Time Series Prediction with Kusto Query Language and Azure Data Explorer

Charunthon Limseelo

Microsoft Learn Student Ambassadors

Global AI Bootcamp 2024 on

March 1st 2024, at Microsoft Thailand (45-min workshop/demo) and Rerun (TBA)

About Me

Charunthon Limseelo

Field of interests

- Machine Learning and Data Analysis
- Time Series Data Prediction and Analysis
- Rookie Full-Stack Development with Cloud Architecture

Work experiences

- Beta Microsoft Learn Student Ambassadors/Google Developer Student Lead as a tech contributor
- Data Science and .NET community contributor at Seven Peaks Software (current)
- Collaborate with CreatorsGarten

You can contact me and look my GitHub Projects by this QR Code

global AI community

For full biography, please visit at boatchrnthn.vercel.app or my LinkedIn page

Slides Collaborators

Charunthon Limseelo

Microsoft Learn Student Ambassador, TH

Sirisa Kornnawawat

Google Developer Student, Poland

Asst. Prof. Ph.D. Peerapon Siripongwuttikorn

King Mongkut's University of Technology Thonburi

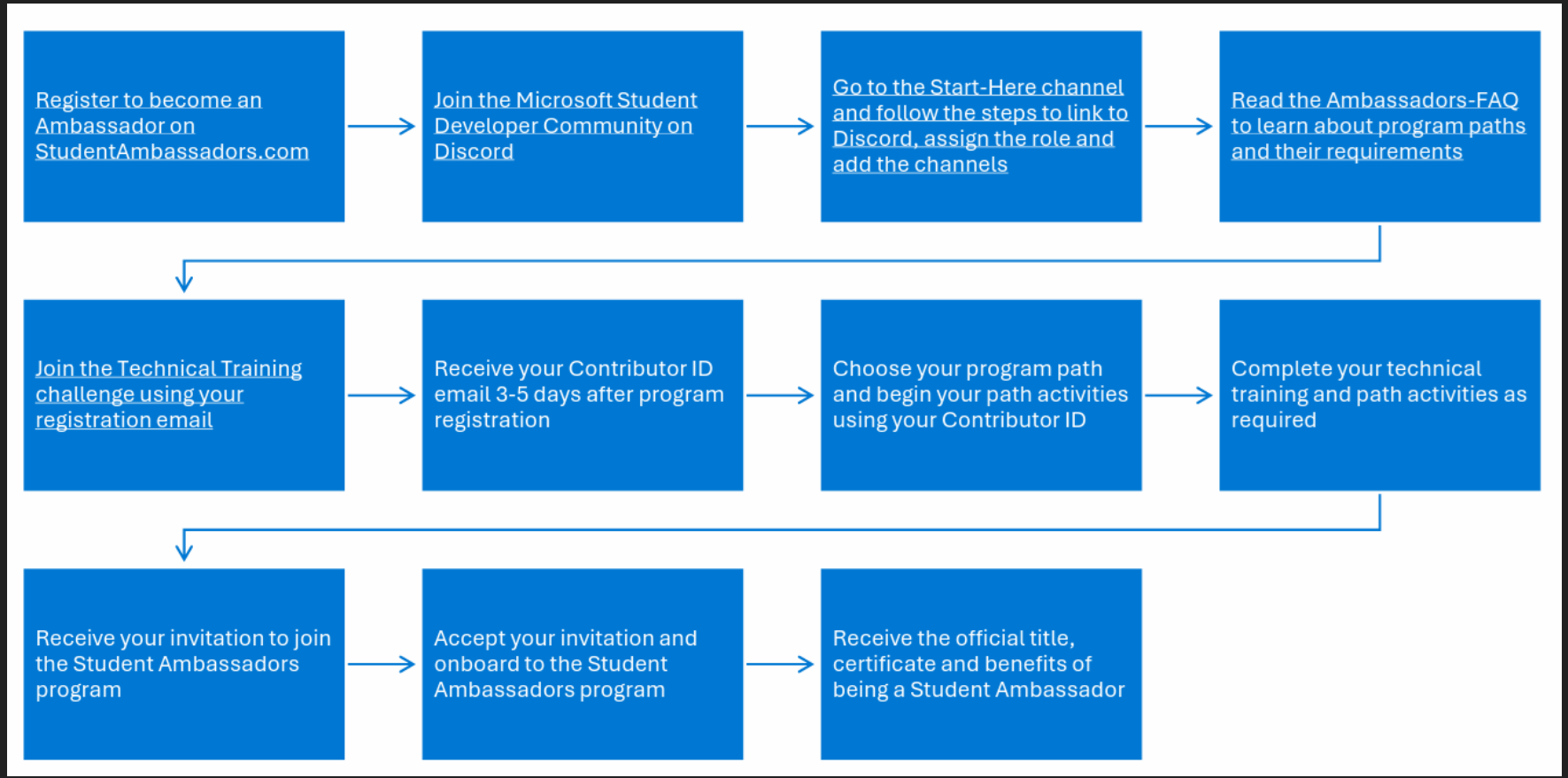
Ph.D. Sansiri Tarnpradab

King Mongkut's University of Technology Thonburi

global AI community

Register as Microsoft Learn Student Ambassadors today!!

At [MVP Communities (microsoft.com)]


Register to become an Ambassador on StudentAmbassadors.com
Join the Microsoft Student Developer Community on Discord
Go to the Start-Here channel and follow the steps to link to Discord, assign the role and add the channels
Read the Ambassadors-FAQ to learn about program paths and their requirements
Join the Technical Training challenge using your registration email
Receive your Contributor ID email 3-5 days after program registration
Choose your program path and begin your path activities using your Contributor ID
Complete your technical training and path activities as required
Receive your invitation to join the Student Ambassadors program
Accept your invitation and onboard to the Student Ambassadors program
Receive the official title, certificate and benefits of being a Student Ambassador

