

Data engineering projects we've delivered

Industry	Description	Data extraction	Data operations	Data storage	Data analysis	Data size
Advertising	Marketing campaigns & Attribution solution	Web scraping (custom Python scripts), internal data streams, internal databases (Aerospike, SQL database, Elasticsearch)	Data streaming: FoundationDB, Flink, Kafka, Airflow Batch processing: Airflow, FoundationDB, Spark	FoundationDB, Elasticsearch, HIVE, Parquet	Looker, Jupyter Notebook, custom Java services	Billions of records per day
Advertising	Managing ad campaigns	Custom Java code	Custom Java code	MySQL database	Custom Java code	1 million records per day
E-commerce	New Data Warehouse implementation	Domain systems, Excel and flat files, External application rest and SOAP API	Apache Airflow, Talend Open Studio for Data Integration, SQL Code, Python Scripts	PostgreSQL	Tableau	250 k records increment per day
Financial	Offloading Splunk log-streams to Hadoop for long-term storage	Apache Kafka, Apache Spark, Splunk	Apache Spark	Apache Hive, Impala, Kudu	Apache Hue (Hadoop-User-Experience)	NA
Financial	Processing data from a Bank application system in PL/CZ/SK region with fully dynamic approach, loading entire DWH model based on configuration and mappings.	Apache Kafka, Apache Spark, MSSQL	Apache Spark, Python, OracleDB, Scala	Apache Hive, Apache HDFS, Impala, Kudu	Apache Hue (Hadoop-User-Experience)	700 different target business tables 1.5 million application states daily
Financial	Real-time machine learning inference platform	Apache Kafka, Apache Spark	Apache Spark, Apache Airflow	Apache HBase, Apache Hive	NA	Tens of millions of records stored and utilized in FeatureStore Tens of thousands of events served daily in real-time

Industry	Description	Data extraction	Data operations	Data storage	Data analysis	Data size
FinTech	Identification of the latest contact details for a large group of people	Microsoft SQL Server Integration Services	Microsoft SQL Server Integration Services, Custom SQL code	Microsoft SQL Server	Microsoft SQL Reporting Services	10 TB of data
FinTech	Migration of an on-premises SQL server-based warehouse to cloud data lake	SQL databases, internal applications, flat files	Databricks, Spark, Spark streaming	Azure Data Lake House, Azure SQL	PowerBI	20 GB of data per day
Food	Monitoring fish conditions for fish farms	Multiple custom APIs of IoT devices (REST, GraphQL, SOAP)	Databricks, Spark	Azure Data Lake House	PowerBI, Azure Machine Learning Services	10 GB of data per day
Insurance	Support of integration platform	Domain systems, external application rest and soap API	IBM InfoSphere DataStage	IBM Informix, Oracle DB	NA	NA
IoT	Generic solution for gathering data from IoT devices	Custom IoT APIs, Azure IoT Hub	Databricks, Spark, Spark streaming	Azure Data Lake, Azure CosmosDB, Azure SQL	PowerBI, Azure Machine Learning	Hundreds of GB of data per day
Labor market	Analysis of labor market data	External systems	AWS Lambda	Snowflake, MySQL, S3	NA	Millions of records per day
M&E	Migration of on-premise Data Warehouse (SQL Server based) into cloud data platform	Domain systems, Excel and flat files, External application rest API	Apache Airflow, AWS Lambda, SQL Code	Snowflake	PowerBI	35 milion records increment per day
M&E	New application for regional management team reporting	Domain systems, Excel files (Sharepoint site)	Apache Airflow, AWS Lambda, SQL Code	Snowflake	PowerBI, PowerApps	NA

