



Vertrouwt u uw
gegevens?

BI wordt volwassen

Kasper de Graaf

31 maart 2009

De kracht van BI en Architectuur in de praktijk - Centraal Boekhuis

- **Business Intelligence:** Het proces dat zorgdraagt voor het transformeren van gegevens tot informatie en kennis en derhalve ondersteuning biedt aan effectieve besluitvorming van het management.
- **Data Warehousing:** Het proces van integratie, het toevoegen van een historisch perspectief en herstructurering van gegevens om het business intelligence proces in een organisatie optimaal te ondersteunen.





Volwassenheid ...

... een aantal waarnemingen

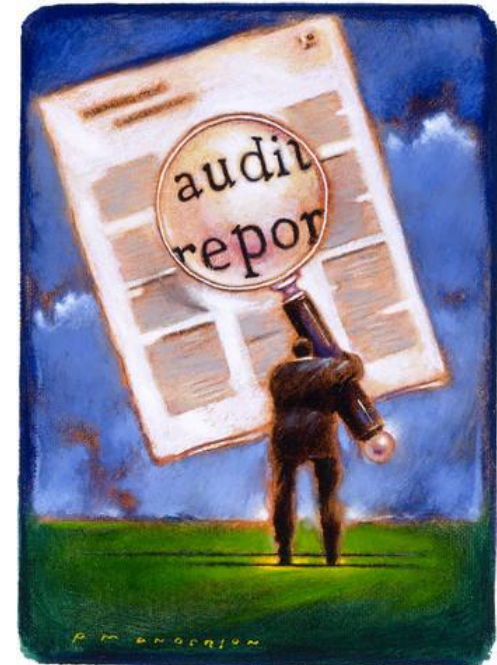
- Veel organisaties worstelen met hun BI initiatieven op o.a. de volgende vlakken:

- Gegevenskwaliteit
- Compliance
- Traceerbaarheid (audits)
- Schaalbaarheid
- Duurzaamheid / beheersbaarheid



- De bestaande methodieken en technologieën voor data warehousing zijn ruim 15 jaar geleden ontwikkeld, zonder veel aandacht voor compliance of traceerbaarheid
- De business heeft zich de laatste jaren sterk ontwikkeld op deze vlakken
- Veel van de bestaande BI-oplossingen zullen een audit op traceerbaarheid en / of compliance momenteel niet doorstaan

- Hoe nerveus denkt u dat een ETL- of rapport-ontwikkelaar wordt wanneer daadwerkelijk mensen ontslagen worden op basis van hun rapportages?
- Bovenstaande waarnemingen brengen grote risico's mee voor organisaties en hun leiders





Volwassenheid ...

... stelt aanvullende eisen en verantwoordelijkheden

Wat vraagt de business van DW/BI?

- *ECHTE* integratie
- Historisch correct (en herhaalbaar)
- Met acceptabele responsetijden
- Eén versie van de waarheid
- Compliant en traceerbaar
- Een 'lenige' IT-afdeling
- Hoge kwaliteit
- Voldoende begrip van de gegevens



■ Integratie

- Integreren betekent dat alle gegevens uit de verschillende bronnen bij elkaar gevoegd dienen te worden om een volledig beeld te kunnen presenteren (zelfs als bronnen elkaar tegenspreken ...)

■ Historie

- Rapportage X over periode Y op moment Z



■ Dimensioneel modelleren

- Het sterschema is nog steeds de beste techniek om 'per-vragen' met acceptabele performance te beantwoorden
 - Omzet in EUR, per week, per maand, per business unit, per klantsoort, per product, per regio, ...

■ Aggregaten

- De meest effectieve manier om de performance van een BI-oplossing te verbeteren
- Gebruik het snowflake schema om feitentabellen van verschillende granulariteit te koppelen

