RANDOM FOREST

BI CLOUD

Technical Reference

SUMMARY

BI Cloud is a packaged solution by Random Forest AB to kickstart your data and analytical platform in Microsoft Azure. This document describes the general architecture and the parts included in the start package

Richard Lautmann

Random Forest AB

RANDOM FOREST

Introduction

Going from a traditional server structure to a PaaS based solution is a game changer for BI. This dramatically reduces the need for setting up and maintaining the platform and outsources the infrastructure. Usage of PaaS is paid by the hour and enables on demand scaling and disaster recovery through geo-redundancy. The service ensures you are always on the latest version and the licensing is included which minimizes the total cost. Moving to Azure opens up new possibilities which raises the need for a different design pattern. The cloud challenges old architectures and requires a new way of thinking. Storing data is no longer a cost driver, data streaming is now preferred over batch loading and the toolkit is vastly expanded.

Top Customer BI Challenges	Ideal Solution	Desired Outcomes
Data Discovery and Visualization Empower employees with data they need **Self Service BI** Cut time to market for reports **Data Quality** Ensure central data quality to enable Self Service BI **Future Ready?** Having an extendable platform	**Central control over data** A data warehouse responsible for data quality, consistency and system logics. Providing a data model to enable Self Service BI and system integrations. **A Data Visualization Tool** Enabling non-IT employees to analyze and build and share high quality reports build on a central secure model. **Build internal knowledge** Reducing the dependency on external experts to provide insights to the business	**Data Orchestration with Data Factory** Bring data from internal and external data sources into a central database in Azure. **A Data Warehouse in Azure Db/DW** To transform, clean, integrate and store the data **Analysis Services** Having a centrally controlled data model **Power BI** To enable Data discovery, self service BI and mobile access

BI Cloud

We have been creating and accumulating best practices for building data and analytical platforms in Microsoft Azure for many years and have packaged and automated many parts of the process in what we call BI Cloud. We are continuously adding reusable knowledge and automating processes in the package for our customers to utilize. The goal is to do less ELT and more of the data analytics on top of the platform.

In short this is what BI Cloud does for you:

- A first iteration of your own Azure BI data platform with standard components (PaaS)
- Delivers a small but working system from data source to a simple report
- Delivers a meta data driven framework for the ETL process
- Enforces and sets non-functional requirements such as naming conventions, logging, traceability and security
- Leaves you with your own proprietary cloud data platform built on standard components, prepared to extend and build upon

Apart from choosing and configuring standard components in Azure the package adds the following parts:

Random Forest AB
Birger Jarlsgatan 55
SE-111 45 Stockholm

info@randomforest.se
+46 70 281 01 68

- ARM templates and scripts for setting up services (prod, dev, test environments)
- Security out-of-the box using Azure Key Vault and best practices
- SQL Database framework for the ELT process including views, tables, functions and procedures
- A standard Data Factory Pipeline driven by metadata
- Forcing naming conventions
- Enables unified logging and maintenance reporting

Setup Procedure

The package BI Cloud is standardized but can easily be customized for the customer specific needs. Setting up the platform is done by the following steps:

1. Create Azure Subscription and resource group (CSP or existing subscription)
2. Deploying Azure PaaS services using ARM template and Power Shell
3. Installing general SQL Db Framework
4. Setting up general Data Factory V2 Framework
5. Installing on-premise gateway if needed

Initial Project – First Iteration

First Iteration Data Integration 1-2 weeks.

a. Initial Workshop with Customer to decide on initial scope, 2-5 tables in one source database, 1 transactional and some dimensional
b. Customer fills in the Data Integration Excel Template
c. Random Forest Implements first integration into Data Layer I SQL Db using the Framework in Azure Data Factory and SQL Database.
d. Random Forest build simple report with automated refresh to prove the system works.
e. Handover workshop is conducted with the customer and next steps are planned

Next Step

After the first iteration the customer can decide on the next steps. Normally there are much more sources needed, need for advanced analytics or better reports. A data and analytics platform requires continuous development and maintenance to be able to give the insights and competitive advantage. A typical second iteration is normally between 3-6 months.

