

SQLintersection

SQL Keynote

The Future of the RDBMS Professional

Buck Woody

Data and Applied Scientist, Microsoft



SQL
intersection



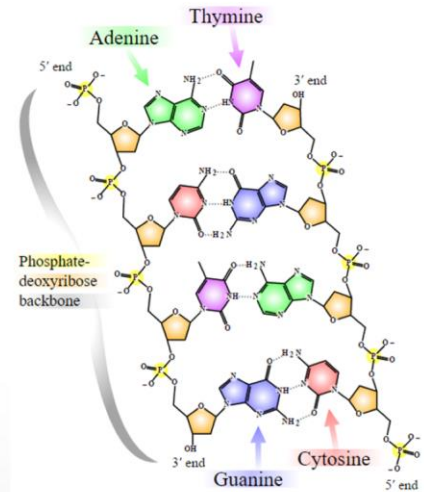
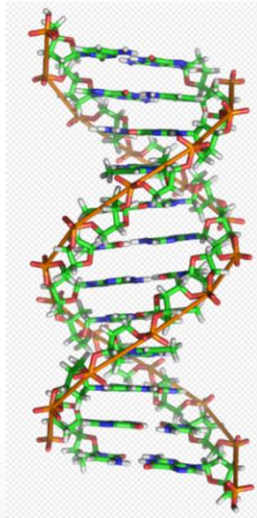
Reminder: Intersect with Speakers and Attendees

- **Tweet tips and tricks that you learn and follow tweets posted by your peers!**
 - Follow: #SQLIntersection and/or #DEVIntersection
- **Join us – Tuesday Evening – for SQLafterDark**
 - Doors open at **7:00 pm**
 - Trivia game starts at **7:30 pm**
 - Winning team receives something fun!*
 - Raffle at the end of the night
 - Lots of great items to win including a seat in a SQLskills Immersion Event!*
 - The first round of drinks is sponsored by SentryOne and SQLskills



© SQLIntersection. All rights reserved.
<http://www.SQLIntersection.com>

All computing is simply re-arranging data



© SQLintersection. All rights reserved.
<http://www.SQLintersection.com>

This will be a “Mixed Technology” session – a state of the industry, along with technical discussions

Let’s examine a little history – it doesn’t repeat, but it does rhyme. This will help us understand what is coming. We keep solving the same problems

<http://dna-structure.com/history.htm>

A Swiss medical student discovered DNA in 1868. Others contributed to understanding DNA, and in 1953 James Watson and Francis Crick created a diagram of how they thought DNA was structured and how it combined.

(O) And what does DNA have to do with databases?

<https://www.technologyreview.com/s/607880/microsoft-has-a-goal-to-store-data-as-dna/>

“Formatted in DNA, every movie ever made would fit inside a volume smaller than a sugar cube.”

These four simple nucleobases solve the problem of all living matter’s growth, development, function and recreation. Let’s look at the problems we’ve solved in data technologies – and the issues we’ll face in the future.

Unrecorded Data



The first issue in data processing was in Unrecorded data.

Writing – earliest is around 3200 BC – Cuneiform (Sales of beer, taxes on same). Chief scribe was called a sangu – “The Accountant”

The medium was more expensive than the data -

<http://abcnews.go.com/International/leonardo-da-vincis-lost-painting-discovered-florence/story?id=15902573>

File originally meant the line used to hang up documents to order them by monks

(O) 1856 – Florence Nightingale should be remembered not only for Nursing, but data analysis: <http://www.atlasobscura.com/articles/florence-nightingale-infographic>

1890 - Herman-Hollerith, US Census (eventually became IBM) -

https://en.wikipedia.org/wiki/Herman_Hollerith

1959 – DoD had 225 computers, and had spent 200 million dollars implementing programs to run on them.

Based on work by Grace Hopper (FLOW-MATIC) - COBOL code is split into four divisions (identification, environment, data and procedure) containing a rigid hierarchy of sections, paragraphs and sentences. Prior to this, scientists cared largely only about numbers and small strings

(O) By the way, COBOL is far from dead - <https://thenewstack.io/cobol-everywhere-will-maintain/>

Atomicity - Atomicity requires that each transaction be "all or nothing": if one part of the transaction fails, then the entire transaction fails, and the database state is left unchanged.

Consistency - ensures that any transaction will bring the database from one valid state to another. Any data written to the database must be valid according to all defined rules, including constraints, cascades, triggers, and any combination.

Isolation - ensures that the concurrent execution of transactions results in a system state that would be obtained if transactions were executed one after the other.

