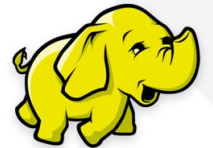


Vad döljer sig bakom begreppet Big Data?



Per Grosskopf
Affärsområdesansvarig
BI & Analys

Surveyföreningen Big Data
20151130



Vi har en stark digitaliseringstrend!



Digitaliseringen driver på tillväxten av data!



90% av all data har producerats senaste 2 åren

80% av all data som produceras är ostrukturerad

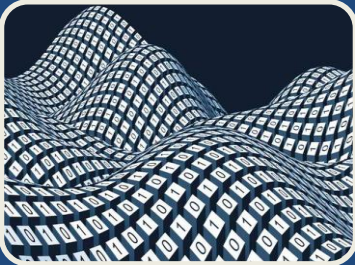
40 Zettabyte (10^{21}) data år 2020!
(50x ökning sedan 2010)

Denna trend skapar nya drivkrafter & utmaningar



Affärsperspektiv

- Nya affärsmöjligheter och affärsmodeller
- IT inte längre bara ett stöd utan en allt viktigare drivare av affären



IT-perspektiv

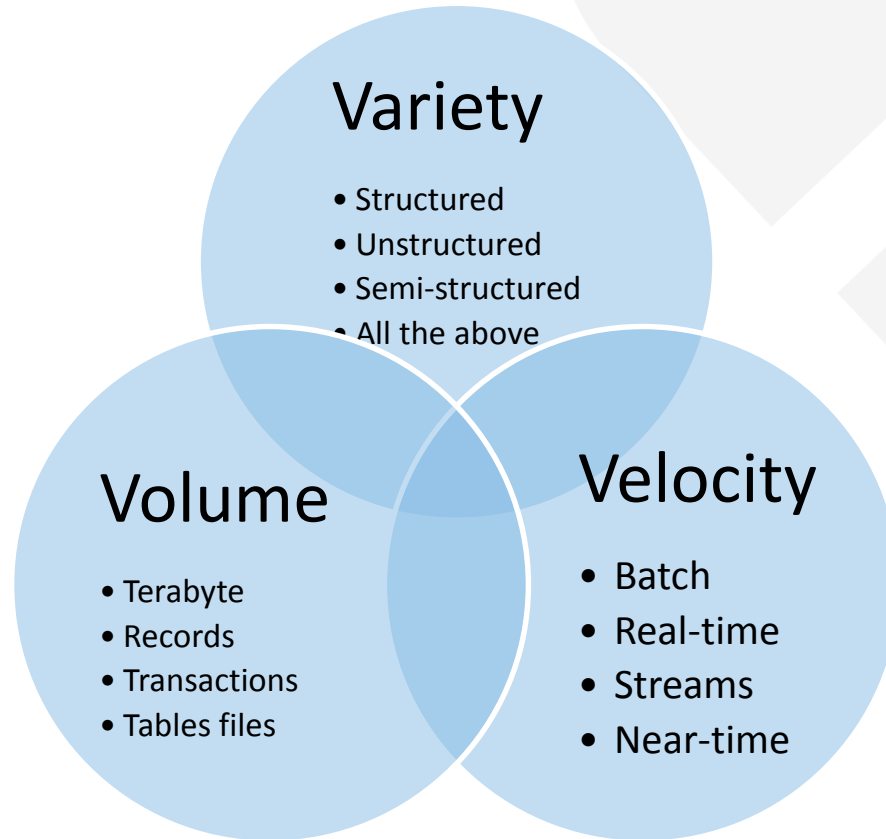
- Befintliga och framtida utmaningar kring data: **Volume, Variety & Velocity**



Kostnadsperspektiv

- Datalagringskostnad allt större del av budget
- Dyrt med licenser & hårdvara

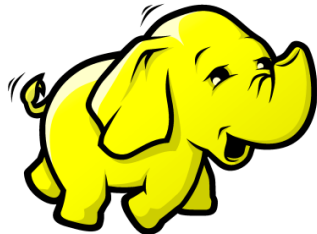
Tre delar av Big Data, 3V



Big data is *high-volume, high-velocity* and/or *high-variety* information assets that demand *cost-effective, innovative* forms of *information processing* that enable enhanced insight, decision making, and process automation.

