ITC-M1

Data Warehouse Using Talend ETL **Tool**

Instructors: Dr. PHAUK Sokkhey and Mr. Chan Sophal

Presented by KHUN Dararith

Content

Talend ETL Tool

What is **Talend?**

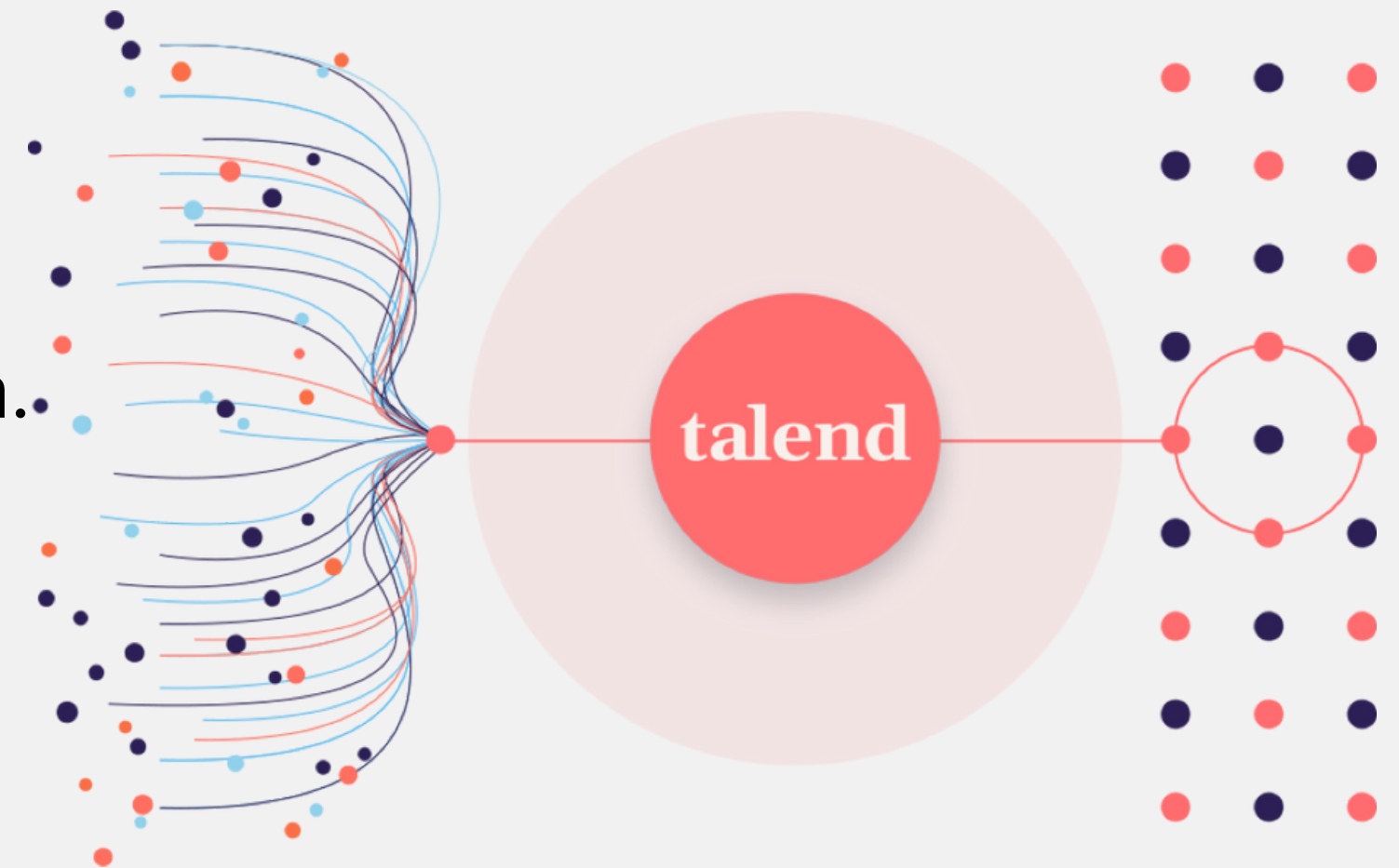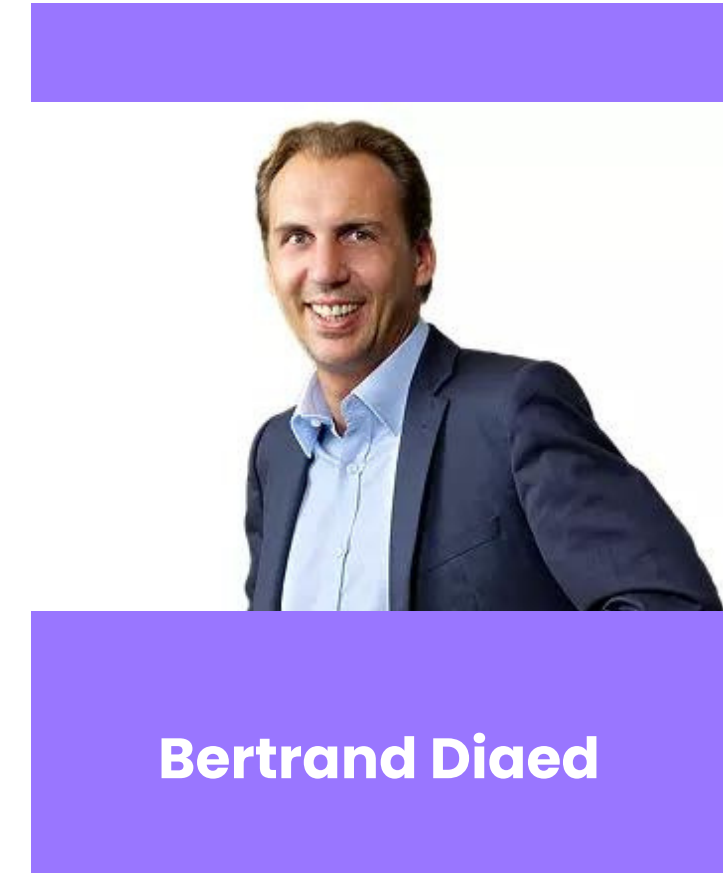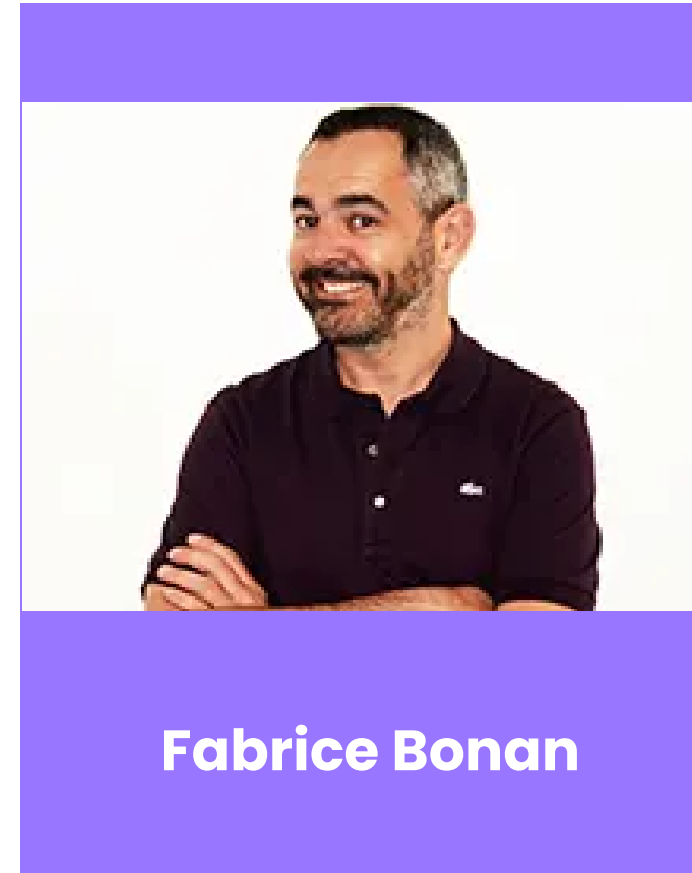Talend