

BEST PRACTICE

ProData Consult magazine about IT, projects and the people behind them. 4, 2014

NoSQL & Big Data



12

At arbejde med
NoSQL er lidt
cowboy-agtigt



18

 Norwegian standards
based on NoSQL

Contents

Pages 4–7

Introduction to 'Not always SQL'

Page 8-11

Big Data: Challenges and opportunities

Big Data & Analytics expert Kim Gregers Petersen explains the challenges and opportunities that Big Data offers consultants and businesses

Page 12-17

"Working with NoSQL is like being a pioneer – for tough guys"

Senior Developer and ProData consultant Thomas Brask Jørgensen has almost 15 years of experience working with NoSQL, and in this issue he offers some of his best advice for consultants and businesses

Page 18-20

Norwegian standard based on NoSQL

Good advice and experiences from software architect consultant Vidar Ingebretsen of Standard Norge, about switching from a traditional SQL database to NoSQL

Page 22-23

The polyvalent database design of the future

NoSQL guru Martin Fowler focuses on NoSQL

22

 Martin Fowler
on the polyvalent
database
design of the
future

8

 Big Data:
Challenges and
opportunities

Best Practice 4 2014

Customer magazine

ISSN nr.: 2245-6090

Legally responsible editor:

Søren Rode

Editors:

Jeanett Skou

redaktion@prodata.dk

Graphics and design:

Jeanett Skou

jmw@prodata.dk

Journalist:

Jakob Lund

Photography:

Jacob Djurhuus

Publisher:

ProData Consult A/S

Circulation: 7.000

Printing:

LaserTryk.dk

Adresses:

ProData Consult Copenhagen

Stamholmen 157

DK - 2650 Hvidovre

ProData Consult Aarhus

Hasselager Centervej 9

DK - 8260 Viby J

ProData Consult Aalborg

Promenaden

Vestre Havnepromenade 5

DK-9000 Aalborg

Tel. +45 43 43 11 71

info@prodata.dk

www.prodata.dk

www.konsulenter.dk

Cover:

NoSQL means NOT ONLY SQL



New consolidated ProData Consult

ProData Consult and Interim Competence merged operations as of 1 January 2014. Together we are now Denmark's largest consultancy within our segment, giving us an even better platform to serve our clients.

Merging the two companies has been challenging, educational and a lot of hard work. We are pleased and proud that we have not experienced any 'churn' in connection with the merger (churn = the number of lost customers in a specific period), and I would like to thank all the clients involved for their patience and cooperation

during this transition.

This issue of *BestPractice* is focused on NoSQL and Big Data. These are two very hyped technology concepts that we are seeing more and more clients implementing. These are technologies of which anyone with an interest in the IT industry should be aware.

ProData Consult has participated in a large number of NoSQL implementations in Denmark, Norway and our Nearshore Center in Warsaw. We have built a solid in-house knowledge of the technologies - and are familiar with the pros and cons,

dos and don'ts - so if you or your business face challenges in these areas, we are ready to put our full weight and expertise into helping you solve them, both in the overall advisory and business models but also the hard core technical knowledge and practical implementation.

Enjoy the magazine.

Søren Rode
CEO



Introduction to

NOT ALWAYS...



THE NOSQL CONCEPT

... was first used in 1998 by Carlo Strozzi – though with a slightly different meaning. It was only when a group of professionals met for an IT conference in San Francisco in 2009 that the term was given the same meaning it has today. The professionals needed a brief and unique hashtag for Twitter so that people would all be referring to the same thing. Eric Evans – then employed at Rackspace – suggested NoSQL (not only sql). It was accepted by the community and since then has been the regularly used common denominator for a group of databases that differ from relational databases.

IN JUST A FEW YEARS, A NEW HETEROGENEOUS GROUP OF DATABASE SYSTEMS HAS CAPTURED THE WORLD'S ATTENTION AS A MORE AND MORE OBVIOUS ALTERNATIVE TO THE RELATIONAL DATABASE SYSTEMS. THIS ATTENTION IS FULLY DESERVED. NOSQL DATABASES REPRESENT AN APPROACH TO DATABASE TECHNOLOGY THAT HAS THE POTENTIAL TO FUNDAMENTALLY CHANGE ENTERPRISE IT ARCHITECTURE.

Analysts, businesses and IT experts speculate from time to time on how the world's data volumes are increasing by the second. For example, data management company AIS recently assessed that if you collected all the world's data records from 2012 and burned them on DVDs, the stack would reach to the moon and back five times. In the same vein, a Cisco analysis has estimated that in 2016 overall the world's data centres will handle 6.6 zettabytes of data per year. This is equivalent to every person on earth streaming approx. 2.5 hours of HD video each day. But regardless of such musings and guesswork, it is a fact that the world's data volumes have increased explosively in recent years. Storing this data is a challenge that we are capable of meeting. But when it comes to extracting meaning from these enormous amounts of data – particularly in real time – the challenge is of a completely different nature.

In terms of databases, SQL has since the mid-1980s been the preferred default language when it came to storing and retrieving data from a relational database system – such as MySQL or Oracle. Data was divided into tables, enabling it to be stored and accessed according to a defined structure. But starting in the early 2000s, the Internet and businesses began generating a different type of data. The nature of this data was much more unstructured, and it resisted being fitted into traditional tables. It included data from Web 2.0 applications and social media, images, geographic information, chats, etc. This change in both data volume and data types led to the emergence of a group of databases, known as NoSQL, which in recent years has gained wider and wider acceptance. Google and Amazon were among the first to use NoSQL databases, and many others have since followed in their wake. Today, major companies such as Facebook, Mozilla, Adobe, Foursquare, LinkedIn and Digg all use NoSQL databases. As a sign that NoSQL databases are not just for Internet giants, but increasingly also for 'ordinary' companies, IT research company Gartner included NoSQL databases in its report entitled 'Emerging Technologies Hype Cycle 2012'.

