

# LEANalyser

## Tredje generationens analysverktyg

*Många frågar sig: Vad är nästa steg i utvecklingen av analysverktyg för Self Service Analytics?*

*För några år sedan var IBM's Watson Analytics mycket hett. Ett verktyg där man ställer frågor på skriven engelska. Denna produkt är inte längre tillgänglig för köp<sup>1</sup>.*

*Nu är det AI, data mining och machine learning som är hajpat - Är det nästa steg i möjliggörandet av Self Service Analytics eller Self Service BI?*

### Bakgrund

Många organisationer investerar stora summor i Self Service Analytics. Syftet är att utnyttja tillgängligt data för att vinna konkurrensfördelar utan att ytterligare betunga organisationens utvecklare av enterprise BI.

Dessa satsningar har i huvudsak bestått av att ge analysintresserade användare och deltidanalytiker tillgång till visualiseringsverktyg som hämtar data från färdigpreparerade data marts, olap-kuber, data warehouses, mm.

Under de senaste åren har tillgången på externa data ökat exponentiellt. Det finns en stor potential i att kombinera interna och externa data för att effektivisera verksamheten. I denna process handlar det mycket om att testa olika kombinationer av data, vilket ställer krav på användarna och deras verktyg. Bland annat måste de olika datamängderna prepareras så att de kan samköras och analyseras, vilket helst ska gå snabbt och smidigt.

### Utmaningen

Dagens visualiseringsverktyg, ibland kallade "Den andra generationens analysverktyg" har långt ifrån tillräcklig funktionalitet för att preparera data. Detta har lett till att nya verktyg för datapreparering eller data-wrangling lanserats under de senaste åren. Emellertid så leder detta till att deltidanalytikerna måste lära sig två verktyg: ett för datapreparering och ett för analys/visualisering. Betänker man att de vanligaste visualiseringsverktygen kräver att man arbetar med dessa flera timmar varje dag för att upprätthålla nödvändig kompetens, så inser man att deltidanalytikerna inte har en lätt uppgift och att många satsningar inom Self Service Analytics får problem.

Analysprocessen med preparering och analys av data är i högsta grad iterativ, dvs man hoppar fram och tillbaka kanske 10—20 gånger mellan dessa faser för att uppnå ett användbart resultat. Använder man olika verktyg så innebär detta att dataprepareringsverktyget läser in data, preparerar data och producerar en utdata-fil som sedan läses in i analysverktyget, varefter analytikern verifierar korrektheten. Man förlorar därmed timmar eller ibland dagar i denna del av analysprocessen. Det ska tilläggas att

---

<sup>1</sup> Enligt IBM's hemsida: <https://www.ibm.com/watson-analytics>

# LEANalyser

priserna för dataprepareringsverktyg är klart högre än priserna för analysverktyg. De är dessutom komplexa att bemästra, vilket i sig inte snabbar upp processen.

## Lösningen

För att komma tillrätta med detta problem har ett antal nya leverantörer lanserat en ny typ av analysverktyg som hanterar både datapreparering och analys/visualisering, den tredje generationens analysverktyg.

Med tredje generationens analysverktyg avser vi ett verktyg som gör det möjligt att hoppa fram och tillbaka i analysprocessens olika faser och göra detta visuellt i realtid, utan att behöva ladda om data. För analysprocessen är ju i högsta grad iterativ.

Den tredje generationens analysverktyg för Self Service Analytics bör:

- Stödja hela analysprocessen med endast ett verktyg
- Vara lätt att använda tack vare ett intuitivt användargränssnitt
- Ha hög prestanda även vid mycket stora datamängder
- Inte kräva programmeringskunskaper eller komplicerad syntax
- Erbjuder en anpassad uppsättning funktioner för just Self Service Analytics

De främsta fördelarna med tredje generationens analysverktyg är att:

- Du inte behöver lära dig något separat dataprepareringsverktyg
- Du blir mycket mer effektiv då du bara arbetar med ett enda verktyg
- Du kan under analysarbetet alltid tvätta, modifiera och finjustera data

## Analysverktygens utveckling och kännetecken

1:a generationen: Dessa verktyg är utvecklade för att bygga färdigpreparerade analyser för BI-konsumenter med liten analyserfarenhet. Analyserna utvecklas av professionella BI-utvecklare. Data hämtas från datalager med förkalkylerade värden.

2:a generationen: Dessa verktyg gör det möjligt att ta fram eleganta visualiseringar av många olika slag för att skapa insiktsfulla analyser. Det är lätt att skapa enklare analyser men alltför ofta tvingas man lära sig avancerad syntax för att nå ända fram. För att uppnå nöjaktig prestanda bör man inte läsa in alla transaktioner i den interna analysdatabasen, vilket innebär att databasen nästan alltid måste prepareras och optimeras av BI-utvecklare.

3:e generationen: Dessa verktyg stöder förutom analys och visualisering även datapreparering i realtid. Med en mycket kort introduktion kan en mindre erfaren analytiker själv skapa en analysdatabas och snabbt genomföra analyser. Med en visuell dataprepareringsfunktion kan användaren på ett intuitivt sätt preparera data i ett fönster och samtidigt visualisera samma data i ett annat.

# LEANalyser

## Vad bör man satsa på?

AI, data mining och predictive analytics är avancerade analystekniker som rätt använda kan ge värdefulla insikter. Användningen av dessa ökar, men inte så snabbt man kan tro. Orsaken bakom detta är att dessa tekniker kräver djup förståelse av de algoritmer som används och därmed specialutbildade användare, ofta data scientists som det råder brist på. Dessutom kräver även dessa verktyg preparering av data. Med rätt typ av analys och visualisering kan en deltidsanalytiker ofta själv komma till samma insikter som data scientists gör via data mining. Därmed kan man konstatera att man bör behärska enklare analysmetoder innan man tar till de mer avancerade.

Exempel på den tredje generationens analysverktyg finner du här: [leanalyser.com](http://leanalyser.com) .

/Tomas Nabel, 2019-01-09