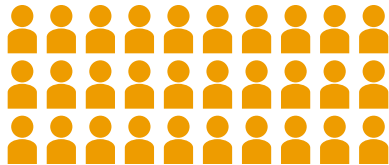


NoSQL: The New Normal

Matt Asay (@mjasay)
VP, Corporate Strategy, 10gen

10gen Overview



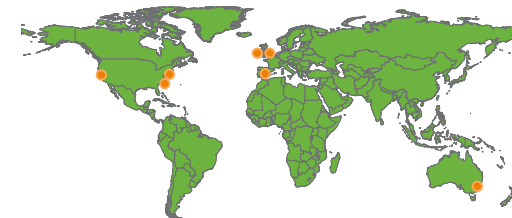
180+ employees



500+ customers



Over **\$81 million** in funding



Offices in New York, Palo Alto, Washington DC, London, Dublin, Barcelona and Sydney

Global Community

3,800,000+

MongoDB Downloads

47,000+

Online Education Registrants

15,000+

MongoDB User Group Members

14,000+

MongoDB Monitoring Service Users (MMS)

10,000+

Annual MongoDB Days Attendees

the past...
is prologue?

Back to the Future?

What has been will be again,
what has been done will be done again;
there is nothing new under the sun.

(Ecclesiastes 1:9, NIV)

What Do You Remember about 1969?

STONEWALL RIOTS
JOE NAMATH
BUTCH CASSIDY AND THE SUNDANCE KID
OCCUPATION OF ALCATRAZ ZIP TO ZAP
EASY RIDER
THE NEW LEFT
HARVARD STUDENT STRIKE
ALTAMONT

1969

THE YEAR EVERYTHING CHANGED
Rob Kirkpatrick

OH! CALCUTTA!
THE WEATHER UNDERGROUND
NEIL YOUNG & CRAZY HORSE
TOMMY
MC5
DAYS OF RAGE
HAMBURGERS
APOLLO 11
BOEING 747
I AM CURIOUS (YELLOW)
PORTNOY'S COMPLAINT
CUYAHOGA RIVER FIRE
BITCHES BREW
LET IT BLEED
CHARLES MANSON
MY LALPHOTOS
THE HORRORUM
LED ZEPPELIN
WOODSTOCK
PEOPLE'S PARK
JIMI HENDRIX
RICHARD NIXON
HELLS ANGELS
THE GODFATHER
IGGY AND JEFFERSON
WOODY ALLEN
THE STOOGES
AIRPLANE
MIDNIGHT COWBOY
NUDITY
THE MIRACLE
BOB & CAROL & TED & ALICE
METS
OPERATION MENU
CROSBY, STILLS & NASH

Daily Mirror The time: 9.18 pm,
July 20, AD 1969

Monday, July 21, 1969 No. 30,111

MAN ON THE MOON



AND THE MESSAGE FROM EARTH: WE'RE BREATHING AGAIN!

Man has landed on the Moon. A new era in his history began at 9.18 last night when the lunar module Eagle settled gently on the dusty surface of the Sea of Tranquility. Inside it—Astronauts Armstrong and Aldrin, destined now for a permanent place in history. They immediately began to prepare for their Moon walk. There are still great perils ahead. But these are truly great achievements. America, the land of frontiersmen, has opened up a new frontier.

The last practice run

THIS was how the astronauts prepared for the great Moon walk... by practising the whole operation inside a Moon simulator at Houston. On the left last run, when this picture was taken, Edwin Aldrin (left) used a pump for collecting samples while Neil Armstrong looked on calmly. Today they know if the rehearsal matched the reality.

FULL STORY—SEE BACK PAGE: THE PIONEERS—CENTRE PAGES

nothing? hold that thought

Back to the Future: PreSQL



- IBM's IMS (1969) – Developed as part of the Apollo Project
- IDS (Integrated Data Store), navigational database, 1973
- High performance but:
 - Forced developers to worry about both query design and schema design upfront
 - Made it hard to change anything mid-stream

Enter SQL

- Designed to overcome “PreSQL” deficiencies
 - Decoupled query design from schema design
 - Allowed developers to focus on schema design
 - Could be confident that you could query the data as you wanted later
- 30 years of dominance later...

...the present...

RDBMS Is Like a Spreadsheet

Person:

Pers_ID	Surname	First_Name	City
0	Miller	Paul	London
1	Ortega	Alvaro	Valencia
2	Huber	Urs	Zurich
3	Blanc	Gaston	Paris
4	Bertolini	Fabrizio	Rom

With “Relations” Between Rows

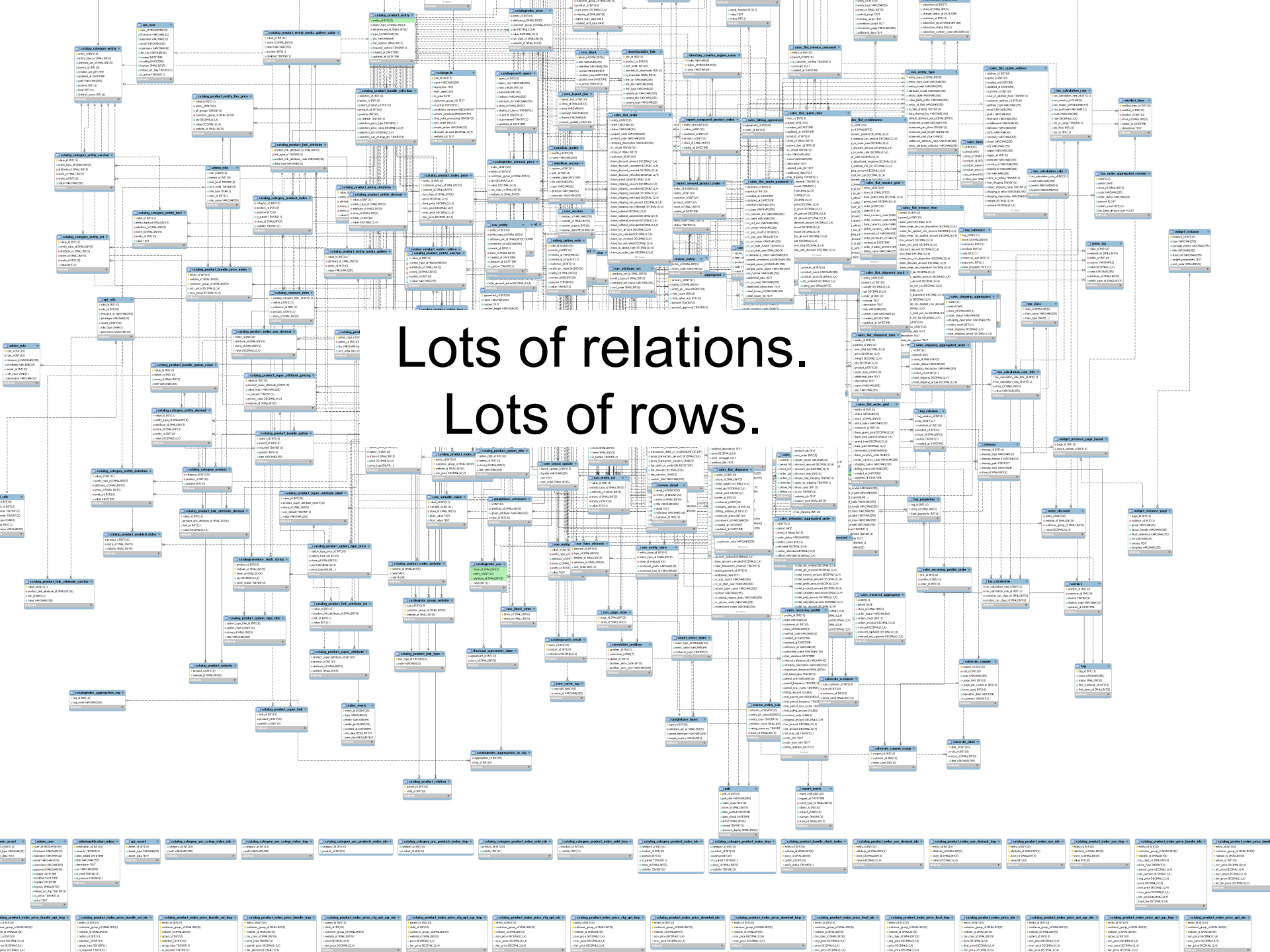
Person:

Pers_ID	Surname	First_Name	City
0	Miller	Paul	London
1	Ortega	Alvaro	Valencia
2	Huber	Urs	Zurich
3	Blanc	Gaston	Paris
4	Bertolini	Fabrizio	Rom

Car:

