Automotive Cloud Data Mesh Platform

Challenges with monolithic data warehouses

Market research

According to IDC (Source: US47771821), the world created or replicated astounding 64.2 ZB of data in 2020. IDC is forecasting 23 percent compound annual growth rate for the global datasphere from 2021 to 2025. But simply having more data is not what matters. Based on IDC´s 2021 Global DataSphere study, less than 3 percent of the data currently created is analyzed to affect enterprise intelligence. 59 percent of mid to upper-level managers in an IDC survey say they are overwhelmed by the amount of information available for decision making; however, 45.6 percent of respondents in the same survey say they don't have enough data and information available to make good decisions. The enterprises that manage to square that circle will end up as the winners of the battle for intelligence.

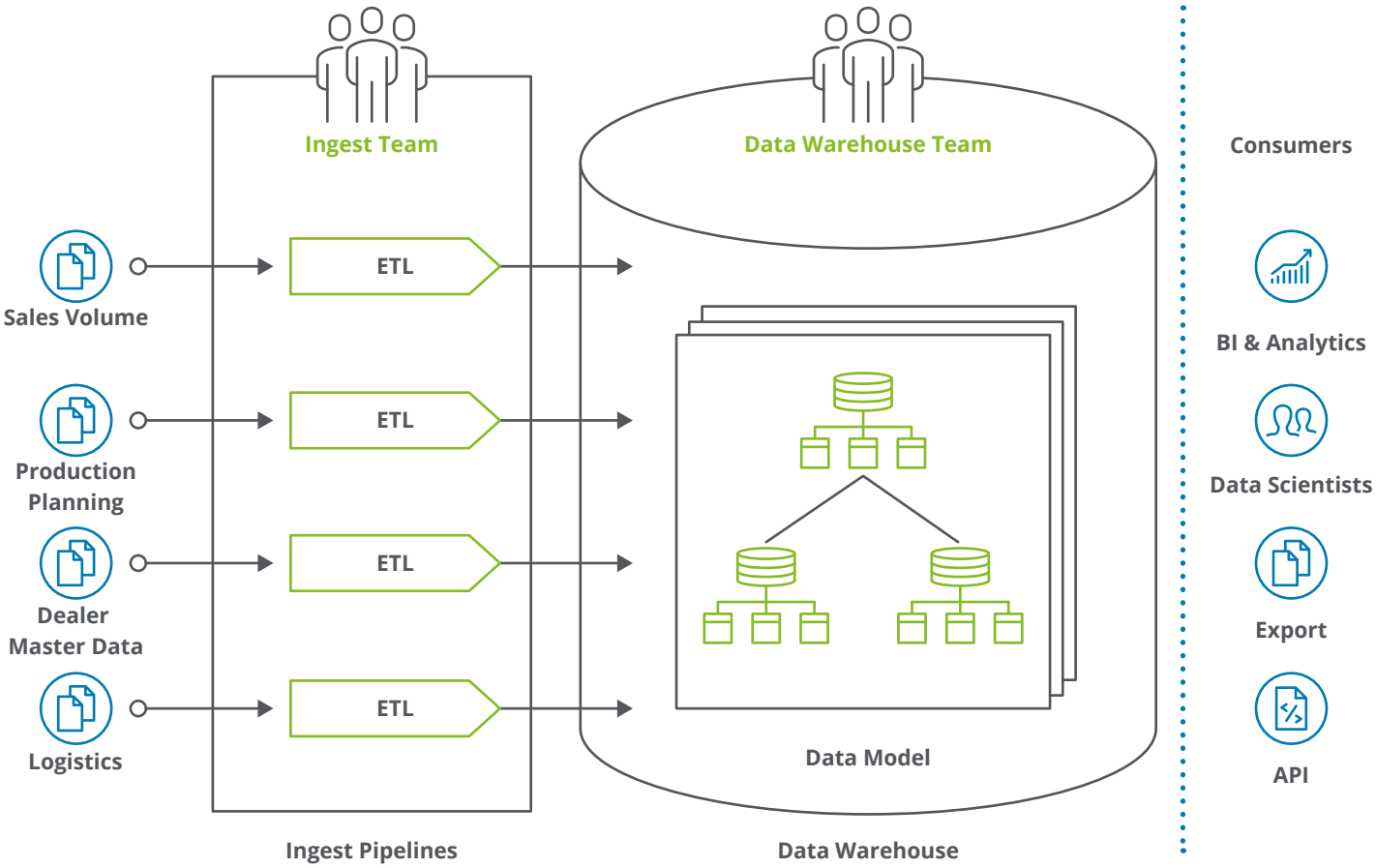
Rapid growth in automotive data
Automotive OEMs capture large amounts of data across the value chain (e.g., from sales, the supply chain, production, distribution, after-sales service and connected vehicles). The manufacturers use these large data sets to analyze and create actionable insights on a variety of use cases such as customer behavior, driving patterns, personalized customer experiences, risk forecasts for the supply chain and production, asset tracking and tracing, giving them the insight they need to make key decisions.

