

# **DATA FEDERATION (GAGNASAMLAG)**

## **HVAÐ ER ÞAÐ OG HVERT ER NOTAGILDIÐ?**

Magnús Guðmundsson – SAS Institute



- Bakgrunnur
- Hvað er Data Federation?
- Hagnýting Data Federation

# BAKGRUNNUR



## AF HVERJU DATA FEDERATION?

- Hefur fengið litla athygli
- Vaxandi þörf og áhugi
- Margar leiðir til hagnýtingar
- Öflugri vélbúnaður eykur notagildi
- Mikilvægur partur af gagnapúslinu

## HVER ER HVATINN?



### Bæta öryggi við gagnaáðgengi

- Meðalkostnaður fyrrtækja vegna öryggisrofa = 3.5 mUSD  
- *Cost of Data Breach: Global Analysis*  
Ponemon Institute, May 2014



### Hraðari afhending gagna til notenda

- 7.9 vikur tekur að meðaltali að bæta nýjum gögnum (source) við vöruhús gagna
- Tekur 7 vikur að búa til flókið stjórnborð eða skýrslu  
- *2013 TDWI BI Benchmark Report*



### Geta nýtt öll gögn sem eru tiltæk

- "...Without data virtualization, you risk knowing less about your customers...fewer real-time business insights, lose your competitive advantage, and spend more to address data challenges..."  
– *Forester Information Fabric (2013)*

# HVAD ER DATA FEDERATION?



- **Data virtualization** is the process of offering data consumers a data access interface that hides the technical aspects of stored data, such as location, storage structure, API, access language, and storage technology.
- **Data federation** is a form of data virtualization where the data stored in a heterogeneous set of autonomous data stores is made accessible to data consumers as one integrated data store by using on-demand data integration.

Clearly Defining Data Virtualization, Data Federation, and Data Integration

Rick van der Lans

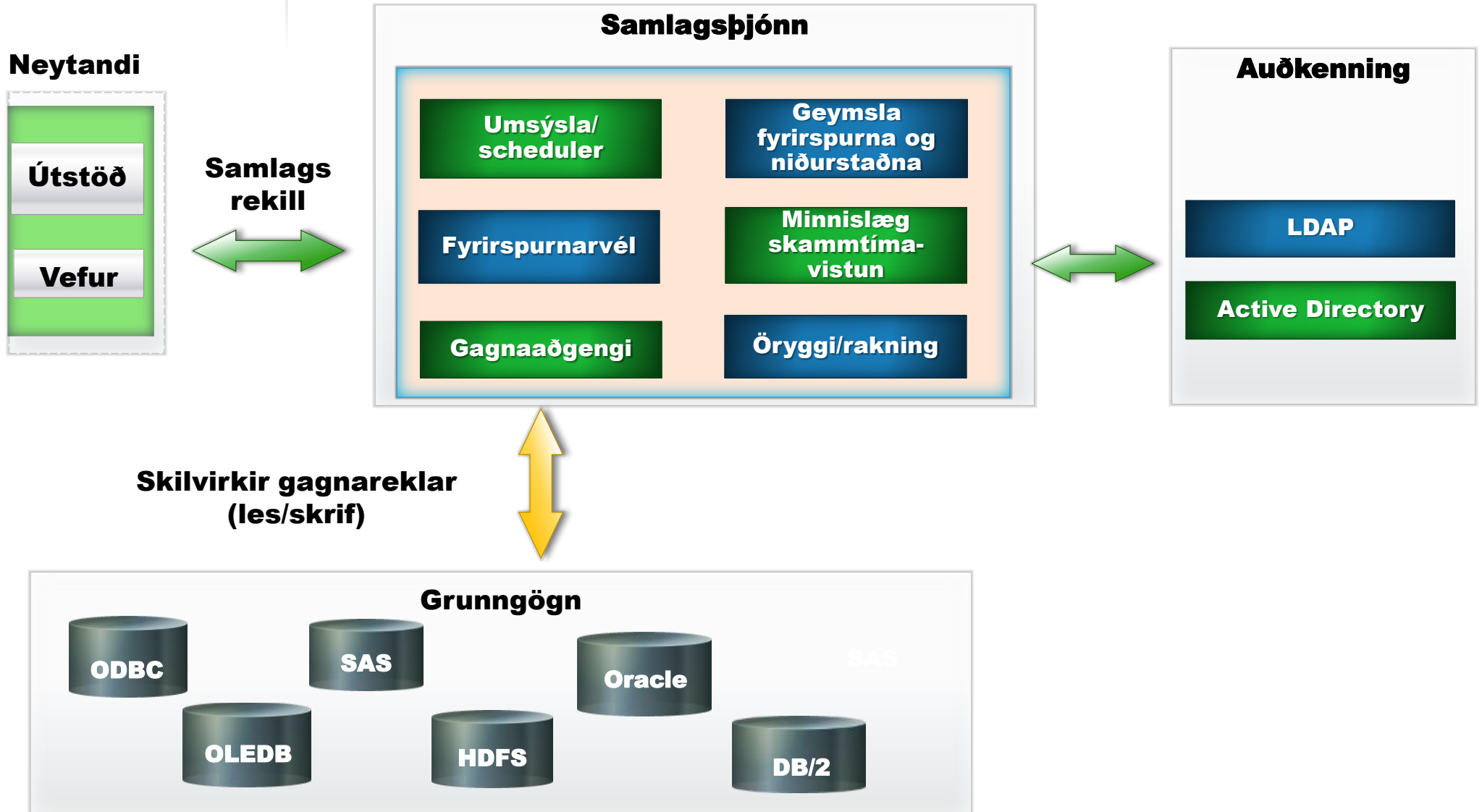
- Humm? Hvað var þetta í miðjunni aftur?



