The Amadeus NoSQL saga from 1992 to 2016: highlights

Abstract

_ In this talk, Jeremy will lead you through the different steps of NoSQL within Amadeus - one of the largest European data processing centers, handling billions of database actions per day over massive critical data updated in real time.

_ With intent to pragmatically assist you to assess what would be the right NoSQL strategy for your business, spotlight will be on showing the successive maturity steps and on sharing experiences with IT enthusiasts to take the most benefit from NoSQL - in case, it fits! Indeed, you will as well receive hints on how to spot if NoSQL actually deserves interest from you.

_ To conclude, Jeremy will share the R&D Amadeus view for the coming years for NoSQL technologies within Amadeus IT group and within the IT industry.

http://www.telecom-valley.fr/pagecmsblog/01-07-big-data-nosql-bi-securite-cloud

http://www.telecom-valley.fr/pagecmsblog/j%C3%A9r%C3%A9my-meyer

AMADEUS

AMADEUS

The Amadeus
NoSQL saga

from 1992 to 2016: highlights

Jeremy Meyer
Architecture, Quality and Governance
Data Store Middleware
Telecom Valley SophiaConf 2014
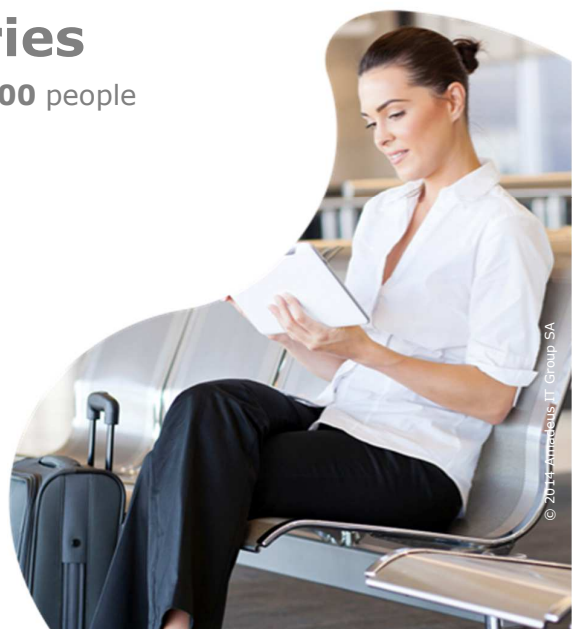Sophia Antipolis, 1st of July 2014

