NoSQL, NewSQL, Big Data… Total Data

The Future of Enterprise Data Management

NoSQL Road Show

Matthew Aslett, The 451 Group
matthew.aslett@the451group.com

NEW YORK · SAN FRANCISCO · LONDON · BOSTON · WASHINGTON DC · SEATTLE · DENVER
WWW.THE451GROUP.COM

© 2011 by The 451 Group. All rights reserved

Overview

- NoSQL and NewSQL databases
  - Adoption and development drivers
- Big Data and Total Data
  - Definition and implications

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

The 451 Group

451 Research is focused on the business of enterprise IT innovation. The company's analysts provide critical and timely insight into the competitive dynamics of innovation in emerging technology segments.

Tier1 Research is a single-source research and advisory firm covering the multi-tenant datacenter, hosting, IT and cloud-computing sectors, blending the best of industry and financial research.

The Uptime Institute is 'The Global Data Center Authority' and a pioneer in the creation and facilitation of end-user knowledge communities to improve reliability and uninterruptible availability in datacenter facilities.

TheInfoPro is a leading IT advisory and research firm that provides real-world perspectives on the customer and market dynamics of the enterprise information technology landscape, harnessing the collective knowledge and insight of leading IT organizations worldwide.

ChangeWave Research is a research firm that identifies and quantifies 'change' in consumer spending behavior, corporate purchasing, and industry, company and technology trends.

© 2011 by The 451 Group. All rights reserved

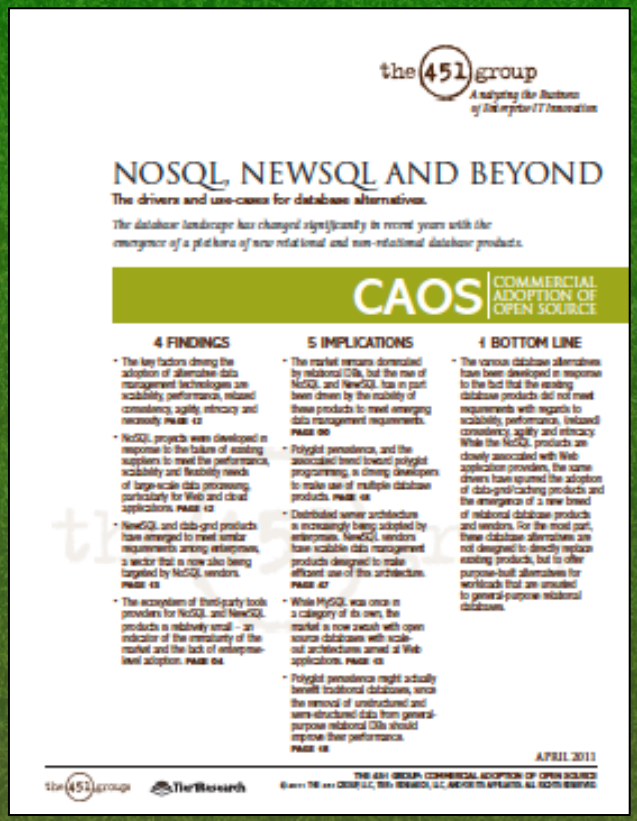
Coverage areas

- Matthew Aslett
  - Senior analyst, enterprise software
  - With The 451 Group since 2007
  - www.about.me/mattaslett
  - www.twitter.com/maslett
- Commercial Adoption of Open Source (CAOS)
  - Adoption by enterprises
  - Adoption by vendors
- Information Management
  - Database
  - Data warehousing
  - Data caching

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

Relevant reports

- NoSQL, NewSQL and Beyond
  - Assessing the drivers behind the development and adoption of NoSQL and NewSQL databases, as well as data grid/ caching technologies
  - Released April 2011
  - Role of open source in driving innovation
  - sales@the451group.com



the 451 group

NOSQL, NEWSQL AND BEYOND

The drivers and use-cases for database alternatives.

The database landscape has changed significantly in recent years with the emergence of a plethora of new relational and non-relational database products.

