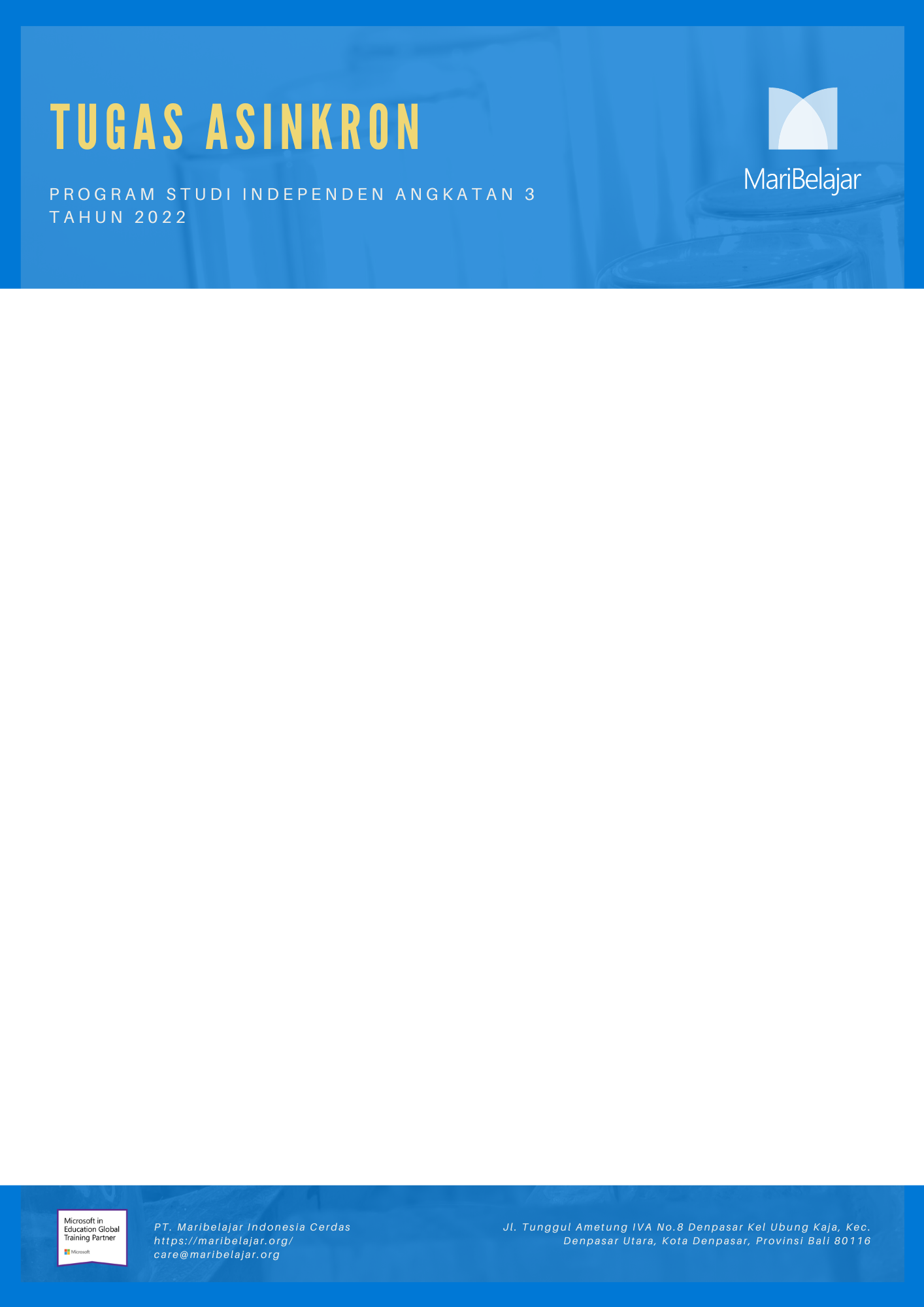
****

**Nama Kursus : Microsoft Azure Data Fundamentals**

**Pertemuan : 6**

**Tim Penyusun : Charles Bernando, S.Si, M.A., Ph.D.**

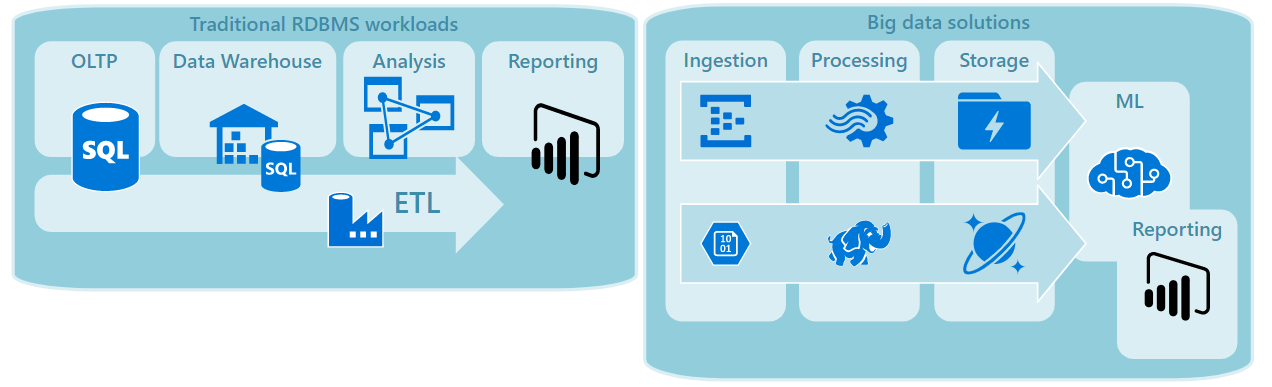
**Total Jam Pembelajaran : 25 JP**

Untuk kasus di dunia perbankan:

1. Buatlah sebuah diagram yang menjelaskan alur data berjalan dari database hingga ke data warehouse sampai akhirnya digunakan untuk visualisasi.
2. Pada kasus dimana diperlukan data social media seperti data dari Twitter, komponen Azure apa yang dapat digunakan untuk memasukkan data social media tersebut ke dalam data warehouse? Jelaskan juga proses dan tahapan masuknya data dari social media ke data warehouse!
3. Buatlah sebuah skema data warehouse dengan menggunakan star schema dimana terdapat satu fact table dan beberapa dimension tables. Berikan juga atribut yang ada pada fact table dan dimension tablesnya (atribut dibebaskan)
4. Buatlah sebuah skema data warehouse dari skema di nomor 3 diatas, namun kali ini dengan menggunakan skema snowflake. Kemudian berikan kelebihan dan kekurangan skema snowflake ini dibandingkan dengan star schema di nomor 3!