1 —— Amadeus

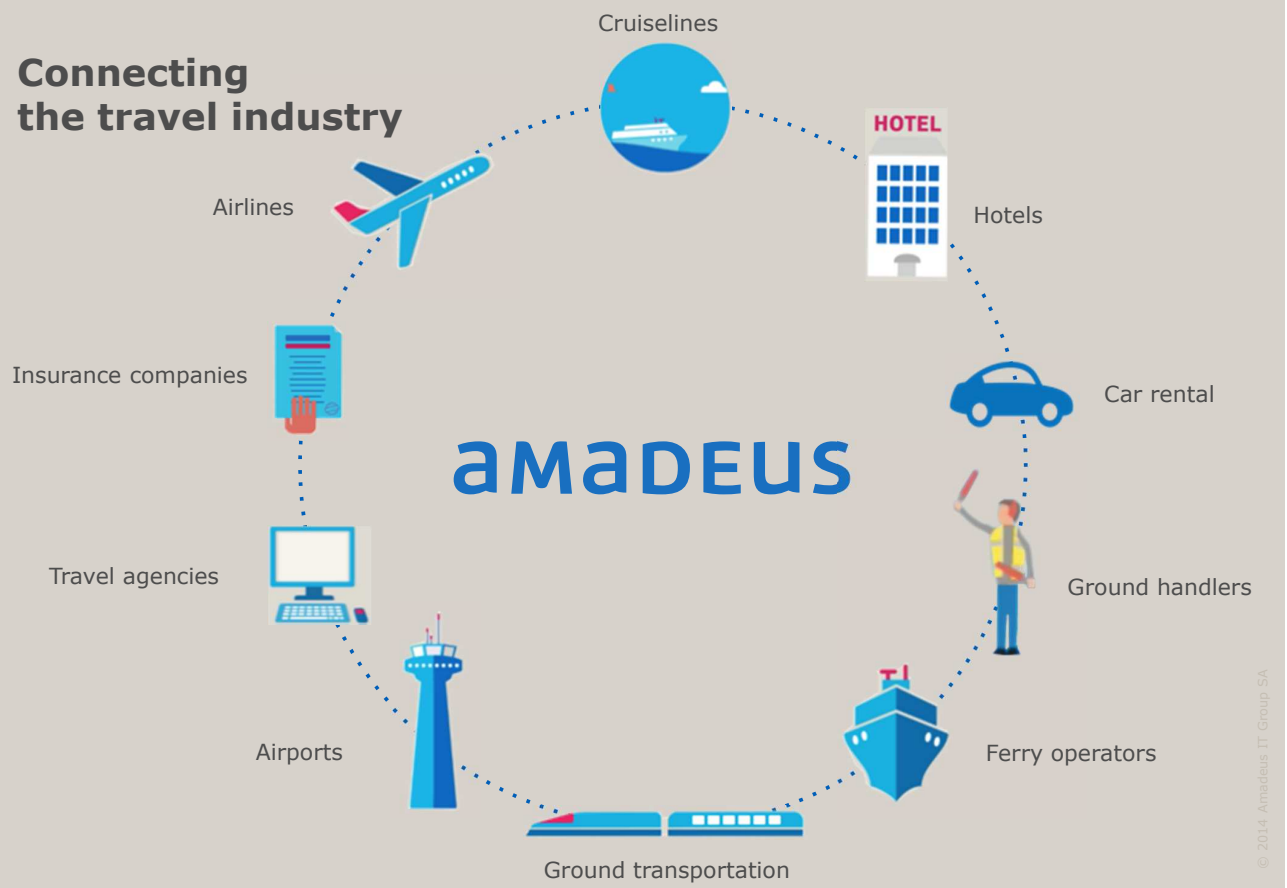**amadeus**

Amadeus in a few words

Amadeus is a technology company dedicated to the
global travel industry

We are present in **195 countries**

with a worldwide team of more than **11,000** people

Connecting the travel industry

- Cruiselines
- Hotels
- Car rental
- Ground handlers
- Ferry operators
- Ground transportation
- Airports
- Travel agencies
- Insurance companies
- Airlines

amadeus

© 2014 Amadeus IT Group SA

Our commitment to innovation

_ Amadeus has invested €2.9bn in **Research & Development** since 2004

Robust global operations

Central Data Centre in Erding, Germany

Erding

1.6+
billion
transactions
processed per day

502+
million
travel agency
bookings
processed in 2013

615+
million
Passengers
Boarded (PBs)
in 2013

95%
of the world's
scheduled
network airline
seats

© 2014 Amadeus IT Group SA

amadeus

2

Our view on NoSQL

© 2014 Amadeus IT Group SA

amadeus

Our definition of NoSQL

NoSQL movement is about
- *promoting alternative storage solutions to complement relational DBs*
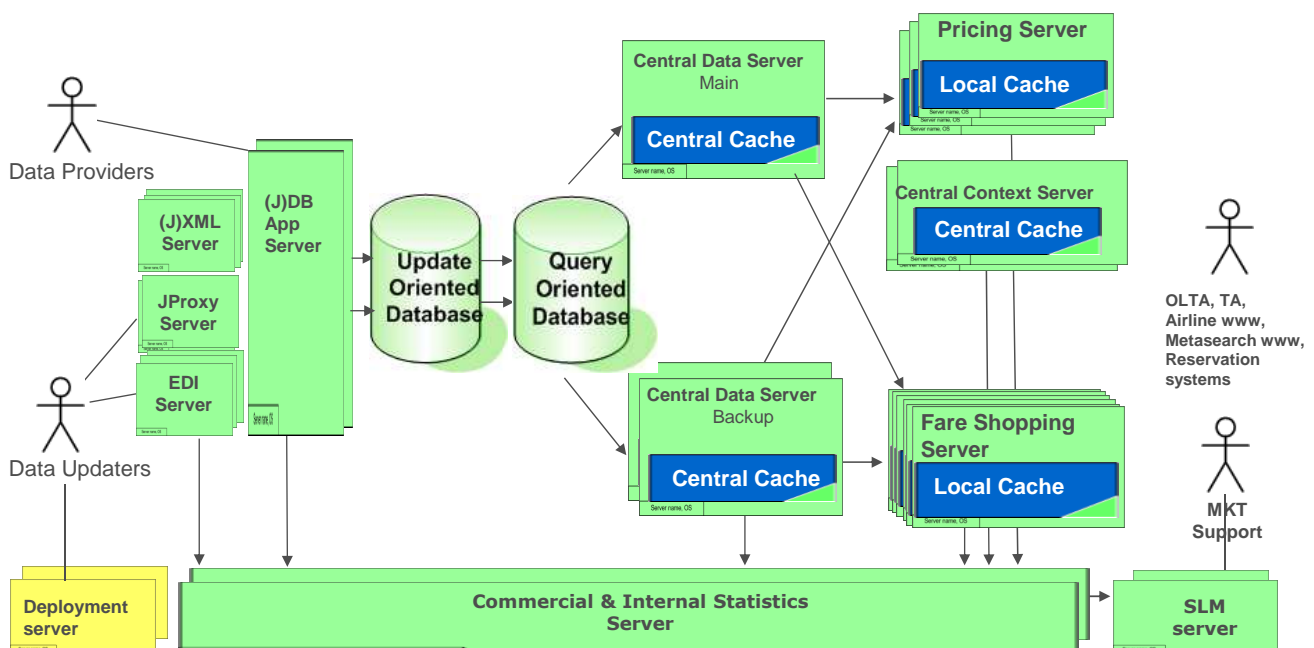- *turning them into commodity products and services*

