panoply

Technology Specifications

PRODUCT SHEET

All-in-One Data Platform

Panoply is the only all-in-one data platform that automates all three key aspects of the data stack: data ingestion, data management, and query performance optimization. Panoply empowers anyone working with data analytics to quickly gain actionable insights on their own – without the need of IT and Engineering. 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. Meet Panoply – the only all-in-one data platform.

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. To ensure efficient query performance, Panoply's algorithms automatically sense frequently queried data and cache it in memory.

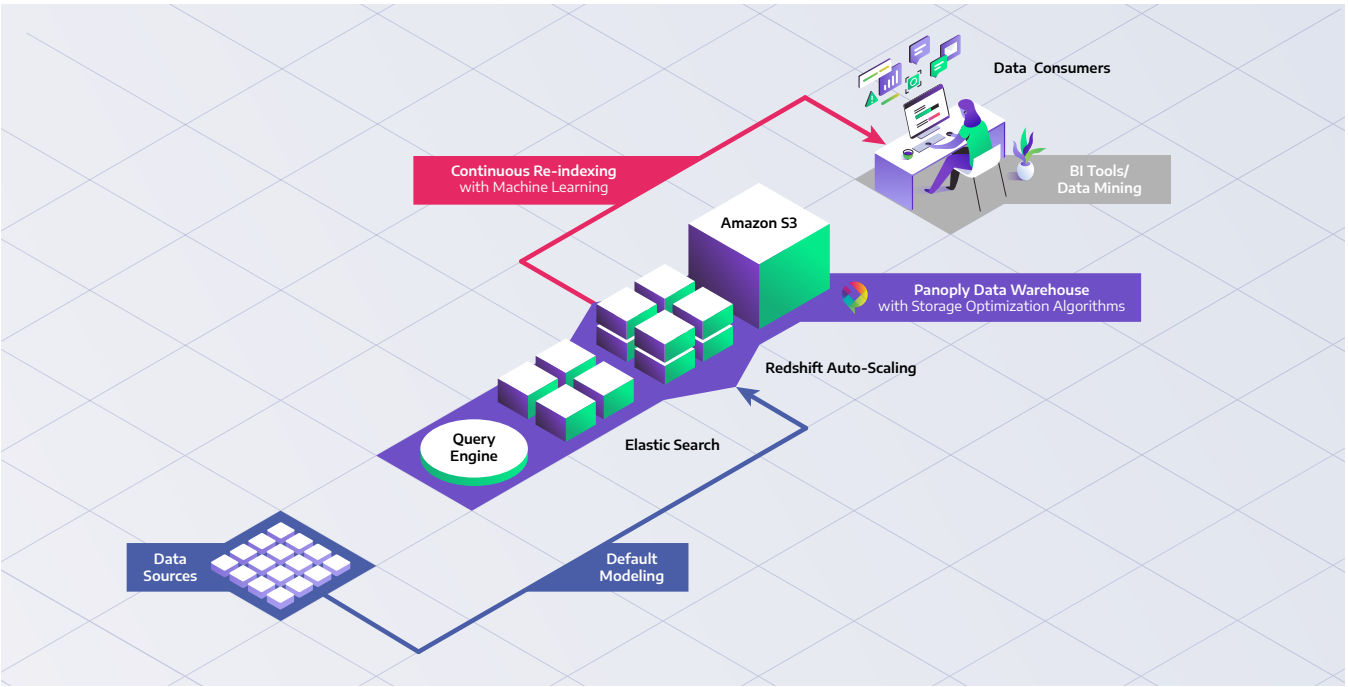
The default architecture is based on Amazon's S3 and Redshift components although the multi-tier stack can be adjusted based on specific organization needs.

Data Sources

Importing data into a traditional on-premises 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 databases 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 compression, nested objects, etc... Making it:

"The only data warehouse management solution with built-in ETL"



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 its maintenance is an ongoing labor intensive task. Panoply offers several tools for automated maintenance of your analytic infrastructure.

- Re-indexing 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 preferred business intelligence, SQL workbench, or

advanced analytics tool and beginning to analyze. Panoply:

- Automates data source connections
- Utilizes AI 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 look 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.