COINCIDING CHARACTERISTICS

While there is broad consensus that there are a number of database systems on the market that differ fundamentally from relational database systems, there is less agreement on a proper definition of NoSQL databases. Not even British database guru Martin Fowler dares to attempt on an actual definition of the concept, since the various NoSQL databases do not share much in common. Instead, Fowler proposes a number of generic characteristics of NoSQL databases:

- They use (as a rule) non-SQL language
- Many of them are designed to run on clusters
- Many are open source
- They do not operate with a fixed schema structure

Although the wording of NoSQL does suggest it, the 'No' in NoSQL

stands for 'Not only', rather than 'No'. This means that NoSQL databases can actually use SQL-type query language, but usually do not do so. Martin Fowler's second characteristic points to the fact that relational databases and SQL are designed to run on one machine, while many NoSQL databases are designed to run on large clusters of machines. This enables NoSQL databases to deliver much higher response times because the system can quickly distribute the load across a large number of computers. The third characteristic is that many of the NoSQL databases are open source. Among other things, this means that, with a limited investment, companies can download, implement, and test whether an application and a specific NoSQL database can communicate with each other. Finally, NoSQL databases do not operate with a fixed schema structure like relational databases. For example, if a company would like to store a customer's phone number, first and last name, address and city, all this data must be defined in a fixed structure in a relational database. This means that the entire structure must be changed if, for example, the company wants to add the customer's preferred product as an additional

Advantages and disadvantages of NoSQL databases:

ADVANTAGES

- High scalability
- High schema flexibility
- Suitable for distributed systems
- Less administration
- Cloud-friendly
- Low costs

DISADVANTAGES

- No standardisation
- Technologies still immature
- Limited tooling possibilities
- Eventual consistency is not intuitive to program

table. This is expensive and bothersome. NoSQL databases are designed to allow the addition of data without a pre-defined structure. This means that the application can be changed in real time without fear of downtime.

FROM ACID TO BASE

As already mentioned, there are several different ways to illustrate the difference between relational and NoSQL databases. In addition to the the above range of characteristics, two database rule sets can also be compared. The first set of rules is known as ACID (Atomic, Consistent, Isolated, Durable), to which relational databases always adhere. This means that a transaction is either carried out completely or not at all (Atomic), that only valid data is added to the database (Consistent), that transactions never affect each other (Isolated), and that transactions are never lost (Durable).

The NoSQL databases operate according to a different set of rules known as BASE (Basic Availability, Soft state, Eventually consistent). BASE is easiest to explain

backwards, beginning with Eventually consistent. An ACID system guarantees data consistency after each transaction; a BASE system guarantees data consistency within a reasonable period of time after each transaction. In other words, there is data consistency in the system – just not immediately. This leads on to the Soft State principle. If the data is not consistent at all times, the system must take a temporary data state - a Soft State - into account. Finally, the sum of both these principles means that data accessibility is given very high priority in a NoSQL system – even if coincident errors occur in the database system, operating system or hardware. If parts of the database do not work, other parts of the database take over, so that data can always be accessed.

A BUSINESS DECISION

Companies, consultants and experts can quickly lose their way in the jungle of technological opportunities within the burgeoning world of database systems. Because the question is, when is a traditional relational database system the best solution

to handle the company's data and when is it better to seek alternatives among the NoSQL systems? The answer to that question should be based on a business decision made in consultation with the IT manager. Take Amazon, for example. Amazon decided very quickly that their business model was to provide prompt shopping to its customers. On the other hand, customers might find that a purchased item was not in stock after all, due to data inconsistency. But never mind, said Amazon. We just want to be known for always giving our customers a shopping opportunity. This goal was out of sync with the rule set in an ACID database system, so Amazon developed its own database system, Dynamo, which better supported the always available approach. An IT decision made on the basis of a business decision. Amazon had the financial muscle to build its own customised database architecture. Many companies do not have those resources. But the point remains the same. Choosing a relational database system as a default reaction is a thing of the past.

**DO YOU HAVE
QUESTIONS
ABOUT BIG DATA
OR NOSQL?**

Please contact us on
prodatabconsult.com/call
or tel: +45 43 43 11 71

Four types of NoSQL databases

Examples of four popular NoSQL databases divided into categories: Document Stores, Wide Column Stores, Key-Value Stores and Graph databases.

CATEGORY	DESCRIPTION	DATABASE SYSTEM
Document Stores	Instead of structuring data in tables and columns, data is structured as document-like information. This is of interest to those who wish to store large amounts of data.	CouchDB, Jackrabbit, MongoDB, OrientDB, SimpleDB, Terrastore etc.
Wide Column Stores	Stores data tables as sections of columns rather than rows of data. Suitable in situations such as managing data warehouses, which gather data from a number of sources.	Apache Hbase, Cassandra, BigTable, HyperTable, Qbase, OpenNeptune, Sqrl Enterprise etc.
Key-Value Stores	Stores information in the form of pairs: a 'key' and the data that belongs to the key. The key is used to quickly retrieve and manage data in the system's internal memory.	Cassandra, Riak, Redis, memcached, BigTable etc.
Graph Databases	Data is presented as graphs – especially in tree structures. Focuses on the relationship between the entities and not the entities themselves. Commonly used in social media.	AllegroGraph, DEX, Neo4j, FlockDB, Sones GraphDB etc.