_ NoSQL is not "new"
- NoSQL in Amadeus since 1992 through IBM TPF mainframes for Bookings
- Home-made
  - two-layer object cache since 2000 for Pricing and Shopping services
  - in-memory GraphDB since 2001 for building routes of the Journeys
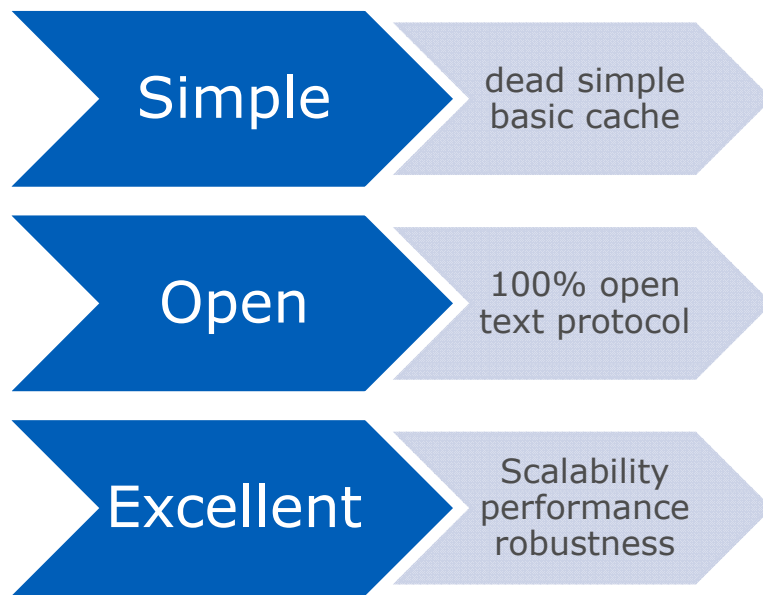- Memcached widely used since 2009 for Flights Inventory

aMaDEUS

2001: Pricing and Shopping
2-level Key/Value object cache

2009: Memcached in Amadeus
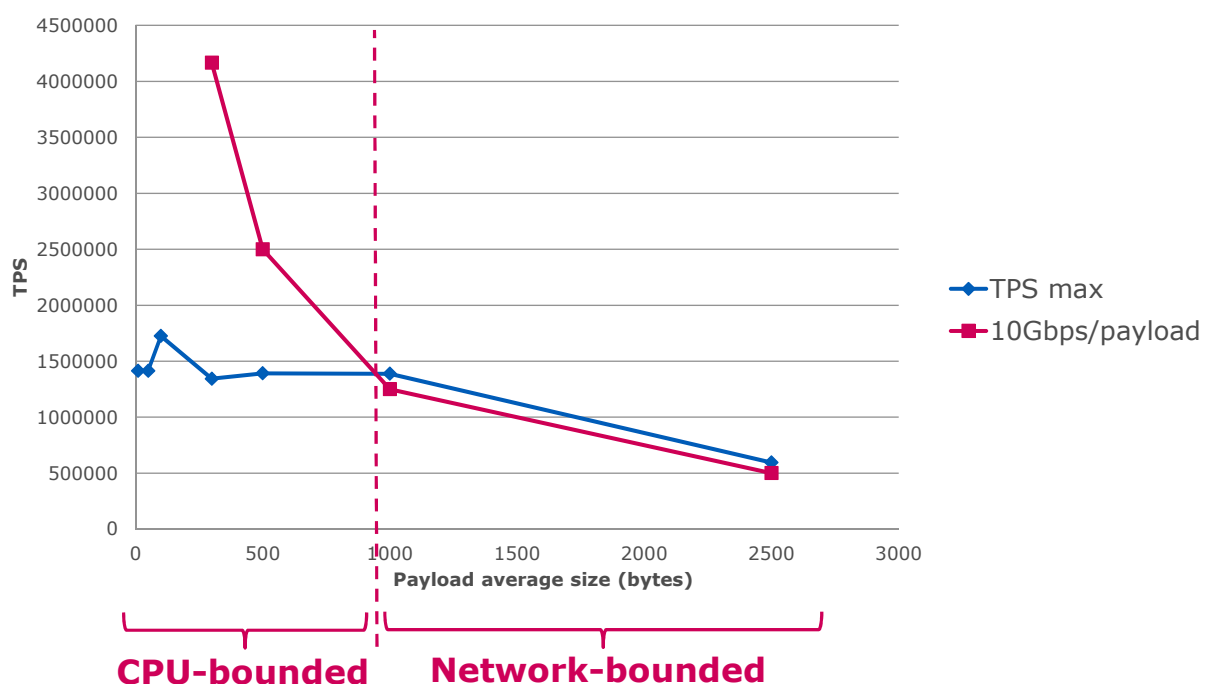## First open-source NoSQL solution in Amadeus

