

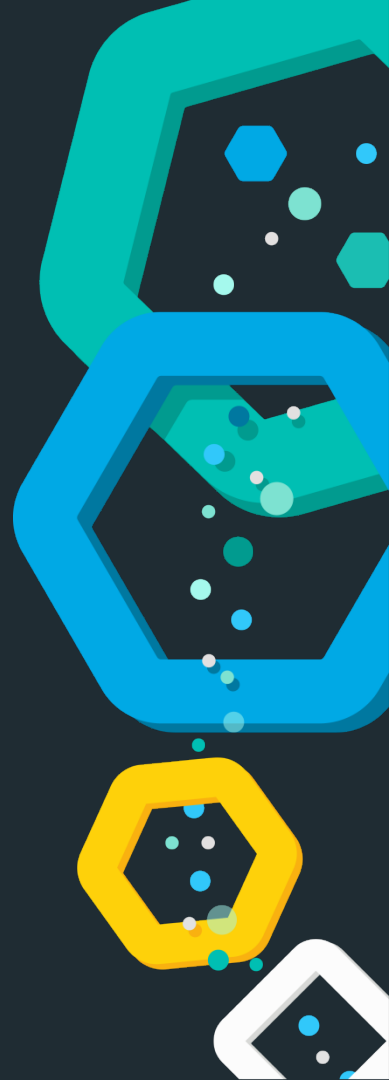


Time Series Data

Estrazione, visualizzazione e analisi di titoli di borsa

Luca Gennari

Elasticsearch Italy User Group Milano, Giugno 05, 2020
Original author: Luca Gennari - *Education Engineer EMEA*





Time Series Data

estrazione, visualizzazione e analisi di titoli di borsa



Una serie temporale è una serie di punti, dati o informazioni indicizzati in ordine temporale.

Più comunemente, una serie temporale è una sequenza di dati presa in punti successivi equamente distanti nel tempo.

Wikipedia

Time Series Data - Introduzione

- Dati associati ad eventi, le informazioni fanno riferimento a un momento nel tempo.
 - logs, cataloghi, metriche, titoli di borsa, dati di produzione, etc
 - in genere crescono molto rapidamente fintanto che l'evento esiste
- Devono essere "*denormalizzati*" e strutturati in maniera corretta prima di poter effettuare la loro analisi.
 - è il modo standard di indicizzare documenti in un motore di ricerca o in qualsiasi NoSQL database.

Un esempio di informazione da serie temporale

```
{
  "runtime_ms" : 99,
  "method" : "GET",
  "host" : "server1",
  "geoip" : {
    "location" : {
      "lon" : -122.4128,
      "lat" : 37.7758
    },
    "country_name" : "United States",
    "country_code2" : "US",
    "continent_code" : "NA",
    "country_code3" : "US",
    "region_name" : "California",
    "city_name" : "San Francisco"
  },
  "level" : "info",
  "http_version" : "1.1",
  "@timestamp" : "2017-04-11T09:04:34.588Z",
  "status_code" : 200,
  "response_size" : 54975,
  "user_agent" : "Google Chrome"
}
```

Server Logs

```
{
  "date" : "2018-10-09",
  "high" : 68.0,
  "low" : 62.56999969482422,
  "open" : 65.58999633789061,
  "close" : 66.44999694824219,
  "volume" : 370400.0,
  "adj_close" : 66.44999694824219,
  "moving_average" : 67.46333312988281,
  "symbol" : "ESTC",
  "name" : "Elastic NV",
  "description" : "Elasticsearch is a search engine based on the Lucene library. It provides distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents. Elasticsearch is developed in Java."
}
```