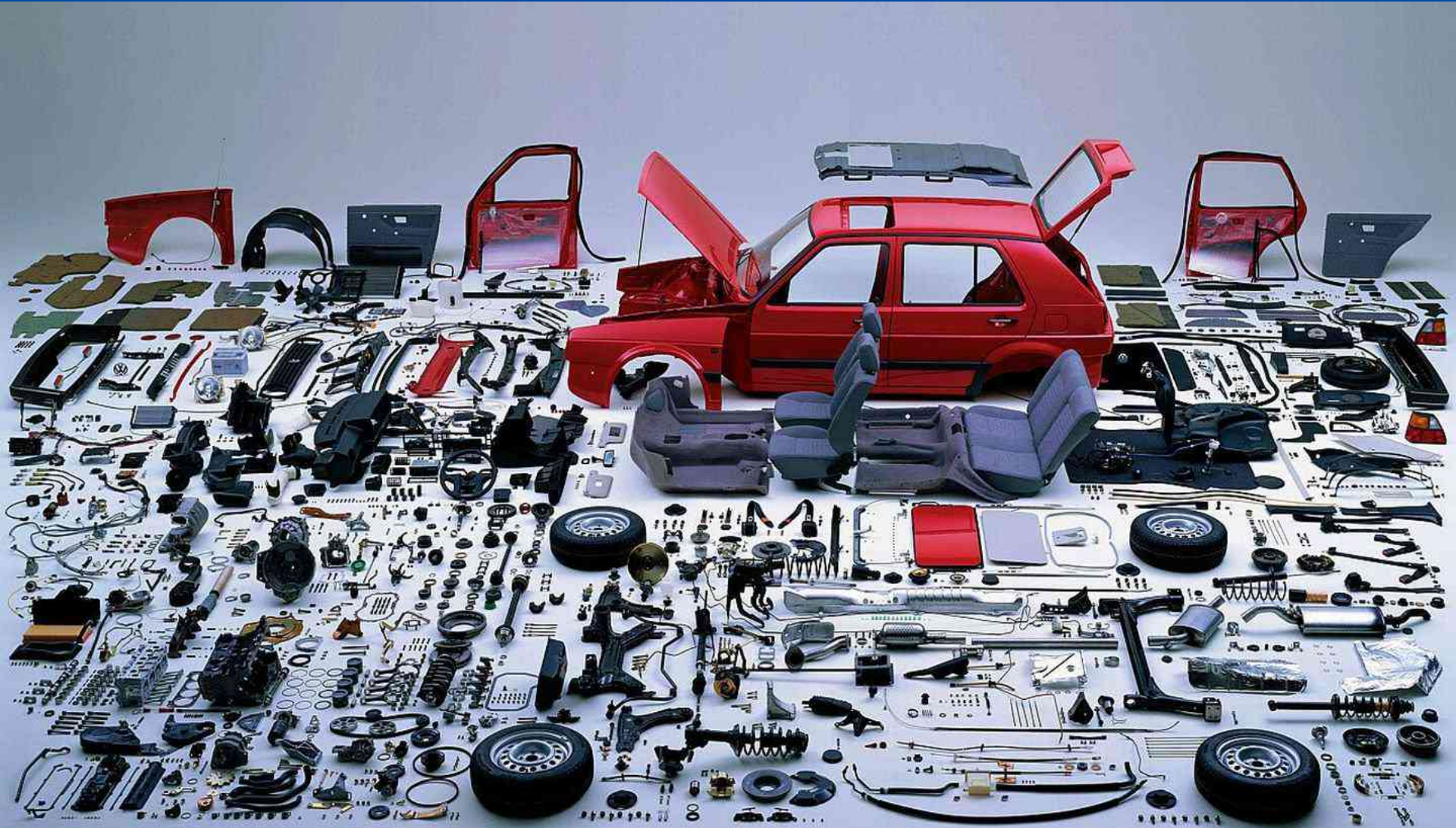
Car_ID	Model	Year	Value	Pers_ID
101	Bentley	1973	100000	0
102	Rolls Royce	1965	330000	0
103	Peugeot	1993	500	3
104	Ferrari	2005	150000	4
105	Renault	1998	2000	3
106	Renault	2001	7000	3
107	Smart	1999	2000	2

no relation

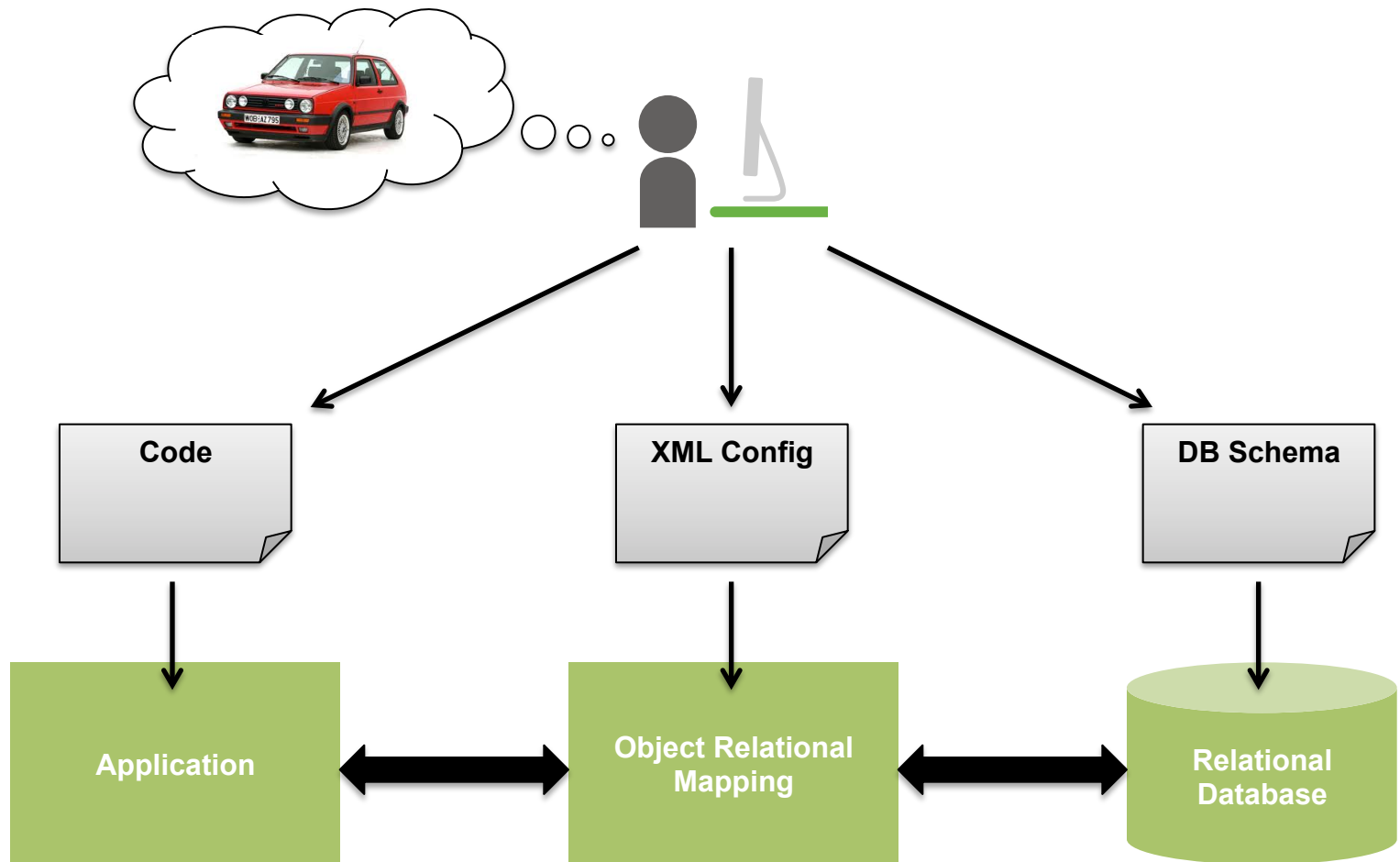


Lots of relations.
Lots of rows.