Source: Gartner

Hadoop: Strategiskt viktigaste komponenten i en Big Data-arkitektur



Hadoop leder nu till ett tekniskt paradigmskifte!

Open source-ramverk för distribuerad datalagring & processing över flera standardmaskiner (Commodity hardware)



Hadoop erbjuder flera fördelar

Billigt

- Billig hårdvara & Licenser
- Avsevärt mycket billigare datalagring

Linjärt skalbar prestanda OCH kostnad

- Skalbart upp till "oändlighet"
- Inga trösklar

Billig datasäkerhet

- Automatisk backup av data
- Automatisk hantering av systemtillgänglighet

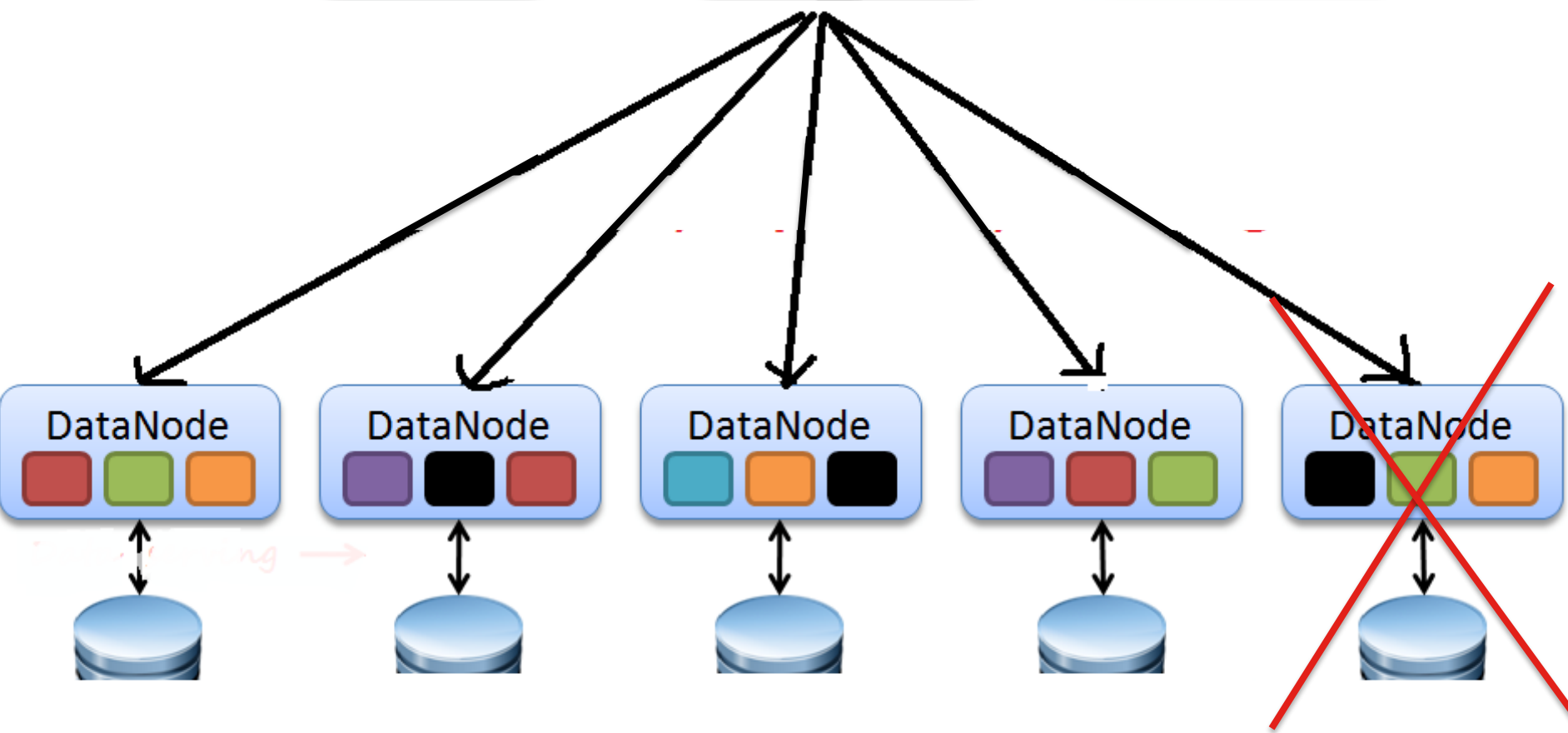
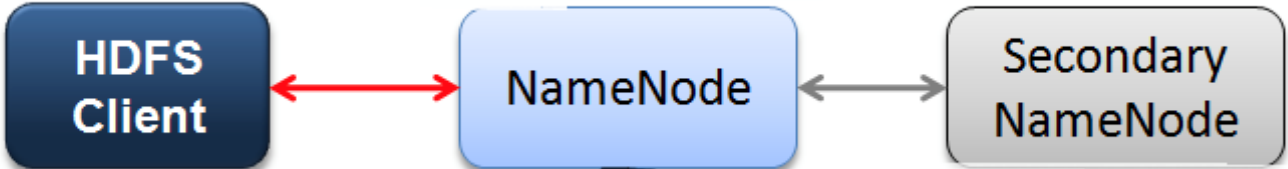
Bred funktionalitet