Stock Exchange Trading Info

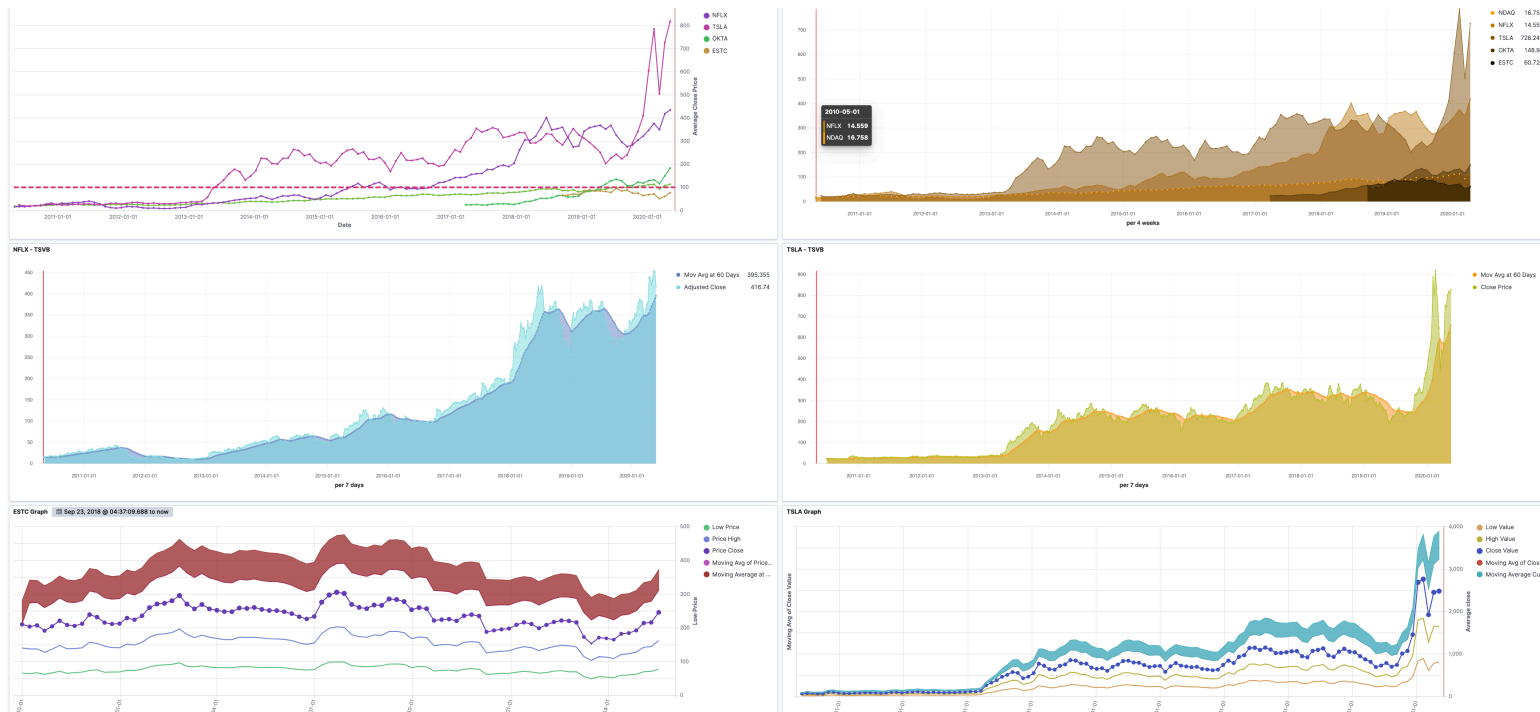
Time Series in Elasticsearch

Indicizzare documenti che rappresentano serie di dati temporali in Elasticsearch è semplice ma richiede che i campi siano *"mappati"* correttamente.

Kibana dispone in maniera nativa di strumenti che permettono di analizzare dati Time Series in ogni sua forma come i TSVB (Time Series Visual Builder)

Vediamone alcuni...

Visualizzazioni di serie temporali standard...



Dashboard with TSVB and Line Graphs

...e con l'uso del Machine Learning.



Machine Learning Analytics



Machine Learning Anomaly Detection

Gentile pubblico, ecco a voi...

Time Series Data

DEFINIZIONE

- Le serie temporali sono un insieme di osservazioni, punti o dati rilevati a specifici intervalli regolari.
- Si possono utilizzare per prevedere valori futuri basati sulle precedenti osservazioni.
- Nelle serie temporali ci deve essere solo ed esclusivamente una variabile, il tempo.

RAPPRESENTAZIONE

- Le serie temporali sono tracciate molto spesso tramite grafici a linee.
- Ci si possono sovrapporre altre informazioni per trovare correlazioni.
- Sono indispensabili nella ricerca di pattern su grafico o numerico.