Security

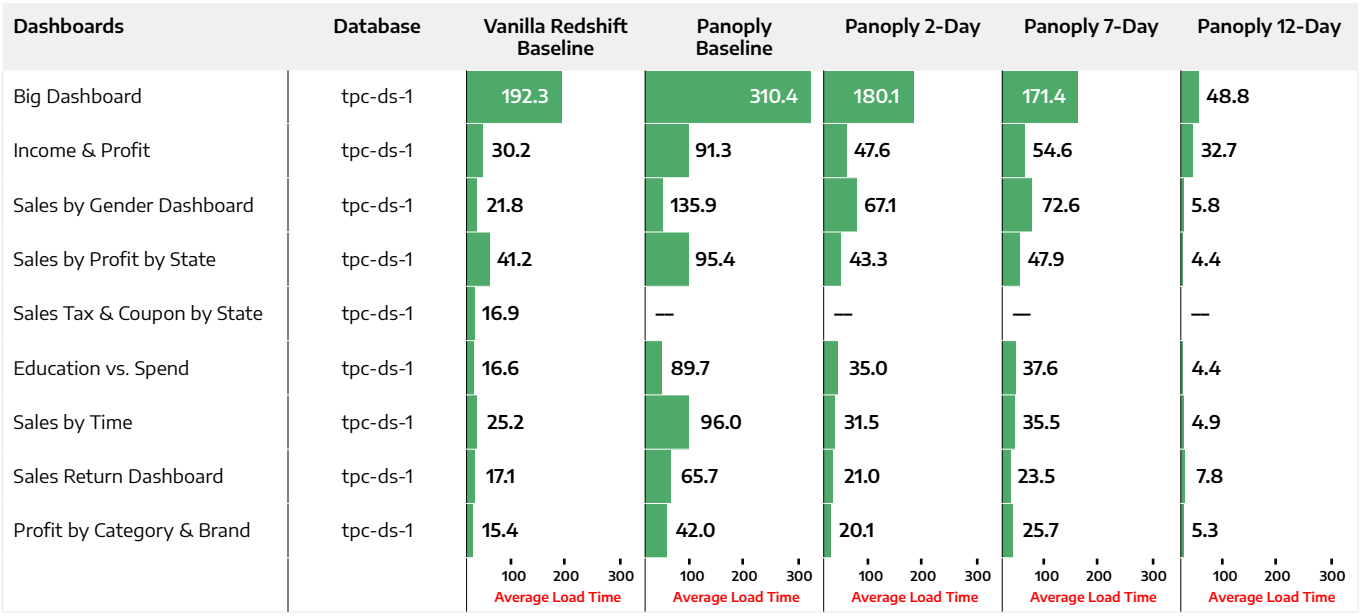
Enhanced Privacy

Panoply offers additional layers of security on top of your general datacenter security, such as columnar encryption that enables you to encrypt your data in a flexible manner with the option of using your own private keys. Keys are rotated on a daily basis.

Access Management

Panoply implements role-based permissions, giving you control over the data that can be accessed by admins, editors, and viewers.

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.

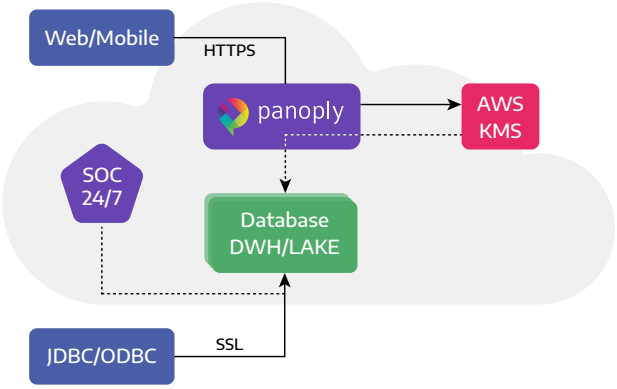
As part of Panoply's ongoing risk assessment process, our security professionals meet with the R&D leaders on a monthly basis to review risk assessments, system security configurations, and policies. In addition, the teams discuss and review new projects and their potential impact on security.

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

Panoply is the only all-in-one data platform built for analytics, that automates all three key aspects of the data stack: data ingestion, data management, and query performance optimization.

Get your Free Trial at Panoply.io

G CROWD HIGH PERFORMER

aws partner network Advanced Technology Partner Big Data Competency

Gartner COOL VENDOR 2019

hello@panoply.io | www.panoply.io | 188 King St. San Francisco, CA 94107