- Hanterar nya datatyper (flexibel datalagring)
- Stödjer real-tidsanalys, interaktiv analys etc
- Extrem utvecklingstakt!

Data lagras i HDFS

- Hadoop Distributed File System HDFS
- Datalagring & klusterhantering (lägga till / ta bort noder)
- Byggt för oplålitlig hårdvara
 - Feltolerans
- Flexibel datalagring, data lagras som vanliga råa filer, Schema on read

Federation





Machine Learning

Mahout
Machine Learning

Distributed Programming

Pig
Scripting

Hive
SQL Query

Spark

Scheduling

Oozie
Workflow

Service Programming



Zookeeper
Coordination

Data Ingestion

Sqoop



Storm

MapReduce Framework-YARN



NoSQL Database

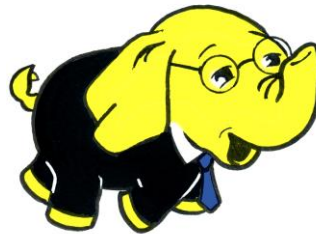


Distributed Filesystem



Hadoop core

Varför är Hadoop intressant för en affärsverksamhet?



Hadoop bemöter “Big Data”s tekniska utmaningar på ett kostnadseffektivt och flexibelt sätt!!

Technical Drivers

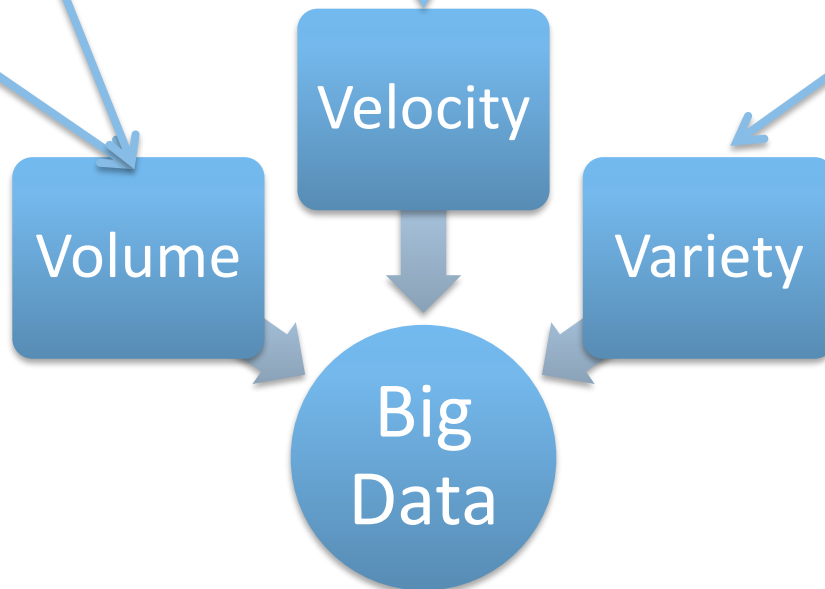
Lower data storage cost

Linear performance & cost scalability

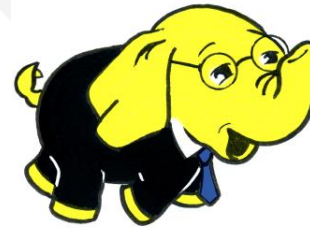
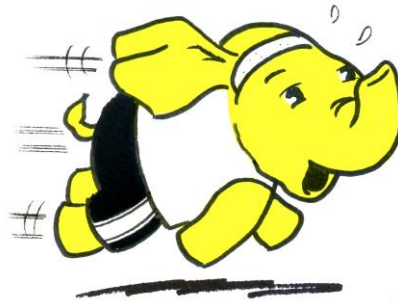
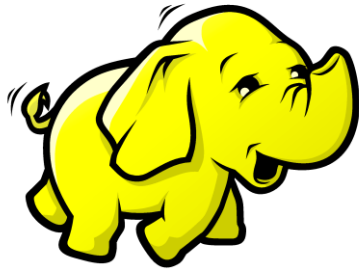
Real-time, interactive, batch processing capabilities

Automatic data backup

Flexibility, possibility of no schema



Hadoop är nu moget för kommersiellt storskaliga lösningar!