CAOS COMMERCIAL ADOPTION OF OPEN SOURCE

4 FINDINGS

- The key factors driving the adoption of alternative data management technologies are scalability, performance, relaxed consistency, agility, intricacy and necessity.
- NoSQL projects were developed in response to the failure of existing suppliers to meet the performance, scalability and flexibility needs of large-scale data processing, particularly for Web and cloud applications.
- NewSQL and data-grid products have emerged to meet similar requirements among enterprises, a sector that is now also being targeted by NoSQL vendors.
- The ecosystem of third-party tools providers for NoSQL and NewSQL products is relatively small – an indicator of the immaturity of the market and the lack of enterprise-level adoption.

5 IMPLICATIONS

- The market remains dominated by relational DBs, but the rise of NoSQL and NewSQL has in part been driven by the inability of these products to meet emerging data management requirements.
- Polyglot persistence, and the associated trend toward polyglot programming, is driving developers to make use of multiple database products.
- Distributed server architecture is increasingly being adopted by enterprises. NewSQL vendors have scalable data management products designed to make efficient use of this architecture.
- While MySQL was once in a category of its own, the market is now awash with open source databases with scale-out architectures aimed at Web applications.
- Polyglot persistence might actually benefit traditional databases, since the removal of unstructured and semi-structured data from general-purpose relational DBs should improve their performance.

1 BOTTOM LINE

- The various database alternatives have been developed in response to the fact that the existing database products did not meet requirements with regards to scalability, performance, (relaxed) consistency, agility and intricacy. While the NoSQL products are closely associated with Web application providers, the same drivers have spurred the adoption of data-grid/caching products and the emergence of a new breed of relational database products and vendors. For the most part, these database alternatives are not designed to directly replace existing products, but to offer purpose-built alternatives for workloads that are unsuited to general-purpose relational databases.

APRIL 2011

the 451 group | TierResearch

THE 451 GROUP: COMMERCIAL ADOPTION OF OPEN SOURCE
© 2011 THE 451 GROUP, TIER1 RESEARCH, LLC AND/OR ITS AFFILIATES. ALL RIGHTS RESERVED

© 2011 by The 451 Group. All rights reserved

NoSQL, NewSQL and Beyond

NoSQL

- New breed of non-relational database products
- Rejection of fixed table schema and join operations
- Designed to meet scalability requirements of distributed architectures
- And/or schema-less data management requirements

NewSQL

- New breed of relational database products
- Retain SQL and ACID
- Designed to meet scalability requirements of distributed architectures
- Or improve performance so horizontal scalability is no longer a necessity

… and Beyond

- In-memory data grid/cache products
- Potential primary platform for distributed data management

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch

© 2011 by The 451 Group. All rights reserved

NoSQL, NewSQL and Beyond

NoSQL

- Big tables – data mapped by row key, column key and time stamp
- Key-value stores - store keys and associated values
- Document store - stores all data as a single document
- Graph databases - use nodes, properties and edges to store data and the relationships between entities

NewSQL

- MySQL storage engines - scale-up and scale-out
- Transparent sharding - reduce to manual effort required to scale
- Appliances - take advantage of improved hardware performance, solid state drives
- New databases - designed specifically for scale-out

Data grid/cache

- spectrum of data management capabilities, from non-persistent data caching to persistent caching, replication, and distributed data and compute grid

© 2011 by The 451 Group. All rights reserved

SPRAINED RELATIONAL DATABASES

Photo credit:
Foxtongue on Flickr
http://www.flickr.com/photos/foxtongue/4844016087/

451 research | Tier1Research | UptimeInstitute | THE INFOPRO | ChangeWaveResearch

© 2011 by The 451 Group. All rights reserved

Scalability

Hardware economics – scale out

- Example project/service/vendor:
  - BigTable, HBase, Riak, MongoDB, Couchbase, Hadoop
  - Amazon RDS, Xeround, SQL Azure, NimbusDB
  - Data grid/cache

- Associated use case:
  - Large-scale distributed data storage
  - Analysis of continuously updated data
  - Multi-tenant PaaS data layer

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

Performance

MySQL limitations, performance at scale

- Example project/service/vendor:
- Hypertable, Couchbase, Riak, Membrain, MongoDB, Redis
- Data grid/cache
- VoltDB, Clustrix