Simple — dead simple basic cache

Open — 100% open text protocol

Excellent — Scalability performance robustness

Accurate representative of NoSQL products !

11

AMADEUS

Max throughput of a single Memcached process
over 10Gig network

CPU-bounded   Network-bounded

Legend:
- TPS max
- 10Gbps/payload

Y-axis: TPS
X-axis: Payload average size (bytes)

AMADEUS

NoSQL roots
CAP theorem, E. Brewer, 2000*
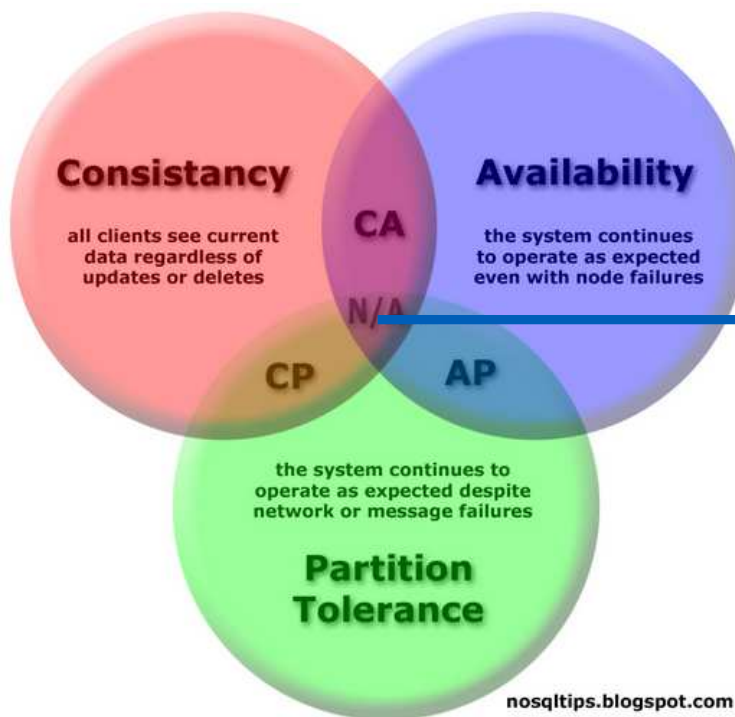
Pick two!

Symposium on Principles of Distributed Computing

aMaDEUS

NoSQL roots

_ From academia
- Mike Stonebraker: *"One size does not fit all"**
  - Ingres, Postgres, Informix Illustra, Streambase, Vertica, VoltDB

_ The big four of the web industry: the "*GAFA"*
- Pragmatic people with immediate & tangible problems
- RDBMSs do not scale-out for massive write operations