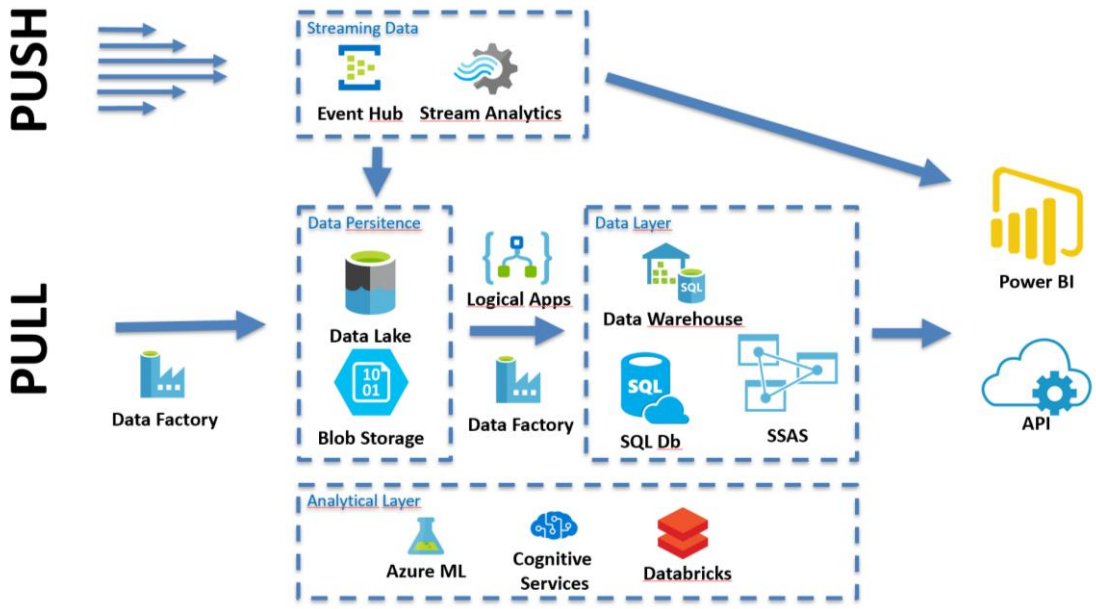
Random Forest AB
Birger Jarlsgatan 55
SE-111 45 Stockholm

info@randomforest.se
+46 70 281 01 68

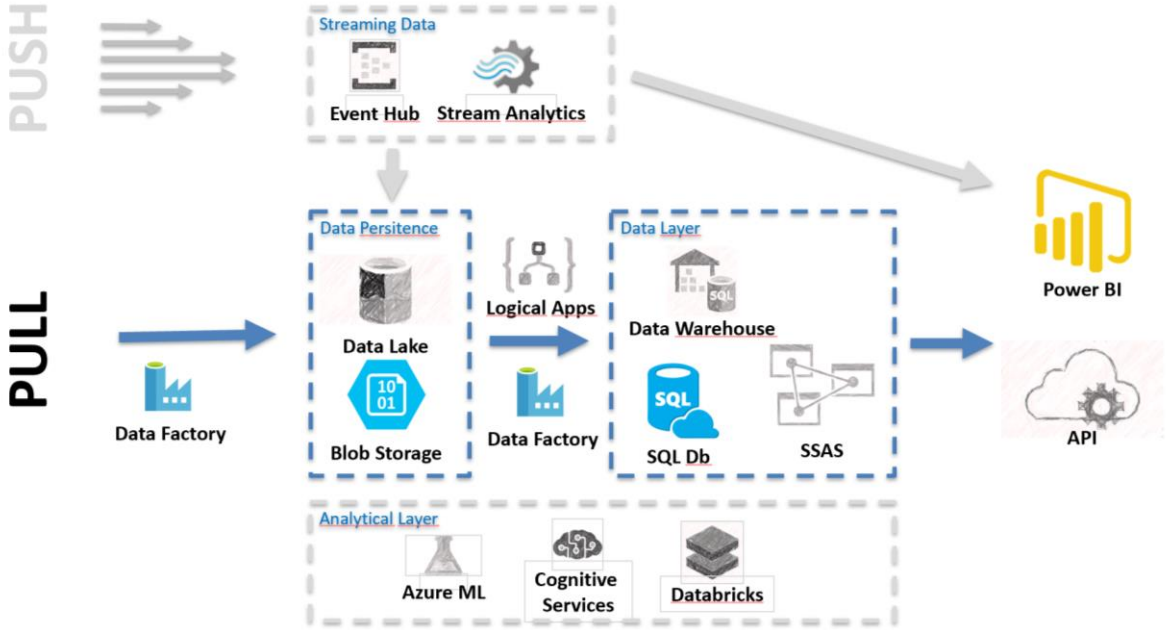
RANDOM FOREST

A data and analytics PaaS architecture

Target Architecture - Azure Services used



BI Cloud – First iteration Architecture



Random Forest AB
Birger Jarlsgatan 55
SE-111 45 Stockholm

info@randomforest.se
+46 70 281 01 68

Data Layer Architecture