



# Big Data en Fast Data technologie voor Business Intelligence

Hadoop, NoSQL, Spark, Kafka, Storm, NewSQL, Streams

Verhelderende sessie met Rick van der Lans

**Adept**Events

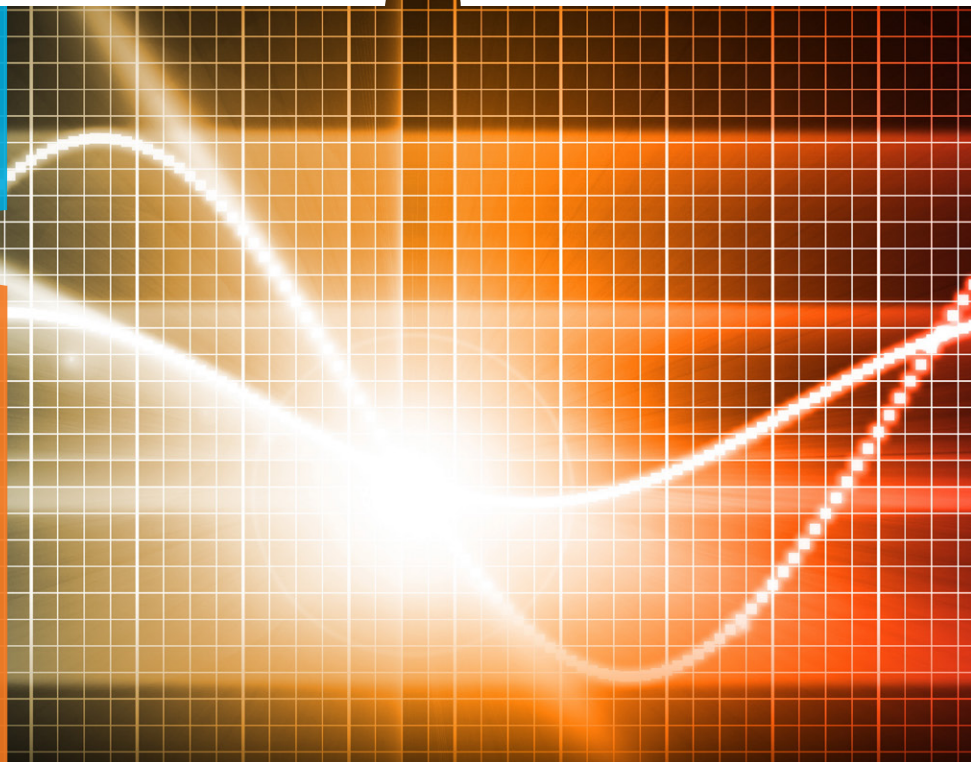
## LOCATIE

Omgeving Utrecht/Hilversum

## REGISTRATIE

[www.adeptevents.nl](http://www.adeptevents.nl)

- Een zeer verhelderende sessie waarin kritisch en onafhankelijk de populaire Big Data en NoSQL ontwikkelingen onder de loep worden genomen
- Fast Data, Streaming Analytics en Data Science
- Wat zijn de toepassingen voor úw organisatie?
- Hoe integreert u met bestaande infrastructuren?
- Aan bod komen onder meer: Hadoop, Mapreduce, MongoDB maar ook Aster Database, Oracle Exalytics, SAP HANA en Teradata





# Big Data en Fast Data technologie voor Business Intelligence

## Hadoop, NoSQL, Spark, Kafka, Storm, NewSQL, Streams

### Inleiding

De wereld van business intelligence en data warehousing wordt de laatste jaren geconfronteerd met een niet aflatende stroom technologieën en daarbij moet een steeds bredere groep gebruikers ondersteund worden. *Big Data* is de grote aanjager van de meeste van deze veranderingen. Big Data heeft geleid tot de introductie van nieuwe gegevensopslag-technologieën, zoals Hadoop, NoSQL en NewSQL, en tot nieuwe platformen voor gegevensverwerking, zoals Spark en Storm. En voor streaming analytics zijn onder andere Kafka en Streams geïntroduceerd. Maar ook de gegevensconsumptie is niet ongewijzigd gebleven. Naast klassieke en interactieve rapporten moeten BI-omgevingen ook ondersteuning bieden voor self-service BI, investigative analytics en data science. De lat voor BI lijkt elk jaar weer hoger gelegd te worden.