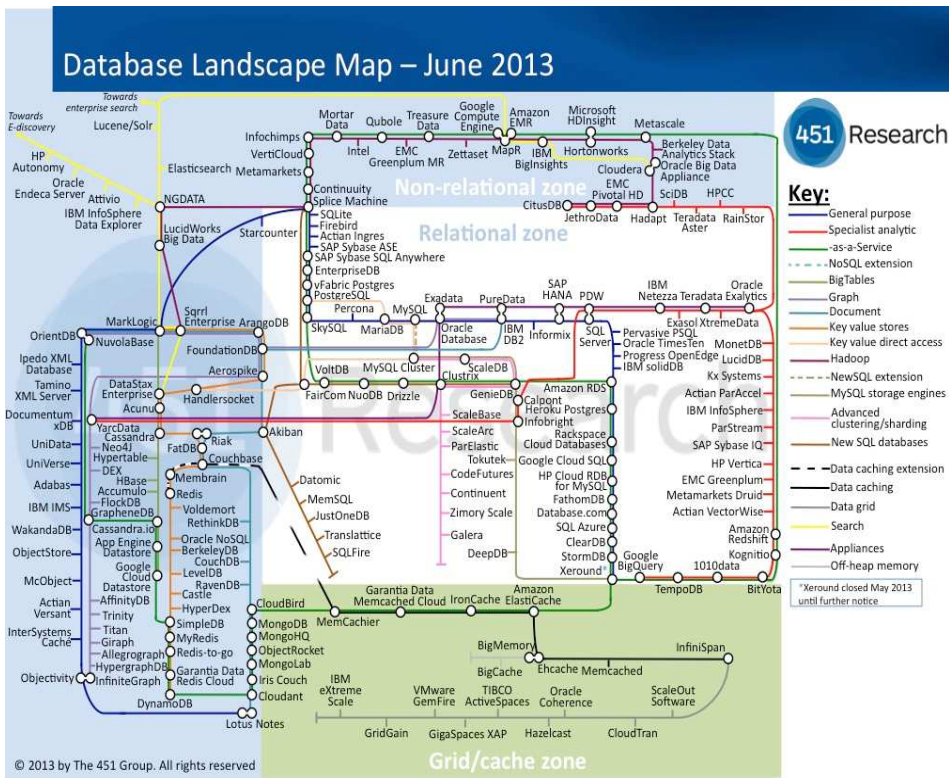*21st International Conference on Data Engineering, 2005*

aMaDEUS

NoSQL roots

150+ products !

aмaрeus

With NoSQL, we got

_ Scalability and Performance
_ Eventual Consistency
  • with freedom of data-persistency choice

_ Many nodes to administrate instead of a pair
  • + network infrastructure to be adapted

_ Access
  • to source code of the solution
  • to active communities of enthusiastic people

_ Technologies with clear can-do/cannot-do boundaries
  • Improve app designs

aмaрeus

With NoSQL, we lost

_ Relational database core abilities
- Large and flexible feature set
- Convenient query language
- ACID transactions

_ Decades of investments spent in enterprise RDBMSs
- Advanced admin/tuning/maintenance facilities
- Mature security features

_ Wealth of trained people all over the world

aMaDEUS

NoSQL

Enabler of web-scale apps

Not suitable for all purposes

For people who know "what they do"

aMaDEUS

3 NoSQL implemented in Amadeus

aMaDEUS

NoSQL & Amadeus: 2013

Setting-up our collaboration models with NoSQL providers

__ Comparison of Key/Value store providers + choice
  • Redis, JBoss DataGrid "Infinispan", Oracle Coherence
  • Couchbase

__ First experimentations with Mongo Document Store

__ Definition of our way-of-working with software providers
  • Major partners, driving how the products evolve
  • Benefit from consulting, training, support services

__ Strong relationship tied with Couchbase
  • Gave talks to Couchbase UK conference of 2013 and 2014, *see youtube*
  • Our Executive Committee members visited them in the SiliconValley

aMaDEUS

Couchbase throughput test

➔ requested and got Couchbase enhancement to multithread write-to-disk
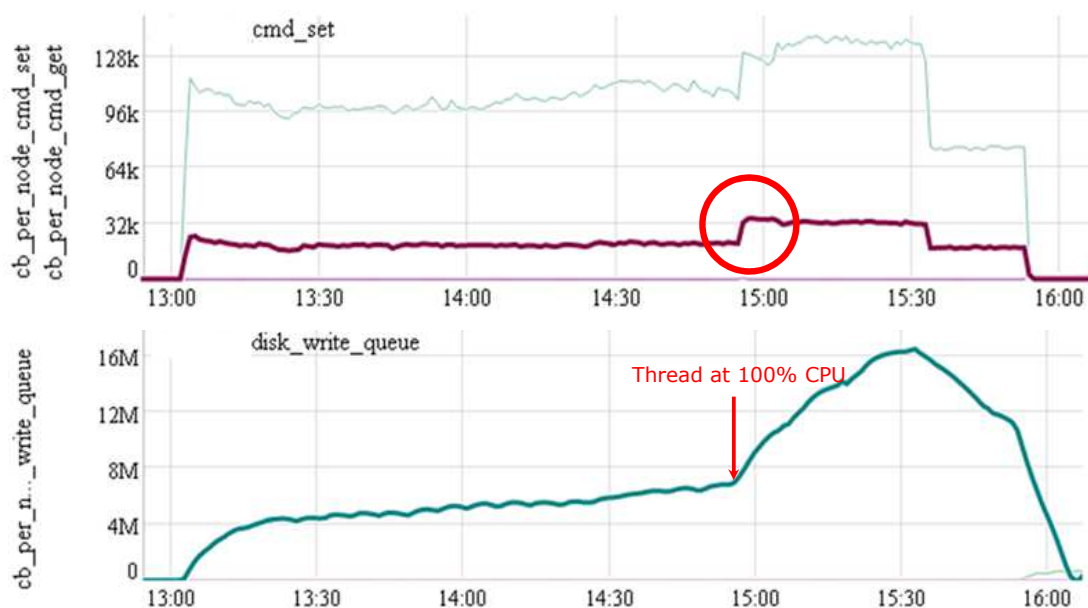
AMADEUS