- Associated use case:
- Real time data processing of mixed read/write workloads
- Data caching
- Large-scale data ingestion

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

Relaxed consistency

CAP Theorem – relax consistency to maintain availability

- Example project/service/vendor:
- Dynamo, Voldemort, Cassandra, Riak
- Amazon SimpleDB

- Associated use case:
- Multi-data center replication
- Service availability
- Non-transactional data off-load

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

Agility

Polyglot persistence, schema-less

- Example project/service/vendor:
  - MongoDB, CouchDB, Cassandra, Riak
  - Google App Engine, SimpleDB, SQL Azure
- Associated use case:
  - Mobile/remote device synchronization
  - Agile development
  - Data caching

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

Intricacy

Big data, total data

- Example project/service/vendor:
  - Neo4j, GraphDB, InfiniteGraph
  - Apache Cassandra, Hadoop, Riak
  - VoltDB, Clustrix
- Associated use case:
  - Social networking applications
  - Geo-locational applications
  - Configuration management database

© 2011 by The 451 Group. All rights reserved

Necessity

Open source

- The failure of existing suppliers to address emerging requirements

- Example projects:
  - BigTable: Google
  - Dynamo: Amazon
  - Cassandra: Facebook
  - HBase: Powerset
  - Voldemort: LinkedIn
  - Hypertable: Zvents
  - Neo4j: Windh Technologies

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

Big Data… Total Data

451 research | Tier1Research | UptimeInstitute | THE INFOPRO | ChangeWaveResearch

© 2011 by The 451 Group. All rights reserved

Current data management trends

The volume, variety and velocity of data is growing rapidly

Data processing capabilities have never been better

The value of data has never been better understood

RISK → **OPPORTUNITY**

- The data deluge problem is also a big data opportunity

451 research | Tier1Research | UptimeInstitute | THE INFOPRO | ChangeWaveResearch

© 2011 by The 451 Group. All rights reserved

What is Big Data?

- More than just rising data volumes
- Big Data ≠ Volume

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

What is Big Data?

- Also variety of data types/sources and velocity of data updates
- Big Data = Volume ± Variety ± Velocity

© 2011 by The 451 Group. All rights reserved

Current data management trends

The volume, variety and velocity of data is growing rapidly

'Big Data' covers a diverse set of products that can be applied to different problems

Data processing capabilities have never been better

The value of data has never been better understood

RISK → **OPPORTUNITY**

- 'Big Data' highlights the problem – volume/variety/velocity,
- and promises a solution – value,
- but doesn't provide a path in between

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

What is Total Data?

- Not just another name for Big Data
- It's about how you deliver value from that data

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

What is Total Data?

- Also the desire of the user to store and process all their data
- Value = (Volume ± Variety ± Velocity) x Totality

© 2011 by The 451 Group. All rights reserved

What is Total Data?

- And the desire to explore their data for new value

- Value = (Volume ± Variety ± Velocity) x (Totality + Exploration)

© 2011 by The 451 Group. All rights reserved

What is Total Data?

- Within tolerable time frames

- Value = $\frac{(\text{Volume} \pm \text{Variety} \pm \text{Velocity}) \times (\text{Totality} + \text{Exploration})}{\text{Time}}$

451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

What is Total Data?

- Within tolerable time frames
- Value = $\frac{(\text{Volume} \pm \text{Variety} \pm \text{Velocity}) \times (\text{Totality} + \text{Exploration})}{\text{Time}}$
- Total Data is making the most efficient use of existing and new data management resources to deliver value from data
- If your data is big, the way you manage it should be total
- Inspired by 'Total Football'

© 2011 by The 451 Group. All rights reserved

TOTAL FOOTBALL

Source: Wikimedia. Attribution: Bundesarchiv, Bild 183-N0716-0314 / Mittelstädt, Rainer / CC-BY-SA
http://commons.wikimedia.org/wiki/File:Bundesarchiv_Bild_183-N0716-0314,_Fu%C3%9Fball-WM,_BRD_-_Niederlande_2-1.jpg