## EININGAR DATA FEDERATION LAUSNA

- Fyrirspurnarvél (Query Processor) er kjarninn
- Reklar til að tengjast undirliggjandi gögnum
- Rekill til að tengjast Data Federation fyrirspurnarvélinni
- Viðmót til umsýslu og skilgreininga á samlagsfyrirspurnum
  - Þ.m.t. skilgreining á abstrakt gagnaskilum
- Auðkenning og aðgangsstýringar
- Aðgerðaloggun til endurskoðunar/rekjanleika
- Skammtí mavistun á niðurstöðum á diskum og/eða í minni
- Scheduler til að keyra fyrirspurnir á tilteknum tímum
- Vefskil fyrir gagnaneytendur
- Birtingarstýring (disclosure control)

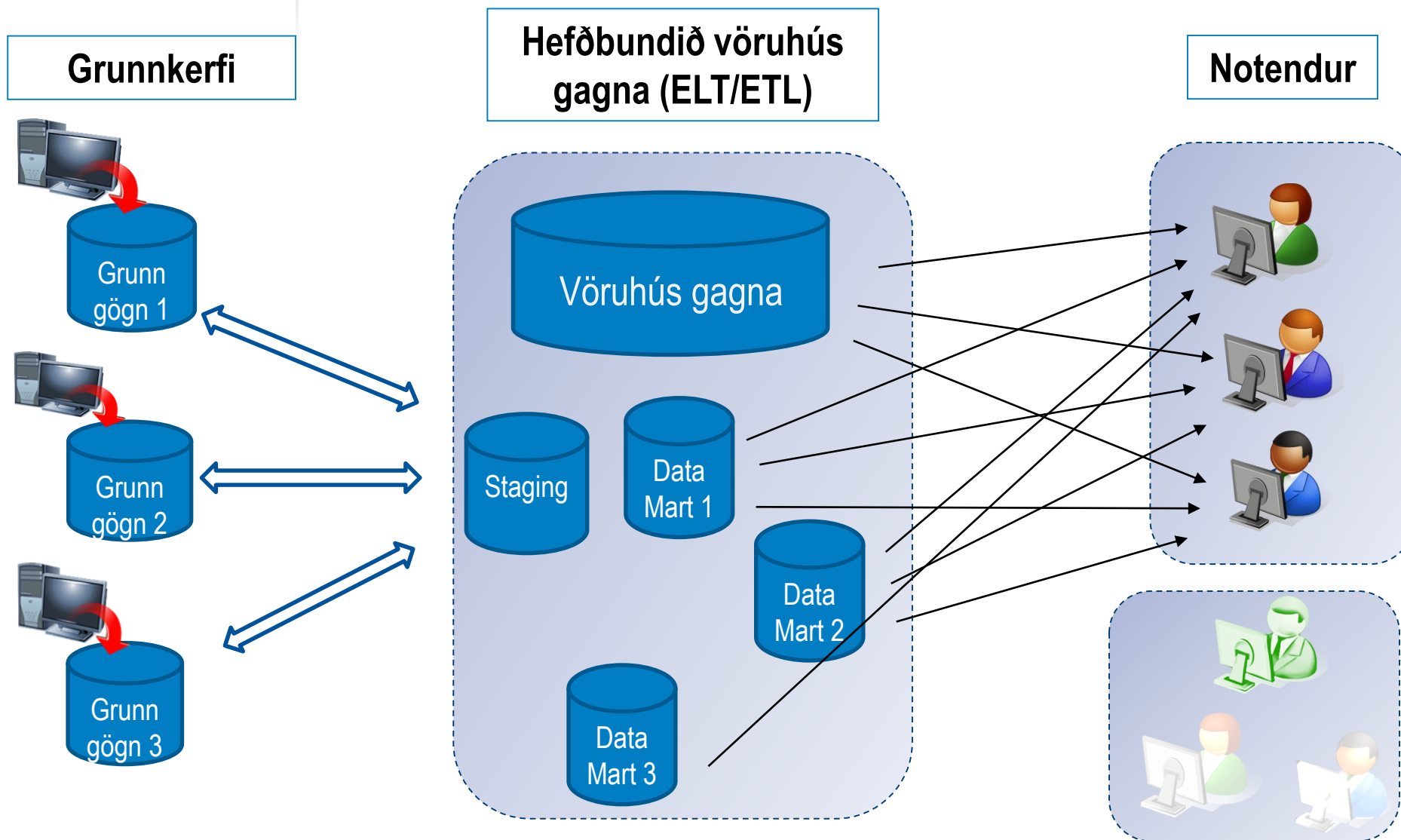
MYND SEGIR MEIRA EN ÞÚSUND ORÐ



