A Relational Model of Data for Large Shared Data Bank

E.F. Codd

Presented by: Changbo Qu

Outline

- Why is this paper important?
- An example of hierarchy data model
- What is wrong with it?
- Relational model
- Conclusion

Why is this paper important?

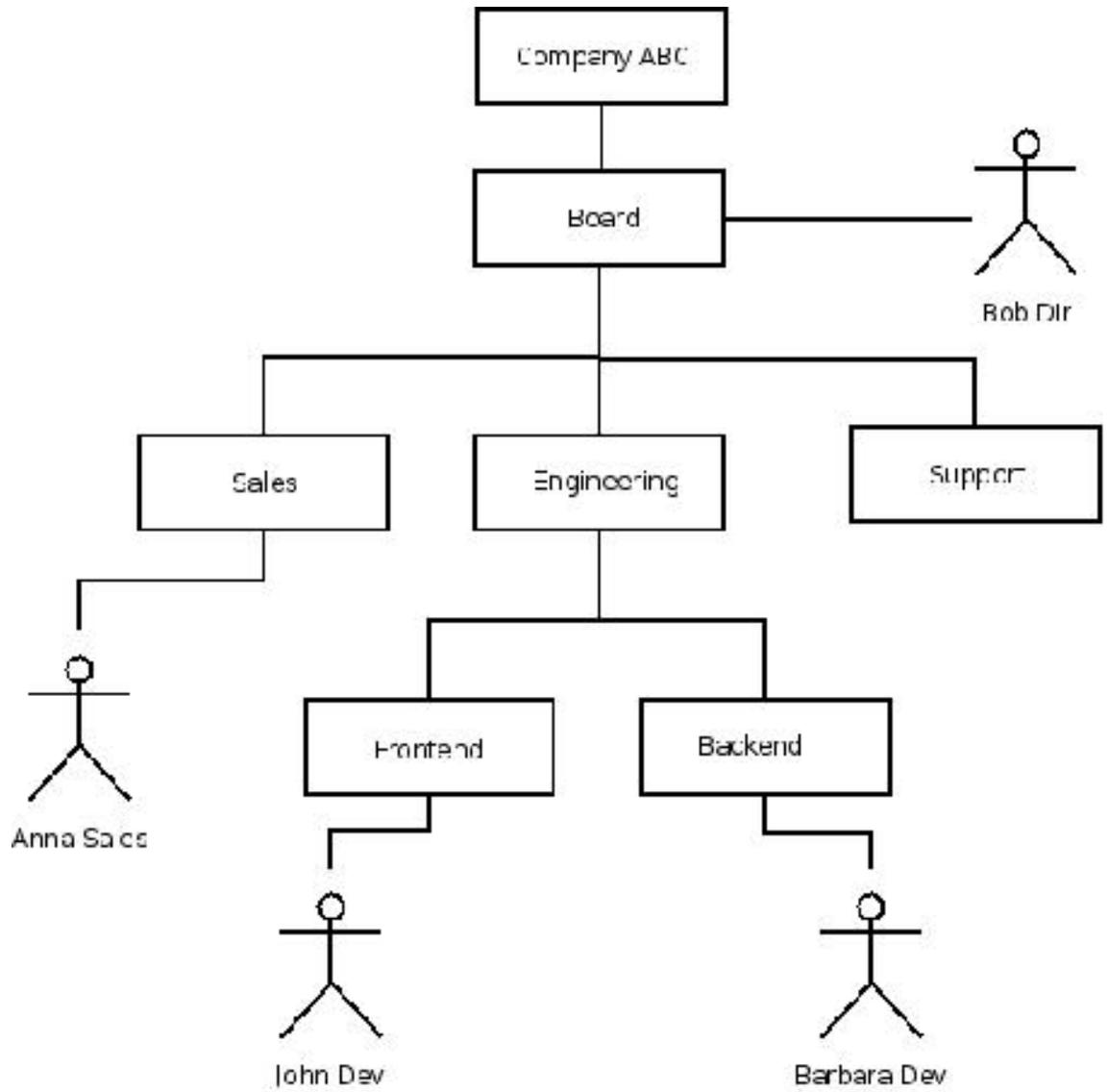
- First proposed the Relational model in 1970.