Maar de grote vraag is: hoe moeten deze nieuwe gebruikerswensen ondersteund worden en wat is de beste manier om al deze nieuwe technologieën te integreren met traditionele datawarehouse systemen? Wanneer dienen Hadoop en NoSQL ingezet te worden? Kan Spark gebruikt worden als datamart? Welke modules van Hadoop dienen voor welke toepassing ingezet te worden? Wat is de relatie tussen een datawarehouse en een op Hadoop-gebaseerd data lake? Wat kan wel en niet met de nieuwe technologieën?

Tijdens dit seminar worden al deze nieuwe Big Data technologieën kritisch besproken. Wat is hun geheim? Hoe kunnen zij opvallend meer verwerken dan de traditionele databaseservers? Voor- en nadelen worden toegelicht en mogelijkheden en technische eigenschappen worden

besproken. Na het seminar heeft u een gedegen en compleet overzicht van deze snel veranderende wereld van Big Data technologie en hoe deze in te zetten bij BI-systemen.

### Doel van de Powersessie

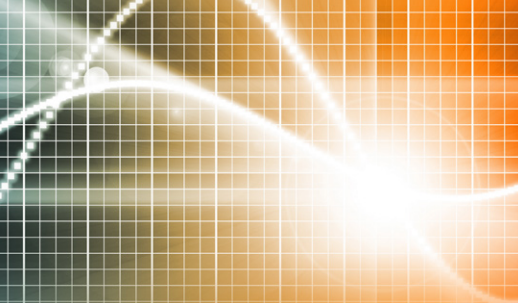
Rick van der Lans geeft in deze intensieve dag antwoord op o.a. de volgende vragen.

- Wat behelst Big Data en waarom is dit belangrijk voor elke organisatie?
- Wat is de relatie tussen Fast Data en streaming analytics?
- Wat is Hadoop en welke concrete producten zijn voor mijn organisatie relevant?
- Wanneer dienen welke Hadoop en NoSQL producten ingezet te worden?
- Wat is de toegevoegde waarde van Spark voor high-end processing?
- Wat zijn de belangrijkste toepassingsgebieden?
- Hoe ontsluiten we Big Data voor Data Science?
- Waarom SQL-on-Hadoop engines de toegankelijkheid van Hadoop verhogen
- Welke richting gaan we op met deze nieuwe producten in bestaande BI-omgevingen?

### Bestemd voor ú

Dit seminar is voor een ieder bestemd die vanuit zijn functie op de hoogte moet blijven van de ontwikkelingen op database- en datawarehousegebied, zoals databasespecialisten, databasebeheerders, datawarehouse- en databaseontwerpers, database-experts, consultants, technology planners, projectleiders en systeemanalisten.





## Programma-overzicht

### 1. De komst van Big Data en Fast Data

- Nieuwe BI-vormen, waaronder data science, investigative analytics en streaming analytics
- Het inzetten van Big Data om een concurrentievoordeel te krijgen
- Verschillen tussen semi-structured, poly-structured, multi-structured en unstructured data
- Voorbeelden van Big Data: sensordata, (micro-) eventdata en clickstream data
- Fast Data = Big Data + fast analytics + fast reactions
- Het belang van meer schaalbaarheid en query performance

### 2. De wereld van Hadoop, NoSQL en Spark toegelicht

- De Hadoop stack: HDFS, MapReduce, Hive, Spark, HBase, YARN, ZooKeeper, Pig, HCatalog, enzovoorts
- Alternatieve implementaties onder andere van MapR, Amazon (Hadoop as a service) en ScaleOut (Hadoop in-memory)
- MapReduce of Spark voor analytics en reporting?
- Classificatie van NoSQL producten: key-value stores, document stores, column-family stores en graph data stores
- Marktoverzicht, inclusief Apache HBase, Cassandra, CouchDB, Cloudera, DataStax, MongoDB, Neo4j en Riak

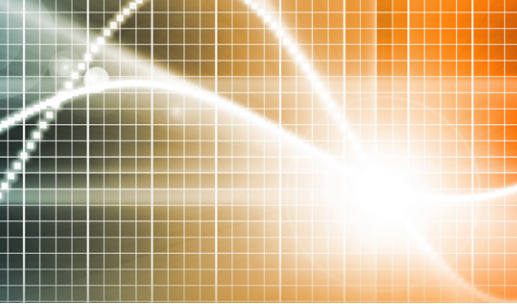
### 3. Big SQL oplossingen: SQL-on-Hadoop, NewSQL en analytical SQL databaseservers