Sources: www.wikipedia.com and www.techworld.idg.se

BIG DATA: CHALLENGES AND



OPPORTUNITIES

BIG DATA IS ON EVERYONE'S LIPS NOWADAYS AS A COLLECTION OF TECHNOLOGIES THAT CAN CHANGE THE USE OF DATA THE WORLD OVER AND IN ALMOST ALL TYPES OF BUSINESSES. CHALLENGES, SAY SOME PEOPLE. OPPORTUNITIES, SAYS A DANISH EXPERT.

Interview with Kim Gregers Petersen, Big Data & Analytics expert.

Recently there was a job ad at SKAT (the Danish tax authority). SKAT was looking for an assistant director to spearhead a brand-new department dedicated to 'Business Intelligence and Analysis'. The new department would, among other things, contribute to SKAT's overall efficiency and further develop data models "where the use of Big Data will be a natural part of activities".

"I read the advertisement as a sign that not only companies, but also the government, have begun to work seriously with Big Data," says Kim Gregers Petersen, Big Data & Analytics expert. As a consultant within Big Data solutions, he notices an interest building up day by day because companies and organisations like SKAT (the Danish Tax authority) are constantly seeing new opportunities to analyse their growing piles of data. "This is an area of explosive growth, and it's all about getting on board right now," says Kim Gregers Petersen.

By way of introduction, and as a starting point for a discussion of Big Data, Kim Gregers Petersen sums up developments in this area over the last 10 years with four facts:

Fact 1. The world's data volumes are increasing at a pace that far exceeds our wildest dreams.

Fact 2. The world's growing data volume is not just a quantitative challenge, since the data originates from new sources, such as video, photography, audio, navigation systems and instant messaging.

Fact 3. The new types of data are often

unstructured and therefore require very different handling technologies to those we are accustomed to.

Fact 4. These technologies are still so new that many in the branch find themselves at a crossroads. On one side, they recognize that much of this new technology will shape their professional future, but on the other, they are not familiar with these technologies as many are so new, and they don't feel it would be possible to gain the needed expertise in these areas.

"In rough terms, this is how things look right now," says Kim Gregers Petersen. "Of course, the question is: What is to be done?" he adds.

Highly interesting for the business

We will get back to the answer to that question. First, Kim Gregers Petersen explains what he defines as Big Data.

"If we take a hypothetical example, a business has data corresponding to 100 %. If you ask the vast majority of companies how much of the data they use in their daily business, they will answer 15-20 %. The remaining 80-85 % of the data is not used, for various reasons. They just store the data, because they have to, or because they do not know how to use it. The whole point of Big Data is to activate as much as possible of the 80-85 % inactive data, so it can contribute to the business," says Kim Gregers Petersen, giving an example.:

"Let's take a business that sells computers. The sales department keeps good track of which computers they sell to which types of customers, their profit margins on the various computers and the price development in the various product categories, etc. In the marketing department, they are good at contacting new and existing customers with offers of promotions, seminars, etc. And in customer service they are good at helping angry customers who call in and complain about a particular product. The point is that the data gathered by the various departments is never combined. It might be interesting for marketing and sales to know that customer service has handled 78 complaints about the same computer within a week. Today, that infor-

mation is lost, because businesses don't have the systems to coordinate this data."

Big Data is screaming for manpower

It is first and foremost corporations who should be concerned about all these many new possibilities and technologies as they don't have the corresponding competences within their organizations.

"This field is screaming for manpower," says Kim Gregers Petersen. "If I were 20 again, I would hurry up and run in that direction. For many years, being a programmer hasn't been very popular, one reason being that ERP solutions and Exchange solutions have been given an elegant administration layer makes them relatively easy for ordinary IT people to handle. In other words, it's become a bit boring to 'just' be a programmer. But with all the new Big Data technologies – most of which come from the open source community – it's suddenly cool to be a programmer again. We do not see the super-hot interfaces in the new products that we know from mature technologies. Big Data is a bit more hardcore."

As Kim Gregers Petersen explains, it is not yet possible to take the formal route if you wish to train in the field of Big Data, since this is not offered at Danish colleges and universities. "This is actually the biggest hurdle preventing the expansion of Big Data right now," says Kim Gregers Petersen. "But I suppose it's related to the fact that technology is so new that the educational system has not been able to keep up."

The Big Data environment

With a generic model of a Big Data environment in front of him, Kim Gregers Petersen outlines the long journey that the data takes, from the first knock on the company's door, such as Twitter, video or telecommunications data, to its final appearance as e.g. BI reports. During the journey, the name Hadoop pops up. According to Wikipedia's definition, Hadoop is 'an open-source software framework for the storage and large-scale processing of data in large clusters that run on commodity hardware'. Kim Gregers Petersen describes Hadoop as a key component of many of the largest Big Data environments in the world.