© 2011 by The 451 Group. All rights reserved

Total Football meets Total Data

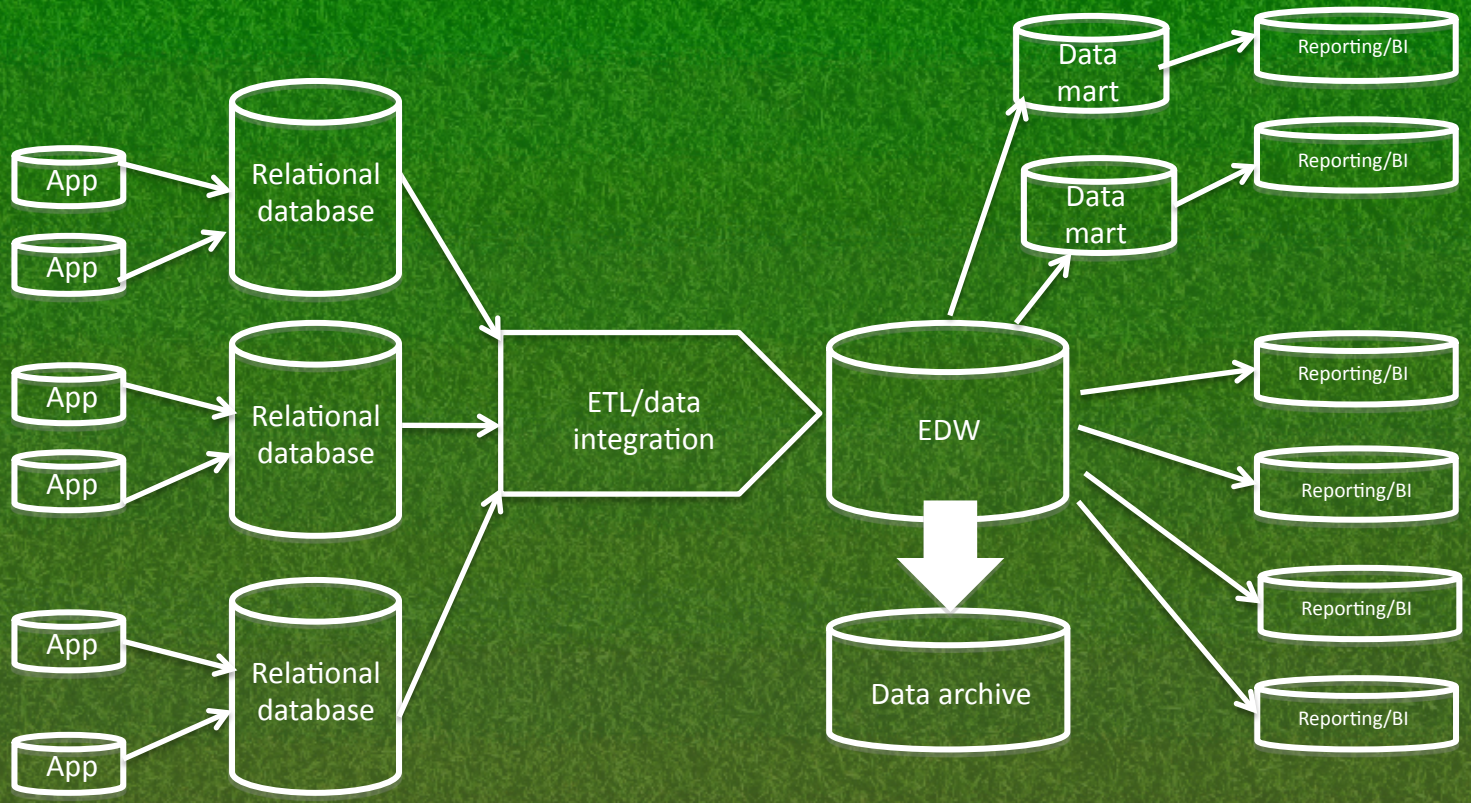
- "You make space, you come into space. And if the ball doesn't come, you leave this space and another player will come into it."