UTILIZZO

- Analisi statistica.
- Previsioni di borsa.
- Elaborazione del segnale.
- Riconoscimento dei modelli.
- Previsioni meteorologiche o dei terremoti.
- Astronomia e in qualsiasi settore della scienza applicata e dell'ingegneria.
- IoT e applicazioni industriali.

Kibana TSVB

Time Series Visual Builder

TSVB - Introduzione

- ◆ Combina un ampio range di aggregazioni, incluse aggregazioni di tipo pipeline.
 - ✓ questo permette di analizzare i dati in diversi modi
- ◆ Supporta diversi tipo di personalizzazioni su come visualizzare i dati
 - ✓ colore di sfondo
 - ✓ informazioni su diversi assi
 - ✓ sovrapposizione di dati aggregati
 - ✓ tipi di grafici differenti in un unica visualizzazione
- ◆ Supporta index patterns multipli nella medesima visualizzazione
 - ✓ permette di valutare le differenze tra informazioni diverse
- ◆ Si possono creare annotazioni a livello grafico
 - ✓ rende più semplice individuare anomalie o punti di interesse.

TSVB - Funzionalità

◆ Aggregazioni supportate

- ✓ operazioni matematiche (max, min, sum, etc...)
- ✓ Valori statici
- ✓ massimi e minimi complessivi
- ✓ aggregazioni su dati aggregati (pipeline aggregations)
- ✓ media mobile, deviazione standard, percentile, somma di quadrati, varianza, etc...
- ✓ ...e molto altro

◆ Offre una maggiore flessibilità dei grafici standard

- ✓ Shifting di serie temporali
- ✓ clonazioni di serie esistenti
- ✓ Stili differenti di visualizzazioni
- ✓ confronti di diversi dati da diversi indici