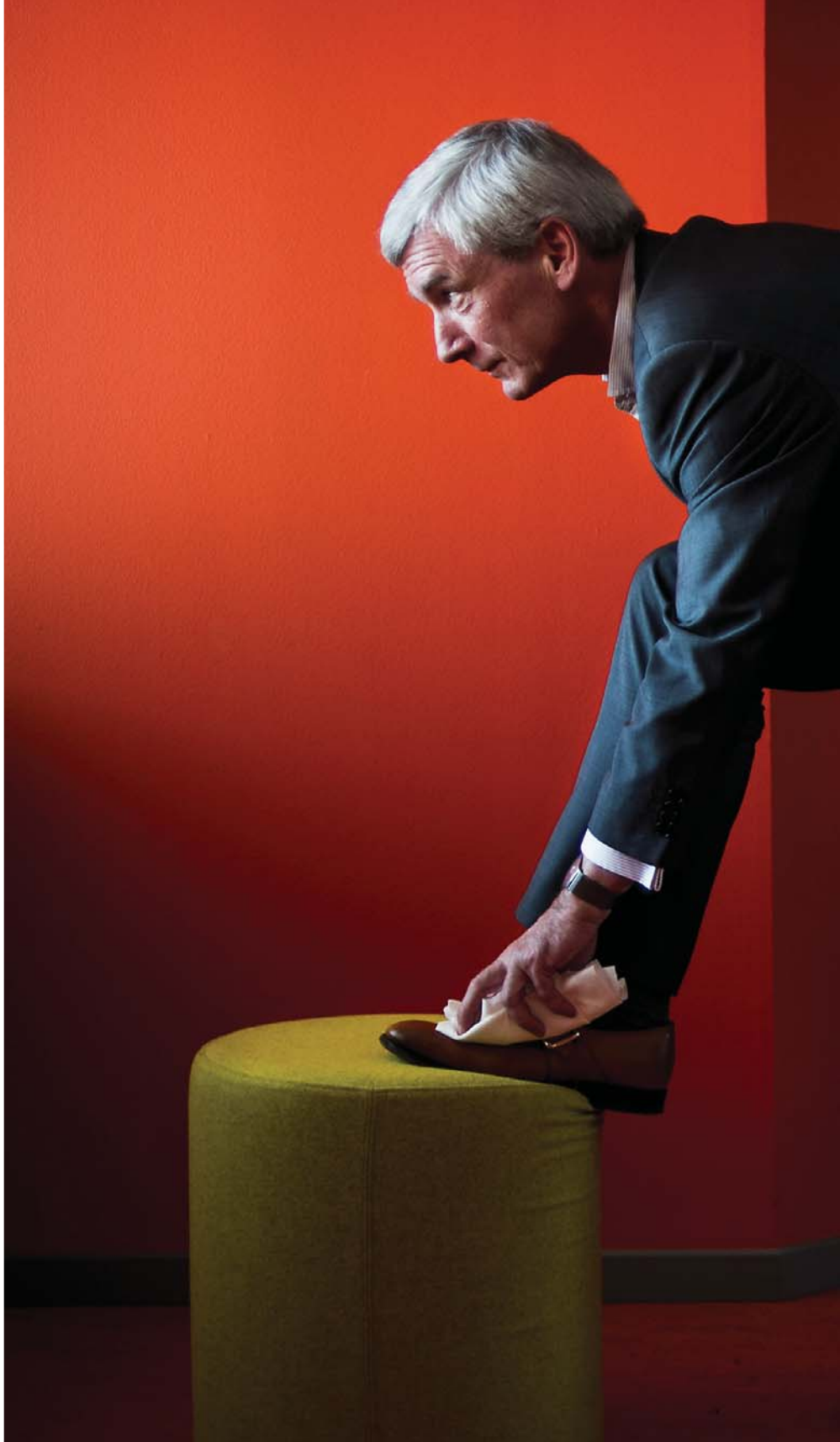
"The great thing about Hadoop is that it acts as an infinite number of buckets into which you can pour both structured and unstructured data. You may wish to analyse some of the data immediately, while other data may not be analysed until after three years, when this is more relevant. Hadoop was created to meet these and many other

requirements," states Kim Gregers Petersen.

"I recommend that you take a closer look at Hadoop and all the technologies that comprise Hadoop. I say this for several reasons, including that never before have large commercial enterprises had so much at stake in an open source environment. For example, Hadoop represents the backbone of the IT systems of Yahoo, Twitter, Netflix and Facebook, and they will do everything to ensure

that Hadoop gets better and better."

He can barely bring himself to mention the case, because it has received so much media attention, but Kim Gregers Petersen mentions in passing Vestas' large Hadoop installation and how they are able to make almost real-time simulations for the location of new wind turbines, in order to demonstrate the potential of Big Data and Hadoop. In another, less known example, Sweden's Royal Institute of Technology (KTH) – is





using IBM's streaming technology STREAMS for traffic monitoring in Stockholm. A variety of data sources, such as vehicles' GPS signals, alarm messages from traffic control, sensors on the roads and weather data, help direct traffic to flow as smoothly as possible.

The logic is that, no matter which industry, any business of a certain size could benefit from Big Data?

"Exactly. But this requires creative thinking, and that you know the technologies," concludes Kim Gregers Petersen.

How to Get Started with Big Data

Big Data is many different things and there are probably no two clients having identical problems and issues. Below however, is an excellent road map for getting started.

1. Try to approach Big Data as a shared vision and task; the business side working together with IT
2. Look at current data and assess how more value can be extracted. Use a tools such as IBM Watson Explorer to do this
3. Do you have performance issues with your SQL databases, and could it make sense to move your data into a scalable NoSQL database?
4. Are you currently using BI, and could make more data available for BI tools with a Big Data environment?
5. Do you already have "data scientist" profiles employed? Do you have existing employees who are enthusiastic about the area who can be "upgraded"?
6. Choose a software package with tools and Hadoop, such as those supplied by IBM so you can get started faster and get/show results
7. Identify consultants (for example ProData Consult), who can both advise on the business side and assist in the implementation and programming

“WORKING WITH NOSQL IS LIKE BEING A PIONEER – FOR TOUGH GUYS”

Almost 15 years of experience with NoSQL and NoSQL-type databases have given ProData consultant Thomas Brask Jørgensen a deep insight into working with the new, fast-growing database technologies. Here he talks about his experience and also offers good advice to both businesses and colleagues.