Our 3 NoSQL picks

- Sub-millisecond latency
- Extreme throughput
- Extreme scalability
- Basic query language

  Recommended use cases
  massive simple read/write
  (millions per second)

- Rich query model
- Rich data modelling abilities
- Developer friendly
- Suitable for asymmetric
  workloads: Write << Read

  Recommended use cases
  Small to medium OLTP apps

- Huge amount of data (>10 TB)
- Strong data-crunching abilities

  Recommended use cases
  Batch processing
  Data analytics
  Long term archiving

Low-Latency Key/Value Store

Document Store

BigData

AMADEUS

NoSQL & Amadeus: 2014
NoSQL to enter Amadeus technology portfolio

_ One pilot app for each selected NoSQL area
- Couchbase pilot: Air Availability
- MongoDB pilot: Accounting
- Hadoop/MapR pilot: Searching and Shopping Logging

_ Learn how to deploy and operate NoSQL
- Non-Functional Requirements are key

_ Deploy NoSQL farms with guidelines defined/validated during pilots
- Ex: we like big large servers, 4 sockets, 12-core CPU, FusionIO PCIE cards

_ 100+ persons involved on NoSQL in Amadeus
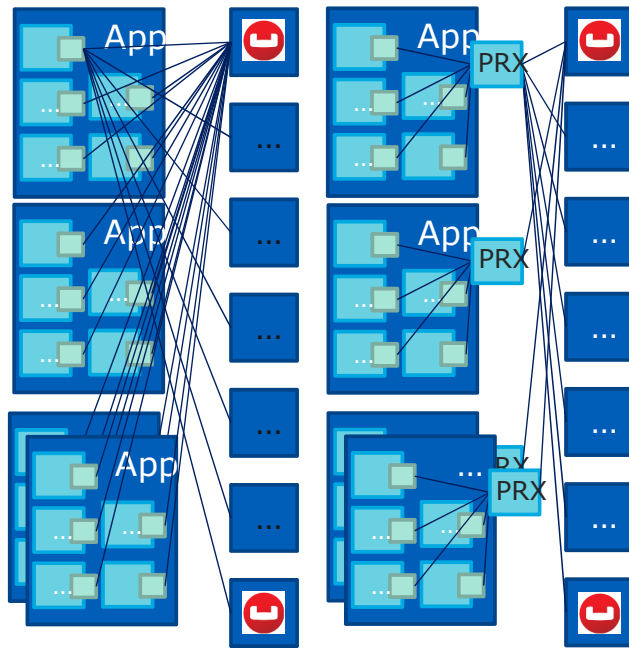- 20+ groups of R&D
- 10+ groups of Operations

AMADEUS

Air Availability
part of Searching and Shopping stack

2M reads/s, 500 000 writes/s

Our latest in-house NoSQL product: ProxyD

Scalability: need to multiplex network connections

_ **100 times** less network connections

_ Based on Twemproxy source code from Twitter

_ We submitted code back to community

aMaDEUS

Accounting: ad-hoc searches

_Datamart: fed in real time, 3 years history, 6TB
_Web UI: chaining queries to drill data
_Interactive results