Durability - ensures that once a transaction has been committed, it will remain so, even in the event of power loss, crashes, or errors. In a relational database, for instance, once a group of SQL statements execute, the results need to be stored permanently (even if the database crashes immediately thereafter). To defend against power loss, transactions (or their effects) must be recorded in a non-volatile memory.

//GO.SYSPR
//

$$R * S * T = \{ (a, b, c, d) : R(a, b) \wedge S(b, c) \wedge T(c, d) \}$$

children' (man#, childname, birthyear)

FIG. 3(b). Normalized set

© SQLintersection. All rights reserved.
<http://www.SQLintersection.com>

Flat files – read into memory in entirety and written entirely back out after modification – quite brittle, complicated and limiting. COBOL is an example of Programmatic Referential Integrity.

1970 – EF Codd introduced a calculus to help solve these issues. It included the ACID properties of databases. The Consistency property is a mechanism to ensure Declarative Referential Integrity.

We see here a “safe” query from tuple calculus

UnCAPtured Data

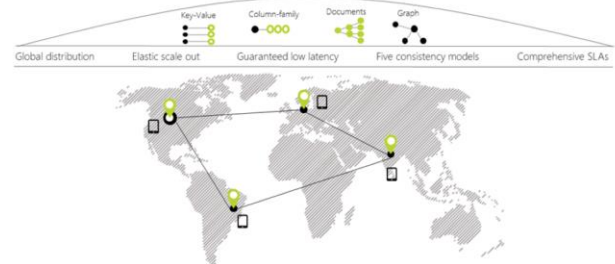
Consistency - Every read receives the most recent write or an error

Availability - Every request receives a response, without guarantee that it contains the most recent version of the information

Partition tolerance - The system continues to operate despite an arbitrary number of messages being dropped by the network between nodes



Azure Cosmos DB



We returned to the past in two ways – first, we left data out because we created so much of it, we couldn't query and store it all in an RDBMS. The General Electric GenX aircraft Engine creates 1 TB of data per flight.

By fixing the first problem, we return to the PRI issue

The term NoSQL was used by Carlo Strozzi in 1998 to name his lightweight, Strozzi NoSQL open-source relational database that did not expose the standard Structured Query Language (SQL) interface, but was still relational

Much "NoSQL" data is cleaned and transformed back into columns and rows, and is searched in a structured way

Uninterpreted Data

Data Acquisition, Storage, and Processing

- Collection hygiene
- Intelligent storage
- Data engines (RDBMS, NoSQL)

Reporting

- Query languages
- Tabular results
- Graphical presentation

Business Intelligence

- Multi-dimensional queries
- Historical aggregations
- Exploration and question generation

Cognitive Intelligence

- Adaptive
- Interactive
- Iterative and Stateful
- Contextual



© SQLintersection. All rights reserved.
<http://www.SQLintersection.com>

The problem we're solving now is deciding what to do, based on the data we have collected. According to a study reported in business insider in 2015, 99.5% of all captured data is never evaluated.

(O) Connectedness Index: <http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/global-flows-in-a-digital-age>

According to Business Insider, More than 90% of the data in the world was created in the last three years. 1 small notebook computer can store the data collected in all the space program shots combined. By 2020, digital recordings are expected to hit 44 zettabytes, from 4.4 zettabytes today. All of this will need to be transformed.

Analytics brings a change from X-Ray to an MRI – Cognitive Intelligence has these attributes (<https://dzone.com/articles/introduction-to-cognitive-computing>):

Adaptive, Interactive

Data Futures

- “Big Data” becomes just “Data”
- Connectedness becomes a primary metric
- Clean data is the lifeblood of Cognitive Computing
- Analytic languages embed into current systems
- Graph databases become part of relational data
- Security specializes into data
- RDBMS becomes the “bookends” of analysis
- “Alert Fatigue” takes hold

“When Information is cheap, attention becomes expensive” – Information Magazine



© SQLintersection. All rights reserved.
<http://www.SQLintersection.com>

Clean data is essential – and the reason the RDBMS was invented.

Alert fatigue is not a UI problem – it’s a data problem

Call to Action

- Understand your platform
- Integrate with other platforms
- Brush up on statistics - use R or Python to learn it
- Understand the business and what they need to act on
- Learn data wrangling
- Focus on security



<http://aka.ms/backyarddatascience>

© SQLintersection. All rights reserved.
<http://www.SQLintersection.com>

Help your business with their data: What data do they need to make an informed decision and What would you do if you knew the answer to your question

Save the Date

Oct 30-Nov 2, 2017
We're back in Vegas baby!

www.SQLintersection.com