Interview with Senior Developer Thomas Brask Jørgensen





THOMAS BRASK

Jørgensen, has both the age and experience to have worked with NoSQL databases before they were even called NoSQL. That was back in 2000 when he was working in the Jubii community department responsible for the chat platform. This was in the early web 2.0 days, when Danish users were beginning to use public instant messaging with other users about all kinds of topics. Thomas Brask Jørgensen remembers it as a fun project, but also as one that, right from the beginning, presented developers with a number of new challenges.

"Back when I started, Jubii Chat was a relatively small Danish product, but you could see that it had great potential. Then Jubii was acquired by Lycos, and their ambition was to introduce the product across Europe, and so all of a sudden we were talking about a completely different size of user base. At that time you used ASP and SQL Server for chat, so it was obvious that there would be scalability problems. It was not remotely fast enough and it would be expensive to solve the problem with the existing technology," explains Thomas Brask Jørgensen.

"So instead, we chose to create our own database. We had a developer in-house for seven years who didn't do much else besides developing the database. It actually ended up becoming a really good product, and one that is still on the market. Today, the database has full SQL, but it didn't back then. It was pure NoSQL. Solely because it had to be capable of scaling across multiple servers and doing so as quickly as possible."

Even to this day, Thomas Brask Jørgensen believes that building their own database to solve the challenge was the right decision. But when they decided to use the database for other types of tasks at Lycos, problems arose.

"We simply moved all of the community products over to the platform and also developed new products that used the same database. This resulted in a few crashes from time to time where everything went down, and then our technician had to sit there for 36 hours straight to restore the system. Downtime might not matter so much when it's just chat and guest books, but it still ended up costing money. This was because chat was based on advertising revenue, and when the chat was down, the users couldn't see the adverts," he explains.

ONLINE BACKGAMMON AND ZYB

When Lycos moved the entire department to Germany in 2005, Thomas Brask Jørgensen did not go along. He was still head consultant for the back-end for a while, but preferred to work from a Danish base.

The next time he worked with NoSQL databases was in 2006, when he was hired for a project for a small Danish entrepreneurship that wanted to create online games. At that time - with top player Gus Hansen as their role model - the Danes had really taken to poker, online poker and other online games in a big way.



"I was asked to handle the back-end for an online backgammon game, and I said yes. One of the databases we used was called Memcached, which is a NoSQL-like database for games and chat functionality. It was fine for that, but it was a bit of a problem that the company also wanted to use it for financial transactions. There are some problems with consistency and concurrency, which means that you should not use NoSQL for the financial transactions; instead, you need to use transactional databases," says Thomas Brask Jørgensen, adding that the small entrepreneurial company never came to market with their online game, but closed the project down prematurely.

"In 2008 I came to ZYB, which was a backup solution for the mobile phone. Before I arrived, ZYB was acquired by Vodafone. At that time, Vodafone had the ambition to build a large community. My job included working on a proof of concept for a new version of the ZYB solution, and in this respect we were also looking at NoSQL databases, including MongoDB and Amazon's SimpleDB," says Thomas Brask Jørgensen, explaining that the developers were working on one overall solution, where MongoDB, SimpleDB and other databases were 'wrapped' in an abstraction layer so you could not see which databases were running underneath. It was not a package solution they would choose in a use scenario; it was only



NEED EXPERT
ASSISTANCE WITHIN
ANY TECHNOLOGY?

WE HAVE THE
IT-CONSULTANTS!

Go to
prodatabconsult.com/it
or call
+45-43 43 11 71

an experiment to test the fitness of the various databases for use against each other.

"Unfortunately, ZYB eventually closed down, so we never completed the project, but it was very exciting to work on. In retrospect, we probably erred when we tried to build a generic SQL interface on top of the NoSQL databases. We took a NoSQL database and tried to turn it into SQL. We probably shouldn't have done that. It is possible to have a query language in this way, but you don't get consistency and atomic transactions," says Thomas Brask Jørgensen, explaining that the requirement for SQL and query opportunities arose from the wish to use the packing solution for data analysis

with business intelligence tools, data mining, etc.

"But it wasn't the right way to go about it.

The right way would have been to move the data and analyse it somewhere else where you would have had more advanced query options. You don't get that kind of functionality in a NoSQL database."

READY YESTERDAY

In 2012, Thomas Brask Jørgensen was brought in to the Atea Tele project, where the company entered the business telephony market. When Thomas Brask Jørgensen was brought in along with

other consultants, the project should – as always – have been completed 'yesterday', he says with a smile. "And after all, that's what

consultants do." Atea wanted tolerating functionality – and the possibility of split billing, allowing the user and not the least the company to separate work calls from private calls. Originally, Atea wanted to perform this rating itself, which would mean millions of records each time a call was

made or a text message was sent. Atea had looked into it, and the project did not suit Microsoft SQL, for which they had their own expertise. So they decided to

"We took a No-SQL database and tried to turn it into SQL. We probably shouldn't have done that."

use RavenDB, which is a very .NET-friendly NoSQL database.

