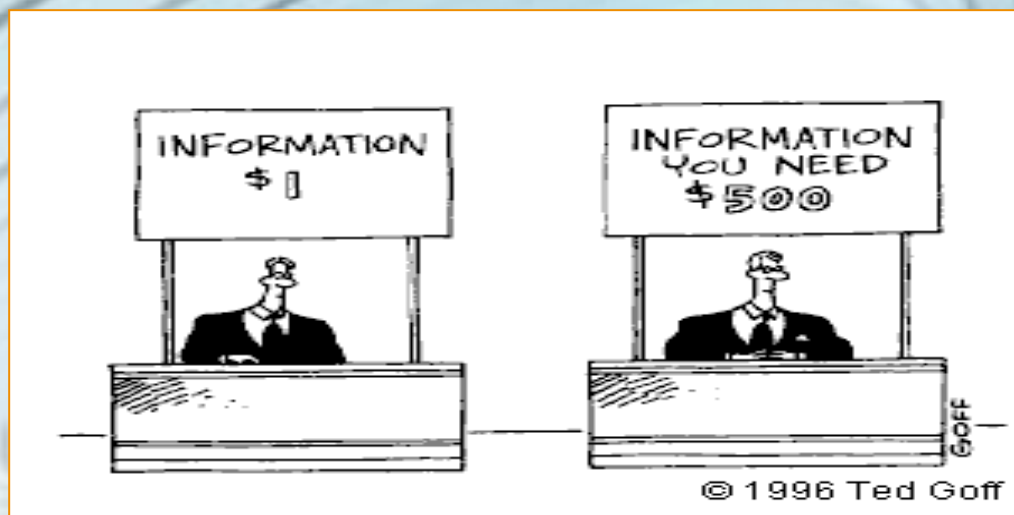


Ad Hoc. Hvordan forberede for det utforutsette?



```

SELECT 'Trade' Type , null Leg_id, Tradeld, drywet, Llpno, Trade_date, From_place_id, To_place_id,
to_number(Via) Via, DWT, Percent, From_id, To_id, To_number(Trading.TRAD.f_Get_Distance( From_id,
To_id, Via, drywet )) Distance, COMMODITY_ID, null as LegType FROM ( SELECT T.TRADE_ID Tradeld,
case when T.Commodity_id = 'A' then 'W' else 'D' end drywet, v.Llpno Llpno, T.Trade_date, T.From_place_id
From_place_id, T.To_place_id To_place_id, TV.Place_id Via, T.Dwt, T.Percent,
To_number(Trading.TRAD.f_Get_area_id(1, T.From_place_id)) From_id,
To_Number(Trading.TRAD.f_Get_area_id(1, T.To_place_id)) To_id, T.COMMODITY_ID as Commodity_id
FROM Trading.v_Ship V, Trading.Trade T, Trading.Trade_via TV WHERE NVL(T.Coasting_code,'') <> 'S'
AND V.Llpno = T.Llpno AND T.Trade_id = TV.Trade_id(+) AND V.LLPNO = '7801673' )
union
select Type, Leg_id, Tradeld, drywet, Llpno, Trade_date, from_place_id, to_place_id, via, dwt, Percent,
From_id, To_id, nvl(to_number(Trading.TRAD.f_Get_Distance( From_id, To_id, Via, drywet )),Distance2)
Distance, Commodity_id, LegType from ( SELECT 'Leg' Type, Leg_id, null Tradeld, case when
T.Commodity_id is null then null when T.Commodity_id = 'A' then 'W' else 'D' end drywet, l.Llpno Llpno,
nvl(T.TRADE_DATE,null) Trade_date, L.from_place_id, L.to_place_id,
To_number(Trading.Trad.f_Get_via_from_leg(leg_id)) Via, V.DWT, nvl(T.PERCENT,0) Percent,
To_number(Trading.TRAD.f_Get_area_id(1, L.From_place_id)) From_id,
To_number(Trading.TRAD.f_Get_area_id(1, L.To_place_id)) To_id, D.Distance Distance2, T.COMMODITY_ID
as Commodity_id, LEG_TYPE as LegType --From_observation_id, Leg_type, L.From_place_date,
--P1.Place_name, L.To_place_date, To_observation_id, P2.Place_name,
--Trad.f_Get_via_from_leg(leg_id), O.Estimated_values, O2.Estimated_values
FROM Trading.LEG L, TRADING.TRADE T, Trading.v_Ship V, TRADING.DISTANCE D --, Trading.PLACE P1,
Trading.PLACE P2, Trading.OBSERVATION O, Trading.OBSERVATION O2
WHERE L.From_observation_id = T.From_Observation_id(+) AND L.To_observation_id =
T.To_Observation_id(+) AND V.Llpno = L.Llpno AND L.LLPNO = 7801673 AND L.FROM_PLACE_ID =
D.FROM_PLACE_ID(+) AND L.TO_PLACE_ID = D.TO_PLACE_ID(+) --AND L.From_place_id = P1.Place_id
--AND L.To_place_id = P2.Place_id
)

```



”Å tillate at sluttbrukere kan gjøre Ad-Hoc spøringer på et datavarehus er som å gi en kasse håndgranater til en stamme apekatter”



Hva mener vi med ad-hoc?