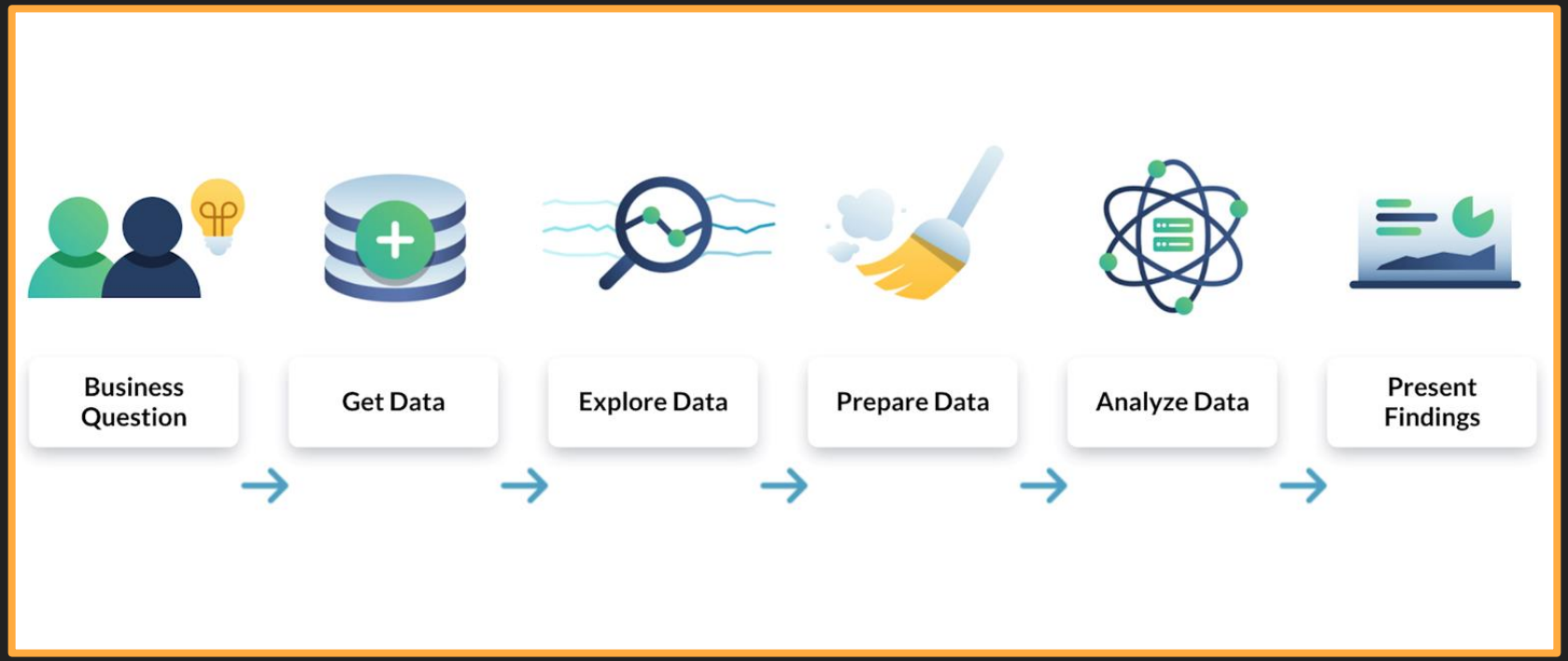
Session Outline

- Time Series Analysis and Forecasting
- Get into Azure Data Explorer
- Kusto Query Language and its structures
- Demo: How to train/forecast data in Azure Data Explorer
  - Time Series Decomposition (with SARIMA-based) model
- Conclusion

global AI community

Are some of you guys a Data Scientist, ML Engineer, Data Engineer, or Analyst?

