# Tipologia di ambienti

* Preproduzione serve per il 20 marzo
* Produzione serve per il 28 marzo

# Descrizione ambiente

|  |  |  |  |
| --- | --- | --- | --- |
| ***Ambiente*** | **Server/Servizio** | **Descrizione** | **Data per cui serve** |
| *Test* | Azure Databricks | Data Analytics Workload, Premium Tier, 2 D8V3 (8 vCPU(s), 32 GB RAM) x 624 Hours, Pay as you go, 3 DBU x 624 Hours | 20/03/2019 |
| SQL Data Warehouse | Tier: Compute Optimized Gen2, Compute: 300 x 480 Hours, Storage: 1 TB | 20/03/2019 |
| Data Lake Gen2 | Data Lake Storage Gen2, Standard, LRS Redundancy, Hot Access Tier, Hierarchical Namespace File Structure, 3 TB Capacity, Write operations: 4 MB x 1,000,000 operations, 1,000,000 List and Create Container Operations, Read operations: 4 MB x 1,000,000 operations, 100,000 Iterative write operations, 1,000,000 Other operations. 1,000 GB Data Retrieval, 1,000 GB Data Write | 20/03/2019 |
| Polybase | Con Installazione di Data Lake Gen2 | 20/03/2019 |
| Analysis Services | Standard S2 (Hours), 1 Instance(s), 360 Hours | 17/04/2019 |
| Data Scientist  Virtual Machine | 1 DS3 v2 (4 vCPU(s), 14 GB RAM) x 720 Hours; Windows – (OS Only); Pay as you go; 0 managed OS disks – S4, 100 transaction units | 20/03/2019 |
| *Produzione* | Azure Databricks | Tipologia come in test ma In attesa specifiche riguardo al dimensionamento | 28/03/2019 |
|  | SQL Data Warehouse |  | 28/03/2019 |
|  | Data Lake Gen2 |  | 28/03/2019 |
|  | Polybase |  | 28/03/2019 |
|  | Analysis Services |  | 14/05/2019 |
|  | Data Scientist  Virtual Machine |  | 28/03/2019 |
| *UAT* | TBD | TBD | TBD |

* Azure Databricks (Fase di ETL)
  + Premium Tier
  + Specifiche
    - 2 D8V3 (8 vCPU(s), 32 GB RAM) x 624 Hours
    - 3 DBU x 624 Hours
  + Da prevedere attività per la connessione con sistemi on-premise
* Azure Data Lake Gen2
  + Hot Access Tier
  + Specifiche
    - Storage
      * 3 TB
    - Write operations
      * 4 MB x 1,000,000 operations
      * 1,000,000 List and Create Container Operations
    - Read operations
      * 4 MB x 1,000,000 operations
      * 100,000 Iterative write operations
      * 1,000,000 Other operations
    - 1,000 GB Data Retrieval
    - 1,000 GB Data Write
    - LRS Redundancy
    - Hierarchical Namespace File Structure
* Azure SQL Data Warehouse
  + Compute Optimized Gen2 Tier
  + Specifiche
    - Computazione
      * 300 x 480 Hours
    - Storage
      * 1 TB
    - Max 12 query in parallelo con il nostro sizing
      * Sizing di produzione – 500 – prevede max 20 query
    - Sizing massimo preveder max 120 query
* Polybase (Interazione tra Azure Data Lake e Azure SQL DWH)
* Azure Analytics Services
  + Non prioritario per Sprint 1
  + Specifiche
    - Standard S2 (Hours)
    - 1 Instance
    - 360 Hours
* Data Scientist Virtual Machine
  + Specifiche
    - 1 DS3 v2 (4 vCPU(s)
    - 14 GB RAM) x 720 Hours
    - Windows – (OS Only)
    - Pay as you go
    - 0 managed OS disks – S4
    - 100 transaction units

# Schema architettura logica

L’interazione tra i programmi presenti nell’ambiente di sviluppo è rappresentata dal seguente diagramma.