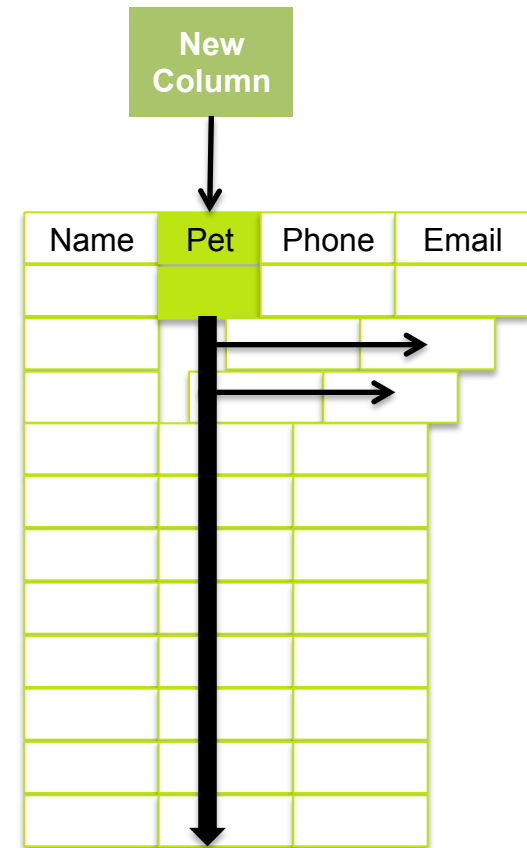
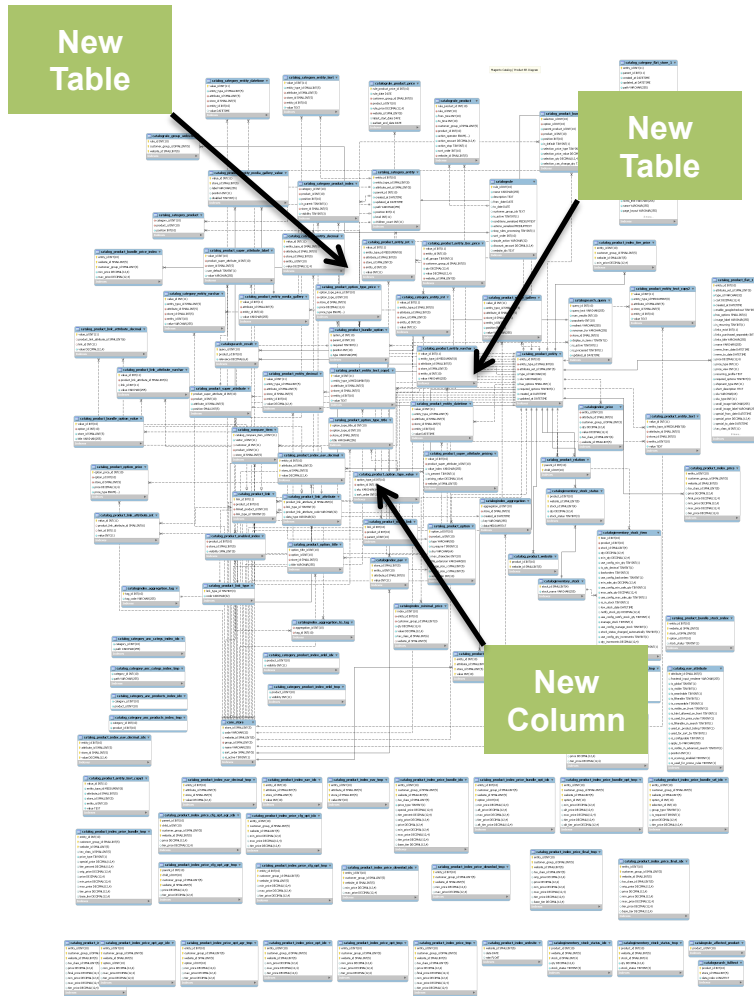
It Hides What You're Really Doing



It Makes Development Hard



And Makes Things Hard to Change



3 months later...

RDBMS Scale = Bigger Computers



“Clients can also opt to run zEC12 without a raised datacenter floor -- a first for high-end IBM mainframes.”

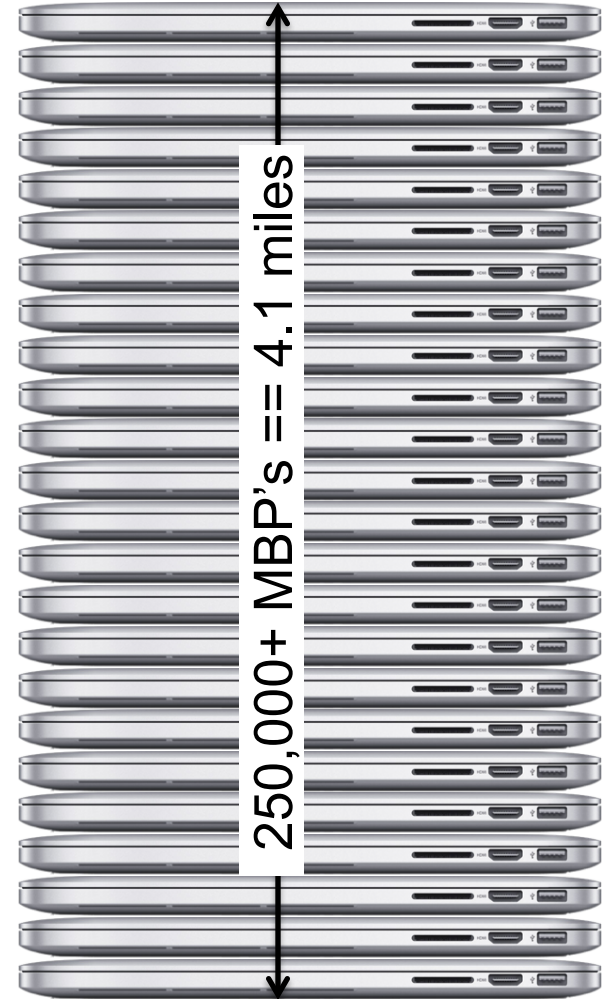
IBM Press Release 28 Aug, 2012

This Was a Problem for Google

2010 Search Index Size:
100,000,000 GB

New data added per day
100,000+ GB

Databases they could use
0



And for Facebook

The Facebook logo, consisting of the word "facebook" in white lowercase letters on a dark blue rectangular background.

2010: 13,000,000 queries per second

And for Facebook



facebook

2010: 13,000,000 queries per second

TPC #1 DB: 504161 tps

TPC Top Results

Rank	Company	System	Performance (tpmC)	Price/tpmC	Watts/KtpmC	System Availability	Database
1	ORACLE	SPARC SuperCluster with T3-4 Servers	30,249,688	1.01 USD	NR	06/01/11	Oracle Database 11g Enterprise Edition w/Partitioning
2	IBM	IBM Power 780 Server Model 9179-MHB	10,366,254	1.38 USD	NR	10/13/10	IBM DB2 9.7
3	ORACLE	Sun SPARC Enterprise T5440 Server Cluster	7,646,486	2.36 USD	NR	03/19/10	Oracle Database 11g Enterprise Edition w/RAC w/Partitioning
4	IBM	IBM Power 595 Server Model 9119-FHA	6,085,166	2.81 USD	NR	12/10/08	IBM DB2 9.5

And for Facebook



facebook

2010: 13,000,000 queries per second

TPC #1 DB: 504161 tps

TPC Top Results

Rank	Company	System	Performance (tpmC)	Price/tpmC	Watts/KtpmC	System Availability	Database
1	ORACLE	SPARC SuperCluster with T3-4 Servers	30,249,688	1.01 USD	NR	06/01/11	Oracle Database 11g Enterprise Edition w/Partitioning
2	IBM	IBM Power 780 Server Model 9119-MHB	10,366,254	1.38 USD	NR	10/13/10	IBM DB2 9.7
3	ORACLE	Sun SPARC Enterprise T5440 Server Cluster	7,646,486	2.36 USD	NR	03/19/10	Oracle Database 11g Edition w/RAC w/Partitioning
4	IBM	IBM Power 595 Server Model 9119-FHA	6,085,166	2.81 USD	NR	12/10/08	IBM DB2 9.5

Top 10 combined: 1,370,368 tps

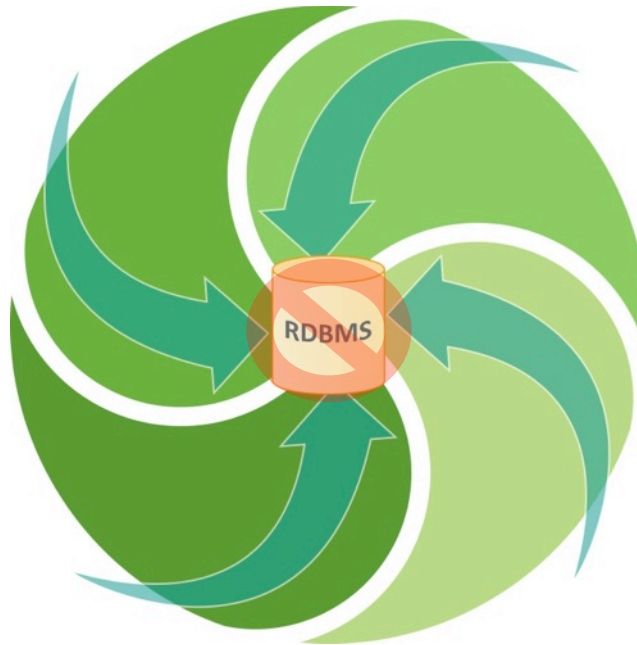
Big Data Changes Everything... for Everyone

