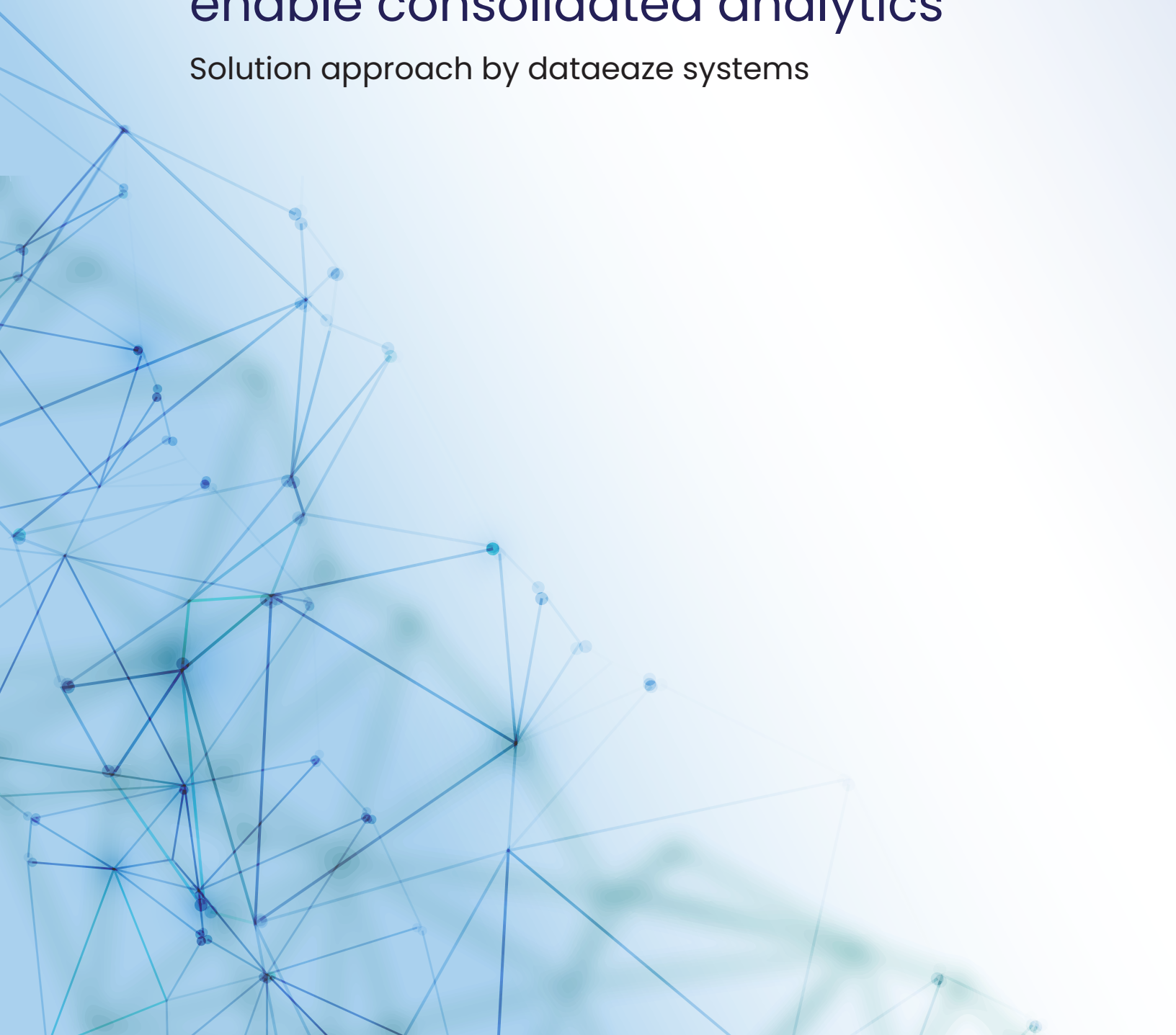


Case Study

Data platform for manufacturing organization to enable consolidated analytics

Solution approach by dataeaze systems



About the client

Client is a billion dollar business conglomerate, is a leading fiber optic and cable manufacturer company having presence across globe and manufacturing units based out of India, Europe and China.

As part of customer offering, customer builds core fiber optic end to end, also provides network installation offering. Customers have large scale contracts of setting up network infrastructure across large telecom operators as well as government agencies.



About data

Being a leading organization in fiber manufacturing and installation, customers have all functions of a manufacturing organization : Purchase, Planning, Operations, HR and a network of Vendors.

