

Case study:

Azure Data Processing Project



Disclaimer: This document is submitted with the explicit understanding that the contents are confidential and would not be divulged to any third party without prior written consent from Dotsquares Ltd.

- UK Office: Unit 2, Albourne Court, Henfield Road, Albourne, West Sussex, BN6 9FF, UK
- US Office (Maryland): 6701 Democracy Blvd. Suite 300, Bethesda, MD 20817
- Australia Office (Melbourne): 56 May Avenue, Altona Meadows VIC 3028, Australia
- India Office: First Floor, 6 Kha 9, Jawahar Nagar, Jaipur, Rajasthan, 302004, India

E-Mail: info@dotsquares.com; **URL:** www.dotsquares.com

Projects are supplied according to our T&C's: <http://www.dotsquares.com/terms>

What is Azure?

A cloud computing service by Microsoft, Azure is used for building, testing, deploying and managing applications and services via the Microsoft-based data centres. There are more than 600 services that Microsoft offers for Azure.

How does Dotsquares help with Azure?

From building, deploying to managing the applications with Azure Services, we at Dotsquares have done it all. With a team of efficient professionals who've got the knowledge of integrating the Microsoft Azure services in the client's project, we provide end-to-end Azure services.

Client Background

Darwin Project is an analytics platform used by the world's brightest and biggest agencies, professional services firms and brand teams to optimise their investments. It has an analytics engine to convert Marketing data from different social sources like Facebook, Instagram, and LinkedIn etc to a meaningful marketing analytical data that can be used by digital marketing agencies or Individual organisation to plan, strategies and bring most of their marketing budget.