Variety of Data

- Unstructured data
- Semi-structured data
- Polymorphic data

Volume/Velocity of Data

- Petabytes of data
- Trillions of records
- Millions of queries per second



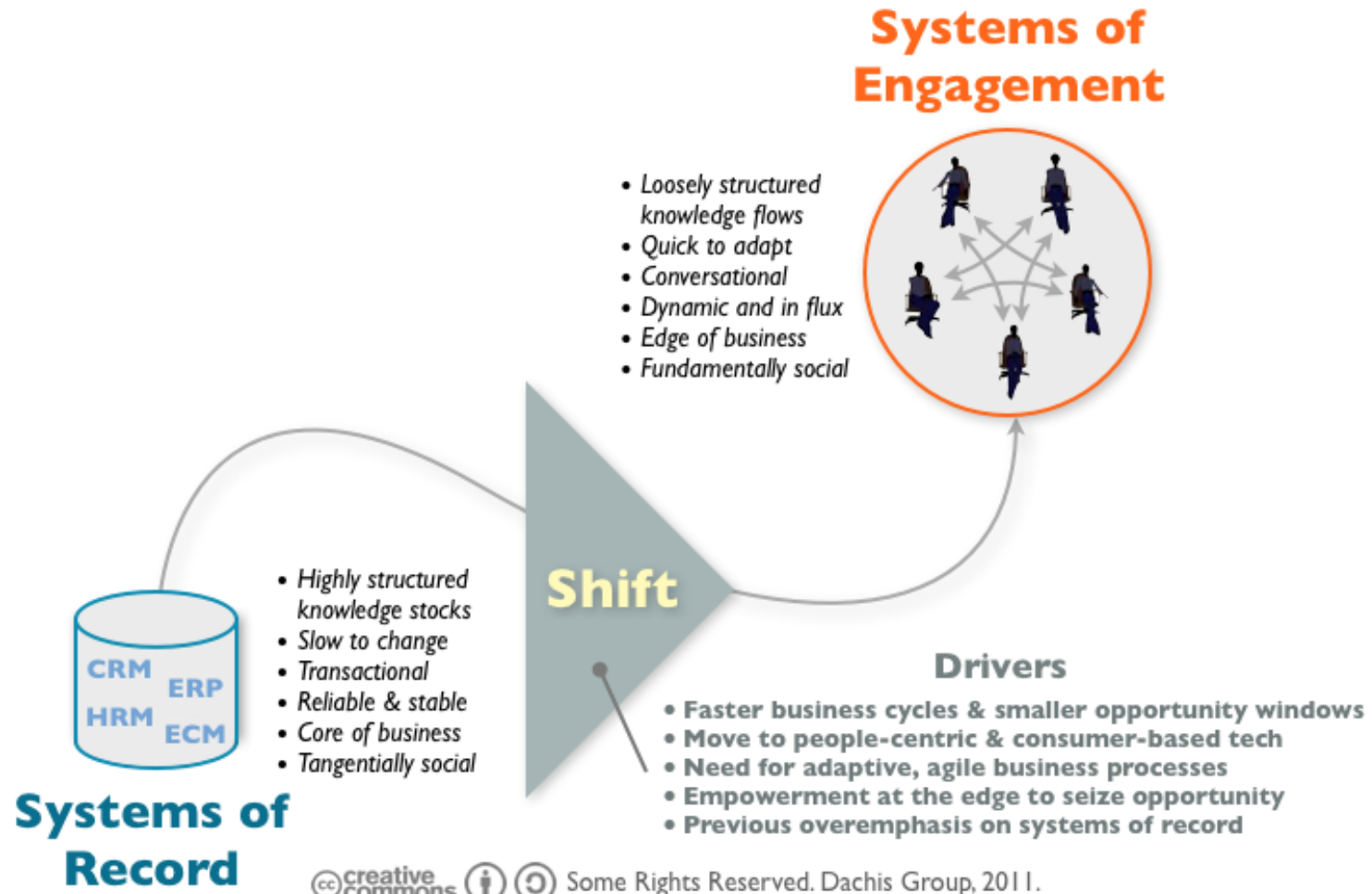
Agile Development

- Iterative
- Short development cycles
- New workloads

New Architectures

- Horizontal scaling
- Commodity servers
- Cloud computing

Shift in What We're Computing



Living in the Post-transactional Future



Order-processing systems largely “done” (RDBMS);
primary focus on better search and recommendations
or adapting prices on the fly (NoSQL)

Vast majority of its engineering is focused on
recommending better movies (NoSQL), not
processing monthly bills (RDBMS)



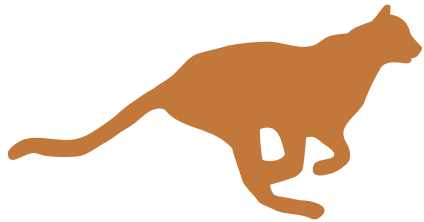
Easy part is processing the credit card (RDBMS).
Hard part is making it location aware, so it knows
where you are and what you’re buying (NoSQL)

Shift in How We Develop Applications

“Systems of Engagement are built by front-line developers using modern languages who are driven by time to market, the need for rapid deployment and iteration....They value solutions that make it easy for them to deploy their application code with as little friction as possible.” (Forrester 2013)



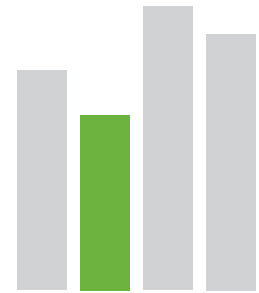
Why NoSQL?



