its code.

Finally, Panoply produces an annual security audit report with all of the findings for our customers. For more information, please review our full Security Overview document.

Panoply ensures end-to-end encrypted security of data and access



Panoply's mission is to bring
the time from data to value to none