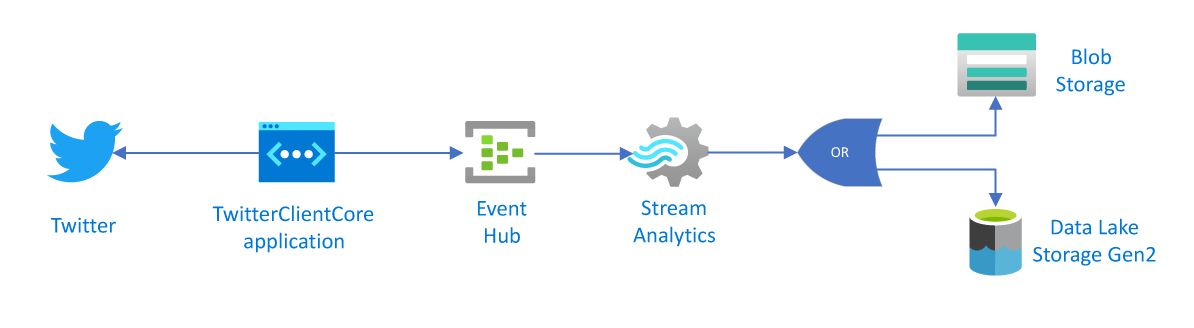
**Nama : Hanifah Al Humaira**

1. Berikut adalah data flow diagram untuk proses data berjalan dari awal hingga visualisasi data.



1. Layanan yang dapat digunakan untuk memasukkan data social media ke dalam data warehouse adalah Azure Stream Analytics. Dalam layanan ini terdapat social media analytics tools yang dapat membantu sebuah perusahaan untuk memahami apa saja yang sedang tren. Topik yang sedang tren adalah subjek dan sikap yang memiliki volume kiriman yang tinggi di media sosial. Analisis sentimen yang juga disebut penambangan opini , menggunakan alat analitik media sosial untuk menentukan sikap terhadap suatu produk atau ide. Contoh yang bagus adalah Real-Time Twitter Analytics Tools, layanan ini memungkinkan kita untuk menganalisis kata kunci (hashtag) tertentu dan mengembangkan analisis sentimen pada umpan.

Berikut arsitektur yang akan diimplementasikan untuk proses dan tahapan masuknya data dari social media ke data warehouse



1. **Membuat Sebuah Event Hub Untuk Streaming Input**

Pada bagin ini kita membuat kumpulan nama hub kejadian dan menambahkan hub kejadian ke kumpulan nama tersebut. Kumpulan nama hub kejadian digunakan untuk mengelompokkan bus instances tentang kejadian terkait secara logis.

1. **Memberikan akses ke event hub**

Sebelum sebuah proses dapat mengirim data ke hub kejadian, hub kejadian tersebut memerlukan kebijakan yang mengizinkan akses. Kebijakan akses menghasilkan string koneksi yang menyertakan informasi otorisasi.

1. **Mengonfigurasi dan memulai Twitter Client Application**

Twitter Client Application mendapatkan kejadian tweet langsung dari Twitter. Untuk melakukannya, perlu izin untuk memanggil API Streaming Twitter. Untuk mengonfigurasi izin tersebut, Anda membuat aplikasi di Twitter yang menghasilkan informasi masuk unik (seperti token OAuth). Anda kemudian dapat mengonfigurasi aplikasi klien untuk menggunakan informasi masuk ini saat melakukan panggilan API.

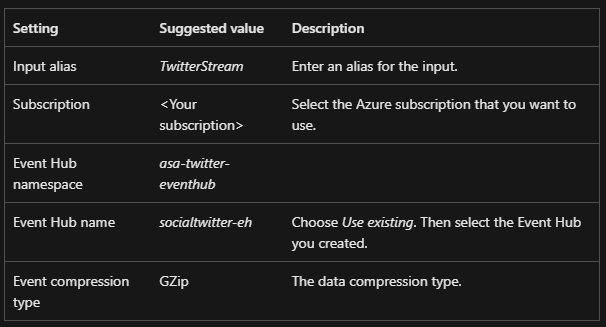
1. **Membuat Aplikasi Twitter**

Langkahnya adalah sebagai berikut:

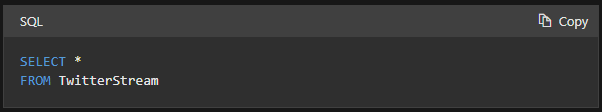
1. Dari browser web, buka Twitter Untuk Pengembang, buat akun pengembang, lalu pilih Buat aplikasi. Anda mungkin melihat pesan yang mengatakan bahwa Anda perlu mengajukan permohonan untuk akun pengembang Twitter. Jangan ragu untuk melakukannya, lalu setelah aplikasi Anda disetujui, Anda akan melihat email konfirmasi. Ini bisa memakan waktu beberapa hari untuk disetujui untuk akun pengembang.
2. Di halaman Buat aplikasi, berikan detail untuk aplikasi baru, lalu pilih Buat aplikasi Twitter Anda.
3. Di halaman aplikasi, pilih tab Kunci dan Token, lalu salin nilai untuk Kunci API Konsumen dan Kunci Rahasia API Konsumen. Selain itu, pilih Buat di bawah Token Akses dan Rahasia Token Akses untuk menghasilkan token akses. Salin nilai untuk Token Akses dan Rahasia Token Akses.
4. Simpan nilai yang kita ambil untuk aplikasi Twitter. Kita membutuhkan nilainya nanti.
5. **Mengonfigurasi aplikasi klien**
6. Pastikan kita telah mengunduh aplikasi TwitterClientCore, seperti yang tercantum dalam prasyarat.
7. Gunakan editor teks untuk membuka file App.config Anda. Lakukan perubahan berikut pada elemen <appSettings>:
   1. Set oauth\_consumer\_key to the Twitter Consumer Key (API key).
   2. Set oauth\_consumer\_secret to the Twitter Consumer Secret (API secret key).
   3. Set oauth\_token to the Twitter Access token.
   4. Set oauth\_token\_secret to the Twitter Access token secret.
   5. Set EventHubNameConnectionString to the connection string.
   6. Set EventHubName to the event hub name (that is the value of the entity path).
8. Buka command line dan navigate ke directory dimana TwitterClientCore berada. Gunakan command dotnet build to build the project. Kemudian gunakan command dotnet run untuk menjalankan aplikasi. Aplikasi akan mengirimkan Tweets ke Event Hub yang telah dibuat sebelumnya.
9. **Membuat Azure Stream Analytics**

Sekarang setelah kejadian tweet streaming secara real time dari Twitter, Anda dapat mengatur pekerjaan Azure Stream Analytics untuk menganalisis kejadian ini secara real time.

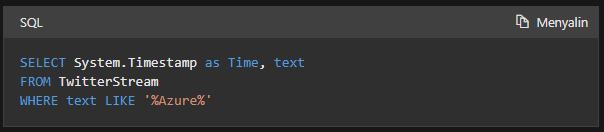
1. Di portal Microsoft Azure, navigasi ke grup sumber daya Anda dan pilih + Add. Lalu cari Stream Analytics job dan pilih Create.
2. Beri nama pekerjaan sebagai socialtwitter-sa-job dan tentukan subscription, resource group, dan location.
3. Pilih Create. Kemudian navigasikan ke pekerjaan Anda ketika penyebaran telah selesai.
4. **Menentukan input pekerjaan**
5. Pada pekerjaan Azure Stream Analytics Anda, pilih Input dari menu sebelah kiri di bawah Job Topology.
6. Pilih + Add stream input > Event Hub. Isi form inputan dengan mengikuti informasi sebagai berikut:



1. **Tentukan job query**
2. Pada bagian **overview**, pilih Edit kueri di dekat kanan atas kotak Kueri. Azure mencantumkan input dan output yang dikonfigurasi untuk pekerjaan tersebut serta memungkinkan Anda membuat kueri untuk mengubah aliran input saat dikirim ke output.
3. Edit Query



1. Data kejadian dari pesan akan muncul di jendela Pratinjau input di bawah kueri Anda. Pastikan Tampilan diatur ke JSON. Jika Anda tidak melihat data apa pun, pastikan generator data Anda mengirimkan kejadian ke hub kejadian dan Anda telah memilih GZip sebagai jenis pemadatan untuk input.
2. Pilih Uji kueri dan perhatikan hasilnya di jendela Hasil uji di bawah kueri Anda.
3. Ubah kueri di editor kode menjadi berikut ini dan pilih Uji kueri:



1. Kueri ini menampilkan semua tweet yang menyertakan kata kunci Azure.
2. **Membuat sink output**

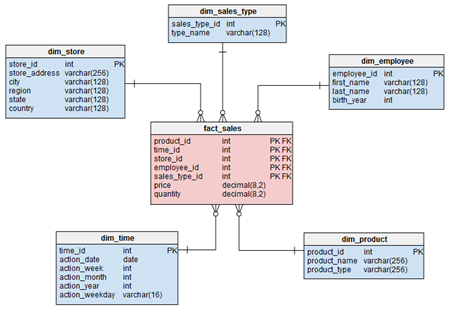
Sekarang telah menentukan aliran kejadian, input hub kejadian untuk menyerap kejadian dan kueri untuk melakukan transformasi melalui aliran. Langkah terakhir adalah menentukan sink output untuk pekerjaan tersebut.

Dalam panduan cara penggunaan ini, kita menulis kejadian tweet yang diagregasi dari kueri pekerjaan ke penyimpanan Azure Blob. Kita juga bisa mendorong hasil kita ke Azure SQL Database, penyimpanan Azure Table, Azure Event Hubs, atau Power BI, tergantung pada kebutuhan aplikasi kita.

1. **Tentukan Job**
2. Di bawah bagian Job Topology di menu navigasi kiri, pilih Output.
3. Di halaman Output, klik + Blob storage/Data Lake Storage Gen2:

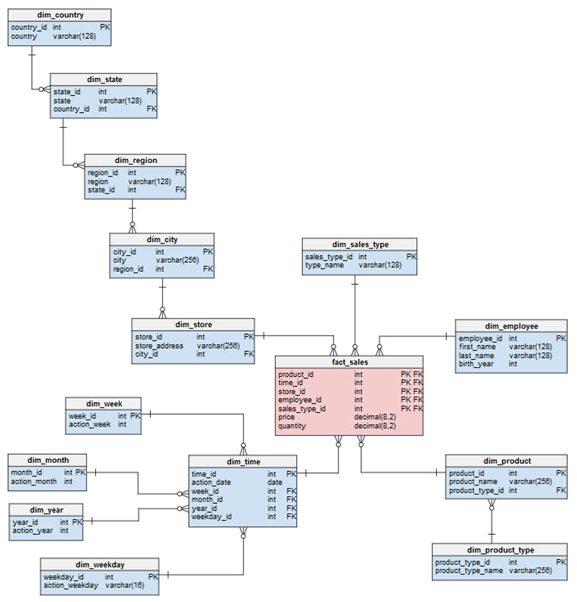
* Output alias: Use the name TwitterStream-Output.
* Import options: Select Select storage from your subscriptions.
* Storage account. Select your storage account.
* Container. Select Create new and enter socialtwitter.

1. Pilih simpan
2. **Memulai Job**
3. Pastikan aplikasi TwitterClientCore berjalan.
4. Dalam halaman overview, pilih Start.
5. Pada halaman Start Job, untuk Job output start time, pilih now lalu pilih Start.
6. Star Schema



Fact table yang berwarna merah muda dan dimension tables yang berwarna biru muda.

1. Snowflake



**Kelebihan dan kekurangan star schema**

Kelebihan model skema bintang:

* Cenderung mudah dipahami karena modelnya yang lebih sederhana
* Memudahkan mencari isi karena kesederhanaannya dengan cara melihat *step by step* dari masing-masing dimensinya
* Proses *query* lebih cepat pada saat proses OLAP

Kekurangan model skema bintang:

* Ukuran data lebih besar karena ada data yang disimpan ulang
* *Maintenance* dan *update* lebih sulit

**Kelebihan dan kekurangan snowflake**

Kelebihan model skema*snowflake*:

* Ukuran data lebih kecil di dalam tempat penyimpanan
* Lebih mudah dilakukan *maintenance* dan *update*
* Proses *query* lebih cepat pada saat proses ETL

Kekurangan model skema*snowflake*:

* Cenderung lebih sulit dipahami karena kompleksitasnya
* Sulit mencari data yang dibutuhkan karena melihat strukturnya yang kompleks dan bercabang-cabang