is an open source data integration platform. It provides various software and services for data integration, data management, enterprise application integration, data quality, cloud storage and Big Data.
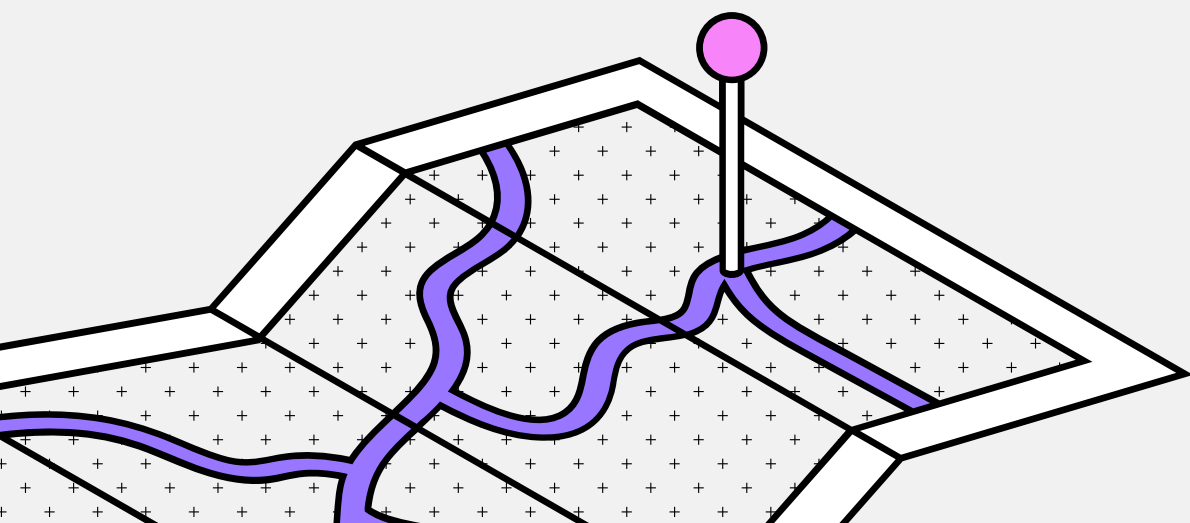
1

