Data is precious. But if it is too time-consuming and expensive to prepare such data, this value can soon be eroded.
Experimenting faster is a trait shared by most innovative organizations. They want to experiment with new products and promotions to see if they can unearth bold new ways of serving the customer. But this experimentation cannot be random: It needs to be guided by data and based on current customer insight.

It was access to this data that was the problem for our client, a large consumer brand. It took a long time to prepare the data to a point where it could be used by business managers. So long, in fact, that the data was no longer relevant; and the moment was often lost. The company needed to be able to experiment faster but was held back by a cumbersome and ineffective analytics infrastructure.
There were a number of reasons it took so long. A heterogenous data landscape and a variety of tools (and not always the right ones) made life difficult, while running analytical models against the Hadoop scale data using traditional means simply took too long.

This mix of systems and the difficulty of extracting and preparing the data every time was not just time-consuming: it was expensive, because it required many skilled data scientists and engineers to make it happen. In fact, 80 percent of the effort went into data preparation; only 20 percent went into the actual analysis. Our job was to turn this equation around and give the business a truly efficient, self-service data analytics model.
HAVEN’T WE BEEN HERE BEFORE?

Yet these problems were quite typical. Because we have seen them before, we already had the answer. We had a proven way of modernizing the analytics infrastructure, so we didn’t have to re-invent it.

We have built one of the few platforms in the market with an integrated capability of discovery, blending, modelling, model management, and model consumption. Our platform leverages a unified metadata framework to provide end-to-end capabilities across the analytics life cycle. It sources the data, cleanses and prepares it, and makes it analytics-ready for business managers to use via a self-service tool.

All we had to do was customize it to work with the particular technology environment that our client had in place. Just three months later, the end-to-end capability was complete.
Hosted on Azure, with both cloud and native support, our implementation was an extreme automation model. Pre-fabricated design patterns reduced data upload time by 70 percent, and minimized the need for human intervention. Instead of multiple gating and handholding, ours was a lean engineering solution that moved us away from a service-based model to capability as a service.

Using our Information Grid platform, it operated on four levels: the physical data repository (the data lake), data preparation, and data modelling. On top of this was our Analytics Workbench platform, which gave business managers the intuitive self-serve tool they needed.
USING THE INFORMATION GRID: A TYPICAL EXAMPLE

NEARLY 70 PERCENT OF AREAS INDUSTRIALIZED BY INFORMATION GRID

OPERATION AND EXECUTION

INCEPTION

START

DEFINE THE DATA REQUIREMENT AND SLA

CREATE JIRA TICKET WITH INTAKE FORM

TEMPLATE REGISTRATION TO CAPTURE SOURCE, TARGET DETAILS, EXECUTE STEPS AND DATA STAGES

REGISTRATION TO CREATE SOURCE AND TARGET CONNECTIONS

JOB CONFIGURATION FOR DATA TYPE MAPPING & SYNTAX VALIDATION CHECKS

SEMANTIC VALIDATION MODULE

AUTO GENERATION OF ARTIFACTS

JENKINS JOB FOR CONTINUOUS INTEGRATION AND DEPLOYMENT OF ARTIFACTS

QA SIGNOFF

STOP

VERIFY OUTPUT

DEPLOY ON AZURE

VERIFICATION

AIRFLOW DAG EXECUTION

PREPARE AND SHARE ORGANIZATIONAL ENVIRONMENT DETAILS

INFRA TEAM

DATA ENGINEERING TEAM

FUNCTIONAL ANALYST

BUSINESS ANALYTICS TEAM

DEPLOYMENT TEAM

INCEPTION

START

DEFINE THE DATA REQUIREMENT AND SLA

CREATE JIRA TICKET WITH INTAKE FORM

TEMPLATE REGISTRATION TO CAPTURE SOURCE, TARGET DETAILS, EXECUTE STEPS AND DATA STAGES

REGISTRATION TO CREATE SOURCE AND TARGET CONNECTIONS

JOB CONFIGURATION FOR DATA TYPE MAPPING & SYNTAX VALIDATION CHECKS

SEMANTIC VALIDATION MODULE

AUTO GENERATION OF ARTIFACTS

JENKINS JOB FOR CONTINUOUS INTEGRATION AND DEPLOYMENT OF ARTIFACTS

QA SIGNOFF

STOP

VERIFY OUTPUT

DEPLOY ON AZURE

VERIFICATION

AIRFLOW DAG EXECUTION

PREPARE AND SHARE ORGANIZATIONAL ENVIRONMENT DETAILS

INFRA TEAM

DATA ENGINEERING TEAM

FUNCTIONAL ANALYST

BUSINESS ANALYTICS TEAM

DEPLOYMENT TEAM
A FOUR-LAYER MODEL

PLATFORM LAYERS

CONSUMPTION, SELF-SERVE ANALYTICS, SELF-SERVICE DATA PREPARATION

MODELING ZONE EXPLORATION, INSIGHT, MODELING

DISCOVERY, ACQUISITION, PREPARATION, DATA INGESTION, DATA CURATION, INFORMATION ENGINEERING

DATA

(DATA LAKE, RESERVOIR, ENTERPRISE SOURCES, ETC)

KEY CAPABILITIES

WORKFLOWS, DEFINING PARAMETERS, PROCESS EXECUTION, DATA EXPORTS, SCHEDULING, VISUALIZATION, KNOWLEDGE CAPTURE (ML)

CREATE, REPURPOSE, REFINE, INPUTS, VALIDATE, AND PUBLISH MODELS

DRAG AND DROP FEATURES, SMART SEARCH, WRANGLING, EXPLORATION

DATA WORKFLOWS, DEFINING PARAMETERS, PROCESS EXECUTION, DATA EXPORTS, SCHEDULING, VISUALIZATION, KNOWLEDGE CAPTURE (ML)

BUSINESS USER, DATA SCIENTIST

ANALYST / DATA SCIENTIST

ENGINEERS

ANALYTICS ROLES

ANALYTICS WORKBENCH

INFORMATION GRID

EXTERNAL DOCUMENT © 2018 INFOSYS LIMITED
WE DID THIS FOR THEM. WE CAN DO IT FOR YOU.

Find out more about how you can spend more time using data and less time preparing it. Reach out to us at askus@infosys.com