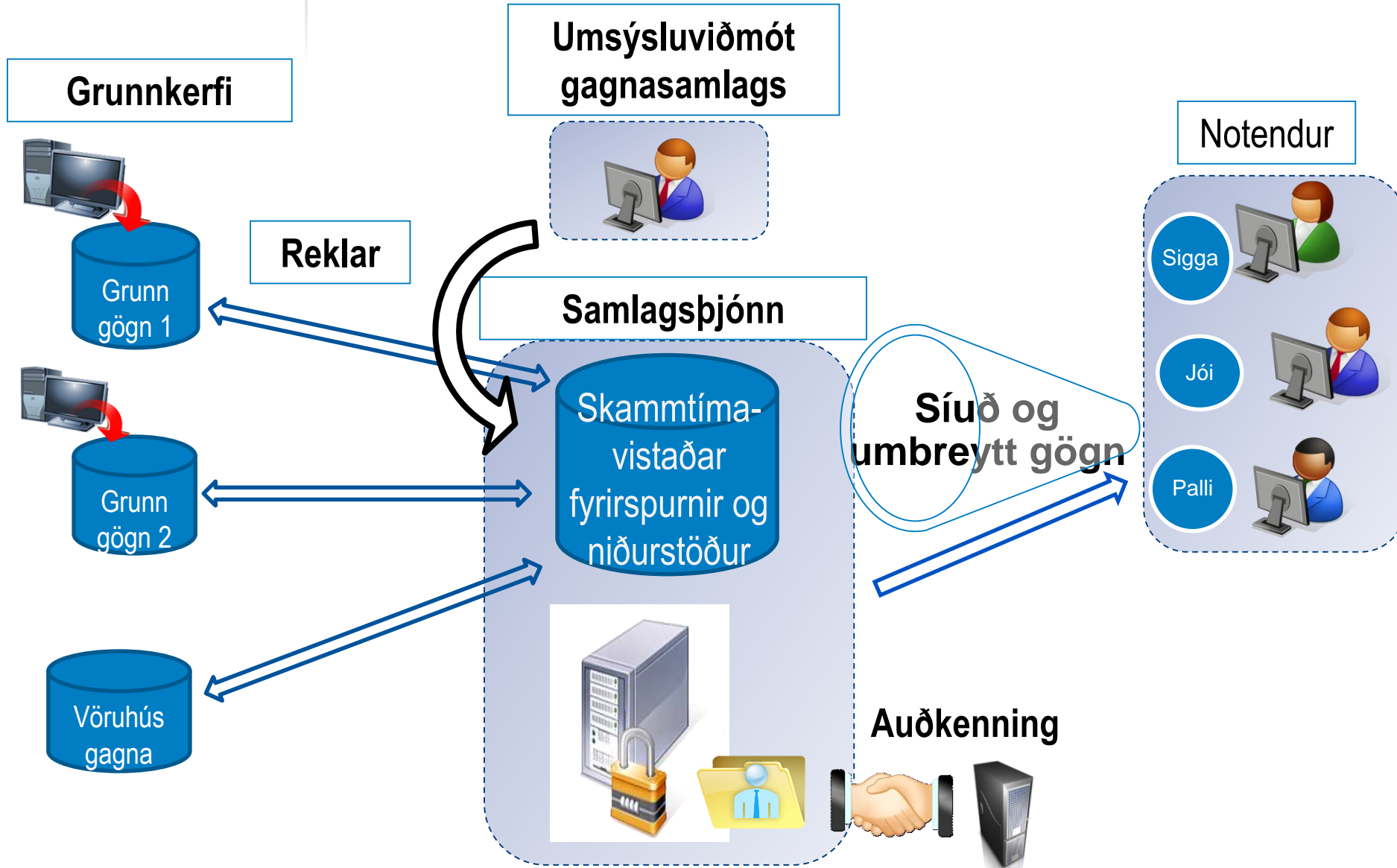
## ÁVINNINGUR AF NOTKUN GAGNASAMLAGSLAUSNA

- Miðlæg stýring á öllu gagnaaðgengi, óháð gagnagrunnum, öryggi og notkunarslóð
- Samandregin sýn á ólík gögn (samlagsfyrirspurnir/sýndargögn)
- Hægt að útfæra fyrirspurnir sem abstraction lag
- Notendur þurfa ekki að skilja samtengingar í gögnum
- Notendur hafa einn aðgangspunkt til að nálgast öll sín gögn
- Rauntímaaðgengi að gögnum með lágmarksálagi á grunnkerfi
- Minni fjölföldun á gögnum, lægra flækjustig en að viðhalda heilindum í vöruhúsi
- Hægt að hylja/dulkóða innihald dálka, því ekki nauðsynlegt að sía raðir frá

# GAGNASAMLAG EÐA HEFÐBUNÐNAR AÐFERÐIR



# GAGNASAMLAG EÐA HEFÐBUNDNAR AÐFERÐIR



## POLLÝANNA TEKUR OFAN GLERAUGUN

- Að sjálfsgöðu ekki besta aðferðafræðin í öllum tilfellum
- Vöruhús gagna nauðsynlegt fyrir dagleg „snapshot“ og gott vöruhús gagna er gulls ígildi
- Leysir ekki gæðavandamál í gögnum
- Ekki nægilegt eitt og sér fyrir t.d. Data Governance
- Afköst og álag á grunnkerfi geta orðið vandamál
- Ekki hægt (praktískt) að setja allar spurningar (sem undirliggjandi gagnagrunnur getur svarað) fram í SQL?
  - SQL er de facto mál fyrir samlagsfyrirspurnir

# HAGNÝTING GAGNASAMLAGSLAUSNA



## TEGUNDIR SÝNDARGAGNA VERKEFNA

Af hverju Data  
Federation/Data  
Virtualization?

