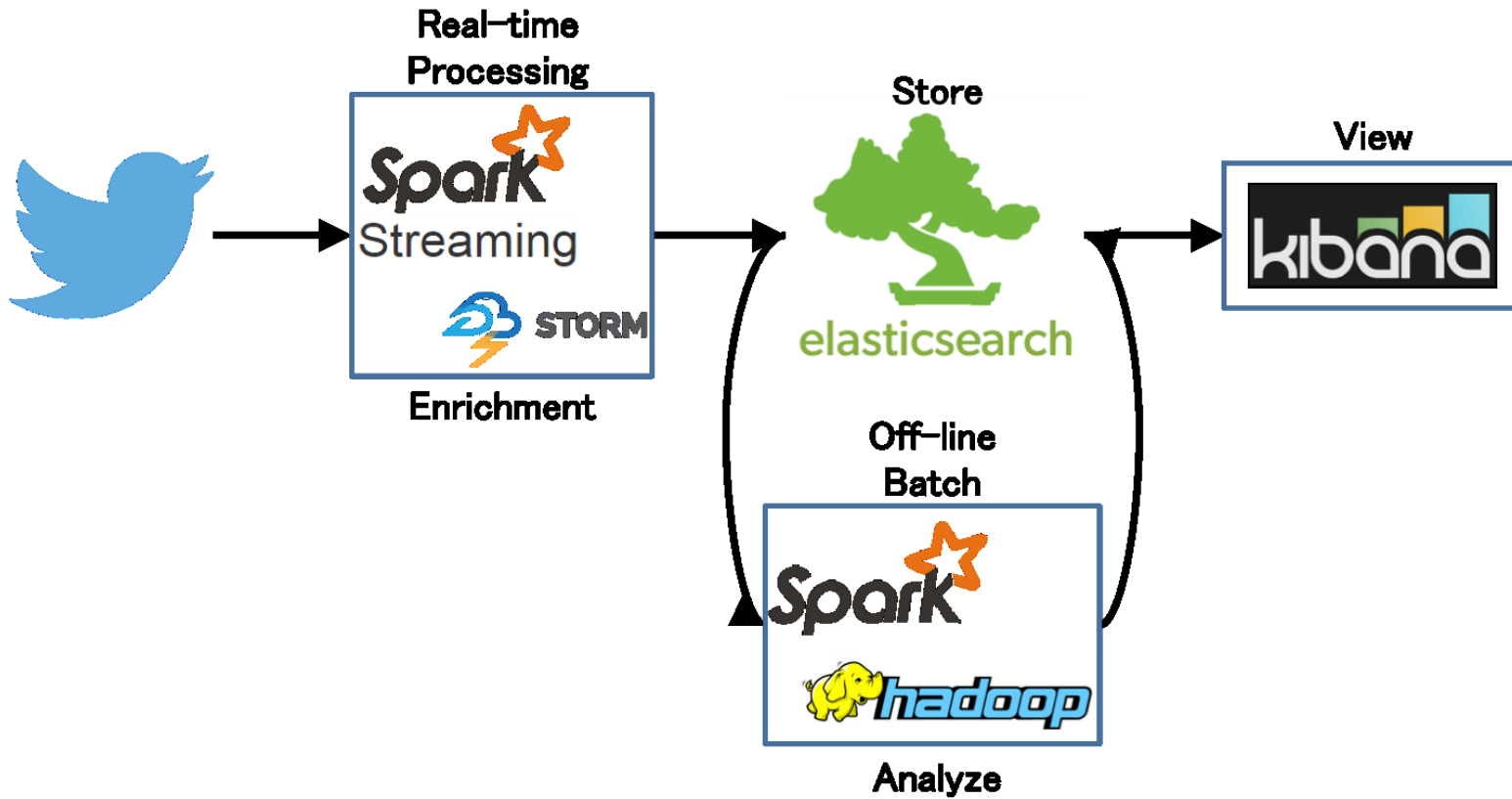


Bu bölümde Apache Spark ile Structured Streaming Mimarisi hakkında bilgiler vereceğiz