  Bernadus Hulshoff, Ajax 1966-77

- Abandonment of restrictive (self-imposed) rules about individual roles and responsibility
- Promotion of individuality within the overall context of the system
- Enabled and relied on fluidity and flexibility to respond to changing requirements
- Reliant on, and exploited, improved performance levels

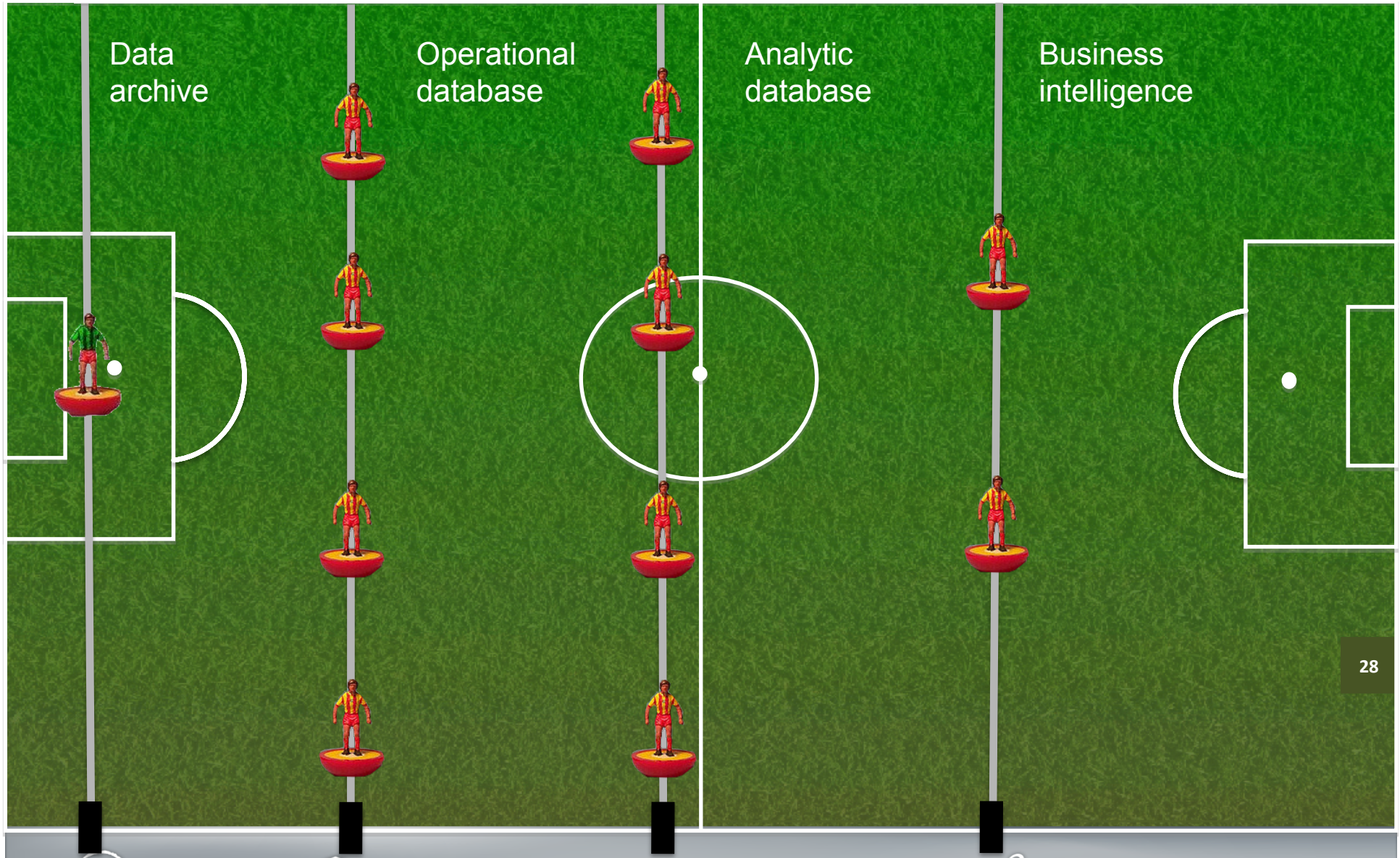
© 2011 by The 451 Group. All rights reserved