About Talend

was founded in 2005,

Where the headquarters is in Redwood City, California, United State.

Fabrice Bonan

Bertrand Diaed

2

Evolution of Talend

2006
2008
2010
2011
2012
2013

Data Integration

Data Quality

Master Data Management

Application Integration

Big Data

Hadoop 2.0

2015
2016
2017
Now!

Spark & Cloud

Data Preparation

Data Fabric

3
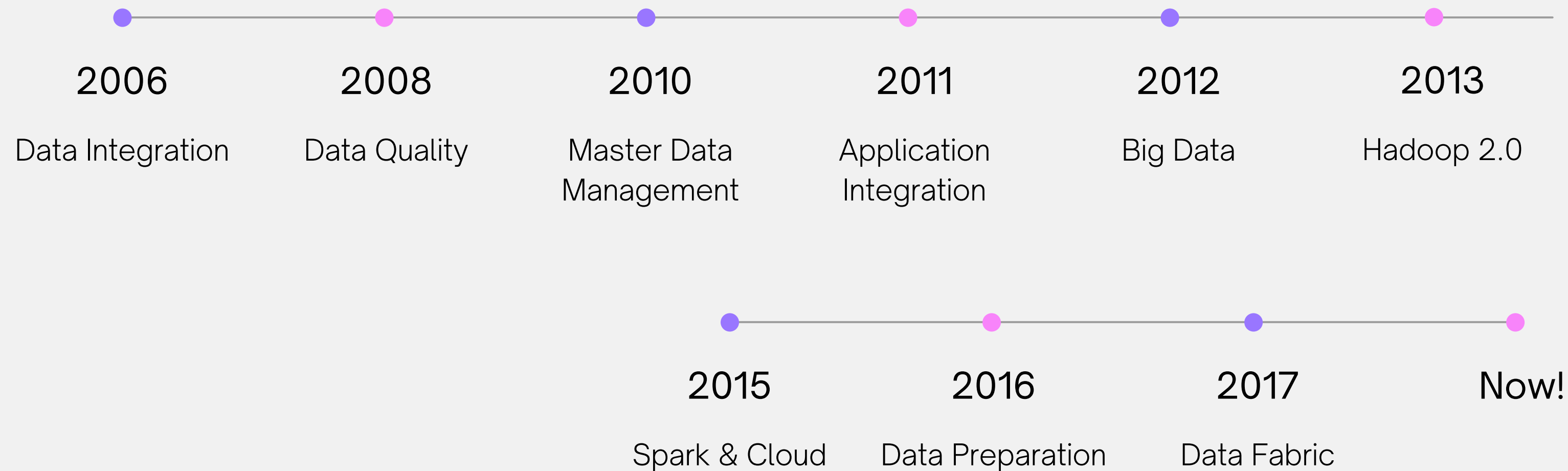
Talend
Products

Talend provides various usage for the user to enhance and handle on data.

Talend Platforms
- Big Data
- Data Management
- MDM
- Enterprise Integration
- Data Services
- Hybrid Cloud

Talend Enterprise
- Big Data
- Data Integration
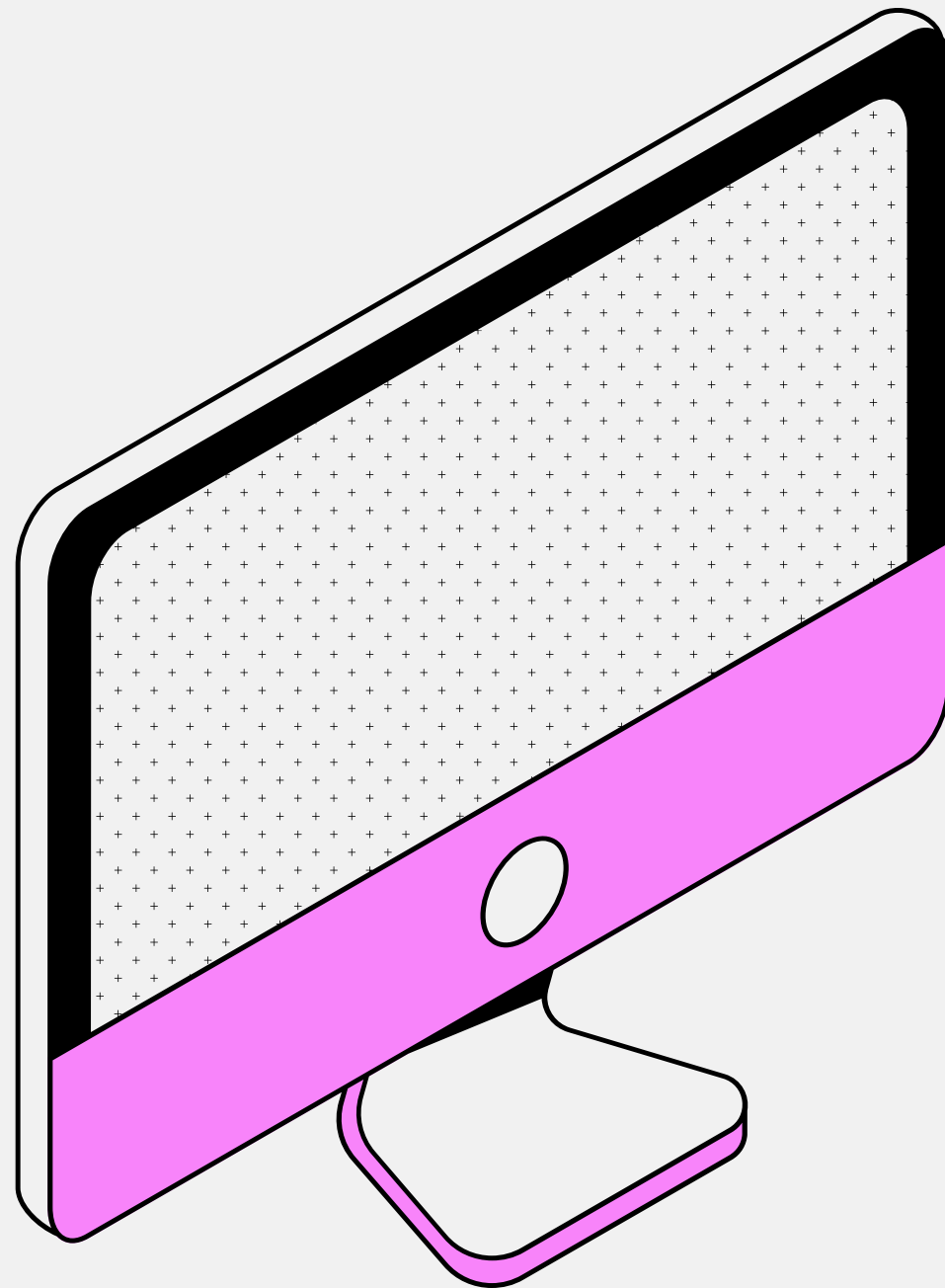- ESB