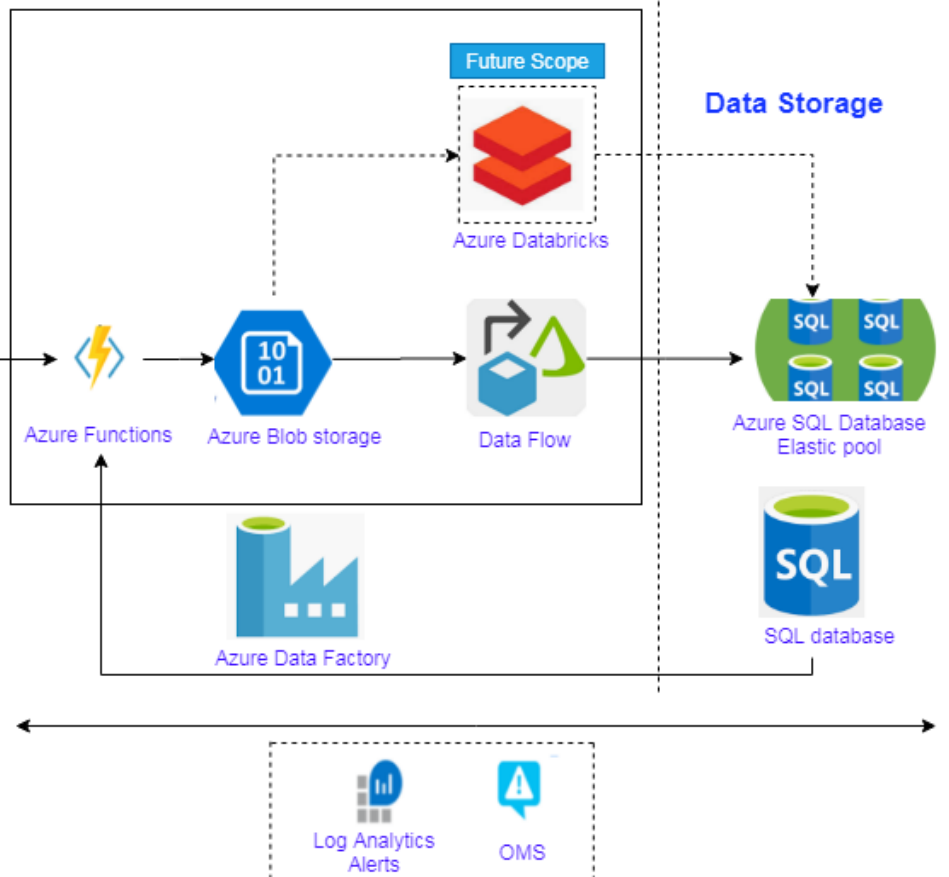
Purpose

Darwin is an ongoing Enterprise project, with new and ongoing requirements. Some of the requirements completed so far include getting data from various data sources where data sources are Marketing Engines e.g. Facebook Ads, LinkedIn Ads, Google Analytics, Google Search Ads 360, Pinterest, Salesforce Marketing Cloud, etc. , merge it and process data according to client needs. The system has multiple web connectors which ingest data to the system and process it as and when the data comes and store it in databases which is used for creating dashboards and other high level analysis tools. Dotsquares has implemented various Azure services including Azure DevOps as per client's needs to Manage, develop, test, monitor and deploy

Architecture

Platform Data Processing Architecture

Data Sources



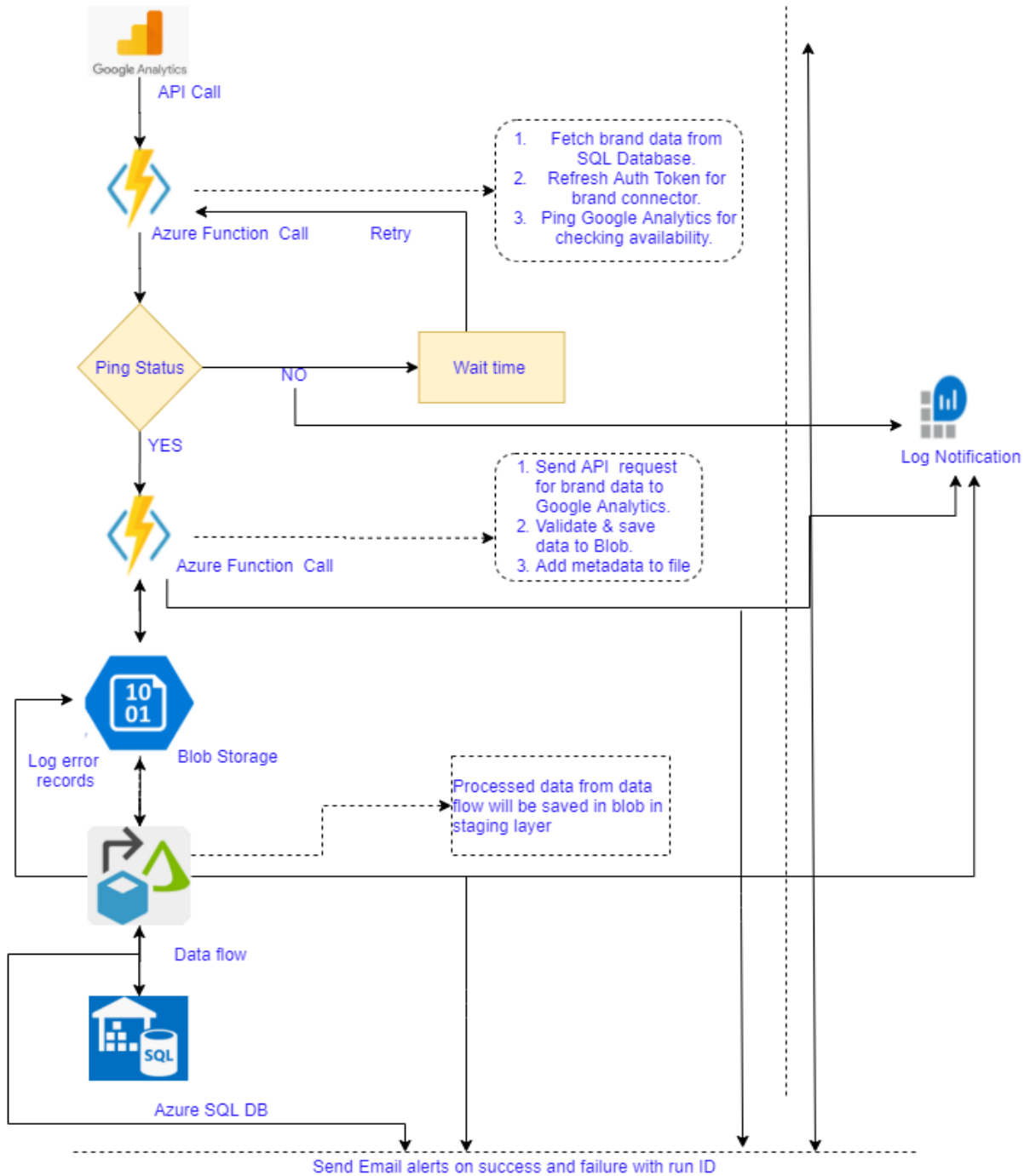
1 What Dotsquares have done so far?

