# 1   Introduction

This document provides a comprehensive architectural overview as well as the activities plan to migrate the application "Sales Facts Scanner" to the Microsoft Azure cloud.

## 1.1   Purpose

This document is meant to provide an architectural map and serve as a guidance, it is intended to depict the big architectural and technology decisions that have been made to successfully migrate the application to the cloud.

## 1.2   Scope

The scope of this document is to explain the new architecture in the cloud of the "Sales Facts Scanner" system, it also describes the modifications that must be made to the system in order to fit and take advantage of the cloud.

This document provides a general roadmap not only to create the proposed architecture in the cloud but also to migrate a system that nowadays works on-premise to the cloud.
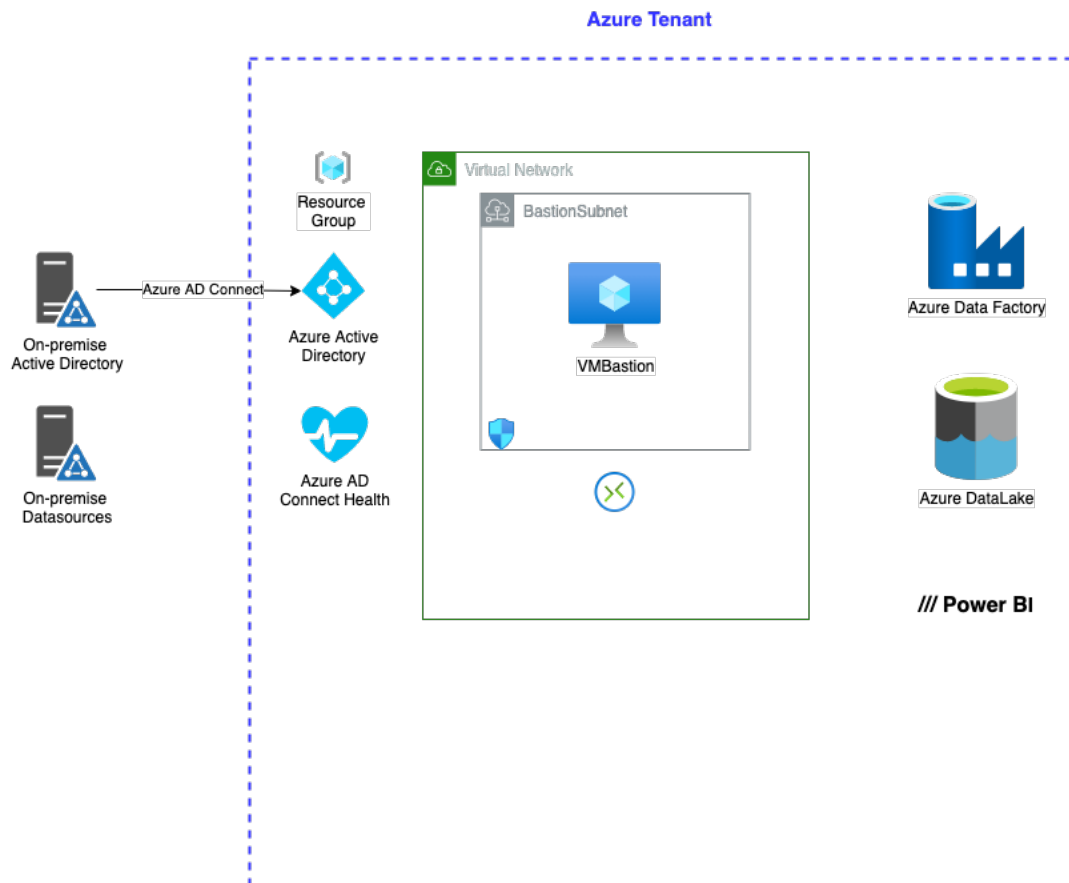
# 2. Architectural goals and constraints

There are some special requirements to consider:
- The application and the data cannot be exposed to the public internet. And some data is sensitive.
- A data scientist group needs to access to the raw and golden data

# 3. Technical Solution

## 3.1 Architectural Model



## 3.2 Activities

Sprint 1. Setting up Azure environment and data migrations

Perform the following activities in the Azure tenant:
a.  Create Resource group
b.  Create Azure Active Directory
c.  Create Azure AD Connect
d.  Connect Active Directory on-premise and sync organizational accounts
e.  Create Azure AD Connect Health and link it to Azure AD Connect
f.  Create Data Factory
g.  Use the Copy Data Tool to migrate the on-premise historical data stored in Oracle and SAP Hana on-premise
h.  Create the pipelines to ingest data every day (7GB approximately)

Sprint 2. Setting up Azure Data Solution

1. Create Data Lake and connect it to Data Factory
2. Code the programs for data transformation, the languages proposed for these data transformation jobs are: Python or R.
3. Connect Power BI to Data Lake for data Visualization

Sprint 3. Setting up access to golden and raw data

1. Configure Azure RBAC and scopes to security principals existing in AD, in this activity there must be crated a DataScientist role which must have access to Power BI Dashboards, Data Lake and Data Factory.
2. Set up a virtual network, a subnet and create a bastion VM for managerial purposes.