Not Only SQL

Data Wrangling 101

John Daily

Basho Software Engineer

@macintux

Highly available & durable

Massive concurrency

Low latency

Operationally simple

Open source

Highly available & durable

Massive concurrency

Low latency

Operationally simple

Open source

Highly available & durable

Massive concurrency

Low latency

Operationally simple

Open source

Highly available & durable

Massive concurrency

Low latency

Operationally simple

Open source

Highly available & durable

Massive concurrency

Low latency

Operationally simple

Open source

Why NoSQL?

Structure tradeoffs

Database categories

Polyglot (hybrid) solutions

User preferences

Hierarchical data

Unstructured data

User preferences

Hierarchical data

Unstructured data

User preferences

Hierarchical data

Unstructured data

Couchbase

RethinkDB

Riak

Redis

Datomic

Aerospike

CouchDB

Cassandra

MongoDB

NuoDB

FoundationDB

Neo4J

Oracle NoSQL

HBase

Structure

Scaling

Structure

Write performance

Reading?

Structure

Reading?

Reading?

Relational

[PK, FK]
[PK, FK]

char(22)	[PK, FK]
bigint	[PK, FK]
char(22)	
char(22)	
char(22)	
char(22)	
char(22)	
char(22)	
char(22)	
char(22)	
char(22)	
char(22)	
char(22)	
int	
int	
int	
int	
int	
char(22)	
int	
char(30)	
char(22)	
int	

[PK, FK]
[PK, FK]
xt

[PK]

adText	char(255)
url	text
richMedia	text
borderColor	char(7)
textColor	char(7)
backgroundColor	char(7)
clicks	int
clicksBought	int
impressions	int
impressionsBought	int
priority	int
nextInPriority	bigint
renderedAd	text

analyticsRule		
ruleId	char(22)	[PK]
sequenceNumber	int	
dateCreated	datetime	
lastUpdated	datetime	
bucketName	char(255)	
regexp	char(255)	

asset		
assetId	char(22)	[PK]
parentId	char(22)	
lineage	char(255)	[U]
state	char(35)	
className	char(255)	
creationDate	bigint	
createdBy	char(22)	
stateChanged	char(22)	
stateChangedBy	char(22)	
isLockedBy	char(22)	
isSystem	int	
lastExportedAs	char(255)	

assetAspectComments		
assetId	char(22)	[PK, FK]
revisionDate	bigint	[PK, FK]
comments	longtext	
averageCommentRating	int	

assetAspectRssFeed		
assetId	char(22)	[PK, FK]
revisionDate	bigint	[PK, FK]
itemsPerFeed	int	
feedCopyright	text	
feedTitle	text	
feedDescription	text	

dateAdded	datetime
options	longtext
configuredTitle	char(255)
shippingAddressId	char(22)
quantity	int

databaseLink		
databaseLinkId	char(22)	[PK]
title	char(255)	
DSN	char(255)	
username	char(255)	
identifier	char(255)	
allowedKeywords	text	
allowMacroAccess	int	
additionalParameters	char(255)	

deltaLog	
userId	varchar(22)
assetId	varchar(22)
delta	int
timeStamp	bigint
url	varchar(255)

donation		
assetId	char(22)	[PK, FK]
revisionDate	bigint	[PK, FK]
defaultPrice	float	
thankYouMessage	mediumtext	
templateId	char(22)	

filePumpBundle		
bundleId	char(22)	[PK]
sequenceNumber	int	
dateCreated	datetime	
lastUpdated	datetime	
bundleName	char(255)	
lastModified	bigint	
lastBuild	bigint	
jsFiles	longtext	
cssFiles	longtext	
otherFiles	longtext	

friendInvitations		
inviteId	char(22)	[PK]
inviterId	char(22)	
friendId	char(22)	
dateSent	datetime	

completedOn	bigint
completedBy	char(22)
userId	char(22)
groupId	char(22)
subject	char(255)
message	mediumtext
sentBy	char(22)

inbox_messageState		
messageId	char(22)	[PK]
userId	char(22)	[PK]
isRead	tinyint	
repliedTo	tinyint	
deleted	tinyint	

incrementer		
incrementerId	char(50)	[PK]
nextValue	int	

karmaLog	
userId	char(22)
amount	int
source	char(255)
description	text
dateModified	bigint

ldapLink		
ldapLinkId	char(22)	[PK]
ldapLinkName	char(255)	
ldapUrl	char(255)	
connectDn	char(255)	
identifier	char(255)	
ldapUserRDN	char(255)	
ldapIdentity	char(255)	
ldapIdentityName	char(255)	
ldapPasswordName	char(255)	
ldapSendWelcomeMessage	char(2)	
ldapWelcomeMessage	text	
ldapAccountTemplate	char(22)	
ldapCreateAccountTemplate	char(22)	
ldapLoginTemplate	char(22)	
ldapGlobalRecursiveFilter	mediumtext	

mailQueue		
messageId	char(22)	[PK]
message	mediumtext	
toGroup	char(22)	