In most cases, the IT system converts this data in a centralized extract/transform/load (ETL) pipeline and stores it in monolithic data warehouses or data lakes that have developed over decades, forcing the OEMs to find a solution to monetize this data at scale.

Some of the challenges automotive OEMs face with the growing volume of data:

- Difficult, slow integration of new data sources as they become increasingly complex
- Problems with data sharing compounded by silo mentality
- Lack of data source domain knowledge in centralized data ingestion and warehousing teams
- Limited ability to address data quality issues at the source or react to change
- Need for a more economic, effective and efficient IT architecture as dispersed data becomes the norm
- Lack of a domain-oriented data catalog to ease data discovery for consumers

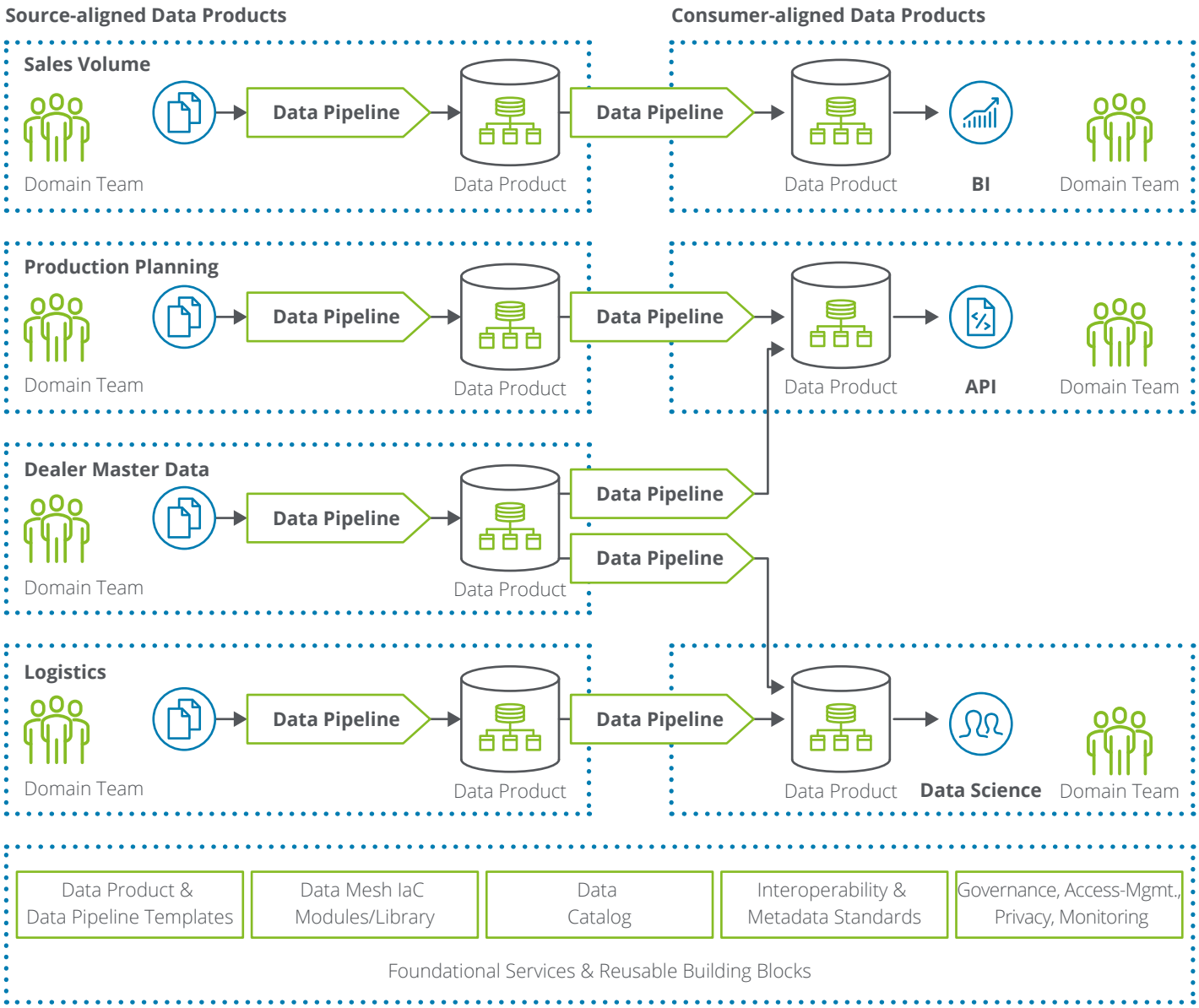
Fig. 1 – Monolithic Data Warehouse (DWH)