- He linked the representation of data with that of mathematical sets.

- First research started at IBM's San Jose Research Laboratory. Prototype was called System R.

- Commercial RDBMSs started to appear in late 1970's and early 1980's. Most well known is Oracle.

What does hierarchy database look like?



What is wrong?

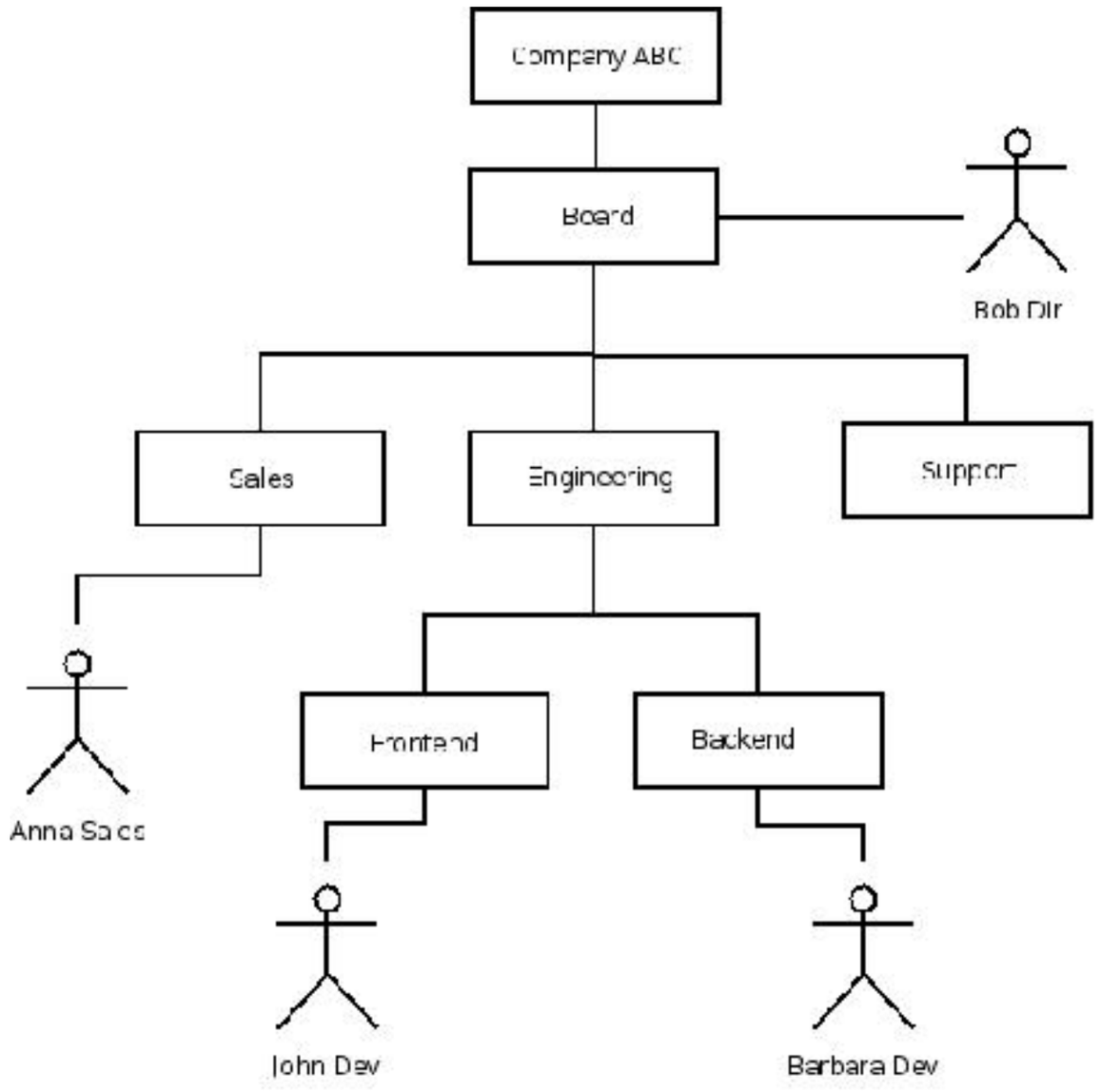
Data dependency between logic and physical storage.