The old way



451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

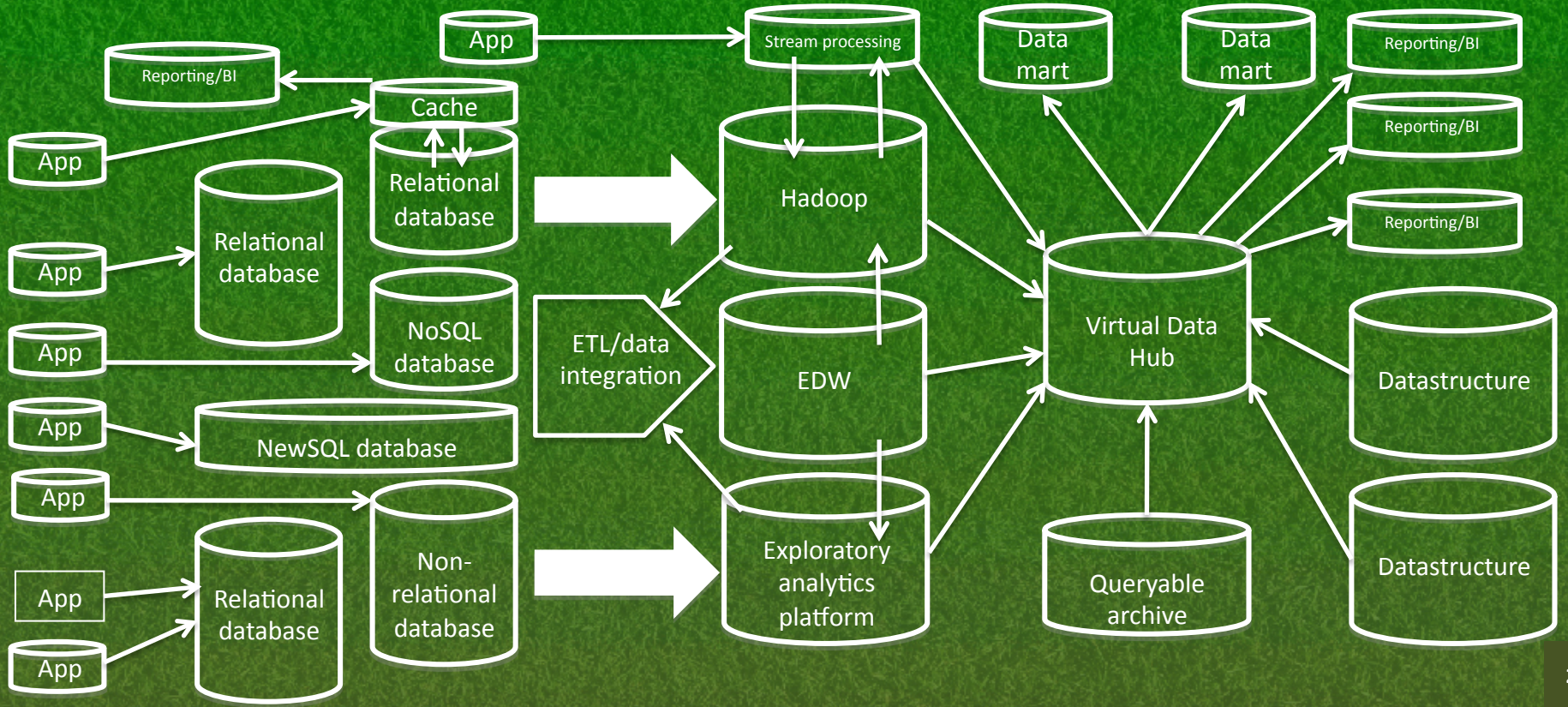
The old way



451 research | Tier1Research | UptimeInstitute | THE INFOPRO | ChangeWaveResearch