Componenti di una analisi su dati temporali



Tendenza

- Direzione nella quale l'informazione si dirige



Stagionalità

- Pattern che i dati possono assumere in particolari periodi



Irregolarità

- Cambi di pattern che non seguono la stagionalità o la ciclicità

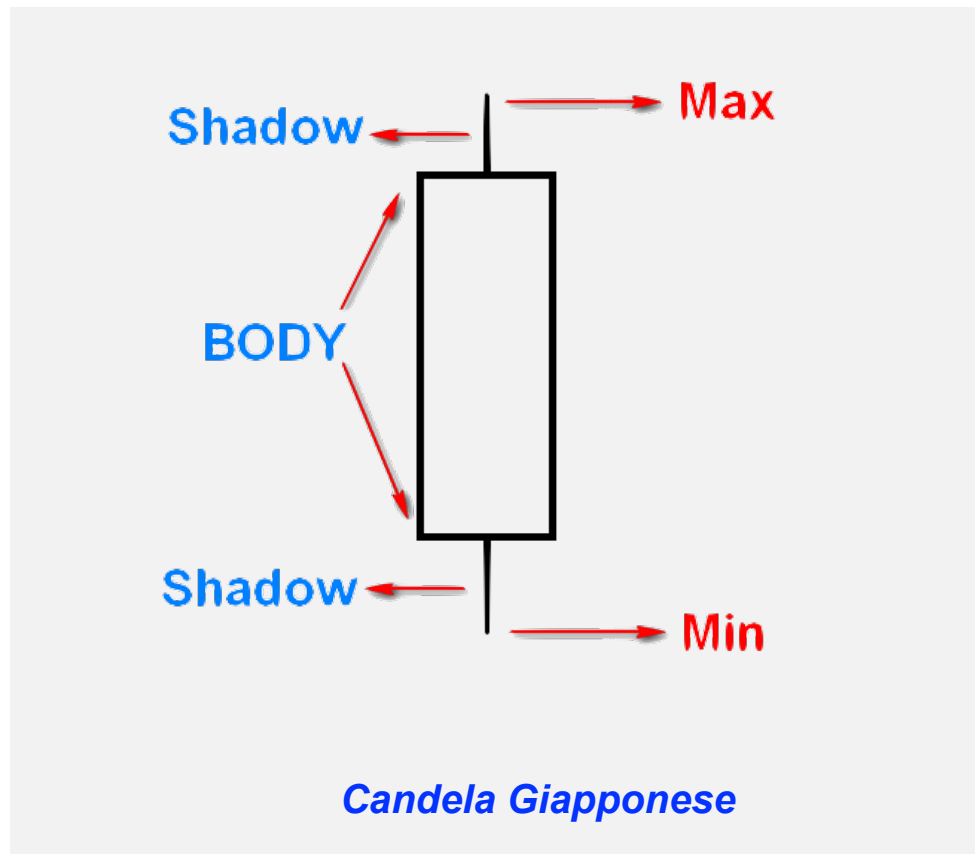


Ciclicità

- Ripetizione di pattern con carattere costante e indipendente dalla stagionalità

Dati di un titolo di borsa

- In genere, per ogni singola giornata di contrattazione si hanno cinque informazioni basilari:
 - Prezzo di apertura
 - Prezzo di chiusura
 - Minimo raggiunto dal prezzo
 - Massimo raggiunto dal prezzo
 - Aggiustamento della chiusura
- Un altro dato non basilare ma utile è il volume scambiato
 - Indica i titoli scambiati sia in vendita che in acquisto
 - Fornisce una idea dell'interesse del mercato su un particolare titolo (alti volumi, alto interesse e vice versa)
- Dati in genere non forniti o che, comunque, non fanno parte della contrattazione ma utili ai fini dell'analisi:
 - Media mobile, tra le più usate:
 - A 30, 60 o 90 giorni
 - Date degli annunci (earnings)
 - *Dato previsto o dato che ci si aspetta (generalmente rilasciato qualche settimana prima della data dell'annuncio)*
 - *Dato reale (lo si ottiene al momento dell'annuncio e viene utilizzato per la previsione successiva)*



Struttura documento

- Un campo data (in questo caso ci interessa solo la data, non il tempo)
- Ci sono sei campi numerici definiti
 - **low** - prezzo minimo raggiunto
 - *tipo: float*
 - **High** - prezzo massimo raggiunto
 - *tipo: float*
 - **Open** - Prezzo di apertura
 - *tipo: float*
 - **Close** - prezzo di chiusura
 - *tipo: float*
 - **Adj_close** - aggiustamento del prezzo di chiusura
 - *tipo: float*
 - **Volume** - contrattazioni scambiate
 - *tipo: Integer*
- Un campo generato in pre-analisi che è la media mobile a 60 giorni.
- Campi descrittivi:
 - Simbolo del titolo
 - Nome del sottostante
 - Descrizione del sottostante

```
{
  "_index" : "indeces_combined",
  "_type" : "_doc",
  "_id" : "mfgmdHIBzFEP8t6vM-TW",
  "_score" : 1.0,
  "_source" : {
    "date" : "2018-10-09",
    "high" : 68.0,
    "low" : 62.56999969482422,
    "open" : 65.58999633789061,
    "close" : 66.44999694824219,
    "volume" : 370400.0,
    "adj_close" : 66.44999694824219,
    "moving_average" : 67.46333312988281,
    "symbol" : "ESTC",
    "name" : "Elastic NV",
    "description" : "Elasticsearch is a search..."
  }
}
```

Struttura documento indicizzato

DEMO TIME

Nota Informativa

Questa presentazione è stata creata a ***scopo puramente informativo***.

Pertanto, le informazioni riportate e/o Elastic, non sono da ritenersi responsabili nei confronti di alcuna parte per controversie legali/civili o amministrative, danni indiretti, specifici, incidentali, punitivi, cauzionali o consequenziali, causati dall'utilizzo o dall'impossibilità di utilizzare le informazioni qui contenute.

Ogni decisione presa privatamente e l'utilizzo delle informazioni qui contenute, è presa a proprio, personale, rischio e pericolo.

Script e dati

Il materiale utilizzato in questa presentazione è liberamente scaricabile e usabile senza restrizione alcuna, al seguente indirizzo github:

<https://github.com/lucagennari-es/time-series-data-meetup.git>

GitHub page:

<https://github.com/lucagennari-es/time-series-data-meetup>

Q&A