

Historique des Bases de données

Sébastien Lardière

Loxodata

mars 2016

Historique des Bases de Données

1 Stockage de données

- Les débuts

2 Théories

- Historique
- Modèle relationnel

- Entité Relation
- Formes Normales

3 Base de données

- SQL
- Postgres

4 Ère Moderne

- CAP

- BASE
- Eventually Consistent

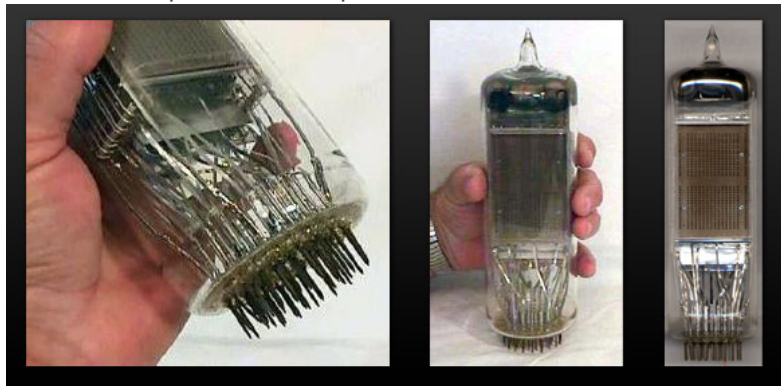
5 NoSQL

- DB Hiérarchique
- NoSQL

6 Perspectives

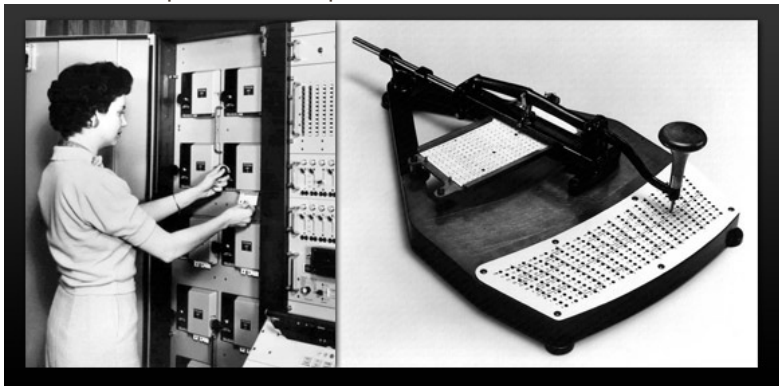
Les débuts

Avant les disques : une lampe de 1024 bit en 1946 !



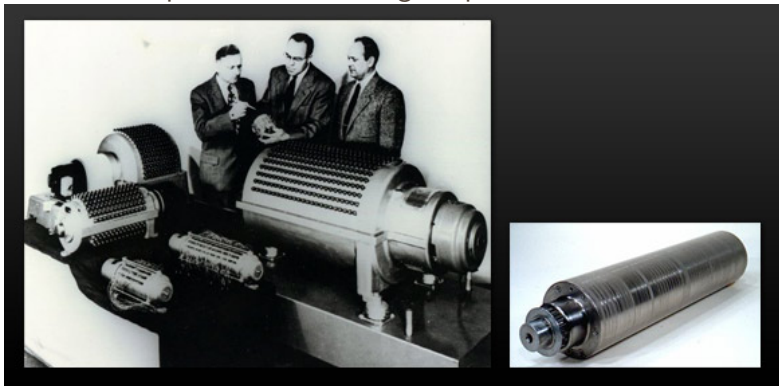
Les débuts

Avant les disques : Cartes perforées



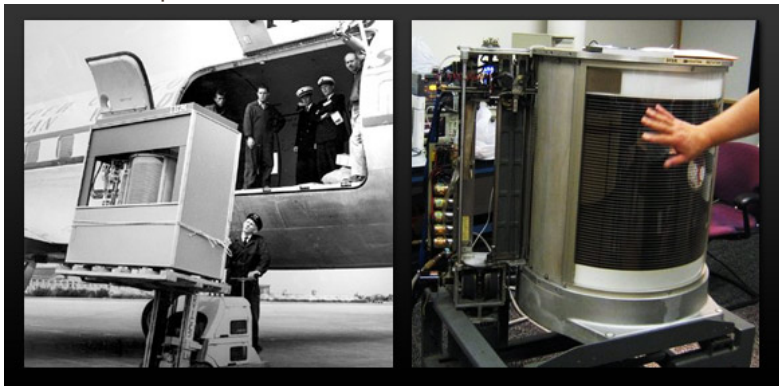
Les débuts

Avant les disques : Tambour magnétique, 10kb en 1950



Les débuts

Premiers disques durs : IBM 350, 5MB en 1956



Les débuts

IBM 3380 : 1GB en 1980

- La taille d'un frigo
- 250kg

Premiers SGBDs

IDS (Général Electric)