"An obvious choice if that was the route you wanted to take," says Thomas Brask Jørgensen. "The problem was that before we were brought in, they had decided that they didn't want to perform the telering after all. They just wanted a user database with profiles, configuration, etc. This meant a standard database with not much data, and the data that was available would fit into a relational database, where you could do advanced queries, etc. The task was not well suited to RavenDB, which they had retained even though they had dropped the idea of telering. After six months on the project, we argued to remove RavenDB and bring in a relational database, and that's how things ended up."

Today, Thomas Brask Jørgensen is at Saxo Bank, where he is currently finalising an Open API project that makes it easier for

"Specifically, I would recommend looking at RavenDB and MongoDB, if you have a Microsoft background. If you're more web-orientated, you could look at CouchDB. And if you have the cloud option, take a look at Microsoft's Azure platform, Amazon's SimpleDB or Google Cloud Datastore."

external companies' banking applications to work with Saxo Bank's internal systems.

"Our system records large volumes of log data. So far we have used log files, but we want to get this into a database. In this case, a NoSQL database such as MongoDB would be a good choice because of the large amount of data and the relatively modest lookup requirements. Saxo Bank already use MongoDB in other contexts, so it's an obvious choice," says Thomas Brask Jørgensen.

LISTEN TO US!

Today, when Thomas Brask Jørgensen begins work on a project in which NoSQL technology is considered, he often experiences that he is asked for advice because of his many years of experience in the field. And so it should be, according to him.

"My experience is that in 9 out of 10 cases the clients listen to us consultants as consultants because we have seen similar projects before. It's very rare that you're told to do things in a certain way. But you do experience this from time to time. I've been in a situation where the choice was between relational databases and NoSQL databases. But now I usually speak my mind if people are going off the deep end with their

solutions," says Thomas Brask Jørgensen, who stresses that the best advice he can give companies facing a database project is to think through their IT environment and future needs very carefully before making the technological decisions.

"You need the right tool for the right task. Making changes along the way is painful. In an ideal world, you should look at the databases' technical features. But the reality is usually that people end up choosing the product they are familiar with," he says.

If Thomas Brask Jørgensen has one piece of good advice for colleagues who may not yet have become acquainted with NoSQL databases, it is that they must first and foremost get an overview of the myriad of NoSQL and NoSQL-like technologies.

"There are well over 100 NoSQL databases on the market. So it's very much about being able to distinguish between the different types of databases and not believing that NoSQL is just one thing. Because it absolutely is not. Specifically, I would recommend looking at RavenDB and MongoDB, if you have a Microsoft background. If you're more web-orientated, you could look at CouchDB. And if you have the cloud option, take a look at Microsoft's Azure platform, Amazon's SimpleDB or Google Cloud Datastore," says Thomas Brask Jørgensen.

DECLARATIVE VERSUS IMPERATIVE

There is a huge difference between working with relational databases based on more than 20 years of accumulated knowledge, proven technologies and a broad palette of tools, and working with the newer and less mature NoSQL databases. But on a personal level, Thomas Brask Jørgensen likes the challenge inherent in NoSQL technologies.

"Working with NoSQL is kind of like being a pioneer – for tough guys. You feel like a pioneer every time, because you don't have the same big toolbox available that you have when working with relational databases. With NoSQL you have to do a lot of the work yourself. You could make some of the same queries as in a relational database, but you have to write this yourself. Most NoSQL databases do not have declarative query options like SQL, where you write what you want and how it should be presented. NoSQL databases are usually imperative. You describe how data is to be found. It's like writing a program. First do this, then do this, then do this, etc. It's a completely different thought process," concludes Thomas Brask Jørgensen.



Bluebook

Name:	Thomas Brask Jørgensen
Age:	39
Position:	Senior Developer
ProData Consult:	Since 2008
Education:	Computer Science and Mathematics from Copenhagen University, 1995–2000

Norwegian standard based on NoSQL

The quest for better search capabilities and XML language led Standard Norway to replace its traditional SQL database with the NoSQL database Mark Logic. Earmark time and resources for skills retraining, advises the Standard Norway project manager. Otherwise you risk getting into trouble.

*Interview with Software Architect consultant
Vidar Ingebretsen, Standard Norway*

Standard Norway – the Norwegian equivalent of Danish Standards – is an organisation devoted to developing and publishing standards. This means that they help to streamline product development across national borders so as not to produce several near-identical versions of the same product. This is not just a good idea for businesses; it also benefits consumers because standards ensure documentation that a product complies with specific requirements – for example in the environmental area. The many thousands of standards are sold by Standard Norway through their sales company Standard Online, either directly to the Norwegian market or as ‘adapted’ variants, where international standards have been published as Norwegian standards. The standards are published online in a solution where for several years the information was stored as PDF files. This means that the users could only search the document metadata when they searched for the desired standard.





Highlights of MarkLogic

- Supports ACID transactions, horizontal scaling, indexing in real-time and disaster recovery
- Supports searches across different data types – text, images, date/time, geospatial data and currencies – from many different data sources
- Ability to run MarkLogic directly on the Hadoop Distributed File System (HDFS) and move data between MarkLogic and Hadoop within the applications
- Supports Analytics and Business Intelligence in real time
- Good tools and APIs for rapid application development

"PDF is a very inflexible format in general," says Vidar Ingebreetsen, a Software Architect consultant at Standard

Norway.

