

What is NoSQL?

Andrew J. Brust

<http://www.bluebadgeinsights.com>

andrew.brust@bluebadgeinsights.com



What is NoSQL?

Customers

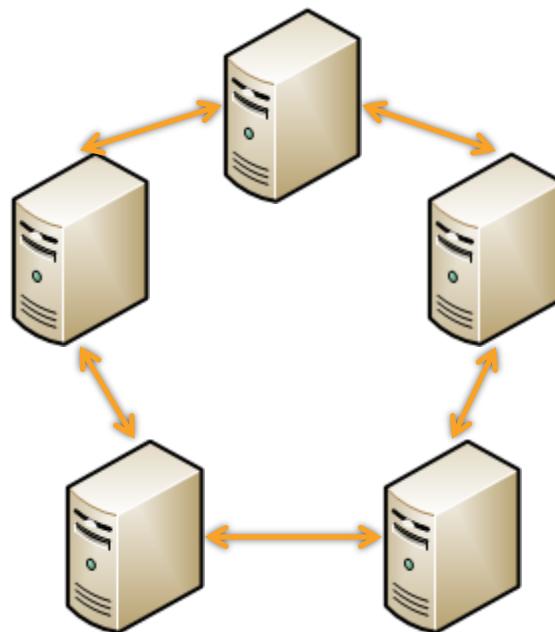
ID: 112 2	Iname : Brust	fname: Andrew	address: 123 Main St.	city: New York	state: NY	zip: 10099
ID: 321 4	Iname : Doe	fname: John	address: 321 Waterford Crescent.	village: Stodday	county: Lancashir e	postal code: LA2 6ET

Orders

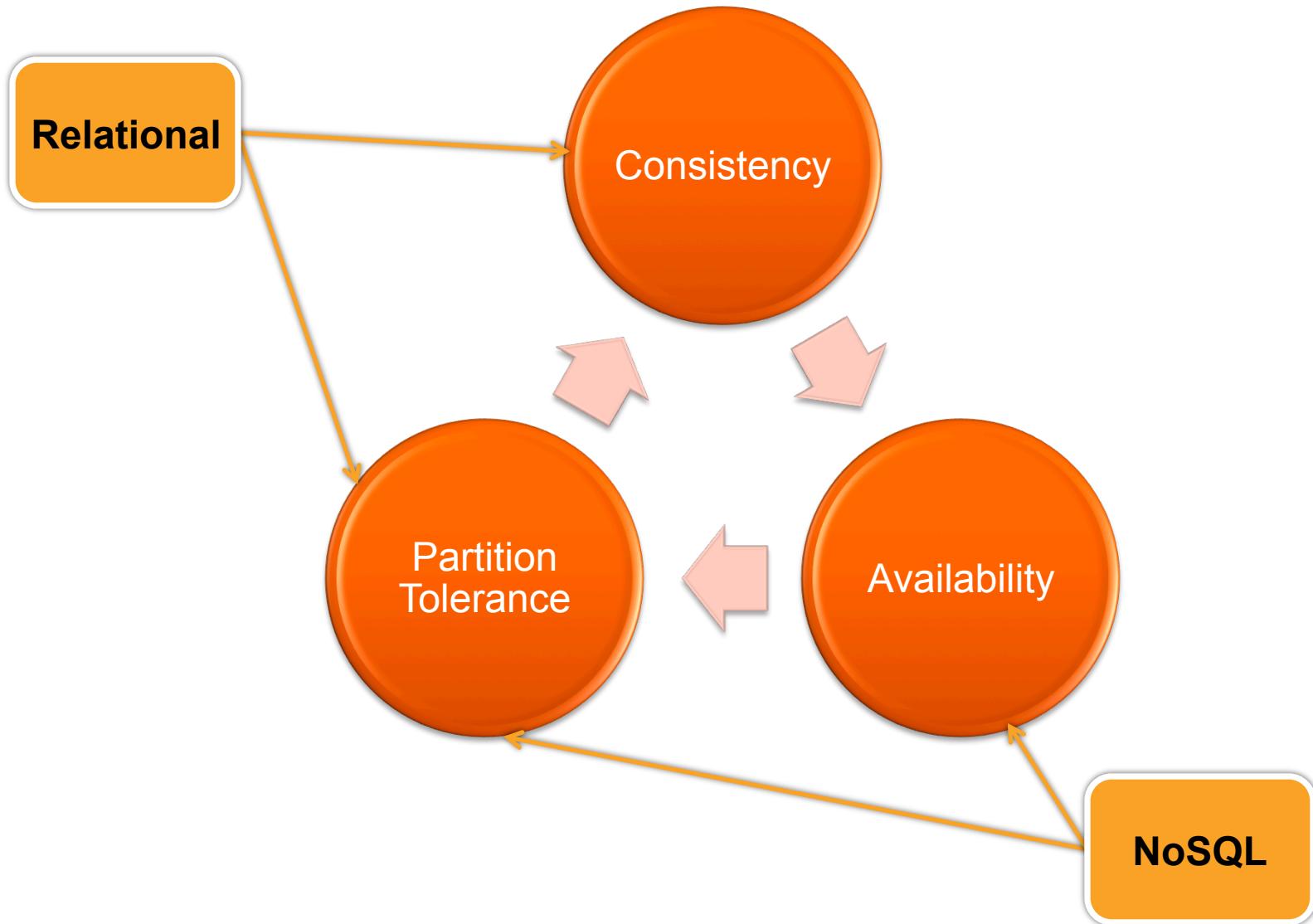
ID: 3001	customerID: 1122	amount: 500	tax: 40	processdat e: 2/20/2012
ID: 3002	customerID: 1122	amount: 250	shipdate : 3/18/201 2	
ID: 3003	amount: 1700	tax: 150		