### Lipurleiki

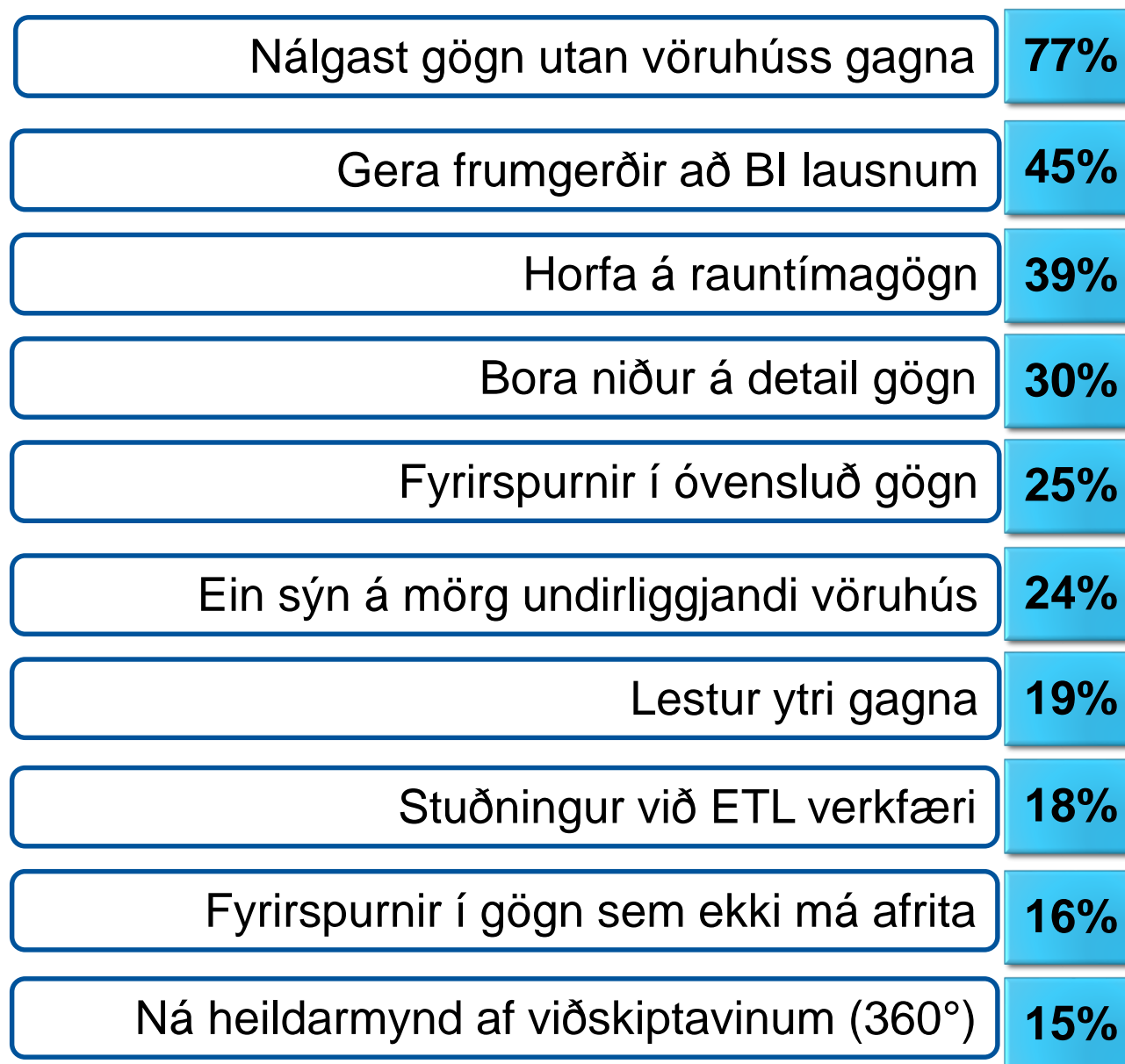
66% nefndu lipurleika  
sem ástæðu  
sýndargagna verkefna