- Gammel definisjon;
 - Folk får lov til å kjøre SQL eller BO Querypanel etter egen lyst
 - Ingen krav til filtere, ingen krav til Join, ingen restriksjon på sortering

Så da får vi kartesiske produkter som skal grupperes på ikke-aggregerte felt og deretter sorteres synkende på uindekserte kolonner ...
- Oppdatert, og mye vanskeligere definisjon;
 - Et uforutsett behov for å sammenstille tall på måter vi ikke har tilrettelagt for i varehuset
 - Det kan innebære at tallene ikke er i varehuset overhode
 - Eller at de heller ikke ligger i noe fagsystem vi råder over

God gammel Ad Hoc

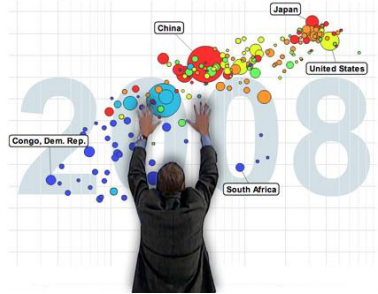
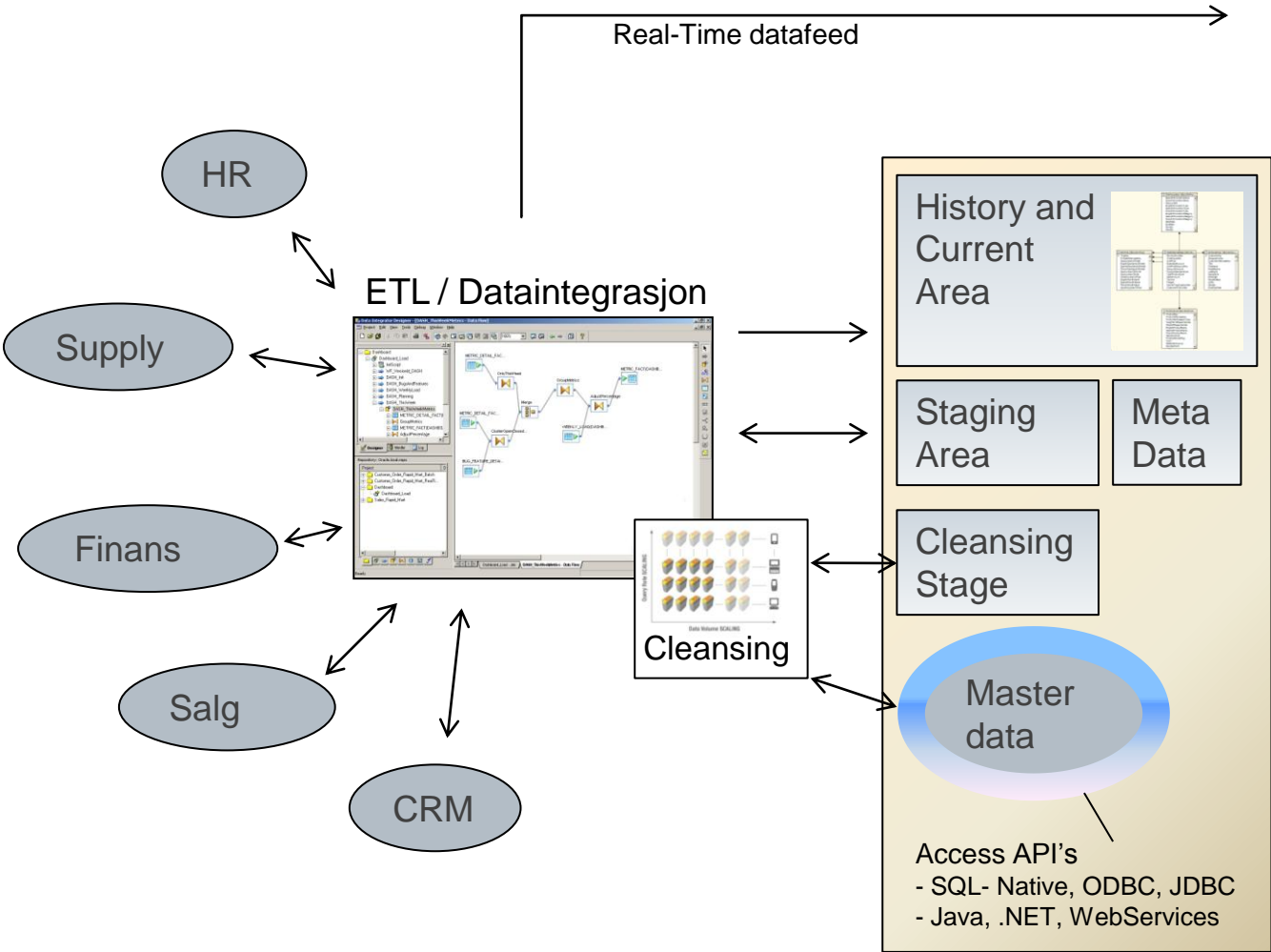
- Det viktigste er å bygge gode indekser i databasen
- Gode indekser betyr å bygge index på mange nok enkeltfelt
- Og det betyr å lage composit-index på felt som naturlig kombineres i spørringer
- Husk at databasens optimizer må velge seg én av disse pr join-operasjon mot hver tabell. Om én tabell opptrer flere ganger i queriet kan optimizeren velge ulik index pr opptreden
- "Update statistics"
- For å hjelpe brukerne tilbyr man gjerne views der joiner er forhåndstestet, slik at ad-hoc mest går mot slike views
- I verktøy som Sybase IQ og Vertica kan man gjøre 'pre-join' av tabeller som 'alltid' skal joines, slik at queriet går fort i 'run-time'