Distributed Architecture

- Many NoSQL Databases federate a number of commodity server together
- Provides redundant storage
- Provides geographic distribution
- Avoids having a “single point of failure.”



CAP Theorem



Consistency

- **Things like inventory, account balances should be consistent**
 - Imagine updating a server in Seattle that stock was depleted
 - Imagine *not* updating the server in NY
 - Customer in NY goes to order 50 pieces of the item
 - Order processed even though no stock
- **Things like catalog information don't have to be, at least not immediately**
 - If a new item is entered into the catalog, it's OK for some customers to see it even before the other customers' server know about it
- **But catalog info must come up quickly**
 - Therefore don't lock data in one location while waiting to update the other
- **Therefore, OK to sacrifice consistency for speed, in some cases**



Compromises

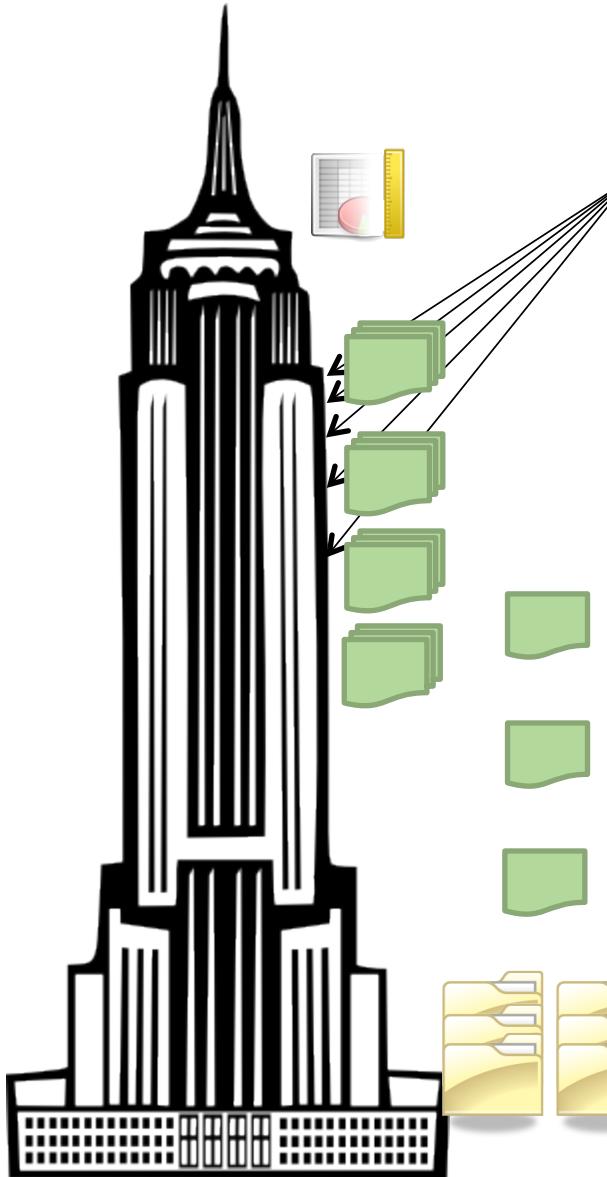
- **Eventual consistency**
- **Write buffering**
- **Only primary keys can be indexed**
- **Queries must be written as programs**
- **Tooling**
 - Productivity (= money)

NoSQL Queries



- **Typically no query language**
- **Instead, create procedural program**
- **Sometimes SQL is supported**
- **But usually MapReduce code is used...**

A MapReduce Example



- Count by suite, on each floor
 - 
 -  Go to lobby
 -  Go to 10th, 20th, 30th floor
 - Tally up each floor
 - Collect the tallies
 - Merge tallies into one spreadsheet

Cloud

- **Microsoft**
 - Azure Tables
 - Hadoop on Azure/Hbase
- **Amazon**
 - SimpleDB
 - DynamoDB
 - Elastic MapReduce
- **DB-specific NoSQL Hosters**
 - E.g. MongoHQ for MongoDB
 - CouchDB on Cloudant
- **DIY on Infrastructure as a Service (IaaS)**