Talend Open Studio
- MDM
- Data Integration
- Big Data
- Data Quality
- ESB
- BPM

4

Introduce to Talend Open Studio

Talend Open Studio is a free open source ETL tool for Data Integration and Big Data. The tool will create the Java code for the job automatically and you need not write a single line of code.
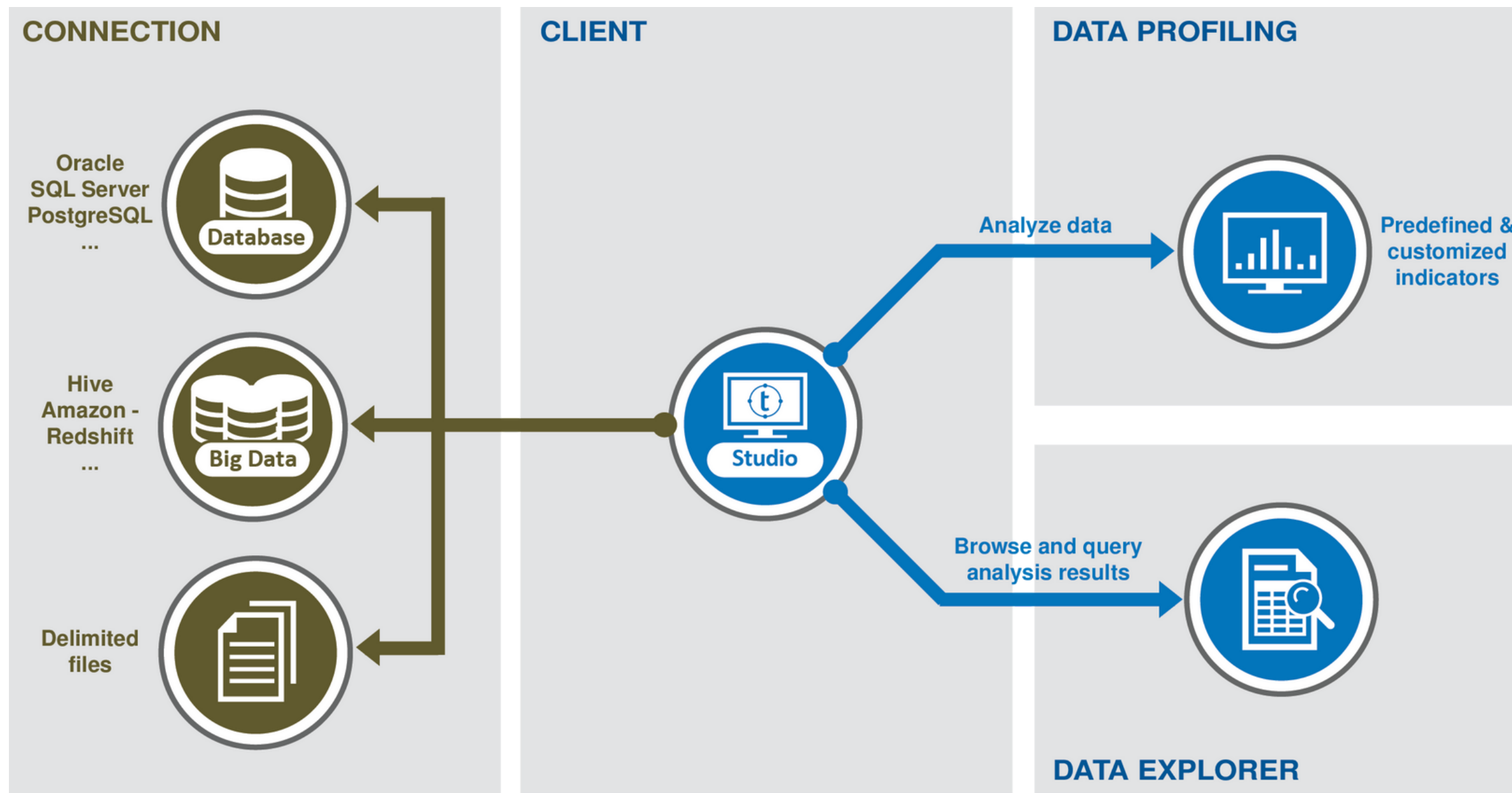
logo of TOS

