Project Documentation: Game Analytics Data Pipeline

Table of Contents

[**Introduction** 1](#_Toc179751172)

[**Architecture Overview** 1](#_Toc179751173)

[**Resource Summary** 2](#_Toc179751174)

[**Data Generation** 4](#_Toc179751175)

[**Azure Event Hubs Configuration** 5](#_Toc179751176)

[**Azure Stream Analytics Job** 5](#_Toc179751177)

[**Azure Cosmos DB Setup** 7](#_Toc179751178)

[**Azure Synapse Analytics Configuration** 7](#_Toc179751179)

[1. Staging Table 8](#_Toc179751180)

[2. Dimension Tables 8](#_Toc179751181)

[3. Fact Table 9](#_Toc179751182)

[**Azure Synapse Pipelines** 10](#_Toc179751183)

[**Creating Views in Synapse** 11](#_Toc179751184)

[1. vw\_TopScoresPerGame 11](#_Toc179751185)

[2. vw\_PlayerPerformance 12](#_Toc179751186)

[3. vw\_GameStatistics 12](#_Toc179751187)

[Power BI Reporting 13](#_Toc179751188)

[Visualizations Created: 13](#_Toc179751189)

[**Conclusion** 15](#_Toc179751190)

# **Introduction**

This documentation outlines the implementation of a real-time data analytics pipeline for game score data using Azure services. The project involves generating gameplay data, ingesting it through Azure Event Hubs, processing it with Azure Stream Analytics, storing it in Azure Cosmos DB and Azure Synapse Analytics, and visualizing insights with Power BI.

# **Architecture Overview**

The data pipeline consists of the following components:

* **Data Generation Script**: Simulates gameplay data and sends it to Azure Event Hubs.
* **Azure Event Hubs**: Acts as the ingestion point for streaming data.
* **Azure Stream Analytics Job**: Processes incoming data and routes it to Azure Cosmos DB and Azure Synapse Analytics.
* **Azure Cosmos DB**: Stores processed data for operational use.
* **Azure Synapse Analytics**: Stores data in a dedicated SQL pool for analytical processing.
* **Azure Synapse Pipelines**: Automates data transformations and loading into dimension and fact tables.
* **Power BI**: Visualizes the data to provide actionable insights.

# **Resource Summary**

Below is a list of Azure resources used in the project:

**1. Azure Event Hubs**

* **Namespace Name**: DataSolutionsEventHub
* **Event Hub Name**: datasolutionshub

**2. Azure Stream Analytics Job**

* **Job Name**: DataSolutionsStreamAnalytics
* **Inputs and Outputs**:
  + **Input**: Event Hub datasolutionshub
  + **Outputs**:
    - Azure Cosmos DB: datasolutionscosmosdb
    - Azure Synapse Analytics: Database datasolutionsdedicated

**3. Azure Cosmos DB**

* **Account Name**: datasolutionscosmosdb
* **Database Name**: GameAnalyticsDB
* **Container Name**: Gameplays

**4. Azure Synapse Analytics**

* **Workspace Name**: datasol-synapse
* **Dedicated SQL Pool Name**: datasolutionsdedicated
* **Database Name**: datasolutionsdedicated

**5. Azure Synapse Pipelines**

* **Pipeline Name**: DataProcessingPipeline
* **Activities**:
  + InsertDimPlayer
  + InsertDimGame
  + InsertFactScores
  + ClearStaging

**6. Azure Synapse Tables (Schema: GameAnalytics)**

* **Staging Table**:
  + StagingPlayerScores
* **Dimension Tables**:
  + DimPlayer
  + DimGame
* **Fact Table**:
  + FactPlayerScores

**7. Azure Synapse Views (Schema: GameAnalytics)**

* **Views Created**:
  + vw\_TopScoresPerGame
  + vw\_PlayerPerformance
  + vw\_GameStatistics

**8. Power BI Report**

* **Report Name**: GameAnalyticsReport

**9. Data Generation Script**

* **Script Name**: score-generator
* **Description**: PowerShell script that generates gameplay data and sends it to Azure Event Hubs.

# **Data Generation**

The score-generator PowerShell script simulates gameplay data, including player names, game names, scores, and timestamps. It sends this data to the Azure Event Hub datasolutionshub.