shippe
shippe
classN
option

shopCr
credit
userId
amount
commer
dateOf

sku
asset
revis
descr
sku
vendor
displa
overri
taxRat
taxCon
shipsS

value

Well-structured data

Ad-hoc queries

Well-structured data

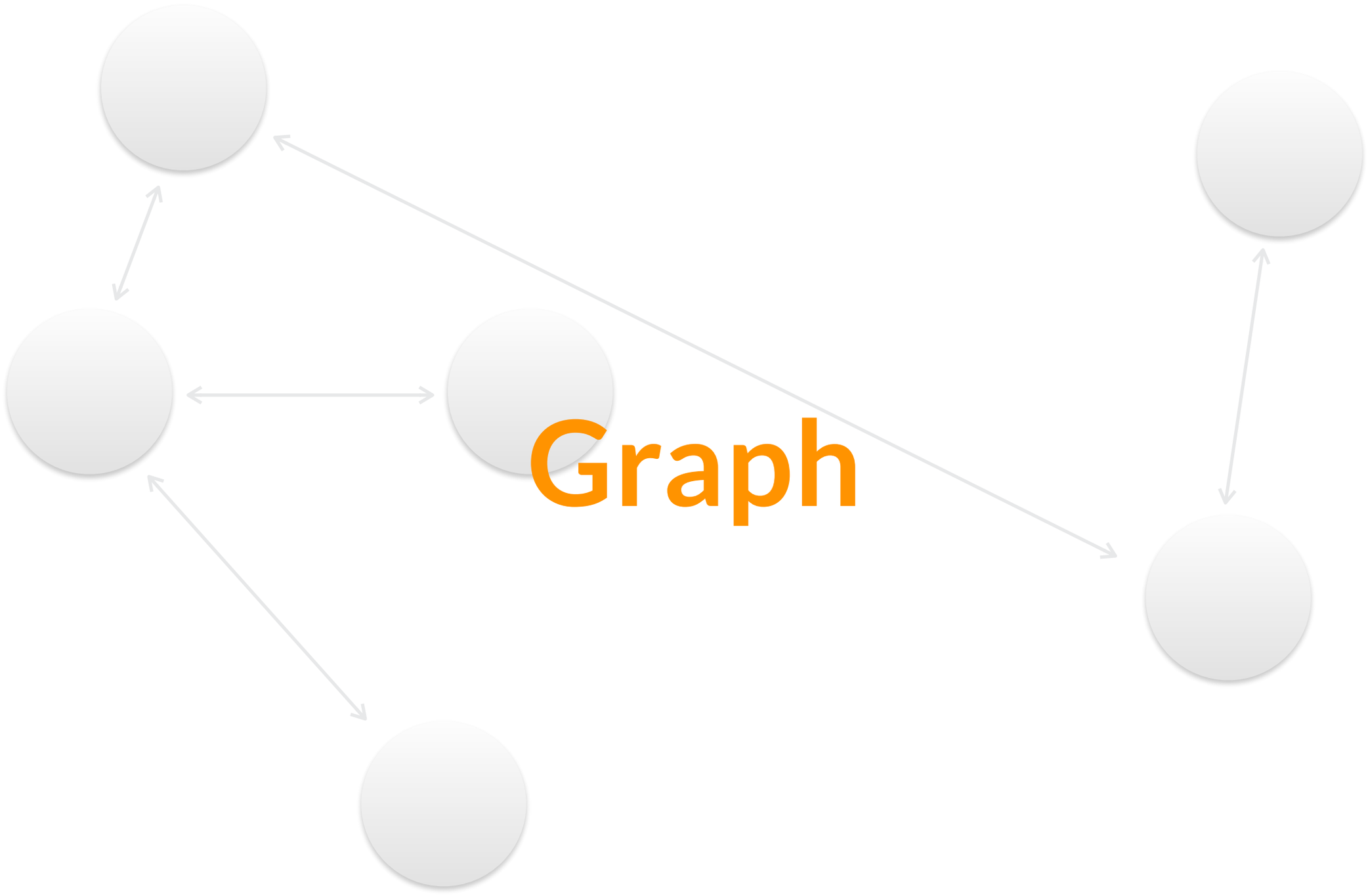
Ad-hoc queries

Well-structured data

35 years of tooling

Ad-hoc queries

Graph



Network modeling

ORM => O

Network modeling

ORM => O

```
{
    "results": [{
        "rsvp_limit": 20,
        "status": "upcoming",
        "visibility": "public",
        "maybe_rsvp_count": 0,
        "venue": {
            "id": 1341113,
            "zip": "32832",
            "phone": "(407)243-4925",
            "lon": -81.237949,
            "repinned": false,
            "name": "Eagle Creek Golf Club",
            "state": "FL",
            "address_1": "10350 Emerson Lake Blvd",
            "lat": 28.370016,
            "city": "Orlando"
            "country": "us"
        },
        "id": "dtxxkcysgbwb",
        "utc_offset": -14400000,
        "distance": 14.251898765563965,
        "time": 1397748600000,
        "waitlist_count": 0,
        "updated": 1337108474000,
        "yes_rsvp_count": 6,
        "created": 1325888238000,
        "event_url": "http://www.meetup.com/OWIB-Orlando-Women-in-Business/events/174542842/",
        "description": "<p><a href="
        http: //photos4.meetupstatic.com/photos/event/b/c/2/2/event_83928162.jpegOWIB"
    }]
}
```


Document

JSON

K/V + indexes

Agility

JSON

K/V + indexes

Agility

JSON

K/V + indexes

Agility

Columnar

GET “amsterdam/sensor123#13-14:00”

Key/value

PUT “magazinesales/foreignpolicy-dec2013”

Schema-less

Name your keys

Reify your views

Schema-less

Name your keys

Reify your views

Schema-less

Name your keys

Reify your views

Distributed Search

Indexing

!DBOR

Indexing

!DBOR

Relational

K/V

Graph

Rethink (allthethings)

Columnar

Document

Ordered vs random access

Consistent vs available

Clustered vs single node

Latency vs durability

Ease of development

vs

ease of operations

Engineering vs marketing

क्या आप अंग्रेज़ी बोलते हैं?

هل تتحدث الانجليزية؟

Polyglot databases

Μιλάτε αγγλικά;

Sprechen Sie Englisch?

Graph + K/V

Datomic

Graph + K/V

Datomic