_MassivelyParallelProcessing way to deploy Mongo
  • Maximization of resource usage, 1 MongoD per CPU core
  • Multiple massive data transformations through aggregation framework

aMaDEUS

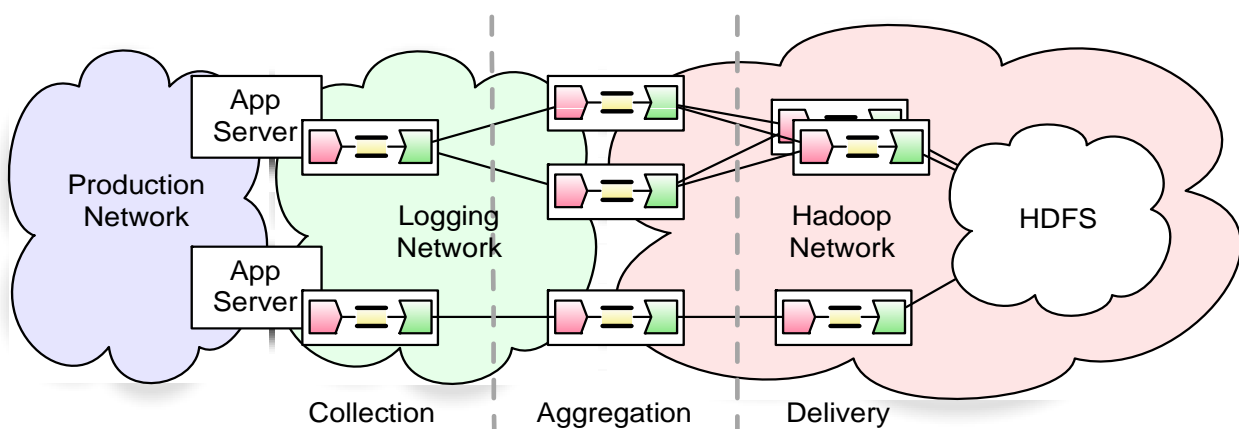# Accounting: ad-hoc searches

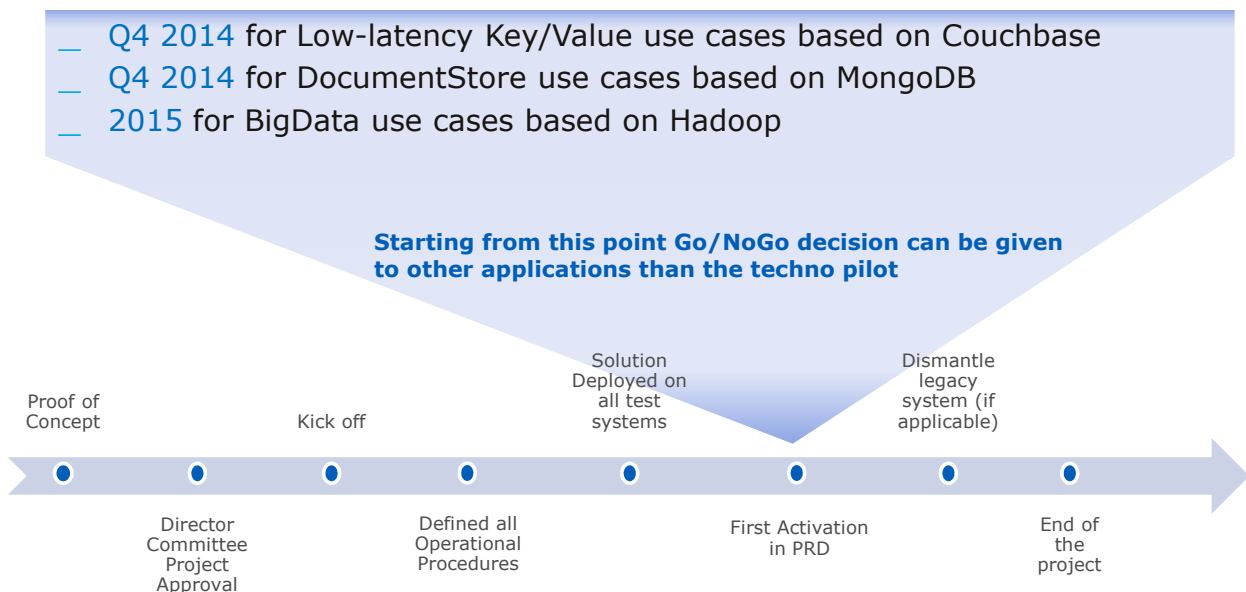## Data model: hierarchical documents
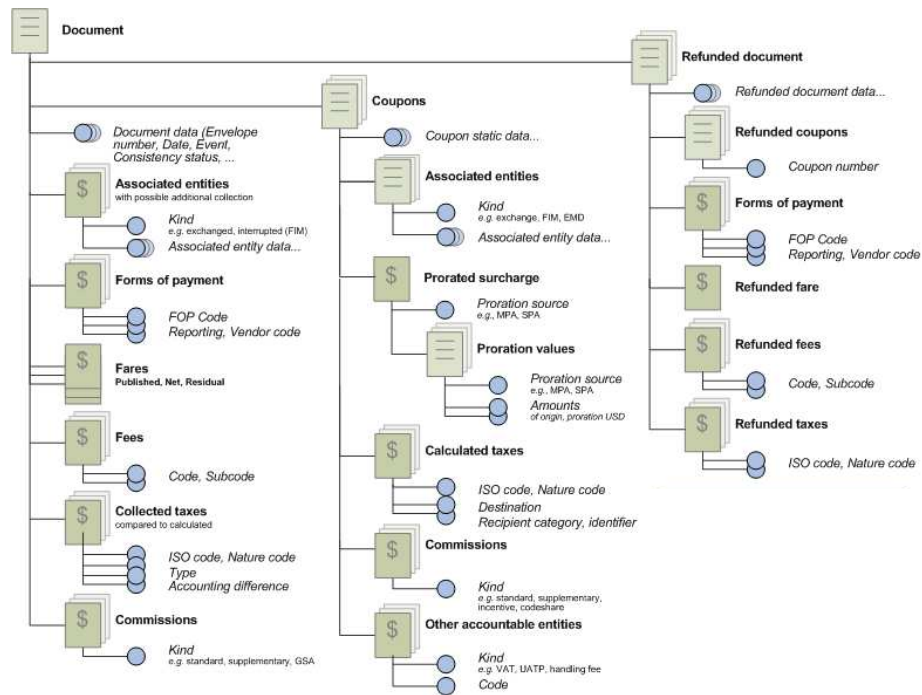
1000+ different document "schemas" !