*Data Virtualization: Perceptions  
and Market Trends*

– Wayne Eckerson

*TechTarget, April 2013*

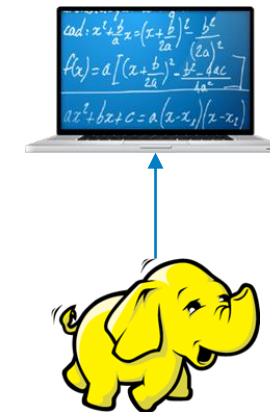
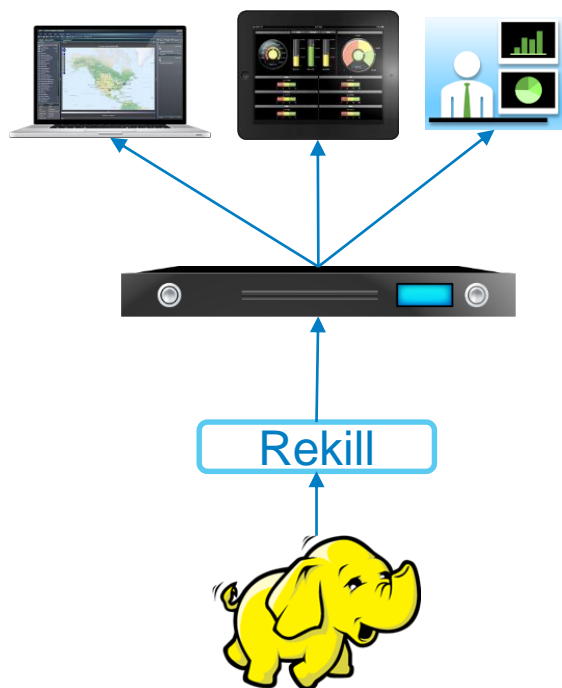
*Könnun meðal BI sérfræðinga*





## GAGNASAMLAG HENTAR ÞEGAR GÖGNIN ERU...

**Ómótuð:** Við viljum bæta nýtingu á gögnum sem eru ómótuð, t.d. Hadoop gögn sem eru t.d. ekki með þekktan töfluþrúktúr þegar þau eru skrifuð niður

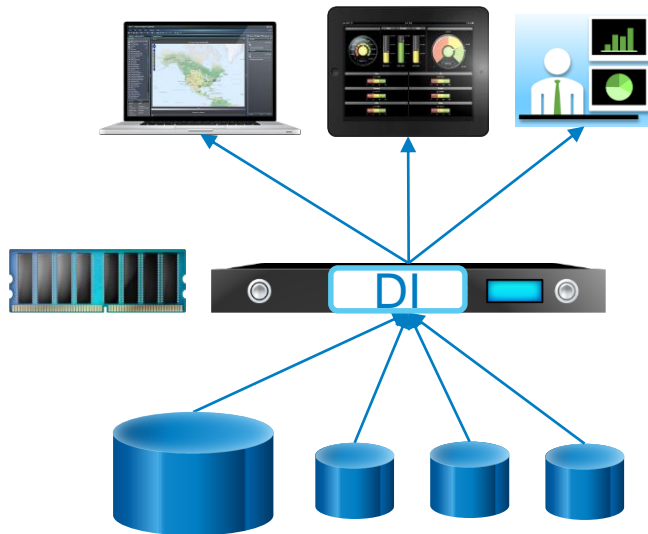
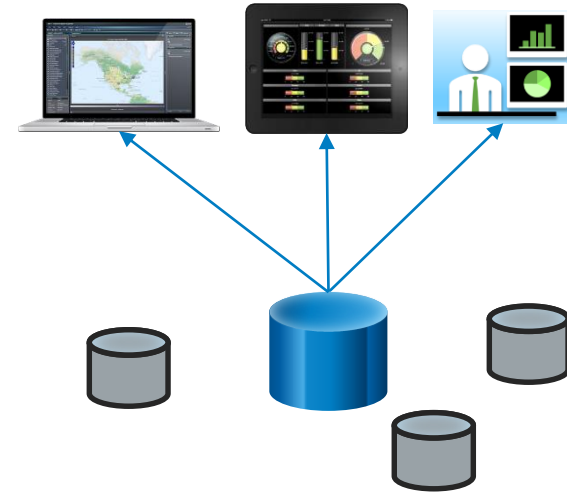