- Hoe volwassen zijn de huidige SQL-on-Hadoop oplossingen?
- Marktoverzicht van SQL-on-Hadoop engines, inclusief Apache Drill, Apache Hive, Apache Phoenix, Cloudera Impala, HP Vertica, JethroData, Spark SQL en Splice Machine
- Classificatie van analytical SQL databaseservers
- De voor- en nadelen van kolomgebaseerde opslagstructuren

- Wat is in-database analytics en wat is de relatie met Google's MapReduce?
- Marktoverzicht, inclusief Apache Greenplum, Exasol, HP Vertica, IBM PureData Systems for Analytics, InfoBright, JustOneDB, Kognitio WX2, Microsoft PDW, Oracle In-Memory, SAP HANA en Sybase IQ, SnowflakeDB, Teradata Appliances en Teradata Aster Database
- NewSQL betekent high-performance transactie-georiënteerde SQL systemen
- Simpelere transactiemechanismen om scale-out te realiseren
- Het voorkomen van multi-table joins die niet geparalleliseerd kunnen worden
- Marktoverzicht inclusief Akiban, Clustrix, GenieDB, NuoDB en VoltDB

### 4. Technologieën voor Fast Data en streaming analytics

- De grootste toepassing van Fast Data: het Internet of Things (IoT)
- IoT betekent streaming data en snel analyseren van die data
- IoT devices: Smartphones (horloges), RFID sensoren, machines, algemene sensoren, cameras, pace makers, enzovoorts
- De uitdaging: real-time reacties op streaming data
- De verschillen tussen Big Data en fast Big Data
- Technologieën voor streaming data: Apache Kafka, Apache ActiveMQ, Amazon Kinesis, Kestrel, RabbitMQ en ZeroMQ
- Verschillen tussen de nieuwe technologieën en traditionele message queuing producten
- Voorbeelden van producten voor Big Data streaming: Apache Storm en Flink, IBM InfoSphere Streams, Informatica for Streaming Analytics, Software AG Apama en Spark Streaming
- Hoe integreren we Fast Data met het enterprise data warehouse?



### 5. Integreren van Big Data technologie in datawarehouse systemen

- Gebruiken van Hadoop en Spark in datawarehouse systemen
- Hadoop als staging area, datawarehouse, data mart of archiveringsgebied
- Datawarehouse gegevens beschikbaar stellen voor rapportage en analyse met SQL-on-Hadoop engines
- Gebruik van datavirtualisatie om gegevens te integreren die opgeslagen zijn in Hadoop, NoSQL- en SQL-systemen
- De mogelijke rol van NoSQL

### 6. Ontwikkelen van data lakes met Big Data technologie

- Wat is een data lake?
- Welke technologieën zijn geschikt voor het ontwikkelen van data lakes?
- Is het realistisch om een fysiek data lake op te bouwen met Big Data?

- Het ontwikkelen van een virtueel of logisch data lake met datavirtualisatie-servers
- Let op de technische en business metadata!

### 7. Data Science, Big Data Technologie en het datawarehouse

- Wat is data science en waarom is het anders dan analytics?
- Wat hebben MapReduce en Spark data scientists te bieden?
- Kunnen we standaard BI tools, zoals QlikView en Tableau, gebruiken in samenwerking met Spark?
- Hadoop als sandbox voor geavanceerde vormen van analytics
- De waarde van graph databases voor data science, zoals AllegroGraph, InfiniteGraph en Neo4J

### 8. Samenvatting en afsluiting

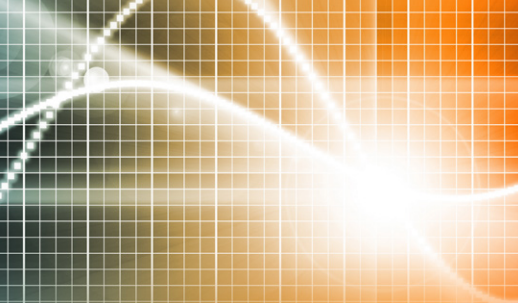


## RICK F. VAN DER LANS

**RICK F. VAN DER LANS** is onafhankelijk adviseur, docent en auteur op het terrein van datawarehousing, business intelligence, applicatie-integratie en databasetechnologie. Als consultant heeft hij door de jaren heen veel grote bedrijven (nationaal en internationaal) geadviseerd bij het ontwerpen van hun datawarehouse-architectuur en bij het maken van een toolkeuze.