*A screenshot of a computer program

Description automatically generated*

# **Azure Event Hubs Configuration**

**Namespace**: DataSolutionsEventHub

**Event Hub**: datasolutionshub

*A screenshot of a computer

Description automatically generated*

# **Azure Stream Analytics Job**

**Job Name**: DataSolutionsStreamAnalytics

* **Inputs**:
  + Configured to read from the Event Hub datasolutionshub.
* **Outputs**:
  + **Azure Cosmos DB**: datasolutionscosmosdb
  + **Azure Synapse Analytics**: Database datasolutionsdedicated
* **Query**:
  + Transforms incoming JSON data into outputs.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

*A screenshot of a computer

Description automatically generated*

# **Azure Cosmos DB Setup**

**Account Name**: datasolutionscosmosdb

**Database**: GameAnalyticsDB

**Container**: Gameplays

* Data from Stream Analytics is written to this container for operational use.

*A screenshot of a computer

Description automatically generated*

# **Azure Synapse Analytics Configuration**

**Workspace Name**: datasol-synapse

**Dedicated SQL Pool**: datasolutionsdedicated

* Created a dedicated SQL pool for storing and querying large volumes of data.

**Database**: datasolutionsdedicated

* Contains the schema GameAnalytics containing staging, dimension, and fact tables.

*A screenshot of a computer

Description automatically generated*

## 1. Staging Table

* **Table Name**: GameAnalytics.StagingPlayerScores
* **Purpose**: Temporarily holds raw data ingested from Azure Stream Analytics before processing.
* **Description**: This table stores the gameplay data as it comes from the Stream Analytics job, serving as a staging area for data transformation.

A screenshot of a computer

Description automatically generated

## 2. Dimension Tables

* **Table Name**: GameAnalytics.DimPlayer
  + **Purpose**: Stores unique player information.
  + **Description**: Contains player IDs and names, serving as a reference for player-related data.

A screenshot of a computer

Description automatically generated

* **Table Name**: GameAnalytics.DimGame
  + **Purpose**: Stores unique game information.
  + **Description**: Contains game IDs and names, serving as a reference for game-related data.

A screenshot of a computer

Description automatically generated

## 3. Fact Table

* **Table Name**: GameAnalytics.FactPlayerScores
* **Purpose**: Stores the core transactional data, linking players, games, and scores.
* **Description**: This table contains foreign keys to the dimension tables (DimPlayer and DimGame), along with the score and timestamp of each gameplay.

A screenshot of a computer

Description automatically generated

# **Azure Synapse Pipelines**

**Pipeline Name**: DataProcessingPipeline

* Automates data movement and transformation from the staging table to the dimension and fact tables.

**Activities**:

1. **InsertDimPlayer**
   * Inserts unique players into the DimPlayer table.

A screenshot of a computer program

Description automatically generated

1. **InsertDimGame**
   * Inserts unique games into the DimGame table.

A screenshot of a computer game

Description automatically generated

1. **InsertFactScores**
   * Joins staging data with dimension tables and inserts records into the FactPlayerScores table.

A screen shot of a computer

Description automatically generated

1. **ClearStaging**
   * Clears the StagingPlayerScores table after processing.

A close up of a sign

Description automatically generated

*A screenshot of a computer

Description automatically generated*

# **Creating Views in Synapse**

Views were created to simplify data access and reporting.

## vw\_TopScoresPerGame

* + Provides a leaderboard of top scores per game with rankings.

A screenshot of a computer

Description automatically generated

## vw\_PlayerPerformance

* + Shows player scores over time for performance analysis.

A screenshot of a computer

Description automatically generated

## vw\_GameStatistics

* + Aggregates game statistics such as total gameplays and average scores.

A screenshot of a computer

Description automatically generated

# Power BI Reporting

**Report Name**: GameAnalyticsReport

* Connected to the Azure Synapse using Direct Query to create interactive reports and dashboards.

## Visualizations Created:

1. **Leaderboard Table**
   * Displays top players per game using data from vw\_TopScoresPerGame.

A screenshot of a computer game

Description automatically generated

1. **Player Performance Line Chart**
   * Shows player scores over time from vw\_PlayerPerformance.

A screenshot of a computer screen

Description automatically generated

1. **Game Statistics** 
   * Visualizes total gameplays and average scores per game from vw\_GameStatistics.

A screenshot of a graph

Description automatically generated

# **Conclusion**

This project demonstrates a complete data pipeline leveraging Azure services for real-time data ingestion, processing, storage, and visualization. The use of Azure Event Hubs, Stream Analytics, Cosmos DB, Synapse Analytics, and Power BI provides a scalable and efficient solution for game analytics.