Various systems being used to handle these functions,

- SAP
- Salesforce
- Custom built ERP
- HR management system
- Optic fiber manufacturing machines generate data

SAP, Salesforce, custom built inhouse ERP is being used for achieving smooth functioning of all departments. These systems generate data points

Machine data is also captured for the purpose of performance measurement, wastage calculation and planning for optimization in fiber optic manufacturing process.

Need of Analytics

Data consolidation

Data is spread across multiple systems,

- ERP data into SAP and custom internal ERP
- Sales and order data into : Salesforce and SAP
- Human resource data : Into HR management system
- Vendor execution data : Into custom vendor tracking system
- Customers data : Into Customers tracking system
- Finance data : Into SAP and other systems
- Company Master Data : Client has plants across the world
- Machine data : Collected from edge devices

In order to enable any meaningful consolidated analytics, it is required to have data from all sources into a single central data lake and warehouse.

Various internal analytics reports

Decision makers need various reports on periodic basis, some of these are mentioned below,

- Order to cash report : Get end to end metrics for every project in execution (ie. Sale order, distribution across vendors, milestone based invoices, payments, vendor payments etc.)
- Sales funnel report : Various metrics to analyze business pipeline and to track performance of business development team
- Human resource Monthly reports : Various metrics covered which analyze and calculate the business need
- Production Monitoring system : batch data processing at very low frequency rate
- There are multiple more such use cases

Analysis of machine data

Fiber optic manufacturing process generates various data points for every batch of cable being manufactured. This process generates various data points. These data points are analyzed for following use cases,

- Measure performance of fiber optic manufacturing process.
- Analyze scrap generation and apply predictive methods to come up with methods to reduce scrap / defective product
- Analyze data to optimize end to end manufacturing process
- and more

Customer pain points

Need to design central data warehouse

- In order to achieve consolidated reporting
- It was decided to build central data warehouse on BigQuery data warehouse.
- Necessary skills to design data lake and reporting specific data warehouse data marts and to achieve it in required time was a challenge

Robust stable data automation to move data

From SAP

- Data onboarding from SAP needs to handle connection with SAP interface
- Need to handle scale of line level data generated at SAP
- Handle incremental pull and support for upsert at destination
- SAP has tables in non human readable format, understand necessary fields and setup robust automation

From Salesforce

- Connect with salesforce source and setup high volume data onboarding
- Mapping of salesforce data with SAP data to decide interconnection mapping

Volume of data

- Total volume of data is high, introduces challenges of first full data move, as well as requires incremental pull for data sync feasibility and cost optimization.
- Daily movement of data is also high, introduces a need of handling complexity for high volume data movement

Machine data being on internal network to move to cloud data warehouse

- Machine data is generated by Scada based system and stored in internal SQLServer database without having direct access to.
- Need to setup robust automation to extracting data and move to cloud.
- Process to translate data from machine format to analyzable format