BigData Logging Architecture

NoSQL standards in Amadeus
Rationale and first prod-activation timeline

_ Q4 2014 for Low-latency Key/Value use cases based on Couchbase
_ Q4 2014 for DocumentStore use cases based on MongoDB
_ 2015 for BigData use cases based on Hadoop

Starting from this point Go/NoGo decision can be given to other applications than the techno pilot

Proof of Concept   Kick off   Solution Deployed on all test systems   Dismantle legacy system (if applicable)

Director Committee Project Approval   Defined all Operational Procedures   First Activation in PRD   End of the project

aMaDEUS

NoSQL: what to look at before embracing?

_ NoSQL Tradeoffs: what to relax?
- Availability, Consistency, Tolerance to NW partitioning
- Transactional needs?

_ Data Storage
- Structure and volume of your data?
- Is NoSQL master? How is NoSQL fed? Migration steps?

_ Data Access Paths: how data is accessed?
- Workload: Read/Update/Insert mix? OLTP? Analytics?
- Latency and throughput?
- How to handle data inconsistencies?

_ Business sensitivity of the data? Security regulations?

_ Business Case: cost vs profit (money, stability, …)
- HW/SW cost of a NoSQL farm can be higher than expected
  Needed to be estimated at the very beginning of your project
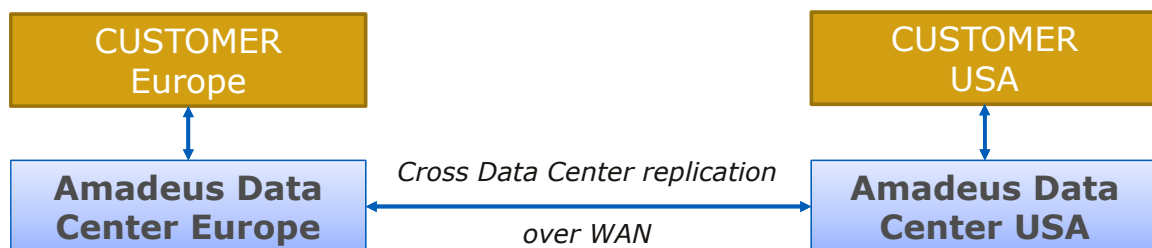
aMaDEUS

4

What's next?

aMaDEUS

Ultra High Availability through NoSQL

Active/Active topology to achieve target of **max** 3mins of outage per year

__Active/Active Data Centers

CUSTOMER Europe		CUSTOMER USA

Amadeus Data Center Europe — *Cross Data Center replication over WAN* — **Amadeus Data Center USA**

__Eventual Consistency issues to be managed

aMaDEUS

BigData

Thomas H. Davenport

Hervé Couturier

At the Big Data Crossroads: turning towards a smarter travel experience report

amaDEUS

---

5 ——— Conclusion

amaDEUS

NoSQL: a major growth enabler

Strong internal **momentum** to **adopt NoSQL**

Wide and controlled NoSQL **standardisation** going-on

NoSQL is a **key enabler** for Amadeus to Shape the **Future of Travel**

AMADEUS

_____ Thank you

amadeus.com/blog
amadeus.com

AMADEUS