Tidigare

Fokus på BATCH

Mer java, MapReduce

Ej moget

First mover disadvantage



Nu

Real-Time, Interactive, Streaming & Batch

Utvecklat SQL-gränssnitt

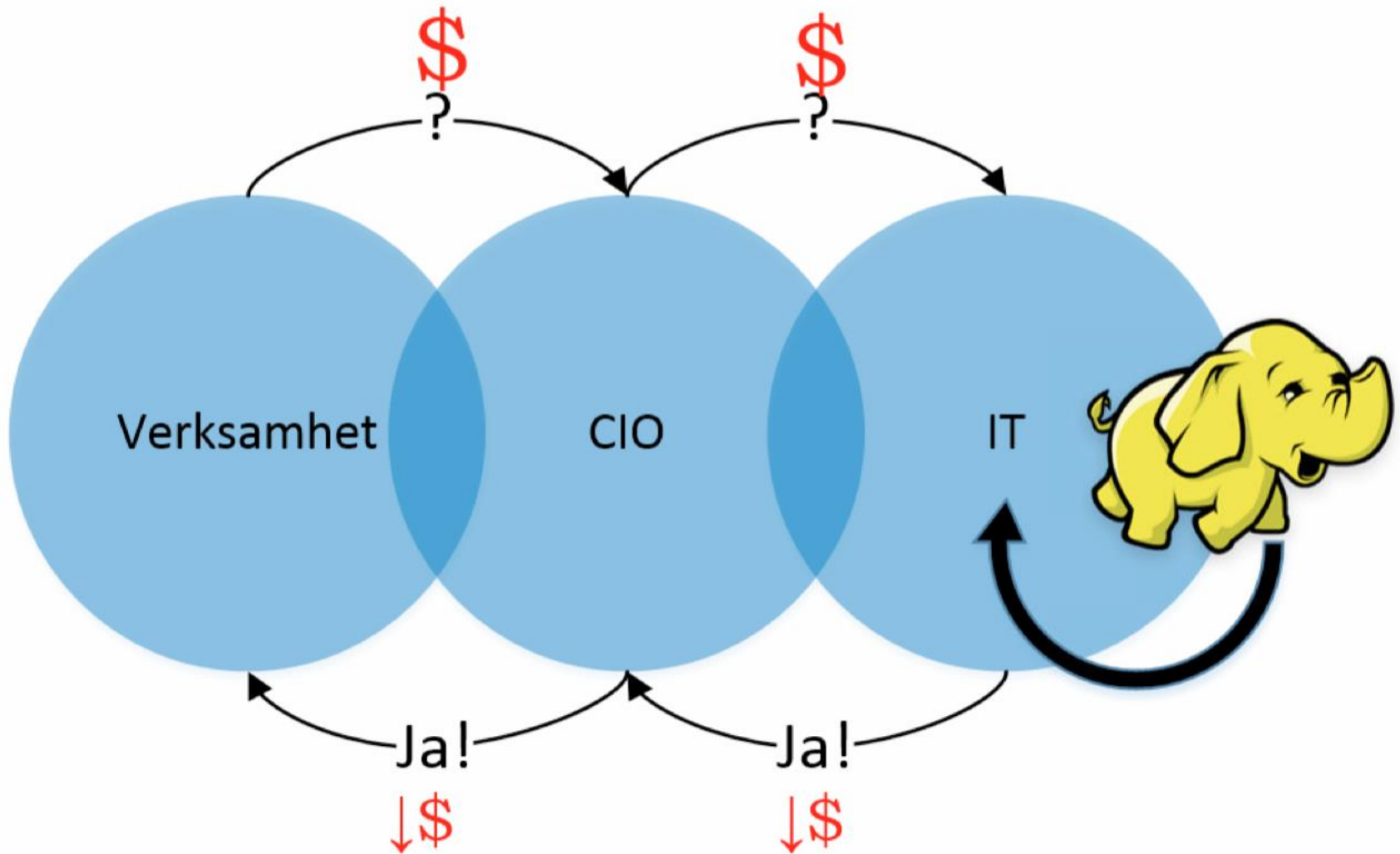
Tekniken **ÄR** mogen

Alla BI-leverantörer har integrerat med Hadoop

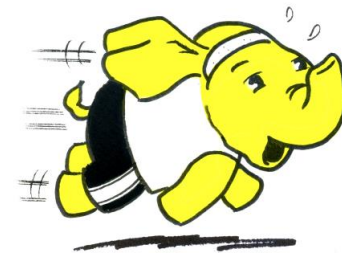
Flera kommersiella Hadoop-paketeringar

Mycket kapital investeras i Hadoop & distributörer

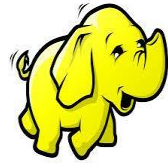
Egentligen inget nytt! Men **sänkt kostnadsbarriär** medför **fler lönsamma Business Case!**



Hur kan man införa Hadoop i sin organisation?



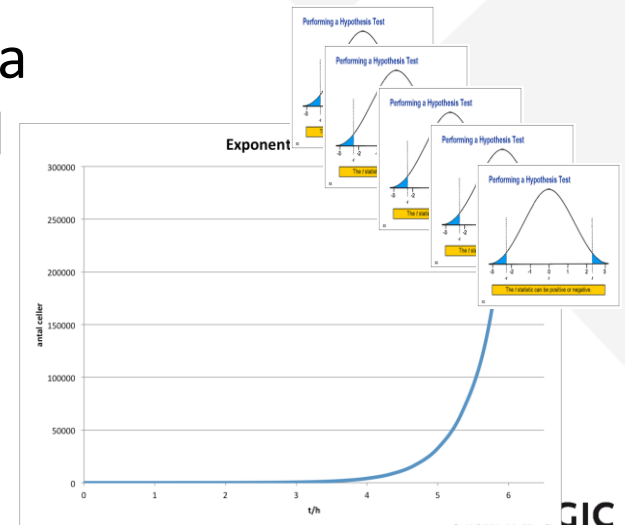
Hur börjar man med



?

-Börja smått, testa och labba

- Testa!
- I en digitaliserad värld är analys viktigare än någonsin
- Hadoop är den teknik som möjliggör det testdrivna företaget
 - Att på ett kostnadseffektivt sätt arbeta med - Variation, Återkoppling och Val
- Tekniken måste hanteras med stöd av:
 - Organisation
 - Kompetens
 - Arkitektur



Digitalisering skapar disruptiva affärsmöjligheter

In 2015 Uber, the world's largest taxi company owns no vehicles, Facebook the world's most popular media owner creates no content, Alibaba, the most valuable retailer has no inventory and Airbnb the world's largest accommodation provider owns no real estate.

Hur kommer surveyområdet att påverkas?

- Omvärldsbevakning
- Marknadsundersökningar
- Webbpanel
- Enkätundersökningar
- Metodik
- Aktörer



FERROLOGIC

Data Science Sweden@LinkedIn



Per Grosskopf

Affärsområdesansvarig

BI&Analys

Kontakt:

per.grosskopf@ferrologic.se

LinkedIn

www.ferrologic.se