Agile



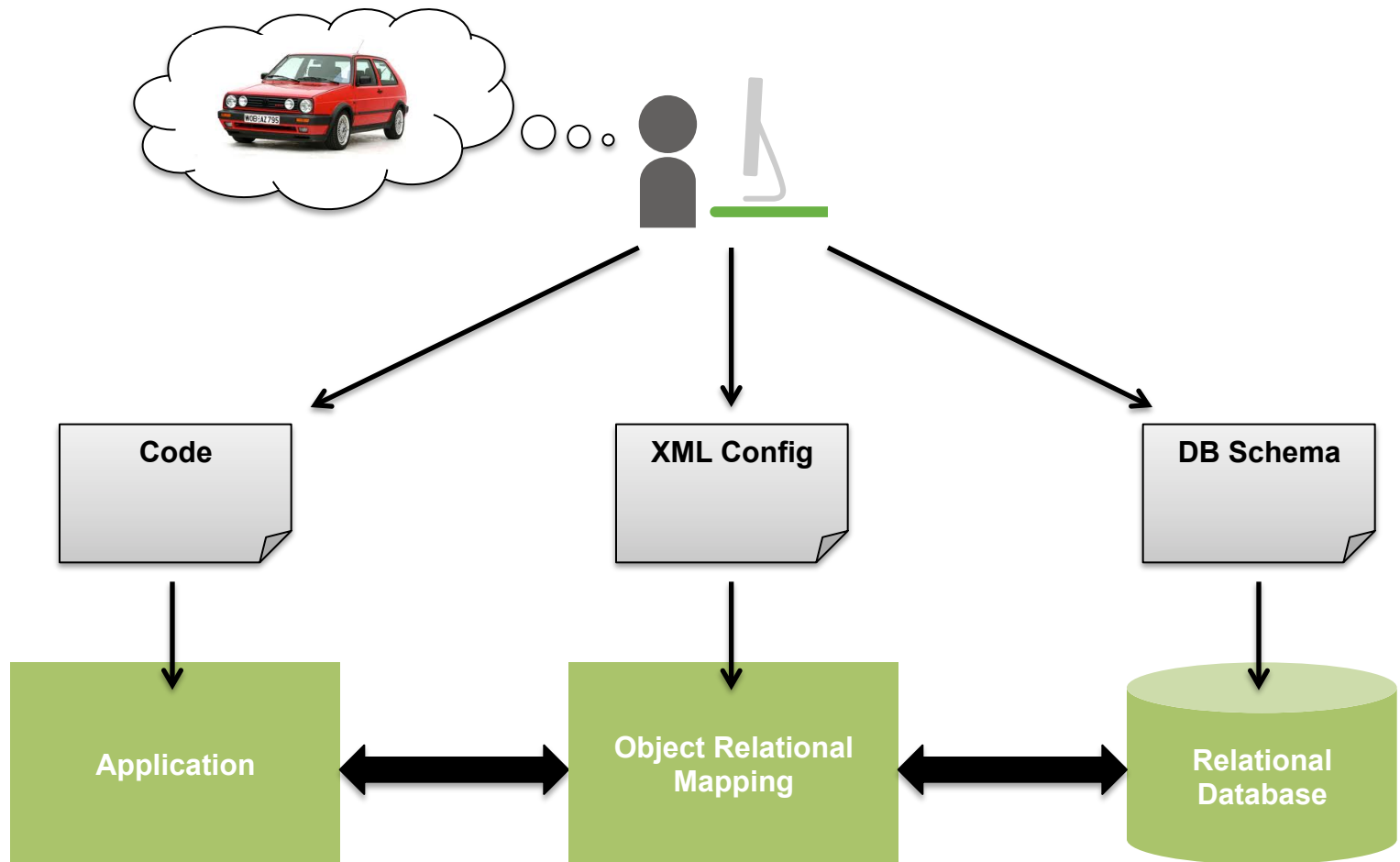
Scalable



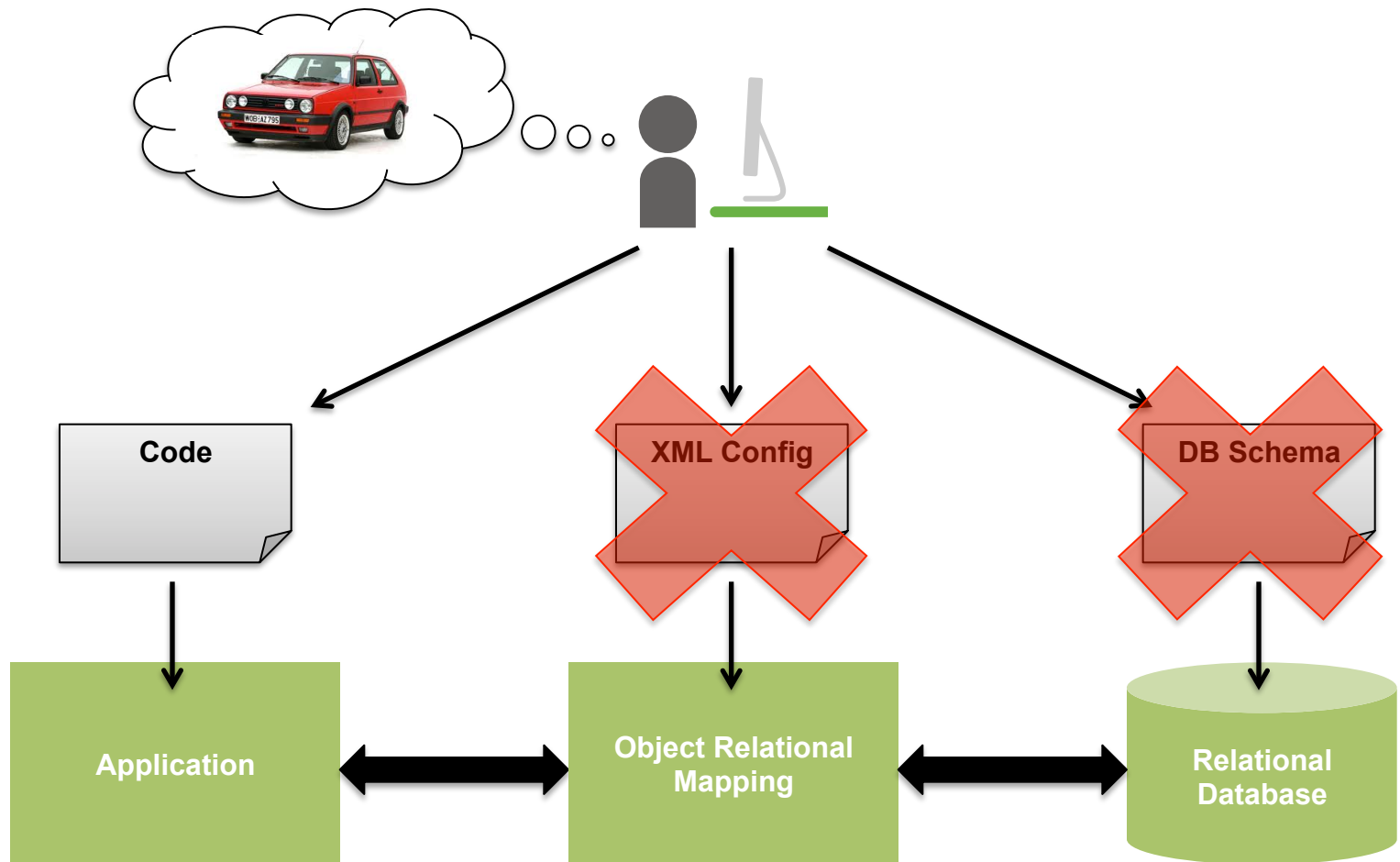
Lower TCO

what does this mean?

Developers Are More Productive



Developers Are More Productive



Developers? They Matter

Stephen O'Grady

The New Kingmakers

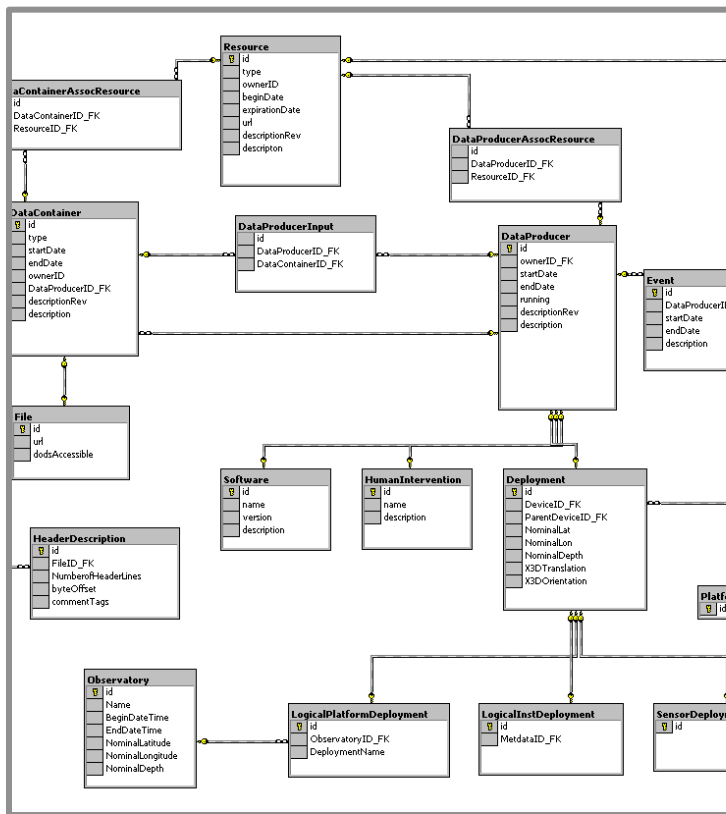
How **Developers**
Conquered the World



some nosql examples

Agility

RDBMS



MongoDB

```
{
  _id : ObjectId("4c4ba5e5e8aabf3"),
  employee_name: "Dunham, Justin",
  department : "Marketing",
  title : "Product Manager, Web",
  report_up: "Neray, Graham",
  pay_band: "C",
  benefits : [
    { type : "Health",
      plan : "PPO Plus" },
    { type : "Dental",
      plan : "Standard" }
  ]
}
```

Serves targeted content to users using MongoDB-powered identity system

Problem	Why MongoDB	Results
<ul style="list-style-type: none">• 20M+ unique visitors per month• Rigid relational schema unable to evolve with changing data types and new features• Slow development cycles	<ul style="list-style-type: none">• Easy-to-manage dynamic data model enables limitless growth, interactive content• Support for ad hoc queries• Highly extensible	<ul style="list-style-type: none">• Rapid rollout of new features• Customized, social conversations throughout site• Tracks user data to increase engagement, revenue

Scalability

Auto-Sharding



- Increase capacity as you go
- Commodity and cloud architectures
- Improved operational simplicity and cost visibility

Manages a wide range of content and services for its web properties using MongoDB

Problem	Why MongoDB	Results
<ul style="list-style-type: none">• Trouble dealing with a huge variety of content• MySQL unable to keep up with performance and scalability requirements• Problems compounded by integrating information from T-Mobile joint venture	<ul style="list-style-type: none">• Move from 6 billion rows in RDBMS to simplicity of 1 document• Automated failover and ability to add nodes without downtime• “Blazingly fast” query performance: “blown away by [MongoDB’s] performance”	<ul style="list-style-type: none">• Significant performance gains despite big increase in volume and variety of data• Greater agility, faster development iteration• Saved £2m in licenses and hardware

Better Total Cost of Ownership (TCO)