Data mesh – an architectural paradigm unlocking analytics at scale

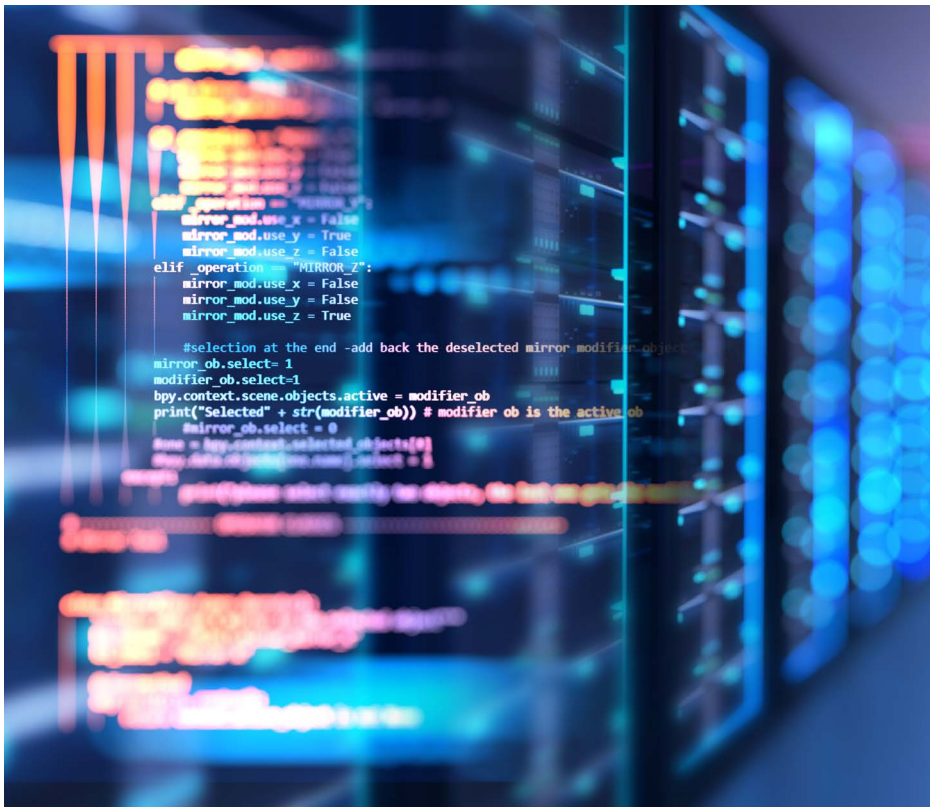
Data mesh is a paradigm shift in modern distributed architecture that embraces the reality of distributed data and treats data as a product. The concept is inspired by domain-driven design, decentralizing the responsibility for data to the corresponding specialist teams within the organization. Cross-functional data domain teams take on responsibility for specific data, from acquiring the data they need and processing it to making it available to the mesh as a data product. Enterprises generally do not explicitly build a data mesh, but rather create it by putting a federated governance structure in place as well as the means of exchange and discovery. Data discovery and exchange requires a set of central tools and policies.