### Svona hjálpar samlagið

- Með góðum rekli er hægt að gera gögn aðgengileg á kunnuglegu formi fyrir notendur
- Þurfum ekki (eins marga) sérhæfða Hadoop forritara

## GAGNASAMLAG HENTAR ÞEGAR GÖGNIN ERU...

**Ekki komin í vöruhúsið:** Algengt er að ekki sé hægt að þróa vöruhús gagna nógu hratt til að uppfylla kröfur notenda, og önnur gögn (t.d. rauntíma) er kannski óraunhæft að setja inn í vöruhúsið

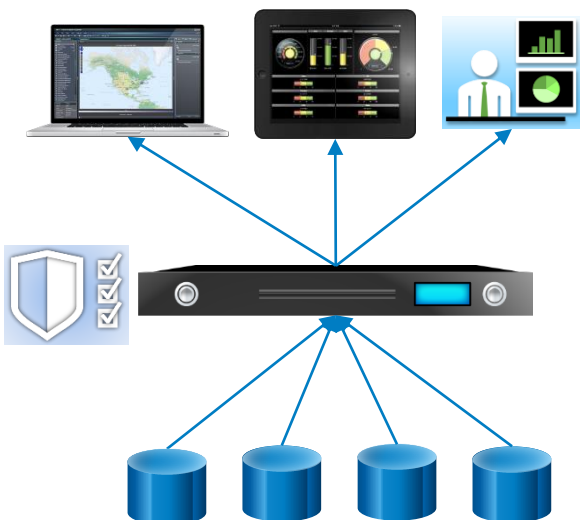
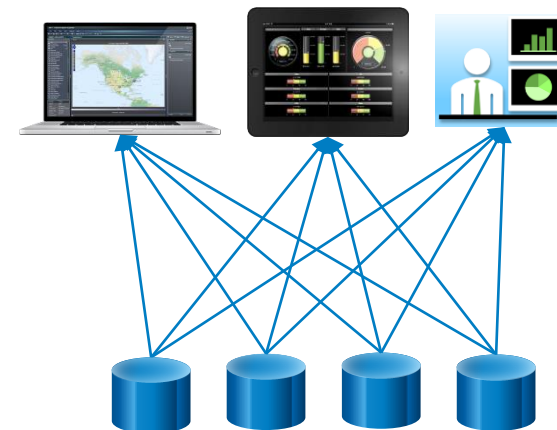


### Svona hjálpar samlagið

- Samlagsþjónn getur sambætt gögn úr vöruhúsi við önnur gögn, t.d. rauntímagögn eða meira niðurbrotin gögn en við höfum í vöruhúsi
- Notandinn upplifir þetta eins og nýtt og betra vöruhús

## GAGNASAMLAG HENTAR ÞEGAR GÖGNIN ERU...

**Of sundurleit:** Við erum með marga gagnagrunna sem hver hefur sitt öryggismódel. Jafnframt getur notendum reynst óyffirstíganlegt verkefni að draga saman gögn úr sundurleitum gagnagrunnum.



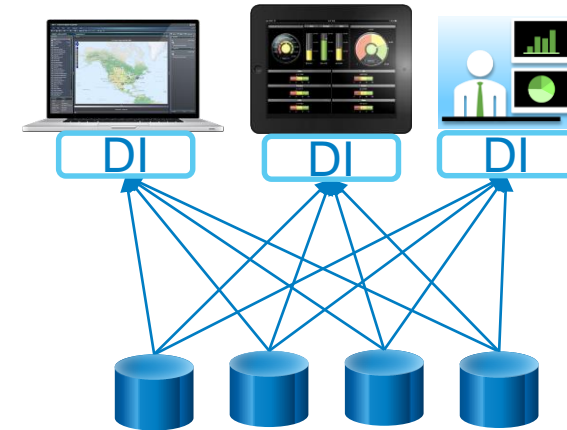
### Svona hjálpar samlagið

- Samlagsþjónn opnar einn aðgangspunkt þar sem notendum, aðgangi og rekjanleika er stýrt
- Drögum gögn saman í eina sýn og gerum þannig notendum lífið auðveldara og aukum lipurleika

iPad® is a trademark of Apple Inc.