global AI community



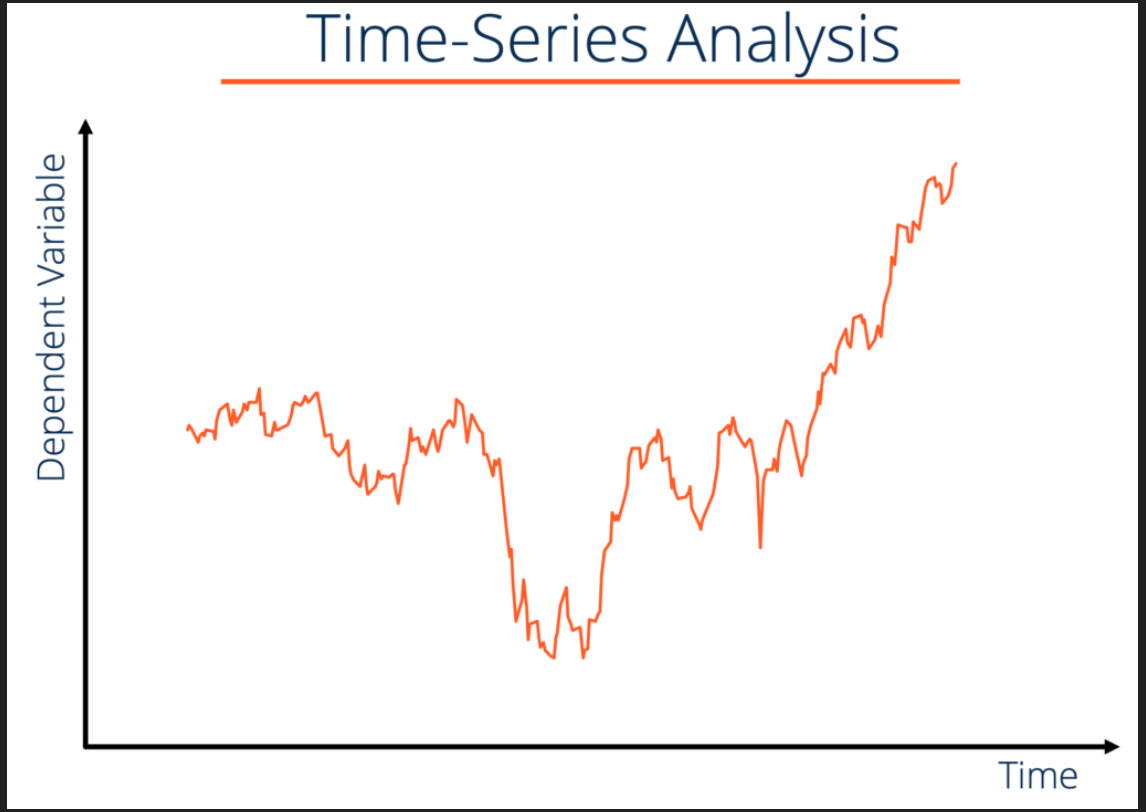
The data analysis process in a nutshell

Source: Data Camp

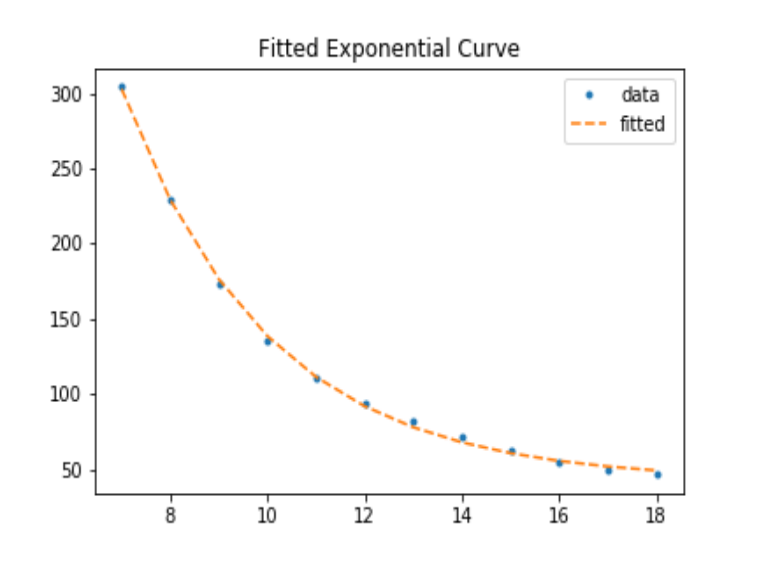
What is Time Series?

Introduction

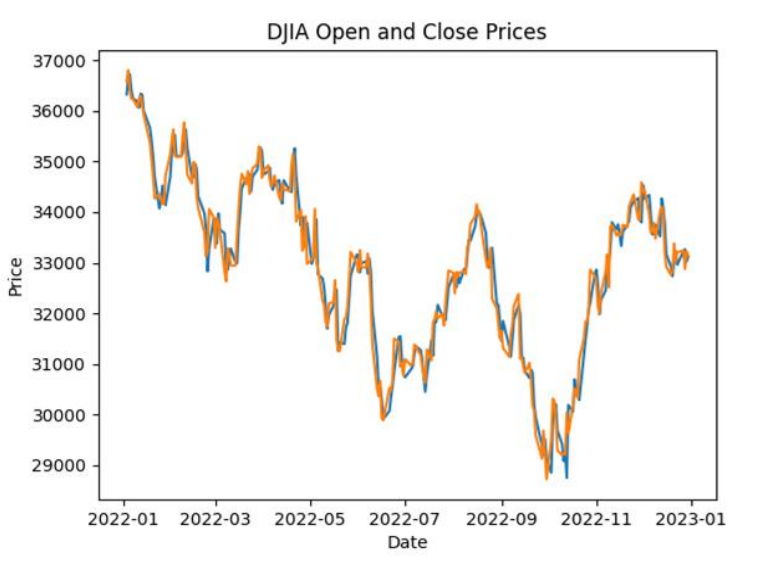
- Studying a sequence of data points collected at **consistent intervals** over a specific period. Unlike sporadic or random data collection, time series data provides insights into how variables change over time.
- Example Cases:
  - Stocks
  - Temperature Anomaly
  - Daily COVID-19 Patients
  - Etc.