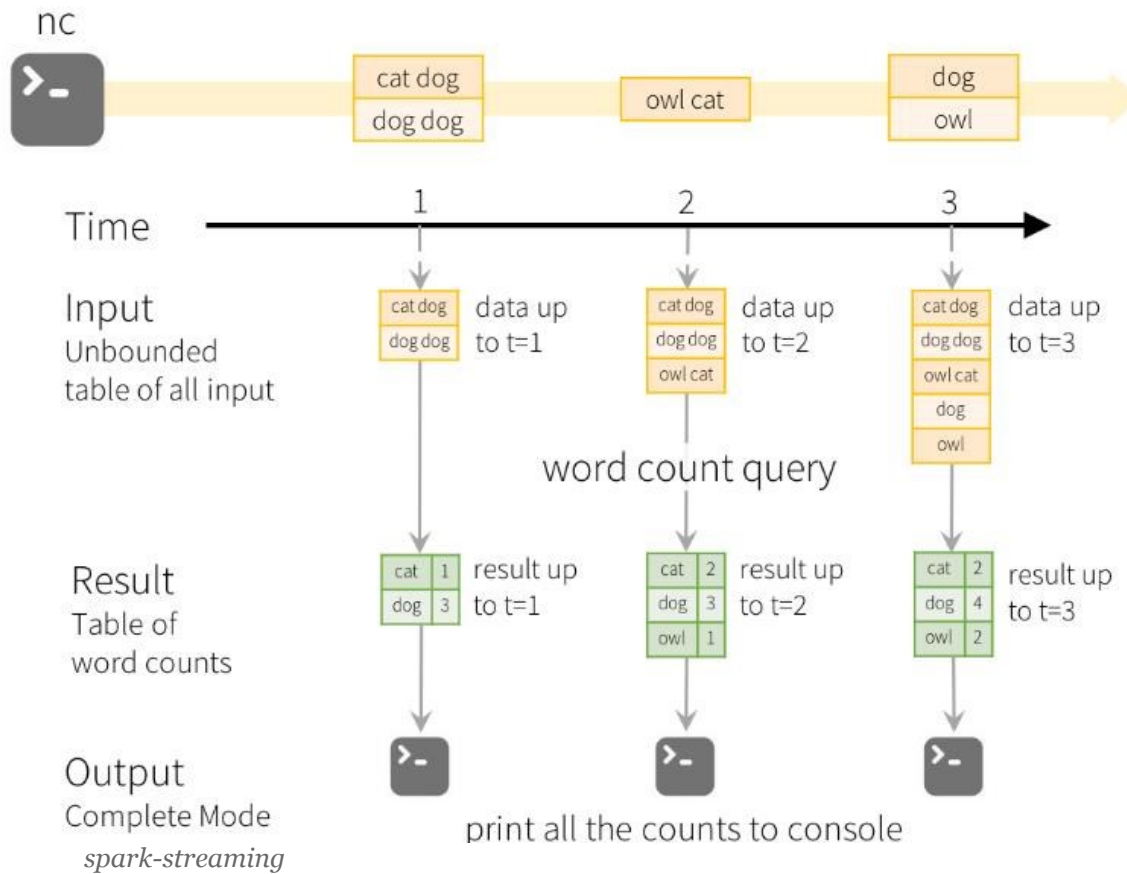
Bir önceki bölümde Structured Streaming ile basit bir örnek yapmıştık. Geliştirdiğimiz uygulamada belirli bir port üzerinden kelimeler gönderip , bu port'u dinleyen Spark uygulamasına word-count analizi yaptırдық

Şimdi ise mimari tarafı inceleyelim



realtime_mimari

Structured Streaming yapısında gelen veriler sürekli olarak belirli sonuçlar üzerine eklenir



Structered Streaming ile veriler şu şekilde analiz edilir

- Veriler dış sistemlerden (dosya, socket, kafka ..) belirli periyotlarda alınır
- Yukardaki örnekte **t=1** zamanında ilk aldığımız cümle "*cat dog dog dog*"
- Gelen verileri kelime sayısına göre analiz ettiğimiz için hangi kelimedenden kaç adet olduğu hesaplanır. Bu aşamaya **result table** denir
- **Result table** ile oluşturduğumuz sonuçlar tekrar dış sistemlere yazdırılabilir (mysql,hdfs..)