Developer/Ops Savings

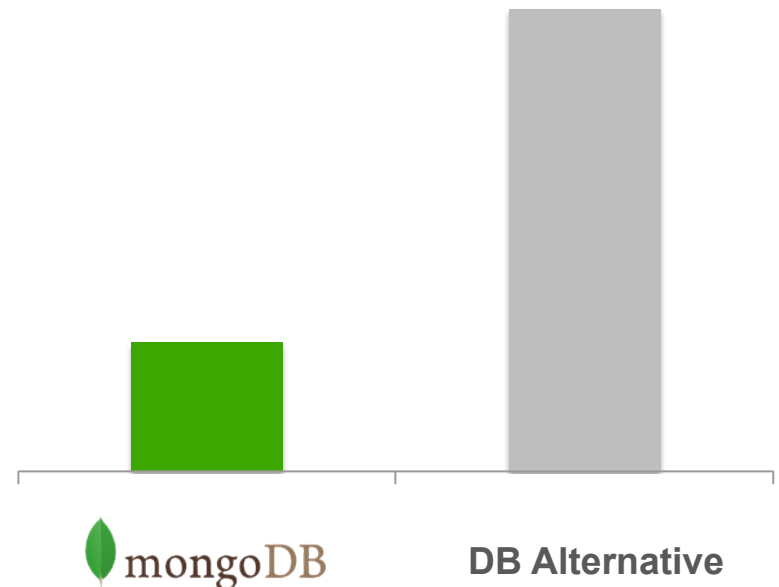
- Ease of Use
- Agile development
- Less maintenance

Hardware Savings

- Commodity servers
- Internal storage (no SAN)
- Scale out, not up

Software/Support Savings

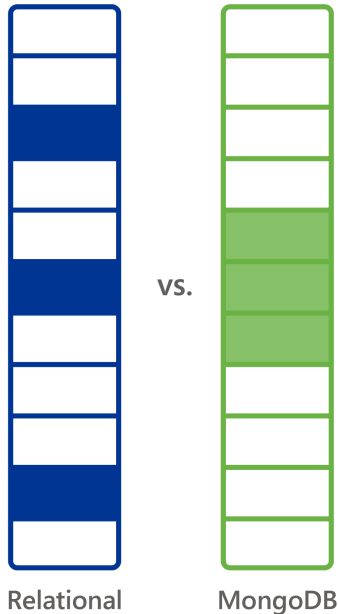
- No upfront license
- Cost visibility for usage growth



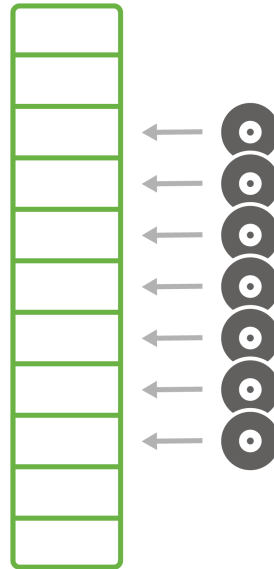
Stores one of world's largest record repositories and searchable catalogues in MongoDB

Problem	Why MongoDB	Results
<ul style="list-style-type: none">• One of world's largest record repositories• Move to SOA required new approach to data store• RDBMS could not support centralized data mgt and federation of information services	<ul style="list-style-type: none">• Fast, easy scalability• Full query language• Complex metadata storage• Delivers high scalability, fast performance, and easy maintenance, while keeping support costs low	<ul style="list-style-type: none">• Will scale to 100s of TB by 2013, PB by 2020• Searchable catalogue of varied data types• Decreased SW and support costs

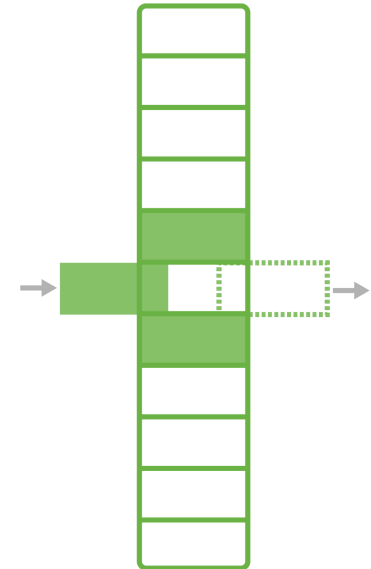
Performance



**Better Data
Locality**



**In-Memory
Caching**



**In-Place
Updates**

Uses MongoDB to safeguard over **6 billion** images served to millions of customers

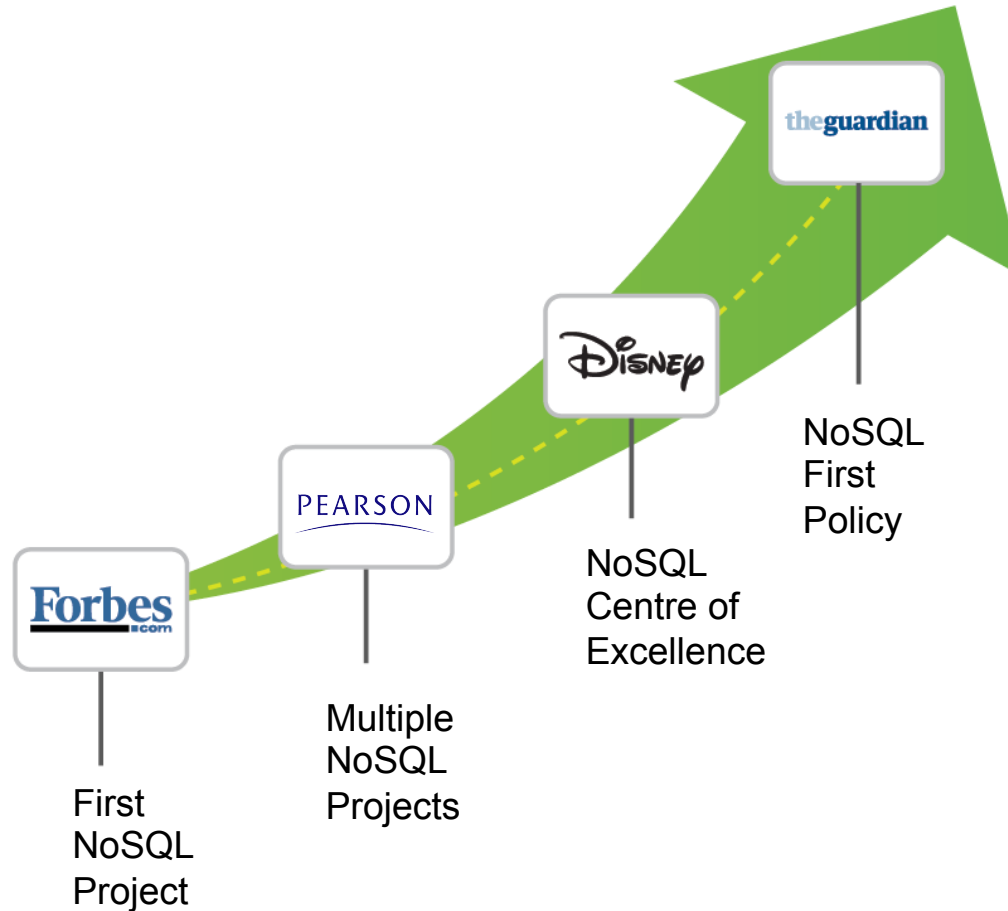
Problem	Why MongoDB	Results
<ul style="list-style-type: none">• 6B images, 20TB of data• Brittle code base on top of Oracle database – hard to scale, add features• High SW and HW costs	<ul style="list-style-type: none">• JSON-based data model• Agile, high performance, scalable• Alignment with Shutterfly's services-based architecture	<ul style="list-style-type: none">• 5x cost reduction• 9x performance improvement• Faster time-to-market• Dev cycles in weeks vs. tens of months