Comparison Between No Noise and Noise Graph



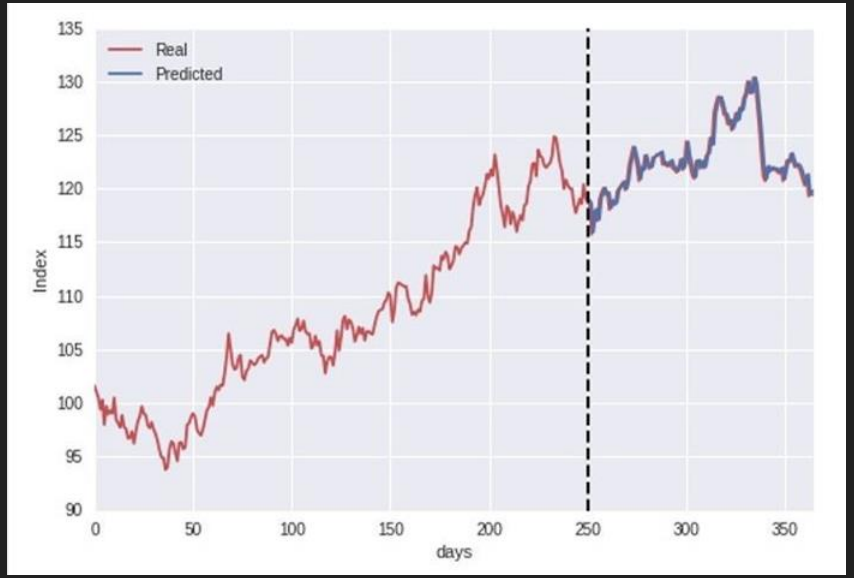
No Noise Graph



Noise/Time Series Graph

Time Series Forecasting

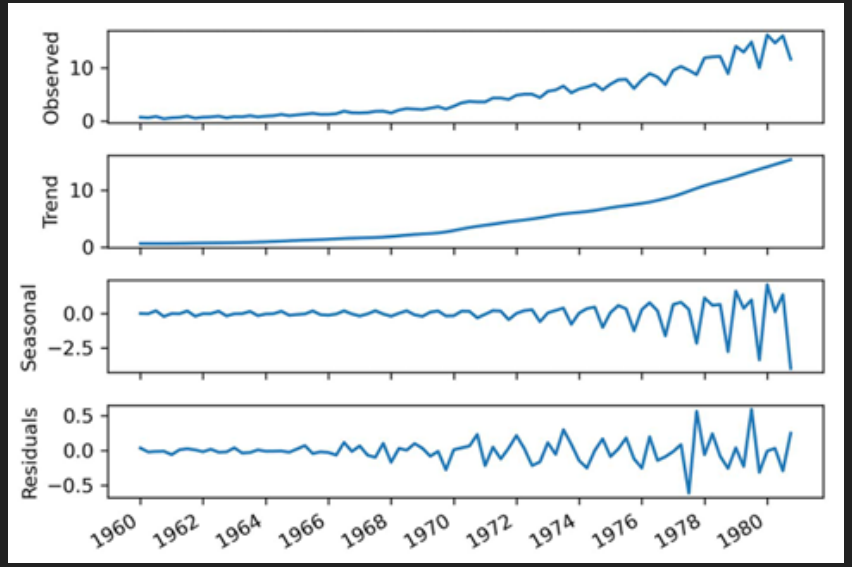
- Sounds like; predicting unknown values
- Involves the collection of historical data, preparing it for algorithms to consume, and then predicting the future values, based on patterns learn from the historical data. **(Calculate only two variables)**
- Some Example Cases
  - Predicting future sales at an SKU (stock keeping unit) level for planning and budgeting
  - Predicting the cumulative COVID vaccinations
  - Forecasting sales by store, so it can schedule the right resources.



global AI community

Components of Time Series

1. Trend Component
    - Represents the slow-moving changes in a time series. It is responsible for making the series gradually increase or decrease over time.
2. Seasonality Component
    - Represents the seasonal pattern in the series. The cycles occur repeatedly over a fixed period.
3. Residuals/Noise
    - Represent the behavior that not able to be explained by the trend and seasonality components. They correspond to random errors, also termed white noise.



Source: Time Series Forecasting in Python – Marco Peixeiro

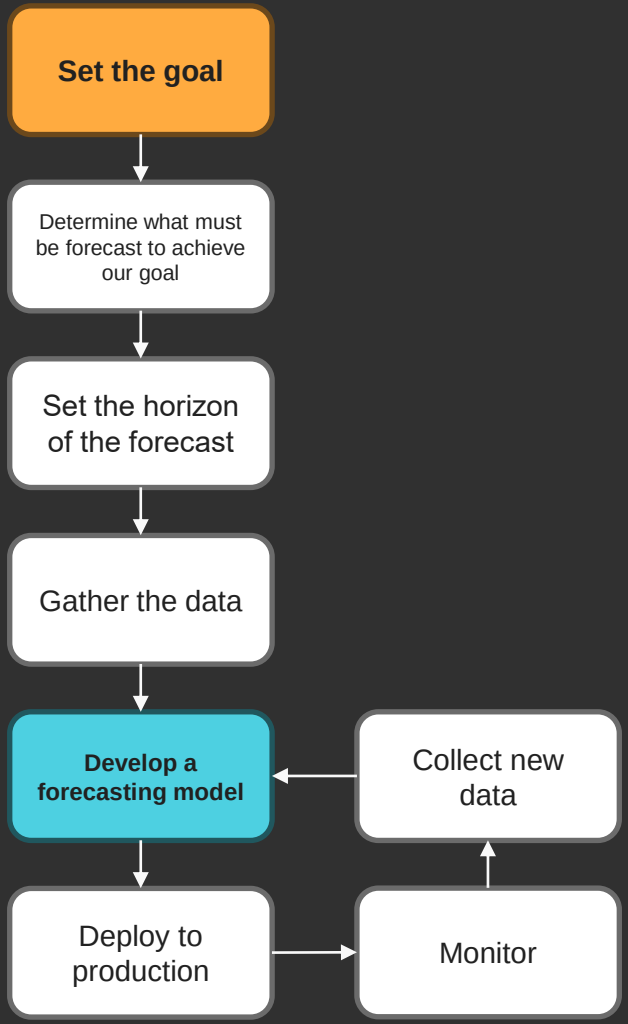
Roadmap

of Making Time Series Analysis
and Forecasting

global AI community

Roadmap

of Making Time Series Analysis and Forecasting



global AI community

Statistical Models For Forecasting

- SARIMA (Seasonal Auto Regressive Integrated Moving Average)***
  - Observe seasonality, and using seasonal effects to produce forecasts
- SARIMAX (adding exogenous variables)
  - Not only calculating with the values and time, but also calculating with another variable.
- ARIMA (Auto-Regressive Integrated Moving Average)
  - Using for stationary or simple time series dataset.

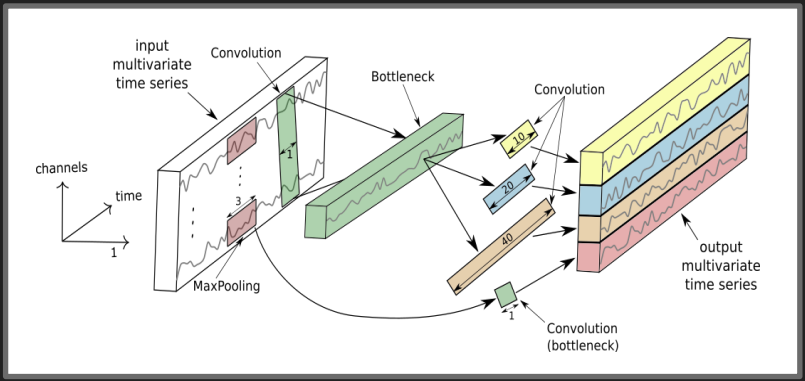
***Most used

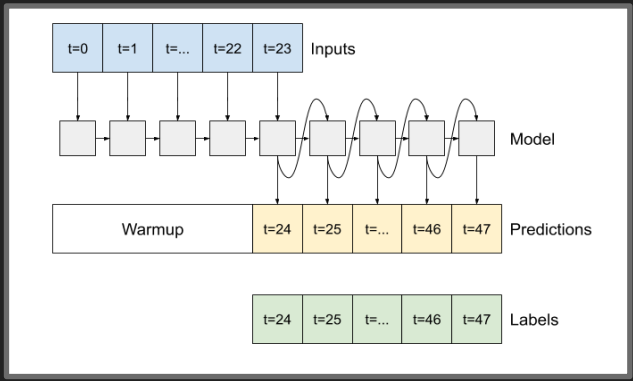
*Never change the order of a time series when modeling. Shuffling the data is not allowed.
*All models are using the concept of time series decomposition

global AI community

Machine Learning Models For Forecasting

- CNN (Convolutional Neural Network)
- LSTM (Long Short-term Memory)





*Never change the order of a time series when modeling. Shuffling the data is not allowed.