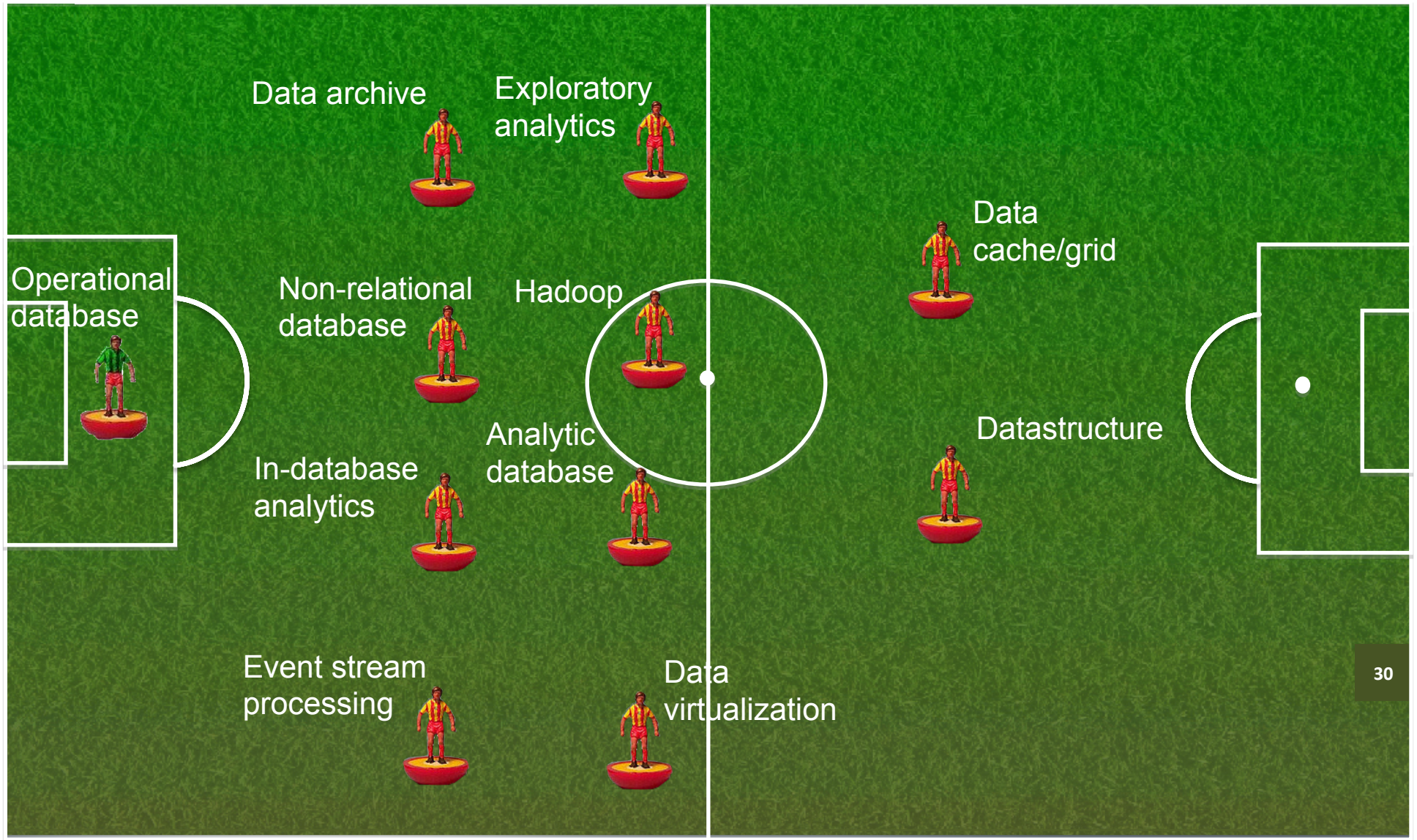
© 2011 by The 451 Group. All rights reserved

The new way



451 research Tier1Research UptimeInstitute THE INFOPRO ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

The new way



451 research | Tier1Research | UptimeInstitute | THE INFOPRO | ChangeWaveResearch
© 2011 by The 451 Group. All rights reserved

Relevant reports

- Total Data
  - Explaining the the total data management approach to dealing with the impact of big data on the data management landscape
  - Coming late 2011
  - sales@the451group.com

- Free copy for completing our Total Data survey:
- www.bit.ly/451data

© 2011 by The 451 Group. All rights reserved

451 research | Tier1Research | UptimeInstitute | THE INFOPRO | ChangeWaveResearch

© 2011 by The 451 Group. All rights reserved