- 1961
- Modèle Réseau

Premiers SGBDs

IMS (IBM)

- 1965-1970
- Modèle Réseau

Premiers SGBDs

Modélisation et recherche : 1970's

- Modèle Relationnel
- Travail universitaire
- Merise
- Premières conférences
- Modèle Entité-Association

Premiers SGBDs

- SYSTEM-R d'IBM
- INGRES de l'Université de Californie

Codd

Modèle relationnel proposé en 1970

- Travaux académiques (IBM)
- Par opposition au monde "industriel"
- Prix Turing pour Codd en 1981

Chen

Modèle “Entité Relation” en 1975

- Besoin d’une méthodologie
- Résultat d’interactions utilisateurs
- Basé sur des maths : Théorie des ensembles, Algèbre, Logique
- ...

Chen

Modèle “Entité Relation” vs. Modèle Relationnel

- “ER” : Sémantique
- Lien Explicite vs. Implicite
- Rapprochement ultérieur des modèles

Chen

Chen évoque l'écriture chinoise :

<u>Original Form</u>	<u>Current Form</u>	<u>Meaning</u>
		Sun
		Moon
		Person

Fig. 5. Chinese Characters that Represent the Real-World Entities

Formes Normales

- Type de relation entre les entités
- Valide le modèle
- 1NF et 2NF systématique
- Recherche académique (70')

SQL

Structured Query Language, auparavant “SEQUEL (Structured English QUery Language)”

- Début 70's, pour IBM System R
- Différences avec la théorie
- Adopté par Oracle, puis DB2

SQL

Norme, mais peu d'interopérabilité

- Norme ANSI : 1986 ; ISO en 1987
- ISO/IEC 9075
- SQL-92, SQL:1999, . . . , SQL-2011
- Norme modulaire (options)

PostgreSQL

Postgres : “Post-Ingres”

- Initié par Michael Stonebraker, Université de Berkeley, 1985
- Idées de Ingres, mais pas le code
- 1994 : arrêt et libération du code
- Yu et Chen remplacent POSTQUEL par SQL : Postgres95
- 1996 : démarrage du projet open-source, et nouveau nom : PostgreSQL

Outils

- Oracle Database
- Sybase, Microsoft SQL Server
- MySQL
- IBM DB2
- IBM Informix
- PostgreSQL

CAP Theorem

Pour un système distribué : 2 propriétés sur les 3 :

Consistency



Availability



Partition Tolerance



CAP Theorem

- Brewer's theorem : Eric Brewer (Berkeley, 2000)
- Preuve Formelle en 2002 : Gilbert et Lynch (MIT)
- Revu en 2012

BASE

Opposé à ACID, informel :

- Basic Availability
- Soft-state
- Eventual consistency

Eventually Consistent

Werner Vogels - CTO Amazon - 2008

- Données Finalement consistente
- Extension au CAP Theorem

DB Hiérarchique ou Réseau

Stockage de données bas-niveau

- CODASYL : Langage de données, 1959
- Hiérarchique : Adabas, IMS
- Réseau : Charles Bachman, 1969

NoSQL

NoSQL recouvre beaucoup de notions

- Pas Seulement SQL
- Pas de DB Relationnel
- Beaucoup de notions et de techniques différentes

NoSQL

Des modèles différents :

- **Column**: Cassandra, Vertica
- **Document**: Apache CouchDB, MongoDB
- **Key-value**: CouchDB, Dynamo, Redis, Riak, Berkeley DB
- **Graph**: Neo4J, Virtuoso
- **Multi-model**: Alchemy Database

Perspectives

De quoi a-t-on besoin ?

- Sharding, DB Distribuée
- Gestion de gros volumes

Conclusions

- Beaucoup de travaux issus d'université
- Réponse à des besoins réels
- Adoption par les acteurs commerciaux
- Validé par les expériences

Questions

Merci ! Des Questions ?