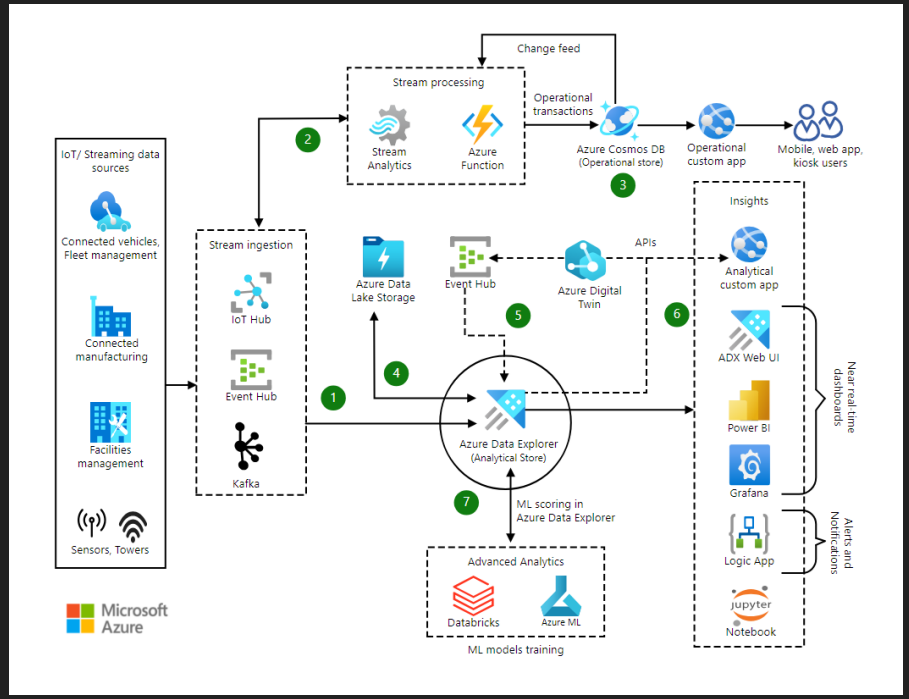
global AI community

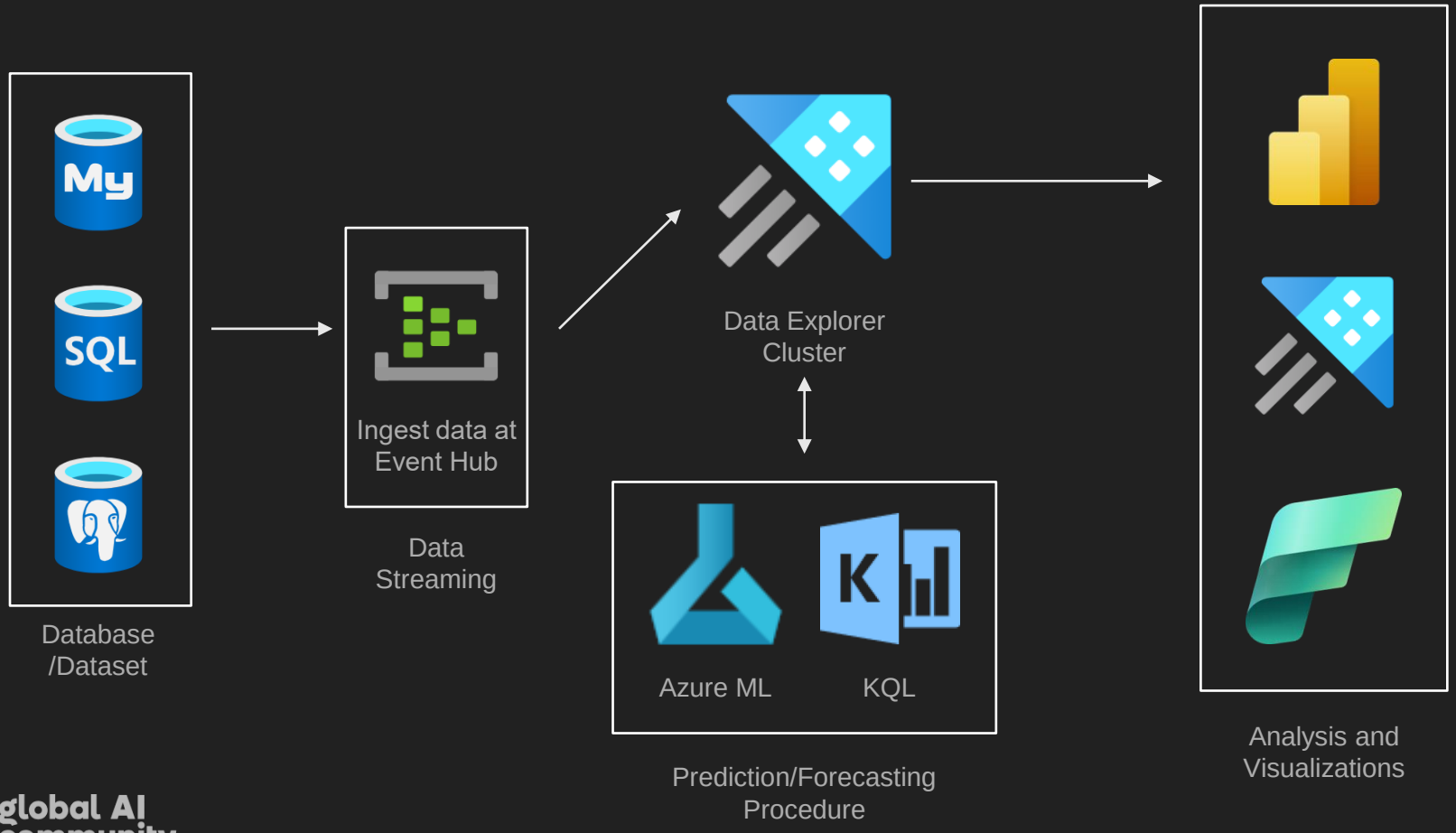
Exploring Azure Data Explorer

global AI community

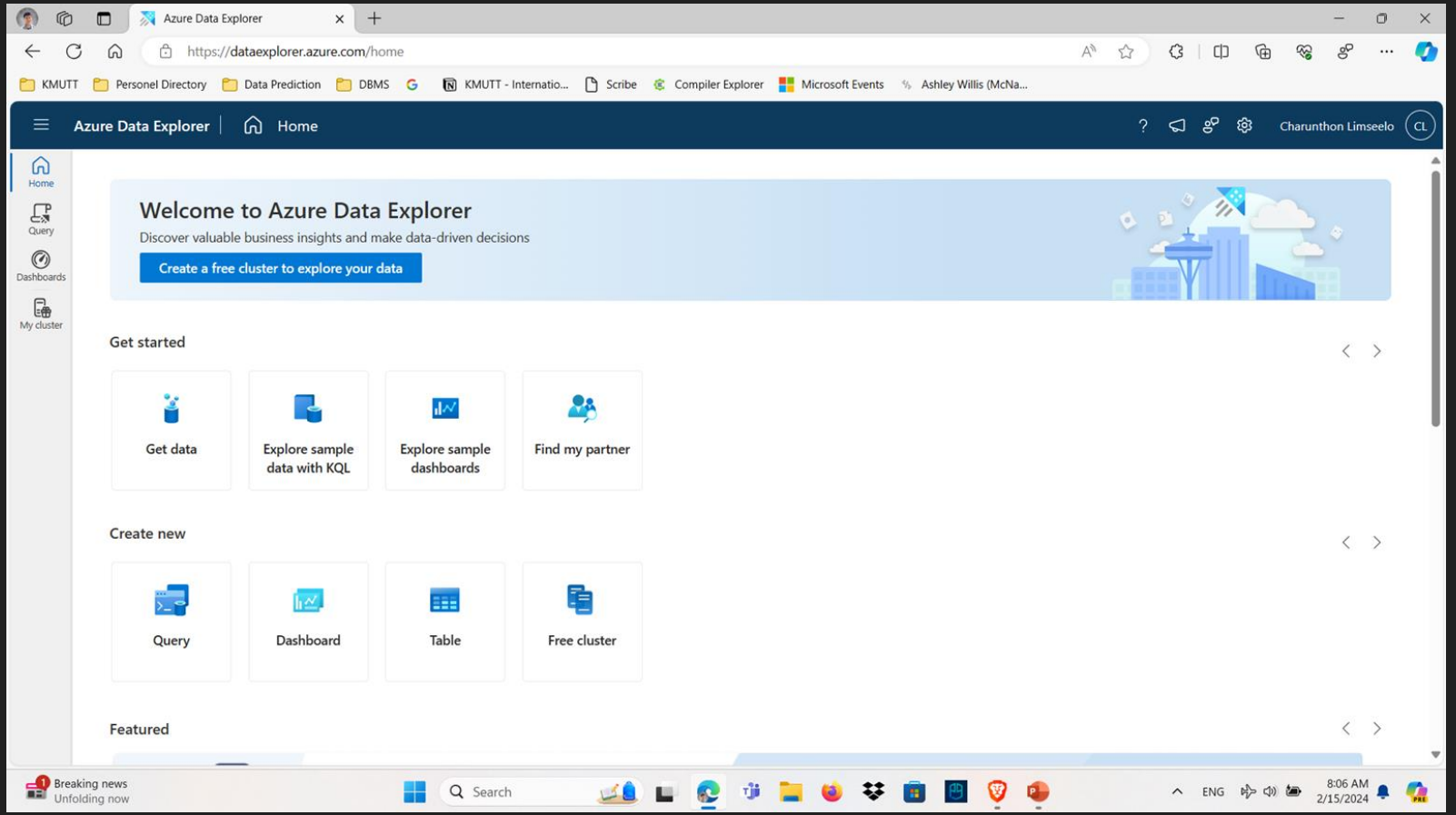
Introduction to Azure Data Explorer

- Fully managed, high-performance, big data analytics platform that makes it easy to analyze high volumes of data in near real time.
- An end-to-end solution for data ingestion, query, visualization, and management.

