



# Azure Data Warehousing Solution for Peer-to-Peer Lending Data

## - A Case Study



## Introduction

Introduction: This case study focuses on a financial peer-to-peer company that successfully implemented an Azure data warehousing solution to address its data management and analytics challenges. With a growing volume of financial data and the need for scalable and efficient processing, the company turned to Azure for a comprehensive solution. By leveraging Azure Data Factory, Azure Data Lake Storage, and Azure Synapse Analytics, the company achieved improved data ingestion, transformation, and modeling.

## Problem Statement

The Client faced the challenge of efficiently storing, managing and analysing their massive and ever-growing dataset. The existing infrastructure struggled to cope with the scale and complexity of the data, impeding timely decision-making, and hindering accurate risk assessment. The client needed a robust and scalable solution to process and analyse their peer-to-peer lending data in near real-time.

## Methodology

- 1. Data Ingestion:** Azure Data Factory was utilized to efficiently ingest data from various sources, including loan applications, credit bureaus, financial institutions, and internal systems. Data pipelines were set up to extract, transform, and load the data into Azure Synapse Analytics.
- 2. Azure Synapse Analytics:** The client leveraged Azure Synapse Analytics as their centralized data warehousing solution. Synapse Analytics enabled seamless integration with other Azure services and provided scalable compute and storage resources to handle the large volume of peer-to-peer lending data.
- 3. Data Modelling and Storage:** The client's data was organized into a logical data model within Azure Synapse Analytics. The data was stored in dedicated SQL pools, ensuring efficient data retrieval and query performance.
- 4. Advanced Analytics:** Azure Synapse Analytics enables the client to perform advanced analytics and derive actionable insights from their lending data.
- 5. Data Visualization:** The data will be presented in an interactive dashboard that will allow users to explore the data and gain insights on Default customers, Bankruptcy customers, Average loan scores, Amount funded, and Amount received.
- 6. Scalability and Cost Optimization:** Azure Synapse Analytics provided clients with the flexibility to scale their computing and storage resources based on demand. This ensured cost optimization by paying only for the resources utilized during peak periods.



# Azure Data Warehousing Solution for Peer-to-Peer Lending Data

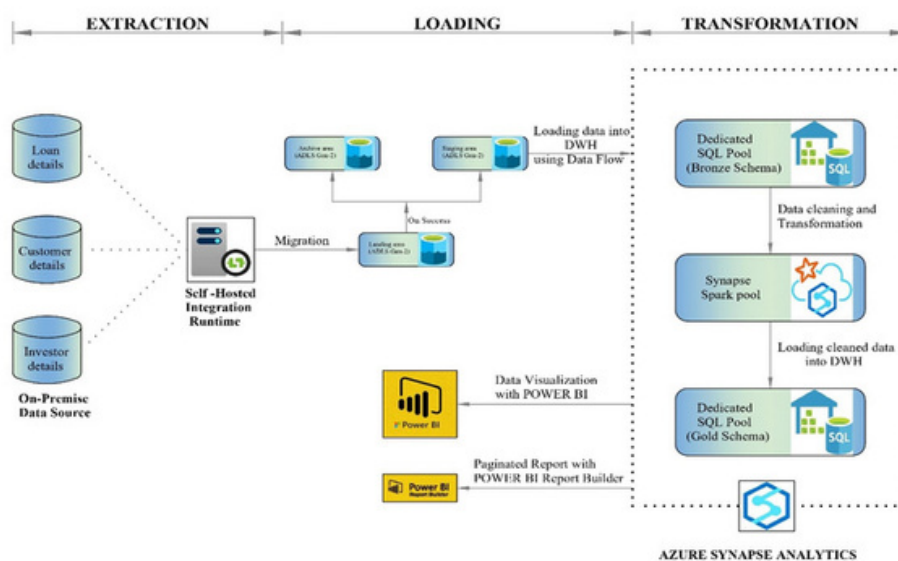
## - A Case Study



## Benefits & Results

- Enhanced Data Processing:** The client achieved faster data processing and analysis, enabling quicker loan application approvals, improved risk assessment, and streamlined business operations.
- Actionable Insights:** The advanced analytics capabilities of Azure Synapse Analytics allowed the client to extract valuable insights from their lending data, enabling data-driven decision-making and improved risk management.
- Scalability and Cost Efficiency:** Azure Synapse Analytics' scalability allowed the client to efficiently handle their growing dataset, while optimized resource allocation resulted in cost savings and improved operational efficiency.
- Future Readiness:** The Client's Azure data warehousing solution provides a solid foundation for future growth and innovation. They can easily incorporate additional data sources, expand analytics capabilities, and explore advanced technologies like AI and machine learning to further enhance their lending services.

## Project Architecture





# Azure Data Warehousing Solution for Peer-to-Peer Lending Data

## - A Case Study



## Our Approach

1. **Extract:** The raw data of Lending Club Loan details is extracted from the website and stored in an on-prem file system which is subsequently migrated to Azure cloud which is connected using self-hosted integration runtime. The extracted data is then staged into an ADLS-Gen2 container.
2. **Load:** Using Synapse analytics pipeline to get metadata and copy activity storing all the data into a dedicated SQL pool in Bronze Schema.
3. **Transform:** Using Spark Pool Cleaning and Transforming the data. Lastly loading clean data into a dedicated SQL pool in Gold Schema using Synapse Spark pool
4. **Visualize:** The data is then visualized using Power BI, which allows for the generation of a paginated report and dashboard for the purpose of gathering insights.