Om Ad-hoc

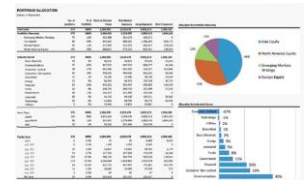
- Drømmescenario
- Et mer reelt scenario
- Et guffent scenario

- Hvordan forberede for det uforutsette?

Ad-Hoc analyse mot et perfekt datavarehus



Analyse
Rapportering
Perf Mgmt



Nirvana

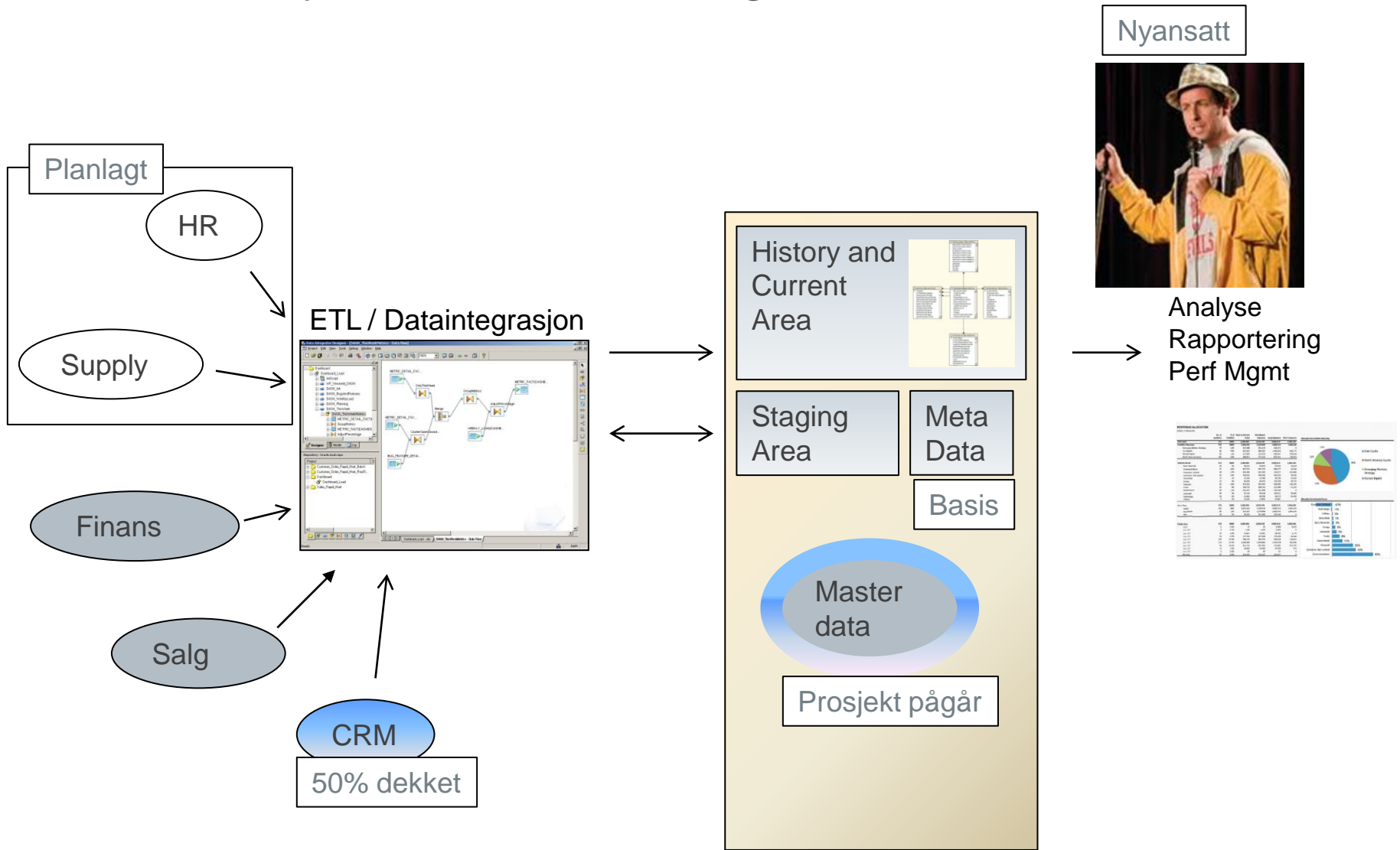
- Alle kildesystemer er komplett dekket, konsistente og up-to-date
- Alle data som kopler har fått på seg veldokumenterte nøkler med raske datatyper (int)
- Metadata inneholder alle tekniske attributter og bruker-relevant informasjon (nytte, kontekst, feller man må unngå)
- Alle kodeverk og støttehierarkier er entydige og komplette (MDM)
- Alle analyseverdige data ligger i stjerneskjema, på detaljnivå (keep it to the grain), men også med forretningsdrevne aggregater
- Vi har oppdaterte og veldesignede univers. Ett pr forretningsområde. Og ett for 'totalen'
- Våre 'brukere' er utdannet på produktene og kjenner egne data godt