...the future

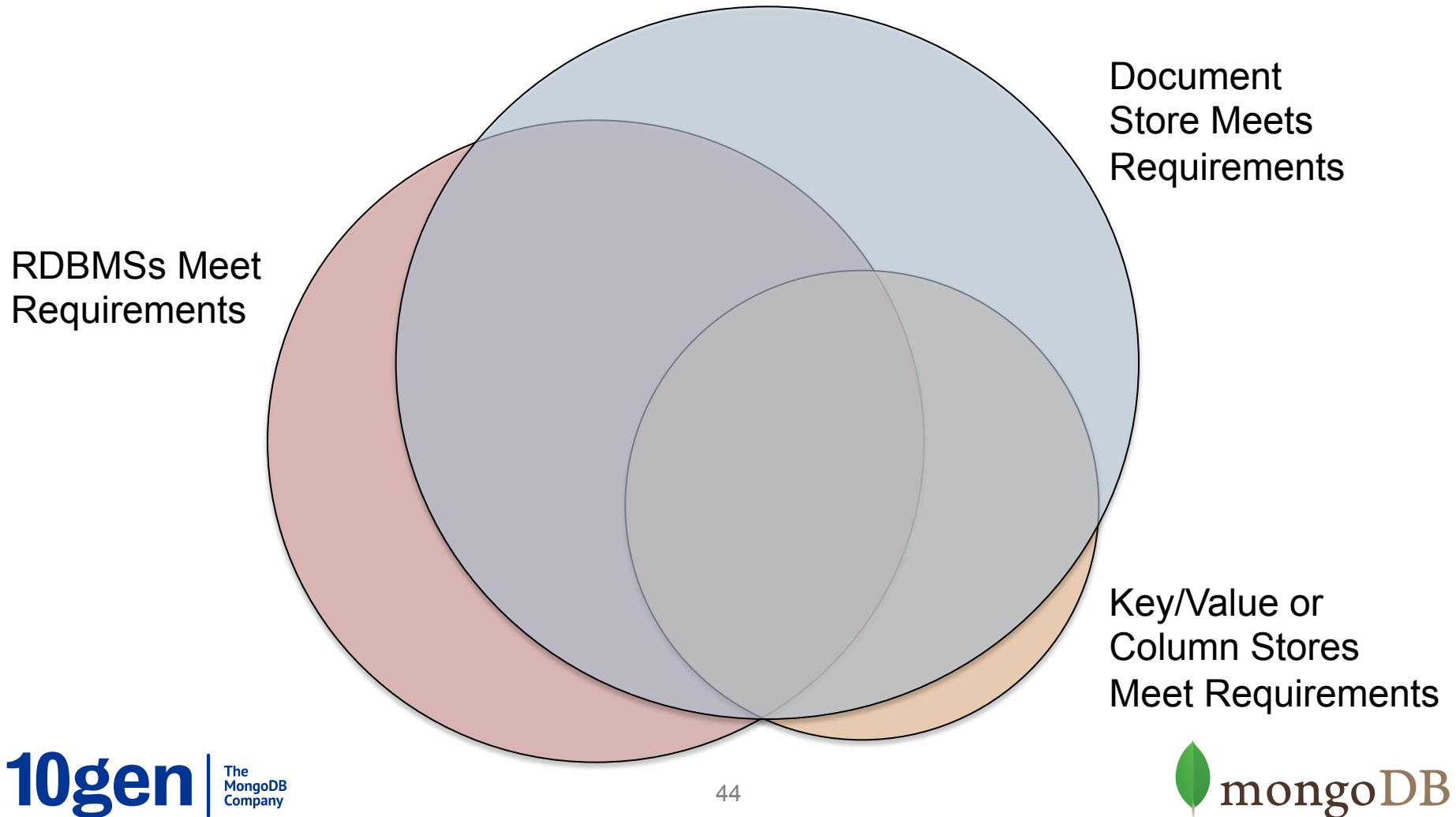
Leading Organizations Rely on NoSQL



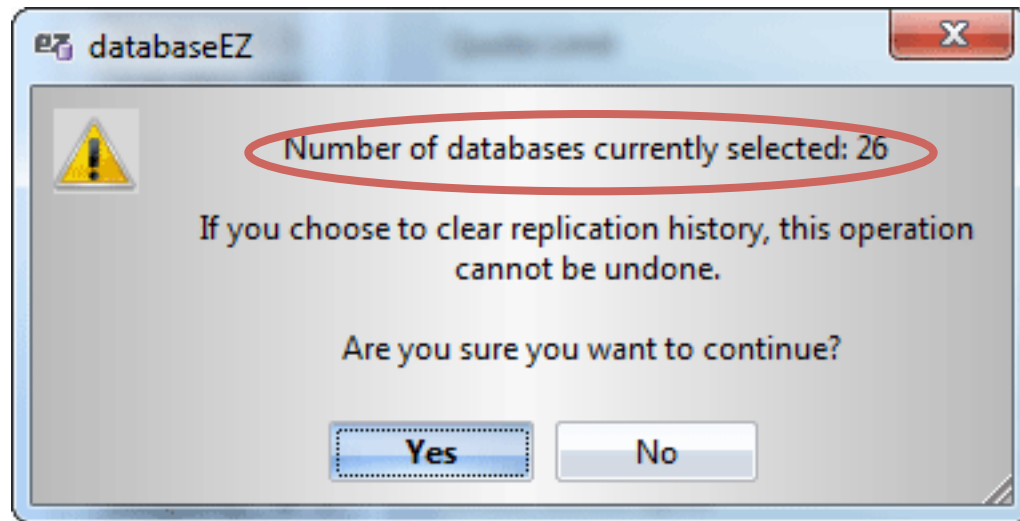
NoSQL Adoption



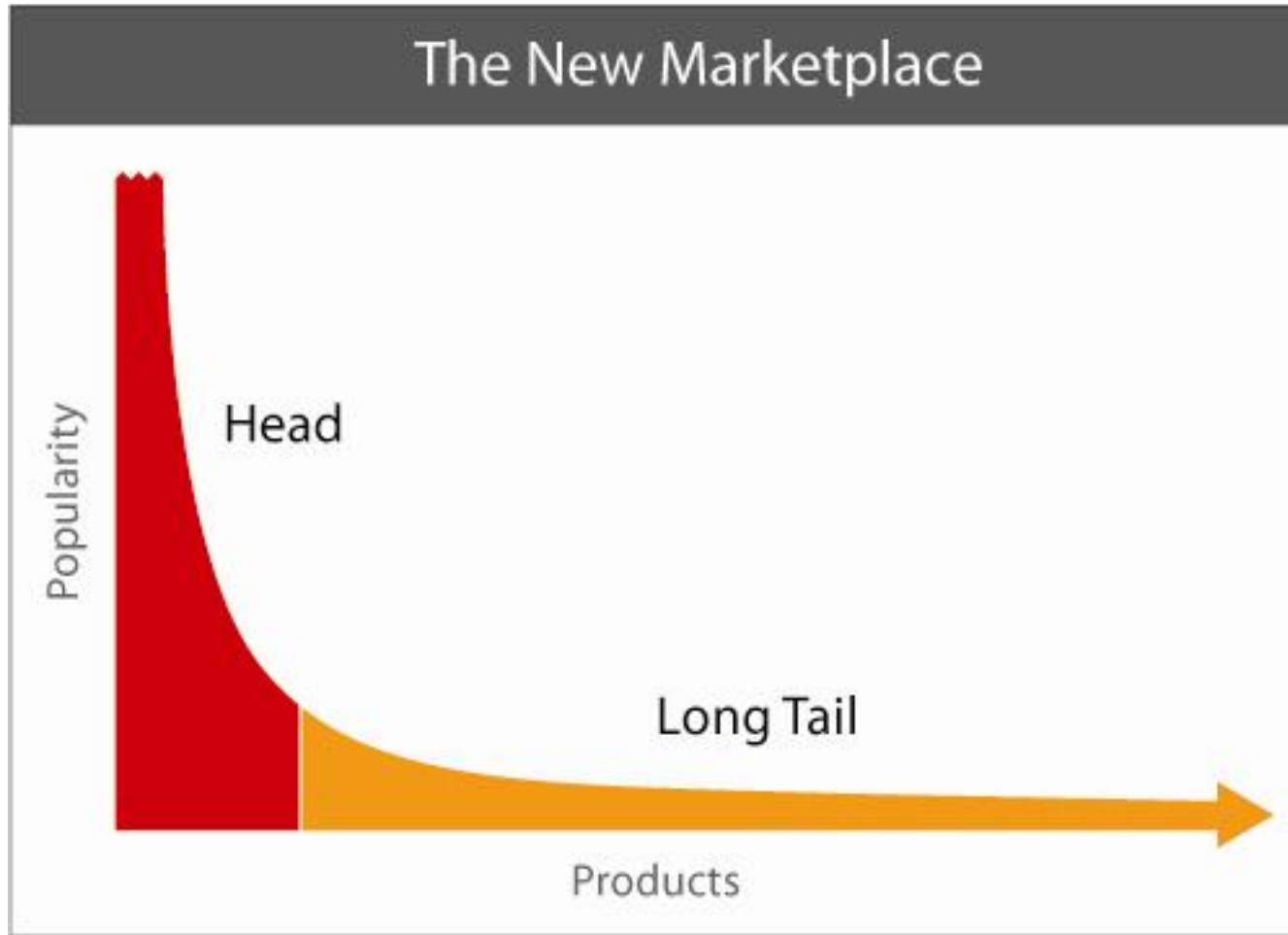
NoSQL: The New Normal



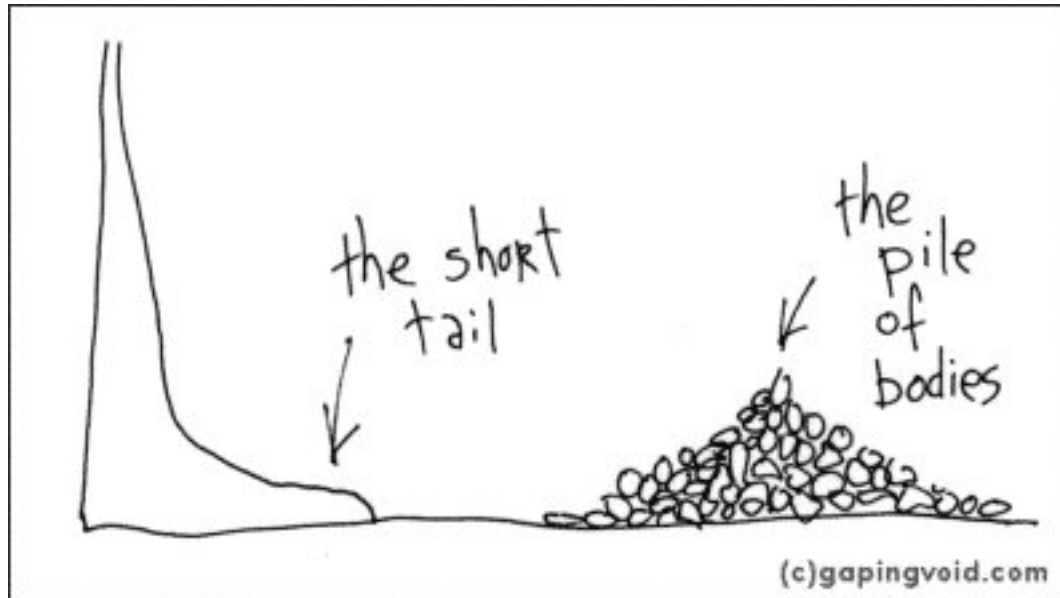
Is It a Polyglot Future?



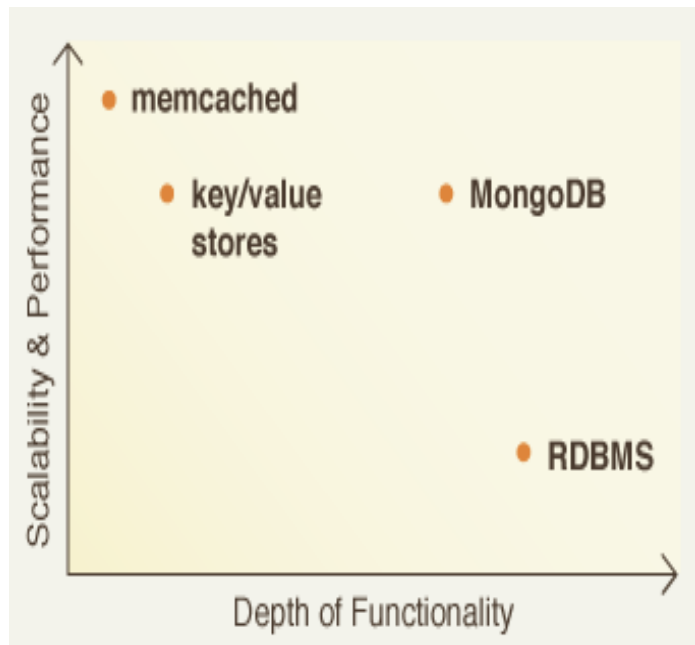
Remember the Long Tail?



It Didn't Work out Well



General Purpose, High Performance

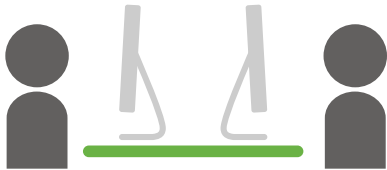


Database Popularity
Jobs, Searches, Mentions, Etc.

Rank	Database	Database Type	Score	% Change
1	Oracle	RDBMS	1567.93	+ 8.6
2	Microsoft SQL Server	RDBMS	1310.61	- 0.73
3	MySQL	RDBMS	1284.78	- 28.89
4	Microsoft Access	RDBMS	186.05	+ 15.12
5	PostgreSQL	RDBMS	183.47	+ 16
6	DB2	RDBMS	162.05	+ 8.39
7	MongoDB	Document store	116.14	+ 20.01