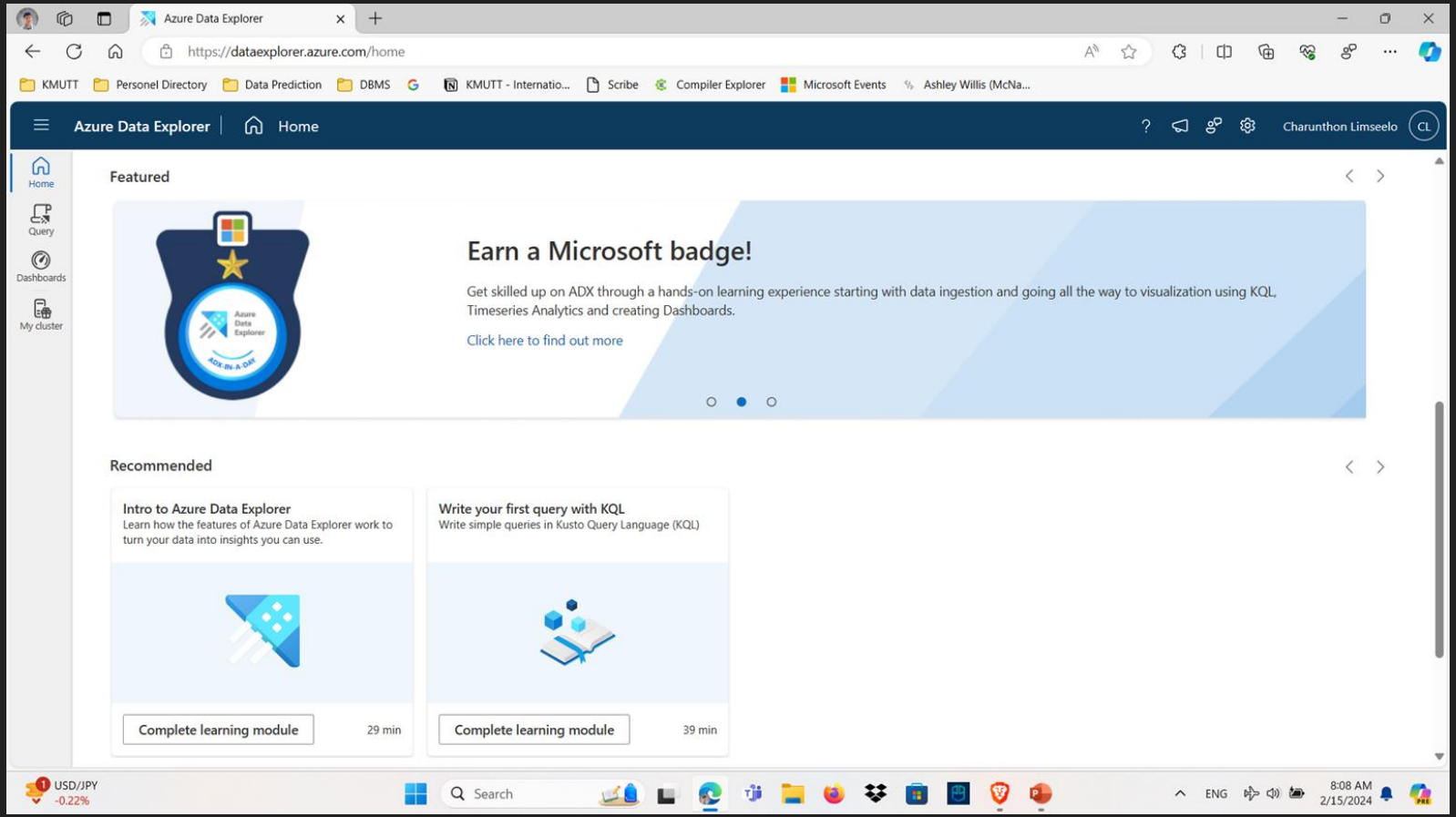




Let’s investigate the interfaces

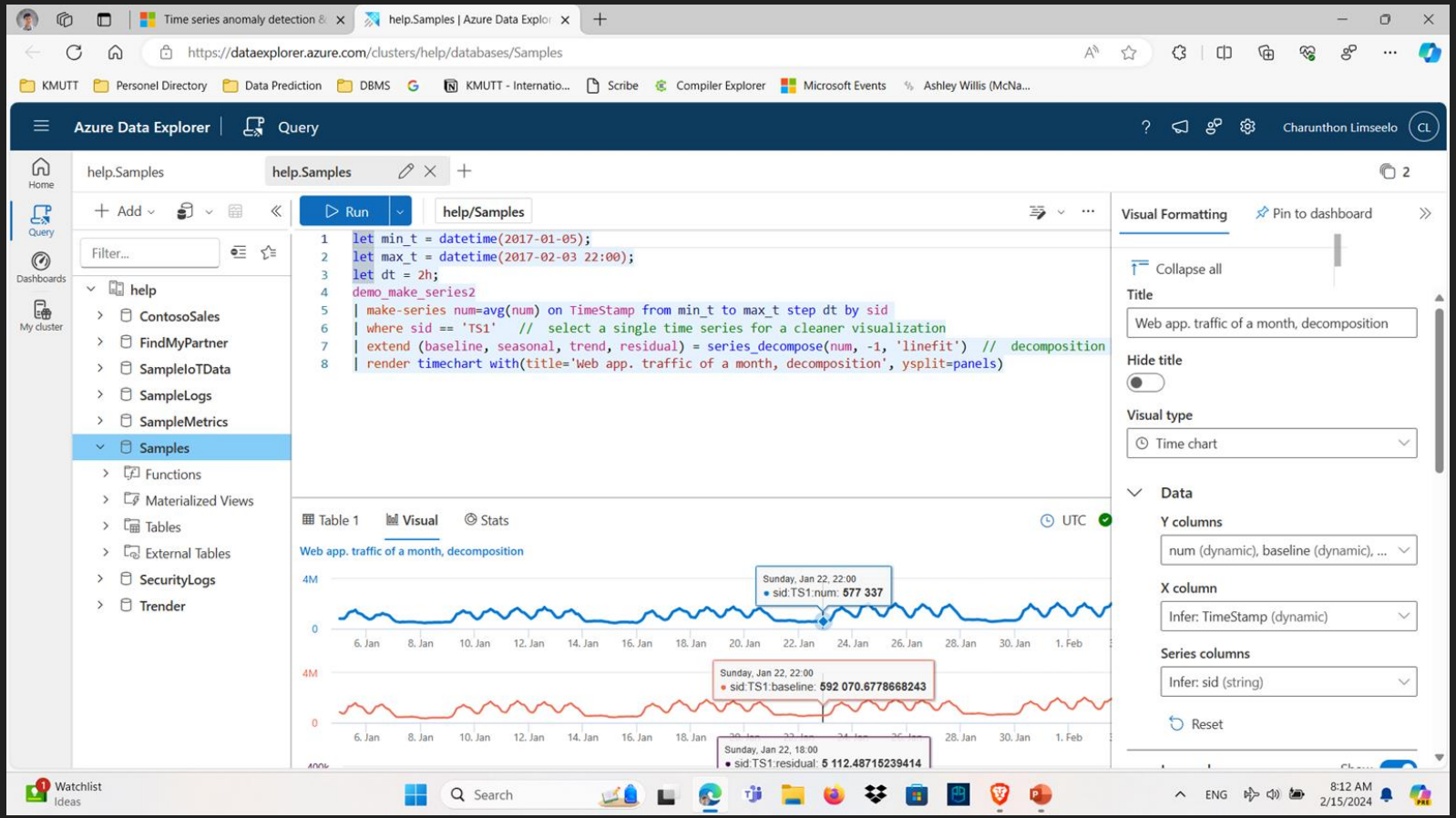


Homepage of Azure Data Explorer



global AI community

Homepage of Azure Data Explorer

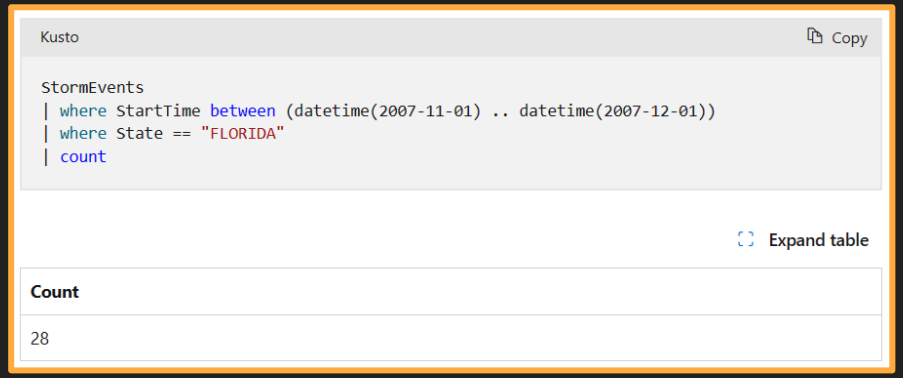

Time series anomaly detection &
help.Samples | Azure Data Explor
https://dataexplorer.azure.com/clusters/help/databases/Samples
KMUTT
Personel Directory
Data Prediction
DBMS
KMUTT - Internatio...
Scribe
Compiler Explorer
Microsoft Events
Ashley Willis (McNa...
Azure Data Explorer
Query
Charunthon Limseelo
CL
Home
Query
Dashboards
My cluster
help.Samples
help.Samples
2
Add
Run
help/Samples
Filter...
help
ContosoSales
FindMyPartner
SampleIoTData
SampleLogs
SampleMetrics
Samples
Functions
Materialized Views
Tables
External Tables
SecurityLogs
Trender
1 let min_t = datetime(2017-01-05);
2 let max_t = datetime(2017-02-03 22:00);
3 let dt = 2h;
4 demo_make_series2
5 | make-series num=avg(num) on TimeStamp from min_t to max_t step dt by sid
6 | where sid == 'TS1' // select a single time series for a cleaner visualization
7 | extend (baseline, seasonal, trend, residual) = series_decompose(num, -1, 'linefit') // decomposition
8 | render timechart with(title='Web app. traffic of a month, decomposition', ysplit=panels)
Table 1
Visual
Stats
UTC
Web app. traffic of a month, decomposition
4M
0
Sunday, Jan 22, 22:00
sid:TS1:num: 577 337
6. Jan
8. Jan
10. Jan
12. Jan
14. Jan
16. Jan
18. Jan
20. Jan
22. Jan
24. Jan
26. Jan
28. Jan
30. Jan
1. Feb
4M
0
Sunday, Jan 22, 22:00
sid:TS1:baseline: 592 070.6778668243
Sunday, Jan 22, 18:00
sid:TS1:residual: 5 112.48715239414
Visual Formatting
Pin to dashboard
Collapse all
Title
Web app. traffic of a month, decomposition
Hide title
Visual type
Time chart
Data
Y columns
num (dynamic), baseline (dynamic), ...
X column
Infer: TimeStamp (dynamic)
Series columns
Infer: sid (string)
Reset
Watchlist
Ideas
Search
ENG
8:12 AM
2/15/2024

Dashboard Page of Azure Data Explorer

Kusto Query Language and its structures

Introduction to Kusto Query Language

- Language for querying and analyzing large-scale data
- Supports various data sources, such as Azure Data Explorer, Azure Monitor, Microsoft Defender, and Microsoft Intra.
- Consists of a set of operators that can be composed together to create complex queries
- Able to perform tasks such as filtering, aggregation, transformation, and visualization of data