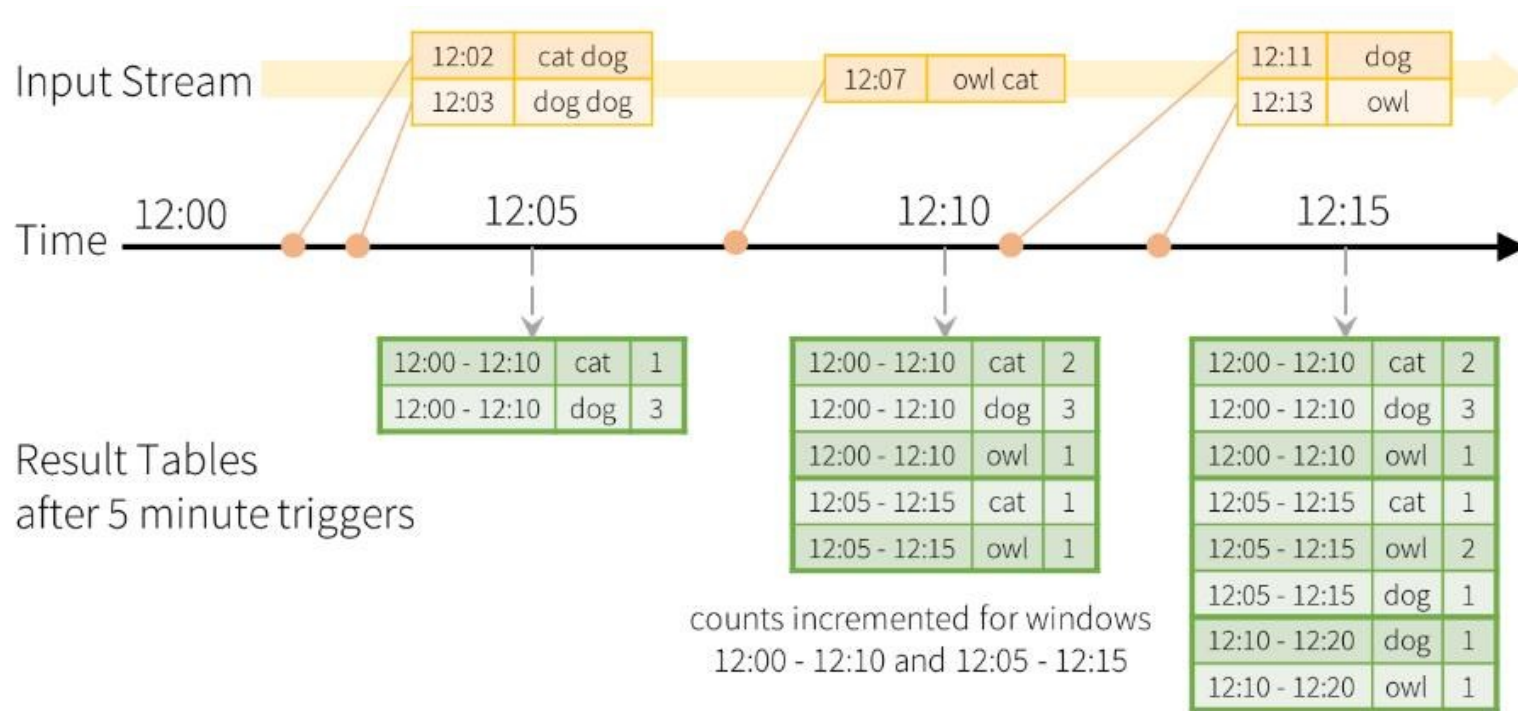
t=2 zamanında ise "*own cat*" kelimesini aldık ve tekrar **word-count** uygulamasını çalıştırarak güncel **result table** elde ettik

Structered Streaming ile veriler zaman gruplarında nasıl işlenir ? (window operations)

Anlık olarak aldığımız verileri zaman bloklarına bölebiliriz. Örnek verirsek bir web sitesinden aldığımız loglarda 5'er dakikalık bölümlerde kaç tane tekil kullanıcının sitemize geldiğini hesaplayabiliriz

Alttaki örnekte ise veriler işlendikten sonra 10 'ar dakikalık bloklara ayrılmıştır . Örnek verirsek **12:00-12:10** zaman dilimi arasında 2 adet **cat** kelimesi , 3 adet **dog** kelimesi , 1 adet **owl** kelimesi gelmiştir

Ayrıca Structered Streaming ile zaman bloklarının belirli periyotlarda kaydırılmasını sağlayabiliriz(sliding). Alttaki örnekte **12:00-12:10** zaman diliminin bir sonraki bloğu **12:05-12:15** şeklindedir .



spark-window

Structered Streaming ile üretilen sonuçlar hangi formatlarda dış sistemlere gönderilir

Structered Streaming ile anlık olarak ürettiğimiz sonuçlar şu formatlarda yazdırılır

- *Complete Mode : Result tablosunun son hali yazdırılır*

```
StreamingQuery query = countApp.writeStream().outputMode("complete")  
                                .format("console").start();
```

```
[2017-01-24 23:31.37] ~  
[Serkan.Serkan-PC] > netcat -l -s 127.0.0.1 -p 9999  
apache spark apache hadoop  
apache oozie
```

spark-window_complete_moba

Sonuçlar

```
+-----+-----+  
| value|count|  
+-----+-----+  
|apache|    3|  
|  spark|    1|  
|  oozie|    1|  
|hadoop|    1|  
+-----+-----+
```

spark-window_complete

-
-
-
-
-

- *Update Mode : Result tablosunda sadece güncellenen kayıtlar yazdırılır*

Bu bölümde spark structured streaming mimarisi hakkında bilgiler verdik .Diğer büyük veri yazıları için sitemizi takip edebilirsiniz