Fig. 2 – Data mesh



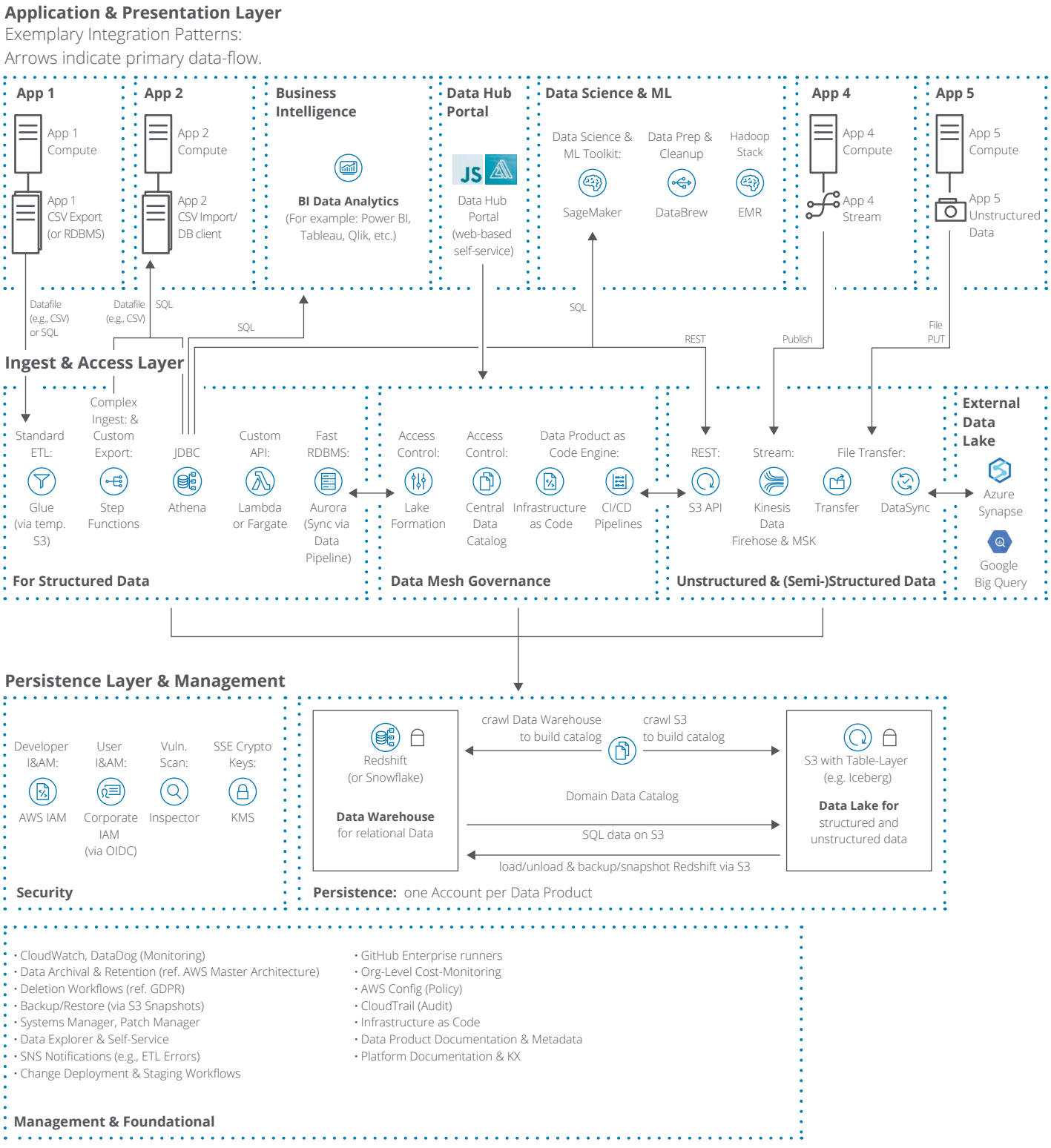
The key benefits of a well designed and implemented data mesh include:

