

Python en la nube: versatilidad y eficiencia

Lidia Sánchez Mérida

Índice

- Conceptos básicos
- Desarrollo en la nube
- Uso eficiente de servicios *cloud*
- Ejemplos prácticos
- Conclusiones
- Bibliografía

Conceptos básicos



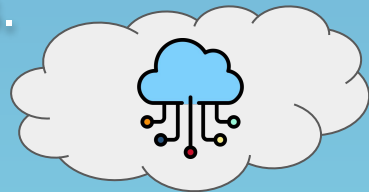
Cloud Computing: conjunto de servicios, principalmente de computación y almacenamiento, accesibles a través de Internet.

Azure: plataforma *cloud* de Microsoft.

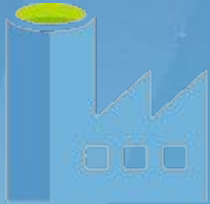
ETL: *Extract-Transform-Load*.

ELT: *Extract-Load-Transform*.

Dataflow: actividad propia del servicio Azure Data Factory capaz de efectuar procesos ETL sobre grandes volúmenes de datos.



Desarrollo en la nube



Azure Data Factory



**Azure Synapse
Analytics**



Azure Functions

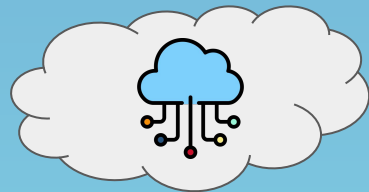


databricks

Uso eficiente de servicios *cloud*



- Azure Data Factory
 - Sin conocimientos en programación
 - Múltiples operaciones de procesamiento
- Azure Synapse Analytics
 - Análisis más complejos
 - Integración con otros servicios: ML
- Azure Functions
 - Procesos personalizados
 - Uso de APIs
- Azure Databricks
 - Big Data
 - Multiproveedor



Ejemplo práctico:



Objetivo: crear un proceso de ingesta, procesamiento y carga de datos procedentes de ficheros situados en servidores SFTP.



Azure Data Factory

- 1 link service por cada servidor SFTP
- 1 pipeline + 1 dataflow
- 1 link service a la base de datos destino
- Tiempo de ejecución: 10 min



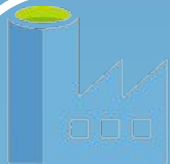
Azure Functions

- Azure SDK Python
- Pandas
- Tiempo de ejecución: 45 segs

Ejemplo práctico:



Objetivo: obtener y almacenar un conjunto de datos financieros de la API de Bloomberg, además de calcular métricas personalizadas.



Azure Data Factory

- *Web Activity* para conectar con la API de Bloomberg
- Problemas de tipos de datos al iterar sobre la respuesta de la API
- No se pudo ejecutar correctamente



Azure Functions

- Azure SDK Python
- Requests
- Pandas
- ELT

Ejemplo práctico:



Objetivo: exponer información situada en bases de datos SQL en la nube controlando el rendimiento y la sensibilidad de los datos.



Microsoft®
SQL Azure™

- Vistas SQL
- No control sobre filtros, ordenación ni rendimiento
- Modificaciones continuas y aumento de la complejidad



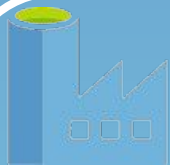
Azure Functions

- Azure SDK Python
- Jerarquía de entidades = Consultas predefinidas
- Filtros y ordenación limitados maximizando el rendimiento

Ejemplo práctico:



Objetivo: crear un recurso generalizado para la transferencia de ficheros entre distintas fuentes y destinos.



Azure Data Factory

- *Copy Activity* parametrizado pero presente en cada pipeline.
- No *switch case* = Tiempo de ejecución en múltiples condiciones



Azure Functions

- Azure SDK Python
- Paramiko
- Fuentes y destinos fuera del código

Conclusiones

Los recursos más sencillos pueden no cumplir con los requisitos mínimos.

Soluciones predefinidas pueden ser demasiado complejas para el problema a resolver.

- Costes adicionales.
- Mantenimiento complicado.
- Sin extensión del servicio.

Bibliografía

What is cloud computing - Azure Training

What is Azure - Microsoft Cloud Services

ETL vs ELT

What are mapping data flows - Azure Data Factory Training

Azure Data Factory

Azure Synapse Analytics

Azure Functions

Azure Databricks

¡Gracias a todxs!



Email: lidia.96.sm@gmail.com

Telegram: @lidiasm06

LinkedIn:

https://www.linkedin.com/public-profile/settings?trk=d_flagship3_profile_self_view_public_profile

GitHub: <https://github.com/lidiasm>