5

TOS Architecture

The following chart illustrates the main architectural functional blocks.

6

Benefits of TOS

Here are some key benifits of using TOS:

Native Code — 01

Better Collaboration — 02

Faster Designing — 03

Early Cleansing — 04

Efficient Management — 05

Easy Scalability — 06

Real Time Statistics — 07

7

Real Life Use Case of TOS

Industry: Retail and E-Commerce

About Company:

Groupon is an American e-commerce marketplace, which connects subscribers with local merchants by offering various activities, travel, goods, and services in 15 countries worldwide.
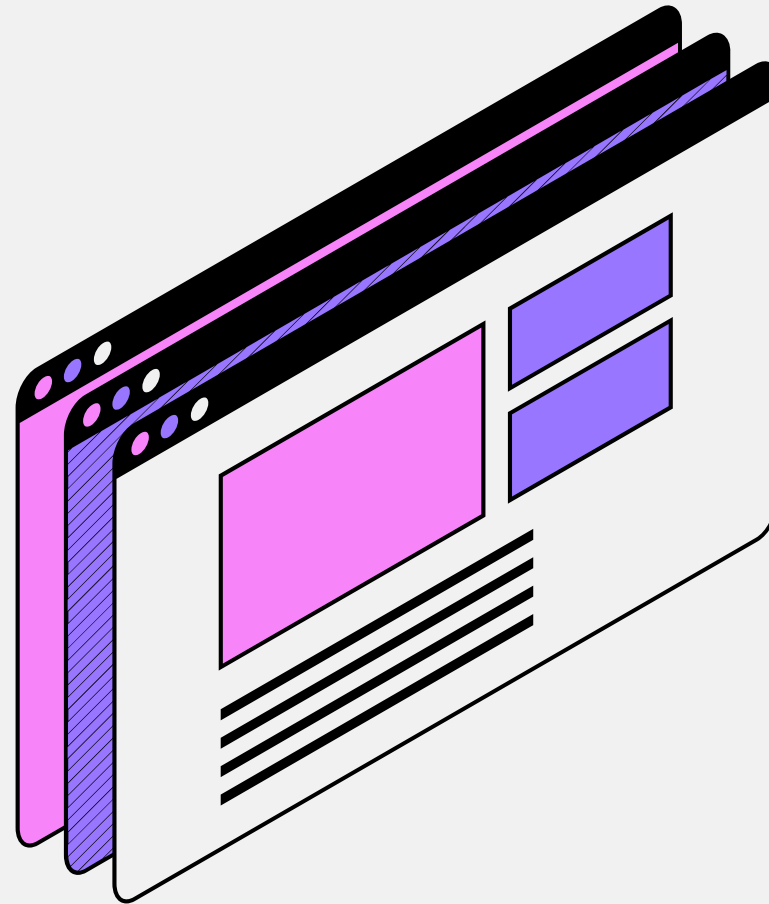
Where has it Helped:
Groupon uses Talend's Data Integration products for importing the raw data from different sources to the data warehouse, then transforming it into the desired format and then exporting it to various target systems.