Rick heeft als spreker op conferenties een zeer goede naam verworven zowel in binnen- als buitenland. Hij weet als geen ander een goede balans te vinden tussen op de praktijk toegesneden technologische en strategische zaken.

Tevens verschenen van zijn hand diverse artikelen in toonaangevende vakbladen. Verschillende van zijn boeken, waaronder het populaire SQL Leerboek, zijn in vele talen gepubliceerd. In 2012 is zijn nieuwe boek verschenen "Data Virtualisatie voor Business Intelligence Systemen".



## Informatie

### DATUM EN TIJD

Het seminar vindt plaats in het voorjaar en het najaar. De exacte data en aanvangstijden vindt u op onze website. Ochtend/middag seminars gaan om 9:30 uur van start en duren tot 17:00 uur terwijl middag/avond seminars starten om 13:30 en duren tot 21:00 uur. Registratie is mogelijk vanaf 1 uur voor aanvang van het seminar.

### PLAATS

Adept Events werkt met verscheidene accommodaties in de omgeving van Utrecht/Hilversum. Zodra de definitieve locatie bekend is, publiceren wij dit op de website. Controleer dit voor uw vertrek.

### AANMELDEN

Aanmelden kan via ons online inschrijfformulier op [www.adeptevents.nl](http://www.adeptevents.nl). Geef u de voorkeur aan schriftelijk aanmelden? Stuur de PDF van uw aanmelding of inkoopopdracht naar [seminars@adeptevents.nl](mailto:seminars@adeptevents.nl). Vermeld altijd duidelijk het e-mailadres van de deelnemer(s) alsmede dat van de crediteurenafdeling. Na ontvangst van uw aanmelding krijgt u de bevestiging en factuur per e-mail toegestuurd.

### KOSTEN

Deelname aan dit seminar kost € 625,50 per persoon indien u zich uiterlijk 30 dagen voor aanvang registreert en daarna € 695,- (exclusief BTW)\*. Documentatie, maaltijd en koffie zijn inbegrepen. Lidmaatschapskortingen kunnen niet worden gecombineerd. Werkt u bij een gemeente of provincie? Dan kunt u BTW terugvorderen via het BTW compensatiefonds.

### AANTREKKELIJKE KORTINGEN

Meldt u tegelijkertijd meerdere personen van één bedrijf aan voor hetzelfde evenement, dan geldt al vanaf de tweede deelnemer een korting van 10% per deelnemer. Vanaf vier deelnemers ontvangen alle deelnemers 15% korting (de deelnemers dienen op dezelfde factuur te staan)\*.

\* Prijzen of kortingen in deze pdf brochure kunnen (tijdelijk) afwijken van de website. In dat geval prevaleert altijd de informatie op de website.

### ANNULEREN

Annuleren dient schriftelijk te geschieden. U kunt annuleren tot drie weken voor het evenement plaatsvindt. Er wordt echter wel € 75,- (excl. BTW) administratiekosten in rekening gebracht. Annuleren is niet meer mogelijk vanaf drie weken voordat het evenement plaatsvindt. Vervanging door een ander dan de aangemelde persoon is te allen tijde mogelijk.

### MEER INFORMATIE



+31(0)172 742680



<http://www.adeptevents.nl/bdn>



[seminars@adeptevents.nl](mailto:seminars@adeptevents.nl)



@AdeptEventsNL / <https://twitter.com/AdeptEventsNL>



<http://www.linkedin.com/company/adept-events>



<https://www.facebook.com/AdeptEventsNL>



<https://google.com/+AdeptEventsNL>



Bezoek onze Business Intelligence en Datawarehousing website [www.biplatform.nl](http://www.biplatform.nl) en download de App



Bezoek onze website over Software Engineering in al zijn facetten: [www.release.nl](http://www.release.nl) en download de App

### IN-HOUSE SESSIES VOOR UW MEDEWERKERS

Wilt u dit seminar binnen uw organisatie aanbieden als in-house sessie voor een groep medewerkers? Neem dan contact op met ons secretariaat of stuur een e-mail via ons contactformulier. Op de [Klantenservicepagina](#) van onze site vindt u meer informatie over de mogelijkheden van In-house seminars en workshops.