7

3

6

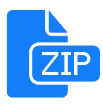
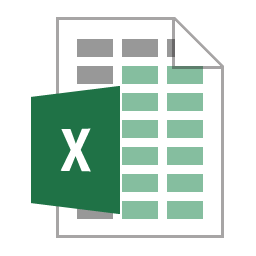
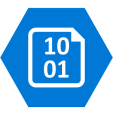
8

5

4

2

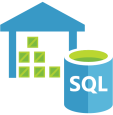
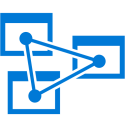
1



polybase

on-premises / azure

Azure Databricks



Azure SQL Data Warehouse

Azure Analysis Services

Azure Data Lake GEN2



solo configurazioni

1. I dati vengono estratti da diverse fonti (descritte nel paragrafo successivo) tramite appositi web crawlers, scritti in Scala, R e Python. Questi crawlers sono eseguiti da Databricks
2. I dati scaricati sono salvati su Azure Data Lake in formato blob (distribuito)
3. Databricks applica diverse operazioni di pulizia e trasformazione sui dati scaricati
4. I dati puliti sono nuovamente memorizzati su Data Lake
5. Una volta terminato il processo di pulizia, i dati vengono caricati sulla DWH.  
   La comunicazione tra il formato blob, distribuito, e la DWH è gestita tramite Polybase.
6. Su Databricks sono anche presenti alcuni scripts che leggono opzioni di configurazione dalla DWH e, sfruttando i dati già contenuti nel Data Lake, producono ulteriori dati per la DWH.
7. PowerBI ed Excel usano le informazioni contenute nella DWH per la produzione di reports.
8. In alternativa, questi strumenti possono operare su un’immagine della DWH prodotta tramite Azure Analytics Services, in modo da non rallentare la DWH stessa.

# Flussi

Occorre che Azure Databricks abbia accesso in lettura alle seguenti fonti (firewall rules):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Priorità | Nome fonte | Tipo di accesso | Tipo IP | Rete | Server | | Porta |
| 1 | **Arkive 1** | Accesso a DB tramite IP IN WHITELIST | Pubblico | Azure AXPO/ARK |  |  | |
| 1 | **Artesian** | API autenticate e autorizzate | Pubblico | Azure AXPO/ARK |  |  | |
| 1 | **axpifotest/axpidata** | Accesso a DB tramite IP in whitelist | Pubblico | Azure AXPO |  |  | |
| 2 | **PFX11** | Accesso a DB | Privato | prod.axponet.ch | VDBDN319 | standard DB MS Sql Server | |
| 3 | **DWH Commerciale** |  | Pubblico | Azure AXPO (new tbd) |  |  | |
| 3 | **DWH Commerciale attuale** | Accesso a DB | Privato | prod.axponet.ch | VDBDN252  SID: axpoitp3 | 1523 DB Oracle | |
| 3 | **IFS** | Accesso a DB | Privato | prod.axponet.ch | VDBDN254 | 1521 DB Oracle | |
| 3 | **IFS** | Accesso a DB | Privato | prod.axponet.ch | subdn666.prod.axponet.ch | 33001 DB MySQL | |
| 3 | **Trimp** | Accesso a DB | Privato | prod.axponet.ch | VDBDN245 | 1433 DB MS Sql Server | |
| 4 | **FTP I4C** | SFTP | Privato | prod.axponet.ch | 10.220.160.164 | 22 | |
| 5 | **XDM Mercati** | Accesso a DB | Privato | prod.axponet.ch | VDBDN253  SID: xdmaxitp | 1523 DB Oracle | |
| 6 | **DWH ETRM** | Accesso a DB | Privato | prod.axponet.ch | VDBDN252  SID: axpoitp1 | 1523 DB Oracle | |
| 6 | **Edigas** | Accesso a DB | Privato | prod.axponet.ch | vdbdn319\axpoit | 1433 DB MS Sql Server | |
| 7 | **TDF** | API | Pubblico | Azure AXPO/ARK  (DB in whitelist) |  |  | |

Inoltre dovrà accedere al servizio sFTP Azure Axpo, per recuperare il mirroring dei file seguenti:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nome fonte | Tipo di accesso | Tipo IP | Rete | Riferimento di rete |
| **SABDN407** | Shared folder | Pubblico | prod.axponet.ch |  |
| **Server misure** | Shared folder | Privato | prod.axponet.ch | \\sfige100.prod.axponet.ch\eglit$ |
| **Server misure** | Shared folder | Privato | prod.axponet.ch | \\prod.axponet.ch\italy\trilance |