Result:
1 TB of raw data is being processed in real time and stored in various target systems every day

1,000 data integration jobs are executed on a daily basis using ETL processes

In every 5 minutes, Talend updates the data warehouse with the OLTP system data

GROUPON®

8

TOS
Installation

Talend Open Studio: Open-source ETL and Free Data Integration

Execute simple ETL and data integration tasks in batch or real-time. Download Talend Open Studio today to start...

t Talend

Recommended Operating system

- Microsoft Windows 10
- Ubuntu 16.04 LTS
- Apple macOS 10.13/High Sierra

Memory Requirement

- Memory - Minimum 4 GB, Recommended 8 GB
- Storage Space - 30 GB

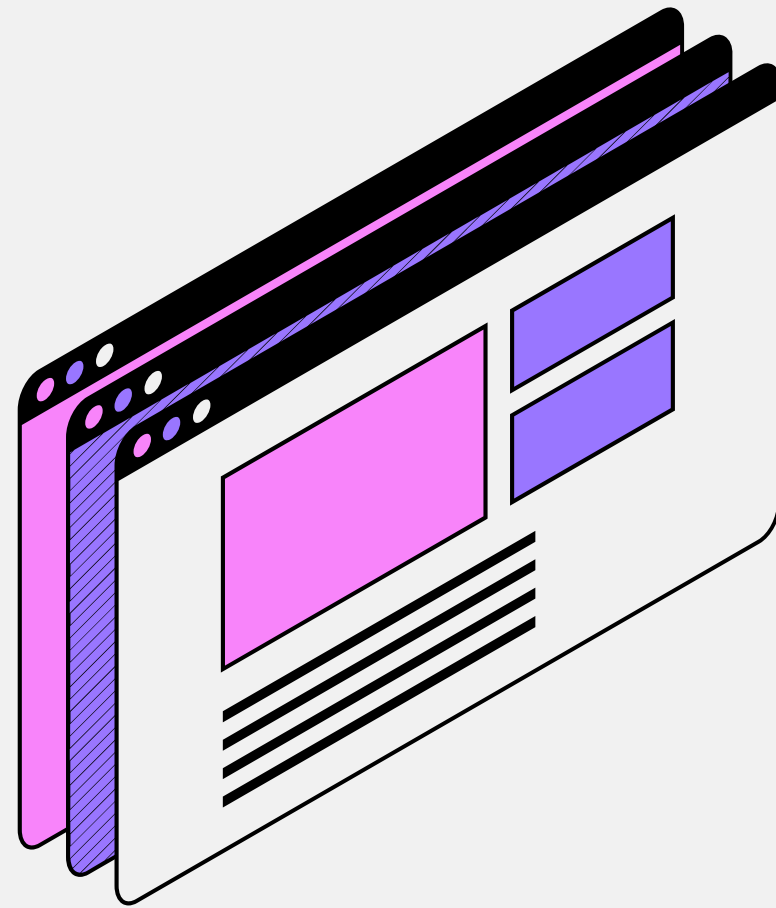also need an up and running Hadoop cluster (preferably Cloudera.

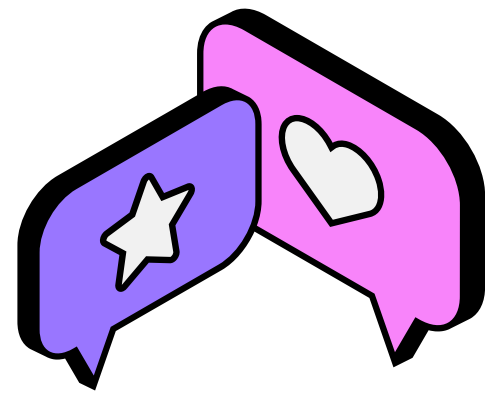Java 8 must be available with environment variables already set.

9

Demonstration

Thank You!