Juhuu 😊.

Utfordringer

- Ytelse / Ventetid hvis man bruker klassisk database
- Kan fikses ved å bytte til en kolonneorientert DB (Sybase IQ, Vertica)
- ??

Ad-Hoc analyse mot et mer vanlig datavarehus



Hvilke krav stiller det?

- Rapporteringsverktøyet må kunne lese data fra flere kilder og sammenstille i én rapport
 - Datavarehuset + fagsystemer
 - Excelmodeller og annet rask som slenger rundt
- Det tar mer tid, blir mer manuelt og feil oppstår med større sannsynlighet
- Det settes store krav til kunnskap

Utfordringer

- Ting koples feil, aggregeres feil, tolkes feil -> Feil resultat
- Ingen tid til testing og avstemming? Rett i 'prod'?
- Avvikende funn – alle får 'hver sine tall'
- Kunnskapen strekker ikke til så utviklere må inn og hjelpe til
- Nå skjer kanskje sammenstillingene i klientverktøyet.
Datavolum? Tid?

Forberede for Et guffent scenario™

- Når man avgjør hvilke data (tabeller/felt) i kilden som skal være med inn i DWH; Ta med 'alt'.
- Del evt. på det man mener er viktig (DWH) og det man tar med for forvaltning av 'corp IQ' / for å være forberedt (Arkiv)
- Ha klart et område i datavarehuset (tablespace el) for uforutsette behov
- Pass på DWH struktur (konforme dimensjoner, detaljnivå)
- Ikke ha Univers 'bare' mot DWH. Lag også Univers mot alle kildene samtidig som man jobber med kilden i ETL. Da koster det lite ekstra.
- God kunnskap om egne Kildesystem er viktig. Dette er ett argument mot 'lukkede' applikasjoner.

Forberede for Et guffent scenario™

- Bli kjent med kunnskapsrike nøkkelpersoner på forretnings siden. Gode kommunikasjonslinjer er nøkkelen til forståelse og presisjon
- Når man oppretter 'små-filer' (Excel ol) må man huske å lagre nøkler også, så informasjonen kan joines inn.
- Vær kreativ på mulige kilder utomhus (Webben? Partnere?)
- Det er mange scenarier man tenker at 'disse var sjokkerende', men det ville kanskje vært mulig å forutse
 - En flodbølge i Asia før eller senere? (Tsunamien)
 - eColi-utbruddet hos Gilde? Ende-til-ende sporing av mat er ikke nye tanker
 - Vulkanutbrudd i Europa? (Med Vesuv, Teide og Island.*)

Hva er en potensiell operasjonell utfordring for din bedrift? Er du klar?

"We are what we repeatedly do.

Excellence, then, is not an act,
but a habit."

- Aristotle