Industry	Description	Data extraction	Data operations	Data storage	Data analysis	Data size
M&E	New operational data store	Domain systems, Excel files	MS Azure Data Factory, SQL Code	MS SQL Server (Azure Managed Instance)	PowerBI	100 k records increment per day
Manufacturing	Analysis of forklift behavior	Custom IoT API	GCP dataflow	GCP Big Table, Timescale, Redis	Custom dashboard	300 million records per day
Manufacturing	Garbage recognition based on provided photos	Custom mobile app that takes photos	Amazon Lambda	Amazon S3, Amazon RDS, DynamoDB	AWS Sage Maker, Custom dashboards	A couple of GB per day
Mining	Accident prevention system in mines	IoT device API	Azure functions	AzureSQL	PowerBI, Custom Python scripts	A couple of GB per day
Real estate	Analysis of real estate data	Web scraping (custom scripts)	Snowflake, Spark, Databricks, Airflow	Snowflake, Postgres, Parquet	Custom BI solution	Billions of records per day
Retail	Streaming CDC ingestion framework	MSSQL Server, Kafka, Schema Registry (Avro)	Apache Spark, Apache NiFi	Apache HBase, Apache Hive	NA	Ingesting 200+ data structures
Retail	Support of integration platform	Domain systems, external application rest and soap API	IBM InfoSphere DataStage	Netezza, Snowflake	NA	NA
Social media	Analysis of social media posts	Custom databases	Kafka, Spark, Spark streaming	Elasticsearch, Cassandra, Postgres	Custom dashboards	2 billion records per week

Industry	Description	Data extraction	Data operations	Data storage	Data analysis	Data size
Telco	Generic stream processing tool used by the decision engine to support marketing campaign processes	Apache Kafka, Apache Nifi	Apache Flink, Apache Spark, Apache Airflow	Apache Ignite, Apache Hive, Apache HDFS, Teradata	NA	20 mld events per day
Telco	Creating Data Preparation Layer as a mediation layer for integrating with external Machine Learning tool	Oracle DB, Hadoop	Apache Spark, Apache Airflow, GreatExpectations, Python, SQL	Apache HDFS, Apache Hive, AWS S3	AWS Sagemaker	30k events calculated daily/ weekly/monthly Volume of a single periodic event dataset scales from hundreds up to tens of millions of records
Telco	Building a CEM Platform based on signalling network data, creating a framework for defining and running customizable reporting	File interface (flat-files), Apache Kafka	Apache Spark, SQL	Apache Hive	NA	Processing ~3 TB of data daily, ~0.8 PB of data in data retention window Processing 28 Billions of network events / day 16 Billions of GTP events 11 Billions of S1 events 307 Millions of VoLTE events 114 Millions of GB events 496 Millions of IUCS events 1,3 Millions of IUPS events Aggregating X0 Billions of network events / day Calculating X00 reports (daily, weekly, monthly)
Telco	Mirror of production environment for data warehouse migration project	SQL databases, internal applications, flat files, XML files	IBM InfoSphere DataStage	Netezza, Horton Works Data Platform	N/A	8 billions records per day
Telco	Anomaly finder in various business flows	NA	NA	Oracle	NA	A couple of terabytes per day

Industry	Description	Data extraction	Data operations	Data storage	Data analysis	Data size
Telco	Low-latency roaming event processor	Custom low-latency code to parse roaming data bit streams	Custom low- latency Java code, communication based on web sockets	Redis (KDB)	NA	30 million events per day (avg. processing time per event: 10 ms)
Telco	Managing mobile network hardware	Proprietary telecom hardware protocols	Hadoop, custom Java services	Redis, HIVE, Postgres	Custom dashboard	150,000 messages per second
Telco	Mobile network alerts monitoring	Apache NIFI	Kafka, Drools, Apache NiFi	Elasticsearch, Cassandra, Postgres	NA	1 million events per day

About Software Mind

Software Mind is a global digital transformation partner with operations throughout Europe, the US and LATAM. Driven by tech and empowered by people, Software Mind provides companies with software engineers and autonomous, cross-functional development teams who manage software life cycles from ideation to release and beyond. For over 20 years they've been enriching organizations with the talent they need to boost scalability, drive dynamic growth and bring disruptive ideas to life.

Let's talk about your data engineering goals

Contact us