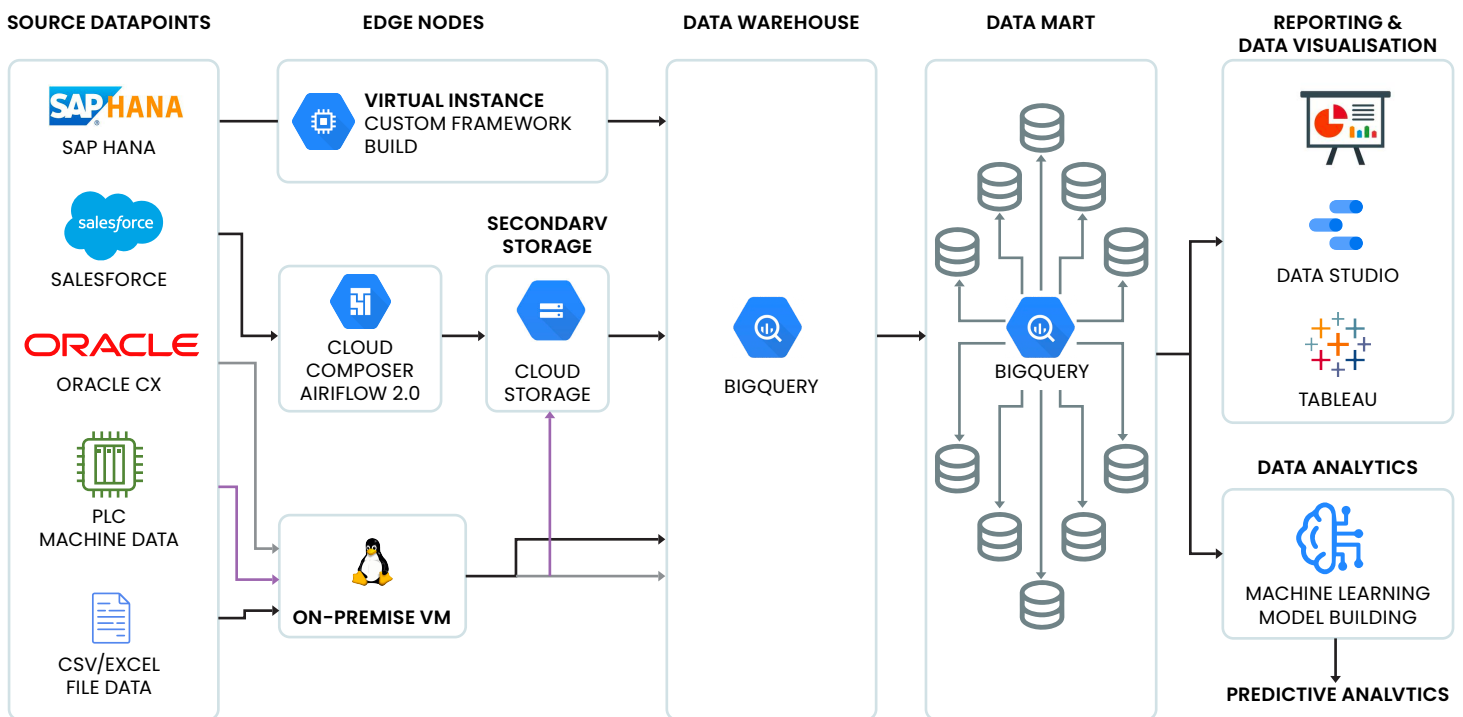
Setup data governance and observability

- Ensure low maintenance data platform
- Setup robust data pipelines which ensure data is always available
- Ensure best data governance practices

Availability of tech experts with required skills to maintain and enhance further

How has Dataeaze assisted the customer?

Built modern data platform



Dataeaze built a modern data platform for customer data with key features,

- Central data lake on Google Compute Storage
- Data onboarding automation (With Spark on Google Data Proc, Python based data pull connectors, Scheduling through Google Cloud Composer (AirFlow))
- Near real-time stream processing capability (With Spark on Google Data Proc)
- Data warehouse with Big Query
- Reporting with Tableau

Set robust data onboarding automation

Set continuous scheduled automation for data onboarding for

- SAP data about ERP and orders (Based on Python and PySpark)
- Salesforce data about lead generation and business development
- Machine data automation to move data from on premise Scada based system to Google Compute Storage.
- Data movement from Oracle and Mysql which are backbone of internal systems

Provided dataeaze experts to build use cases

Dataeaze data experts are a part of the customer data engineering team, with continuous support to develop data processing pipelines for new analytics needs.

Ensured robustness and low maintenance of automation

Setup of

- Alerting of data pipelines
- Monitoring framework to observe robustness of data movement

Benefits to customer

Central data platform

- For data across ERP, CRM, Machine data, HR data.
- Client is now enabled with the central data platform where data is getting on boarded continuously and automatically.
- Always available for analysis for analytics and data science teams.

Easy reporting for analytics team

- The Analytics team now has a single source of data available, making it easy for reporting and analytics.
- They get robust data availability SLAs which are met with full data accuracy. T-1 data is always available in the data warehouse.

Robust low maintenance data automation

- Build provisioning aspects of data platform
- Alerting and monitoring to ensure data platform stability

Complex data onboarding process for Machine Data

- Machine data is getting on boarded to cloud data warehouse
- Complex process, automation and observability is set in order to ensure data is always available for reporting and analytics.

Reporting data warehouse

- Reporting friendly data warehouse schema is built
- Making it easy to build self serve reports through BI tool
- It is a scalable data warehouse capable of handling large volume of data and reporting use cases.

Accessibility to data experts

- Dataeaze data experts are involved to implement data ETL automation and pipelines
- This ensures to achieve necessary speed of development and to mitigate uncertainty of availability of experts

dataeaze

www.dataeaze.io

About dataeaze

Dataeaze helps its customers build an analytics data platform around modern big data ecosystem. Dataeaze systems is focused on making it easy for organisations to work with data. Organisations assisted by Dataeaze get benefit of quick bring up of robust data platform with analytics capabilities brought up as per need.

contactus@dataeaze.io