8	Sybase	RDBMS	84.52	+ 1.07
9	SQLite	RDBMS	81.01	+ 0.65
10	Solr	Search engine	47.35	
11	Cassandra	Wide column store	36.16	+ 1.24
12	Redis	Key-value store	32.03	+ 6.06
13	Informix	RDBMS	25.43	+ 1.52
14	Memcached	Key-value store	23.28	- 1.83
15	Hbase	Wide column store	20.74	+ 1.05

Source: DBEngines, Feb 2013

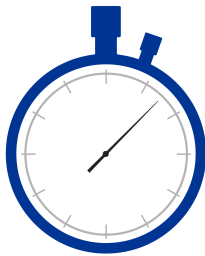
NoSQL Benefits



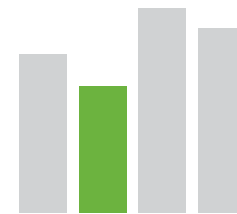
Increased Developer Productivity



Better User Experience



Faster Time to Market



Lower TCO

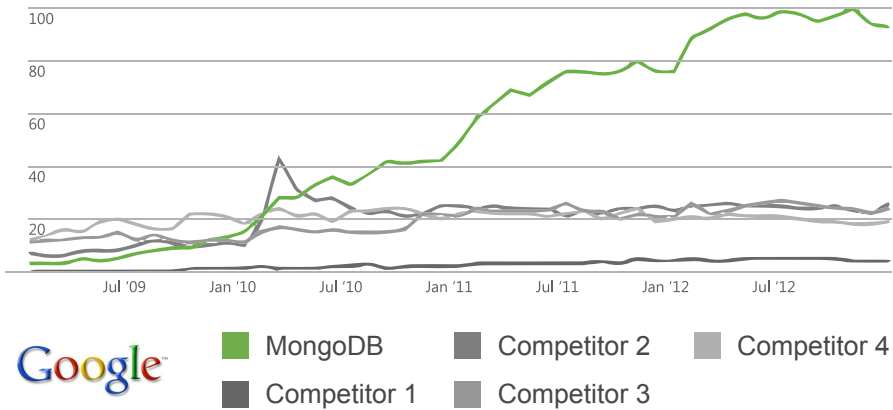
10gen | The
MongoDB
Company



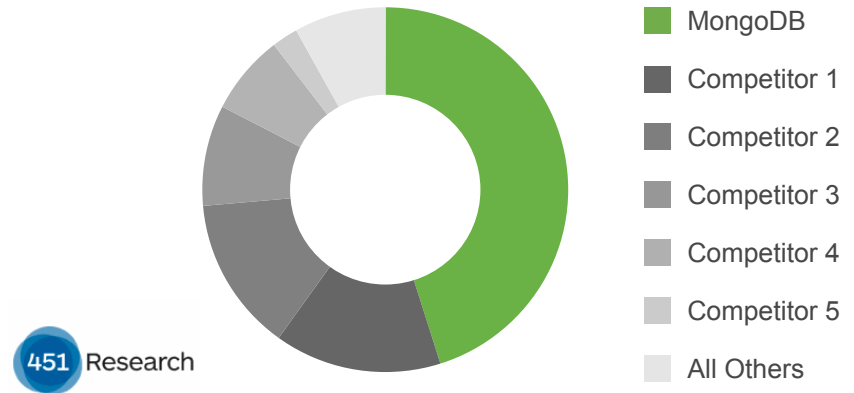
mongoDB

Leading NoSQL Database

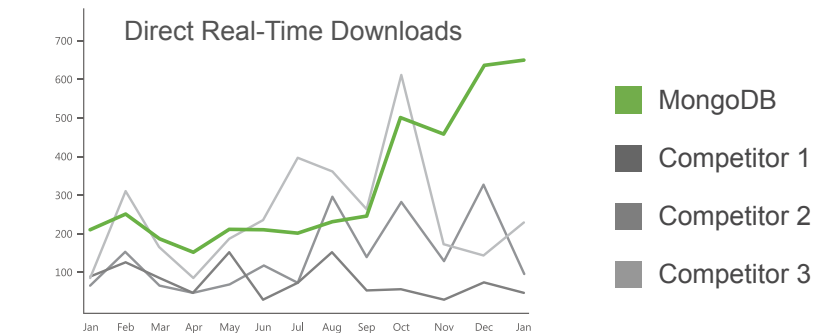
Google Search



LinkedIn Job Skills



Jaspersoft Big Data Index



Indeed.com Trends

