

Element DPS (Data Pipeline Service)

Dexter Taylor, Principal Engineer
Binary Machines LLC

Element is a service offering aimed at SMB's which have significant internal data analytics needs, but which lack the staff and other resources to supply those needs. An Element service instance consists of:

- A custom, cloud-hosted software stack to pull analytical input data from approved sources (e.g., S3 data dumps or replicas of production databases)
- A cloud-hosted data pipeline which performs ETL (Extract, Transform, and Load) operations on that data
- The transformed data written into an OLAP (Online Analytical Processing) database schema.

The Element service combines consulting services and hosted, managed software, with the goal of supplying back-office data in the format decisionmakers need, when they need it.

Process

The Element workflow consists of three phases:

1. **Discovery.** We conduct an intensive data discovery process, where our engagement teams consult directly with the client to define the analytical questions the client wishes to answer and the data sources where the information is located.
2. **Buildout.** We build, test, and deploy a data pipeline which pulls from the approved data sources and populates an analytical database. Data cleaning, reformatting, and correction happens in this pipeline, which can be set to operate at regular intervals (usually on an overnight batch run). The pipeline is hosted in our private cloud data center; the analytical database may be hosted either in our data center, or the client's.
3. **Support.** We follow up with the client to ensure that their data needs are being met, and we provide monitoring and maintenance of the data pipeline for the rest of its operating life.

Cost

- The discovery phase is billed on an hourly basis; clients only pay for the time necessary to fully specify (a) the business facts which the pipeline will supply, and (b) the client data sources necessary to feed the pipeline's input..
- The buildout phase is done on a **fixed price per-pipeline** according to tiered data volume, and pipeline hosting fees are assessed on a monthly, quarterly, or yearly basis.
- Basic support is included in the monthly pipeline hosting fee.

Options

We offer the following services in addition to pipeline hosting:

- **Custom reporting.** We can supply plug-in modules to generate and auto-deliver custom reports from the analytical database, in either Excel, Google Sheets, or CSV format.
- **Streaming data / event capture.** We can provide a software stack that will allow injection of real-time business events into an Element data pipeline.
- **Dashboards / data visualization.** We can build and host custom web-based dashboards to expose commonly used analytical functions, with interactive visualizations to highlight important patterns in your data.
- **Data hosting.** We can provision and host the analytical database in our cloud data center, with 24/7 monitoring and regularly scheduled backups to Amazon S3 storage.

Summary

Data is the lifeblood of the modern enterprise. Collecting, cleaning, transforming, and delivering critical business data is often an intensive -- and expensive -- task, requiring custom software and the time and attention of skilled personnel.

At Binary Machines, we have over a hundred years of combined experience in data architecture, software development, systems administration, and analytics. The Element data pipeline service brings that expertise to bear on the analytical challenges facing small to medium enterprises, helping to maximize performance while minimizing cost and risk.