- Domain-centric ownership of data sources, pipelines and increased data quality
- Data assets offered as products in a “serve & pull” instead of “push & analog” model
- Ownership residing with domain team, ensuring better quality for data consumers
- Faster response from domain teams to source format changes or quality issues



OEMs experience a rapid growth in automotive data.

Fig. 3 – Cloud Data Mesh Platform



Automotive cloud data mesh platform

Designing and implementing a data mesh for an automotive business domain (for example: vehicle diagnosis) requires a highly scalable infrastructure along with the following features on a public cloud platform:

1. Self-service capabilities
2. Integration with other data lakes and interfacing applications within the organization
3. Security and data protection
4. Efficient DevOps tools
5. Mesh governance
6. Automated data pipelines
7. Interoperability among data products
8. Modern big data & AI capabilities

Reference architecture on AWS

Key features of an automotive cloud data mesh platform:

1. Self-service: Enable business users to create their own reports/analyses – and empower users to ingest new data sources using ingest/ETL templates.

5. Interoperability: Rely on standardized APIs to enable easy integration with other applications and data lakes.

2. Scalability: Use auto-scaling mechanisms to easily scale with demand. Unlock new use cases by augmenting the DWH with a data lake.

6. Operative efficiency & reliability: Deploy managed IaaS and PaaS solutions to free up development capacity for client-specific value creation – while offering predictable reliability.

3. Modern features: Use the latest BI & guided analytics tools to provide new features and an attractive look & feel.

7. Security & data protection: Comply with strict security requirements for identity & access management, encryption, audit logging, archival/retention, anonymization/deletion and cyber defense.

4. Performance: Make the most of the latest hardware and passive infrastructure innovation of hyperscalers.

Key technologies & methods

- DWH: AWS Redshift
- Data Lake: S3 with Lake Formation
- Ingest: Glue, Step Functions, DMS, Kinesis Firehose, Transfer, DataSync
- BI Frontend: Qlik Sense, Power BI, Tableau, custom JavaScript
- Custom API: Lambda, Fargate
- Big Data Query: Redshift Spectrum, Athena, Iceberg
- AI: SageMaker & AWS AI Portfolio

Target outcomes & benefits

- Modularized structure of data products with defined interfaces to offer better scalability
- Data ownership with the domain team that has the skills and business knowledge to maintain quality
- Federated governance in the data mesh allowing DP owners to monitor data usage as well as the reason for usage
- Integration with organizational data exchange layer (other external data lakes) to unlock cross-domain data sharing based on predefined governance policies
- Increased data discoverability with central and domain-based data catalog
- Data-product-as-code approach makes data more maintainable and lowers the entry barrier for development of new data products/assets
- Seamless (zero-downtime) onboarding/migration of individual workloads (data assets, reports, dashboards and KPIs as well as custom APIs)
- Benefits due to passive innovation and a large talent pool
- Lower barrier to entry for users due to self-service capabilities
- Pay-per-use pricing structure in public cloud platform helps the OEMs reduce overall storage and computer costs in comparison to legacy solutions

Contacts

Arindam Mandal
Partner
Automotive Cloud Lead, Germany
Tel: +49 89 29036 5733
armandal@deloitte.de

Lukas Kästner
Senior Manager
Automotive Cloud
Tel: +49 89 29036 6725
lkaestner@deloitte.de

Devendar Nandi
Senior Manager
Automotive Cloud
Tel: +49 89 29036 6732
devnandi@deloitte.de

Deloitte.

Issue 11/2022