- Supports user-defined functions, variables, and parameters to make queries more reusable and dynamic



Kusto Query Language (KQL) example snippet

SQL vs KQL (with Table Comparison)

Category	SQL Query	Kusto Query	Learn more
Select data from table	`SELECT * FROM dependencies`	`dependencies`	Tabular expression statements
--	`SELECT name, resultCode FROM dependencies`	`dependencies | project name, resultCode`	project
--	`SELECT TOP 100 * FROM dependencies`	`dependencies | take 100`	take
Null evaluation	`SELECT * FROM dependencies` `WHERE resultCode IS NOT NULL`	`dependencies` `| where isnotnull(resultCode)`	isnotnull()
Comparison operators (date)	`SELECT * FROM dependencies` `WHERE timestamp > getdate()-1`	`dependencies` `| where timestamp > ago(1d)`	ago()
--	`SELECT * FROM dependencies` `WHERE timestamp BETWEEN ... AND ...`	`dependencies` `| where timestamp between (datetime(2016-10-01) .. datetime(2016-11-01))`	
Comparison operators (string)	`SELECT * FROM dependencies` `WHERE type = "Azure blob"`	`dependencies` `| where type == "Azure blob"`	
--	`-- substring` `SELECT * FROM dependencies`	`// substring` `dependencies`	

Kusto Query Language (KQL) with SQL conversion cheat sheet
From https://learn.microsoft.com/en-us/azure/data-explorer/kusto/query/sql-cheat-sheet

global AI community

Demo: How to train/forecast data in Azure Data Explorer

Link : https://bit.ly/tsadx-globalai24

- Train Your Time Series Data with Azure Data Explorer Part 1
- Train Your Time Series Data with Azure Data Explorer Part 2

global AI community

Summary / Key Takeaways

Key Takeaways

- **Kusto Query Language (KQL)** contains native support for creation, manipulation, and analysis of multiple time series. With KQL, you can create and analyze thousands of time series in seconds, enabling near real-time monitoring solutions and workflows.
- The applicable time series functions are based on a robust well-known **decomposition model**, where each original time series is decomposed into seasonal, trend, and residual components.
- **Anomalies** are detected by outliers on the residual component, while **forecasting** is done by extrapolating the seasonal and trend components.
- The KQL implementation significantly enhances the basic decomposition model by automatic seasonality detection, robust outlier analysis, and vectorized implementation to process thousands of time series in seconds.

Thanks! (Q&A)

*Feel free to take away the slide kubb 💕