"So in 2013 we began to look around for a new system that could support XML. XML allows advanced searches that support HTML, so that, for example, a user can read a standard online with hyperlinking – to other standards and also within the same document. And it supports the production of e-books, allowing the creation of an electronic version of a standard. XML makes all this possible," says Vidar Ingebreetsen, adding that traditional databases, such as those Standard Norway originally used, did not support XML particularly well, which is why they went in search of a NoSQL database.

NEW POSSIBILITIES DEMAND NEW SKILLS

They found the NoSQL database by looking at a solution used by the International Organization for Standardization – better known as ISO – which is the global standard-setting organisation with 162 member countries. ISO had recently faced many of the same issues as Standard Norway, and they had chosen a supplier and a NoSQL database solution that they believed was capable of handling the task at hand: its name was MarkLogic.

"So we quickly turned to MarkLogic as well. Their database has excellent search functions, supports XML well, and in addition

we can store all binary data in the system, which means that we can continue to store all PDF files and Word documents in the same database," says Vidar Ingebreetsen. He explains that MarkLogic was responsible for training Standard Norway employees in the use of the MarkLogic database, and also had a fixed team of consultants attached to the project, so that Standard

Norway always had access to

superusers of the system. After several months of analysis, the project was initiated in January 2014 and as of April 2014 it was in full swing handling the migration of data from the old to the new system.

Was there anything that surprised you with regard to the project?

"A lot of scripting has to be done in the database, and there is a lot of very functional programming, which is a different kind of programming to what I'm used to. So the learning curve was somewhat steeper than I would have thought in terms of using NoSQL. But we can also see that it will be good because we have a very close partnership with MarkLogic. Without that, it would have been a huge challenge. There's a great deal that is different to traditional SQL, and although the system offers many options, it requires a certain amount of training to take advantage of them," says Vidar Ingebreetsen.

GOOD BUSINESS CASE

Although the decision to switch from SQL to NoSQL was made before Vidar Ingebreetsen joined Standard Norway, he is convinced it was a business-driven and not an IT-driven decision.

"ISO has set a market trend by switching to NoSQL, and many have followed in their footsteps. With ISO, we've seen better data quality, better search capabilities, e-book capabilities and better link options that give customers a much better user experience. The expectation is, of course, that that this will ultimately lead to increased sales in the

webshop," says Vidar Ingebreetsen, adding: "We also expect that the optimisation of internal business processes can lead to significant savings. Currently, much of the production of the standards is based on manual processes, which can now be automated with the new system. For example, this includes quality assurance of the PDF files, where in future we will use a system called Pitstop to ensure quality control of the data and rectify PDFs by adding fonts that do not exist in the file," explains Vidar Ingebreetsen.

Although he generally describes the NoSQL project as successful, one, there have also been challenges along the way.

"We especially had problems with migrating data from the old system to the new and connecting the data in a good way. This is primarily because the data quality in the old system was worse than we had expected. The complexity of the solution has also proved to be a challenge. It requires excellent XML skills to embark on a project such as this. We had these skills via MarkLogic's consultants, as well as by upgrading and further developing our own resources' expertise. That was a necessity, and you need to be aware of this if you're considering embarking on a similar project," concludes Vidar Ingebreetsen.

NoSQL from Warsaw

ProData Consult Norway are supply a team of five expert consultants for Standard Norway via our Nearshore Center in Warsaw. The team has been the primary technical resource in the building of Standard Norway's NoSQL based online solution. The solution went live and was presented as a reference project for MarkLogic, in mid-May 2014.

About Standard Norway's online solution

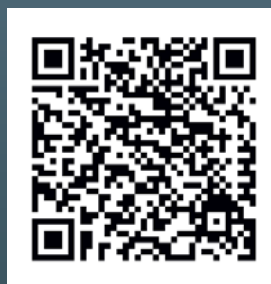
Standard Norway's online solution is based on the e-commerce solution EPiServer Commerce with an advanced product database from MarkLogic as the foundation. The product database supplies both metadata and product content to EPiServer Commerce. In addition to single purchases of standards via www.standard.no, the company is also testing a subscription solution based on the security solution in MarkLogic.

The core business of ProData Consult is to provide IT consultants on Time & Material basis

... and we are doing this right now with over 400
expert IT consultants placed at more than 100
different clients.

We also strive to have the most innovative value
added benefits and services.

Scan the code and discover the benefits of using
a full service provider in IT consulting,
Management Services and Near-shoring in
Poland, as revealed by one of Denmark's most
prominent and expanding companies




or visit
www.prodataconsult.com/nac



prodataconsult.com

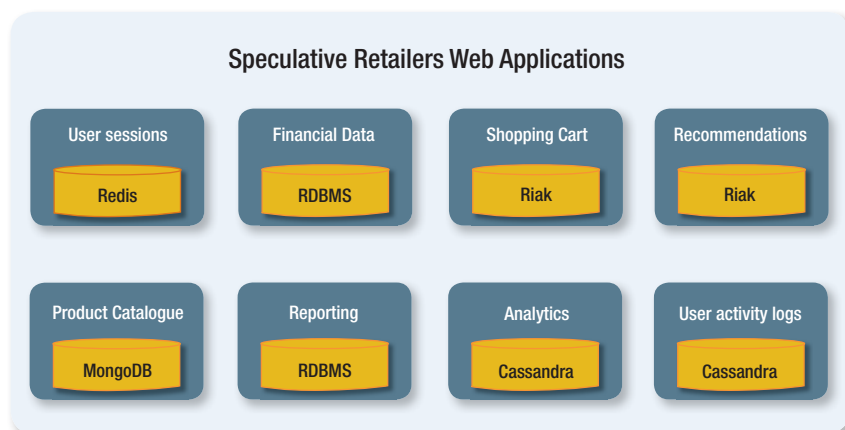
POLYVALENT:
POLY IS GREEK FOR 'MANY'
-VALENT: HAVING ONE OR MORE VALENCES. FROM LATIN
VALERE, 'TO BE WORTH'. SOURCE: WWW.WIKTIONARY.ORG