EPIC 1	New Data processing framework
EPIC 2	LinkedIn connector

EPIC 3	LinkedIn Ads connector
EPIC 4	YouTube connector
EPIC 5	YouTube Ads connector
EPIC 6	Email & Salesforce connector
EPIC 7	Other connectors like Pinterest
EPIC 8	Connector Revamp
EPIC 9	Power BI Embedded
EPIC 10	Power BI Embedded- Advanced
Ongoing	Production Support
New Features	Data reload Automation, Data Gaps Automation, New connectors & More

2 Data Flow Diagram

Data Flow Diagram



3 Epic 1- New Data Processing framework

The first requirement for this project was to improve the data processing architecture. The previous architecture had some loops holes and it required immense production support due to lot of data load failures. These loops holes included lack of planning of data processing, lack of triggers and a planned Architecture. Dotsquares team helped by understanding the requirements by doing brain storming sessions with key stakeholders, further working on architecture to improve the data processing framework, finally changed the framework for data processing initially for 1 data source.

3.1 Outcomes

- Improved framework
- Less data processing time
- No Data failures
- Important alerts and reports
- Time zone based triggers
- New Synchronous and Asynchronous APIs
- Managed Incremental load & Data Pipelines
- Improved Data model
- Process improvement
- Error Logging
- Azure SQL Database
- Tackled Data Concurrency
- AAS refresh Activity
- Log Analytics

3.2 IT Team

- IT Project Manager
- Solution Architect
- ETL Developers
- ADF Developers
- Quality Assurance Engineer
- API Developers
- Tech Lead

- Database developers
- Data Quality analyst

3.3 Azure Services used

- ETL- Databricks
- Azure Data Factory
- Sync Functions
- Service Bus
- ADLS Gen2
- Azure Blob Storage
- Azure DevOps

4 Epic 2-7 New connectors

Dotsquares team helped in development of new connectors (Data source) like LinkedIn, LinkedIn Ads, YouTube, YouTube Ads, Email, and Salesforce while some of them are in pipeline like Pinterest. Team started with requirement gathering followed by data model design to map out any & all required metrics and calculations behind metrics. Further development of required connector as per the discussed metrics and model.

4.1 Outcomes

- New connectors in new data processing framework
- Desired metrics, reports and analytics

4.2 Azure Services used

- ETL- Databricks
- Azure Data Factory
- Sync Functions
- Service Bus
- ADLS Gen2

- Azure Blob Storage
- Azure DevOps

5 Epic 8 Connectors Revamp

This requirement focused to revamp all other connectors into new processing framework.

5.1 Outcomes-

Same as Epic1

5.2 Azure Services used

- ETL- Databricks
- Azure Data Factory
- Sync Functions
- Service Bus
- ADLS Gen2

6 Epic 9 & 10 Power BI Embedded

This requirement focused in giving public access to master database to Darwin clients. Further with access to master data in secured way, customers are able to store brand wise data & see detailed reports on Power BI embedded.

6.1 Outcomes

- Integration with Power BI
- Public DB access in secured way
- Ad hoc Power BI reports
- Advanced Role management & RBAC
- Import Dataset in Power BI workspace
- IP whitelisting

6.2 Azure Services used

- Azure APIs
- Power BI & Power BI Embedded
- Service Principle
- Azure SQL Data warehouse
- Key Vault

7 Production Support

Since this platform is multi-layered including User Interface, Azure Functions, API, ETL, Azure Data Factory and Database, there is a dedicated Production support team from Dotsquares managing ongoing production support needs. Some of the production tasks are data load failure, data gaps, incremental load failures, access token refresh, Data Availability tickets, Data validation and lot more