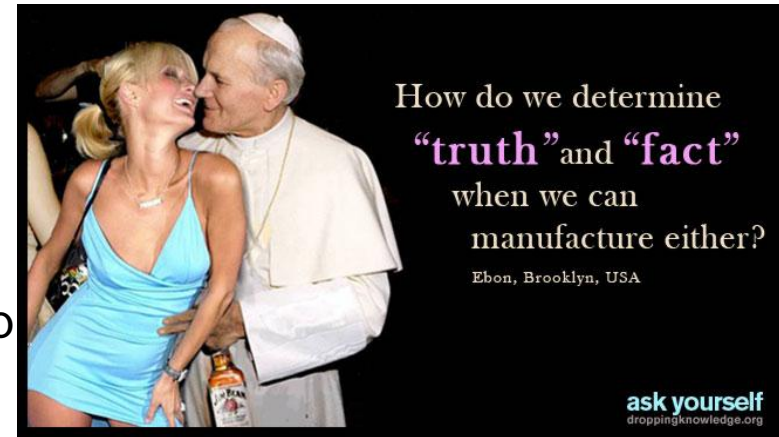
■ Database tuning

- Databases zijn out-of-the-box getuned voor transactieverwerking
- Uw DBA is goud waard!



Eén versie van de waarheid?

- Waarheid is subjectief (wiens waarheid?)
- Waarheid is aan verandering onderhevig
 - Tijdsafhankelijke waarheid?
 - Kunnen we historische gegevens testen op waarheid?
- Waarheid is lastig te definiëren
 - Wat doe je als twee bronstemen elkaar tegenspreken?
- Gezien de eigenschappen van waarheid moeten we de mogelijkheid creëren om de waarheid van zowel gisteren, als vandaag als morgen opnieuw te definiëren. Dit betekent dat de de **feiten** moeten opslaan.



■ Compliance


- Het begrip waarmee wordt aangeduid dat een organisatie werkt in overeenstemming met vigerende wet- en regelgeving (nl.wikipedia.org; 23 maart 2009 11:03 am)
- ISO, Sarbanes-Oxley act, BASEL II, Health Insurance Portability and Accountability Act (HIPAA), etc.

■ Het kunnen *bewijzen* van compliance heeft consequenties voor DW / BI oplossingen, denk aan:

- Interpretatie van gegevens
- Opschonen van gegevens
- Filteren van gegevens van mindere kwaliteit
- Toepassing van business rules



- Het laadproces (ETL) dient **compleet** (100% van de gegevens), **omkeerbaar** (geen onomkeerbare wijzigingen of toepassing van business rules) en **volledig traceerbaar** te zijn naar de bron

- BI requirements zijn zeer dynamisch
 - Vandaag rapporteren we over winst, morgen over kosten, volgende week zijn we geïnteresseerd in kosten en volgende maand willen we alles weten over concurrenten of klantaantallen
 - Bronsystemen wijzigen als gevolg van nieuwe requirements of bugs
 - Business rules wijzigen
 - Een gemiddelde data mart lijkt een levensduur te hebben van 18 maanden (Blog van Bill Inmon in september 2008)
- 
- IT of BI afdelingen moeten lenig genoeg zijn om de business bij te houden, gebrek aan lenigheid zal leiden tot “*the return of the evil spreadmart*” en uiteindelijk tot een gebrek aan vertrouwen in IT

- wikipedia.org (23 maart 2009, 11:17 am)
 - Data are of high quality "if they are fit for their intended uses in operations, decision making and planning" (J. M. Juran);
 - Alternatively, the data are deemed of high quality if they correctly represent the real-world construct to which they refer.

- DW / BI en de spagaat van gegevenskwaliteit
 - De gegevenskwaliteit in bronsystemen is veelal verre van voldoende
 - Eisen met betrekking tot compliance en traceerbaarheid maken het opschonen op zijn minst discutabel
 - De business verwacht gegevens van hoge kwaliteit