THE **POLYVALENT** DATABASE DESIGN OF THE FUTURE



IT IS NOT A QUESTION OF EITHER-OR, BUT BOTH-AND, WHEN IT COMES TO CHOOSING BETWEEN RELATIONAL AND NO-SQL DATABASES. SO SAY DATABASE EXPERT MARTIN FOWLER AND HIS COLLEAGUE PRAMOD J. SADALAGE IN A BOOK THAT INCLUDES FOCUS ON THE 'POLY-GLOTPERSISTENCE' APPROACH.

NoSQL databases have attracted a great deal of attention in a very short time. Hailed by some as the solution to database challenges of the future, and hated by others as unnecessary technological obstructions. Whether you are for or against the NoSQL systems, the huge amount of hype has led to the erroneous conclusion that the relational database, after 20 years of dominance, is on its way to extinction. It certainly is not, says database guru Martin Fowler in his book 'NoSQL Distilled', which he penned together with his colleague Pramod J. Sadalage.



A purely hypothetical example of a future web application, as Martin Fowler and Pramod J. Sadalage envision the design. The diagram demonstrates how different databases match different types of needs.

Martin Fowler and Pramod J. Sadalage describe how there are still many tasks that relational databases solve much better than NoSQL databases. In addition, NoSQL database technologies are still very immature. There are many edges that need to be sanded down before they achieve the same streamlined usability level as relational databases. Just as there are still many more tools for relational databases than for NoSQL databases. In other words, relational databases will continue to be the preferred database type for many tasks. But prompted by the emergence of NoSQL databases, the default approach to database systems has been given a shake-up.

A few years ago, the starting point was always: Which relational database should I use for this specific task? Or an even worse alternative: You must use this relational database for this specific task, because that is the one we use in this company! These boundaries are softening up, argue Fowler and Sadalage, so that the starting point today is more along the lines of: What is the most suitable technology for my specific task? It is only when you know the answer to that question that you are able to select a product regardless of whether it belongs to a group of relational or NoSQL databases. The authors call this new approach the 'PolyglotPersistence' approach.

POLYGLOT SOLUTIONS

'Polyglot' means multilingual, or a person who can speak several languages. And that is exactly what the future within database design offers. Applications should be based on several different languages, say Fowler and Sadalage, so you can take advantage of each database system's ability to solve a challenge. If there is anything we have learned from the past decade and its not only increasing amounts of data, but also more and more complex data structures, it is that there is no one-size-fits-all solution. Complex applications require correspondingly complex solutions to the reciprocally interconnected issues.

This applies not only across the application layer in a business, but also to each individual application. A complex enterprise application currently already uses many different types of data and also integrates data across sources. This reality will be reflected in an increasingly polyglot design of future applications, according to Martin Fowler and Pramod J. Sadalage. (See the above figure.)

COMPLEXITY BRINGS CHALLENGES

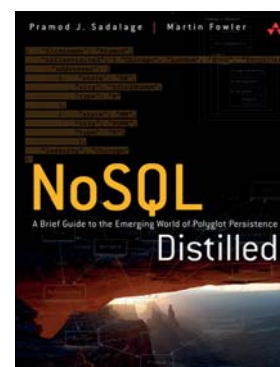
Complexity is synonymous with new opportunities. But complexity is also synonymous with new challenges, warn Fowler and Sadalage. For each completely new unknown NoSQL database you introduce into a business, many human resources and development hours are needed to learn the technology properly. Many of the NoSQL databases are also designed to run on large clusters, which for some companies means that they suddenly have to deal with issues of data consistency and availability in a whole new way.

There are more than enough challenges, according to Martin Fowler and Pramod J. Sadalage, who also offer a word of advice on the types of database projects that lend themselves to a 'PolyglotPersistence' approach.

If a distinction is made between between strategic database projects and routine database projects, NoSQL projects are most suitable for strategic database projects. There is no reason to bring so many unknown factors into play with a routine project that does not have the potential to add value to the business. On the other hand, a strategic database project based on the 'PolyglotPersistence' approach will help to significantly increase the productivity of developers and thereby also increase the time-to-market. A 'PolyglotPersistence' approach may also prove beneficial for very data-intensive projects where there is a lot of data traffic and high availability requirements.

FUMBLING THEIR WAY

What should you take away from this review of the polyvalent database design of the future? You should be aware that many NoSQL databases are still relatively immature technologies compared to relational databases, and therefore we are in a phase where many are still fumbling their way in the dark. You should also be aware that this fundamental uncertainty should be factored in if you are considering using, or advising others to use a mix of relational and NoSQL databases. You must also be aware that the concept of 'PolyglotPersistence' does not stand unchallenged in the industry. Many professionals oppose Martin Fowler and Pramod J. Sadalage's way of thinking and call 'PolyglotPersistence' a short-term approach to the problem. Critics argue that the polyglot design leaves software developers with a bigger headache than they started with. They would rather see developers tackle the challenges of building the necessary data models based on a single architecture, instead of integrating several different ones.



Martin Fowler and Pramod J. Sadalage's 'NoSQL Distilled' provides an introduction to NoSQL databases and includes a discussion of the 'PolyglotPersistence' approach.