## GAGNASAMLAG HENTAR ÞEGAR GÖGNIN ERU...

**Síbreytileg:** Þegar gögn eða kröfur til greininga taka stöðugum breytingum er erfitt að viðhalda samþættum gögnum í vöruhúsi og/eða fjölda mismunandi lausna sem nýta gögnin.

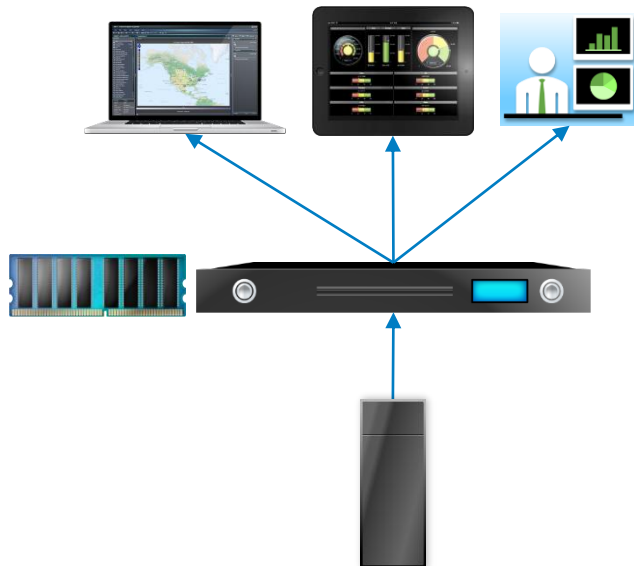
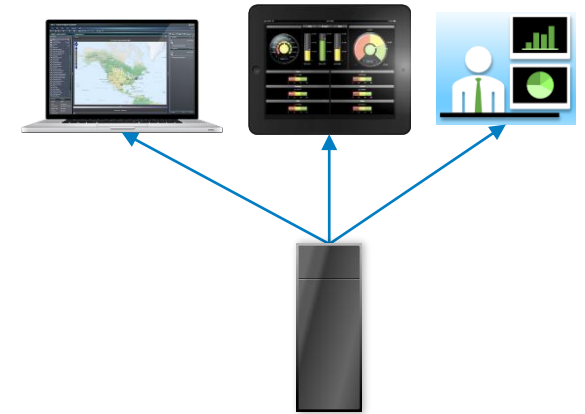


### Svona hjálpar samlagið

- Auðvelt er að búa til nýjar samlagsfyrirspurnir eða breyta þeim til að bregðast við ólíkum og síbreytilegum kröfum
- Notendur gagnanna eru skermaðir frá breytingum í undirliggjandi gögnum

## GAGNASAMLAG HENTAR ÞEGAR GÖGNIN ERU...

**Rauntímagögn:** Við viljum sækja gögn í nær rauntíma beint úr kerfi í rekstri, röng notkun getur valdað læsingum eða töfum í rauntímakerfi. Í sumum tilfellum eru fyrirspurnir verðlagðar, t.d. í stórtölvuumhverfi



### Svona hjálpar samlagið

- Notendur lesa gögn sem eru í skammtí mavistun í stað þess að lesa beint úr undirliggjandi kerfi
- Gögn í skammtí mavistun eru endurnýjuð reglulega (eins oft og þarf, eða hægt er)

## SAMANTEKT

- Data Federation / Data Virtualization lausn er valkostur sem getur nýst við að t.d.:
- Einfalda aðgengi að sundurleitum gögnum
- Hraða þróun á skýrslum eða samþættingarvinnslum
- Vinna með ómótuð gögn
- Nálgast rauntímagögn
- Ná miðlægri stjórn á öllum aðgangi að gögnum og rekjanleika á nýtingu gagna

## SPURNINGAR?

[magnus.gudmundsson@sas.com](mailto:magnus.gudmundsson@sas.com)

Kíkið við á SAS básnum (1.hæð, suðvesturhornið)



THE  
POWER  
TO KNOW®