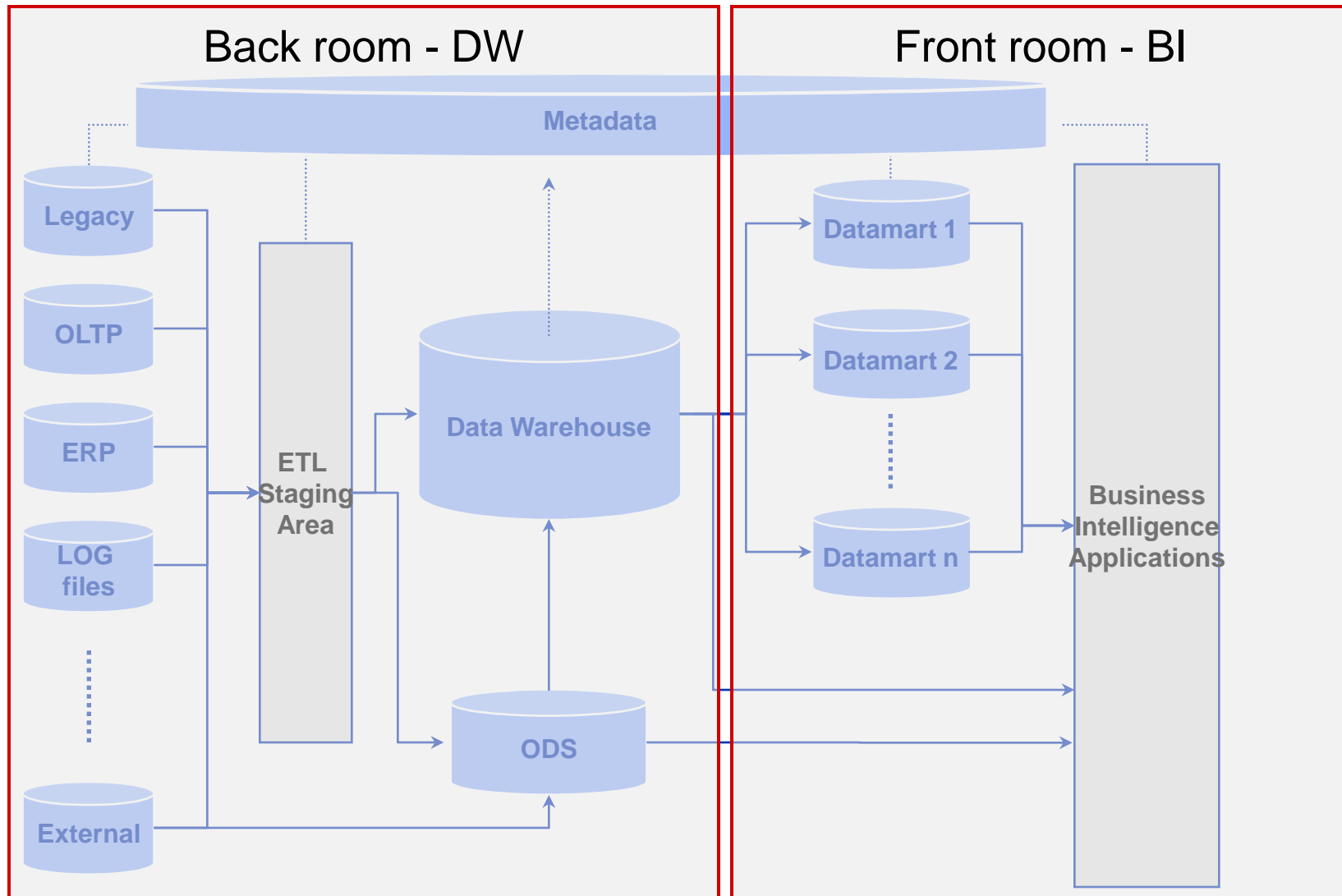
- Om gegevens daadwerkelijk zinvol te kunnen gebruiken (actionable knowledge) moet er voldoende begrip zijn van:
 - De betekenis van de gegevens (definitie)
 - De kwaliteit van de gegevens (incl. betrouwbaarheid)
 - De toegepaste transformaties en business rules





Consequenties voor de architectuur

Back room / Front room



■ Back room

- Integratie
- Historie volledig vastleggen
- Traceerbaarheid bewaken
- Toetsen van gegevenskwaliteit (kwaliteitsaspecten benoemen)

“De **feiten** van gisteren en vandaag, zoals weerspiegeld in de bronnen”

■ Front room

- Consistente reportage
- Toegespitst op het beoogde gebruik
- Opschonen van gegevens (specificatie door de business)
- Toepassing van business rules (specificatie door de business)

“De **waarheid** van vandaag”

- BI wordt volwassen, maar we zijn er nog niet
- Om te kunnen voldoen aan de eisen van vandaag en morgen op het vlak van compliance en traceerbaarheid moet de architectuur van de back room en de front room herzien worden
- Gegevenskwaliteit en meta data management zijn onderbelicht en verdienen onze (hernieuwde) aandacht