- Ordering dependency
- Indexing dependency
- Access Path dependency

Ordering dependency

The data can be stored different forms of orders of the records.

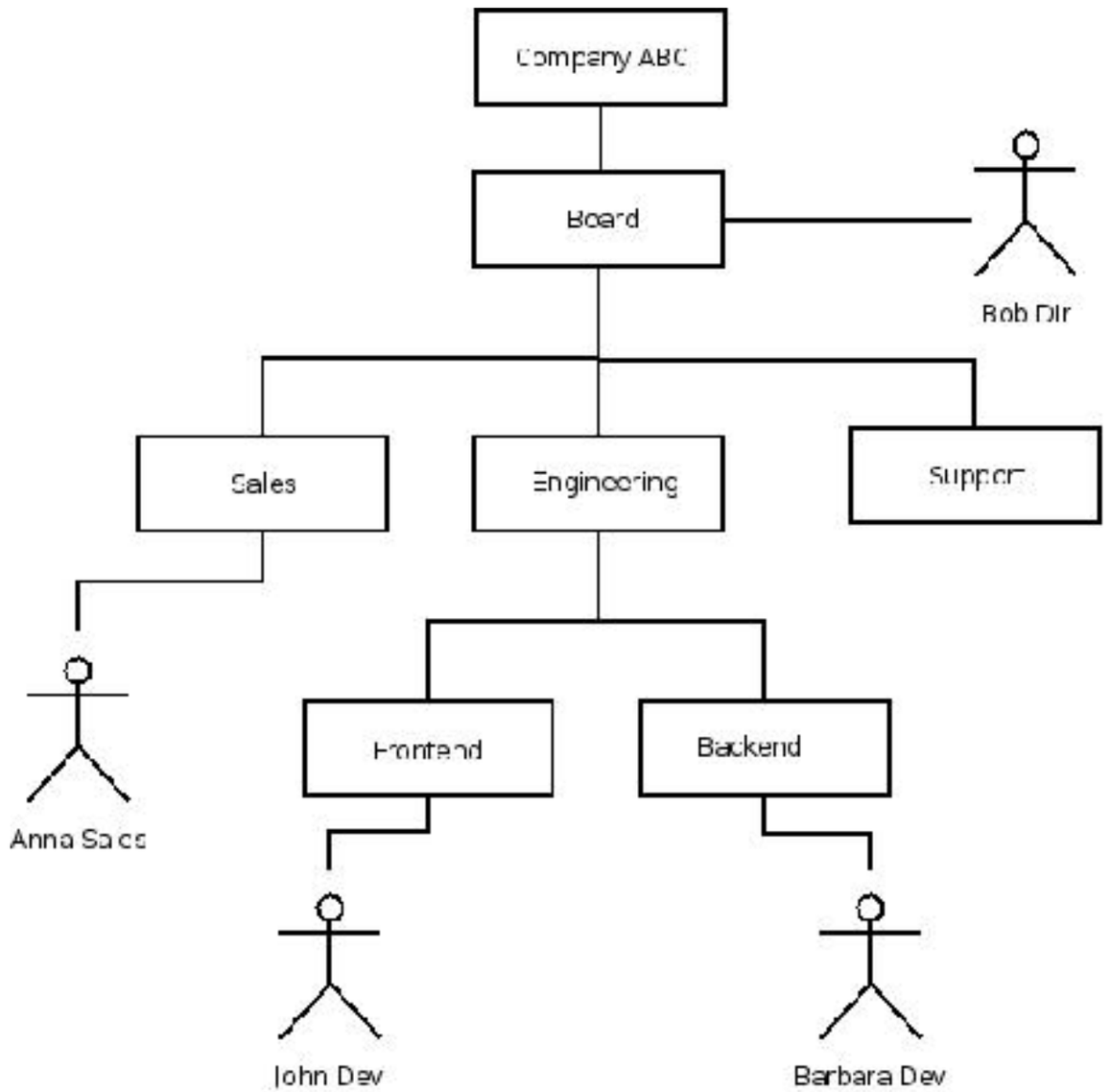
Require the system separates the physical stored order and logic order.



Indexing dependency

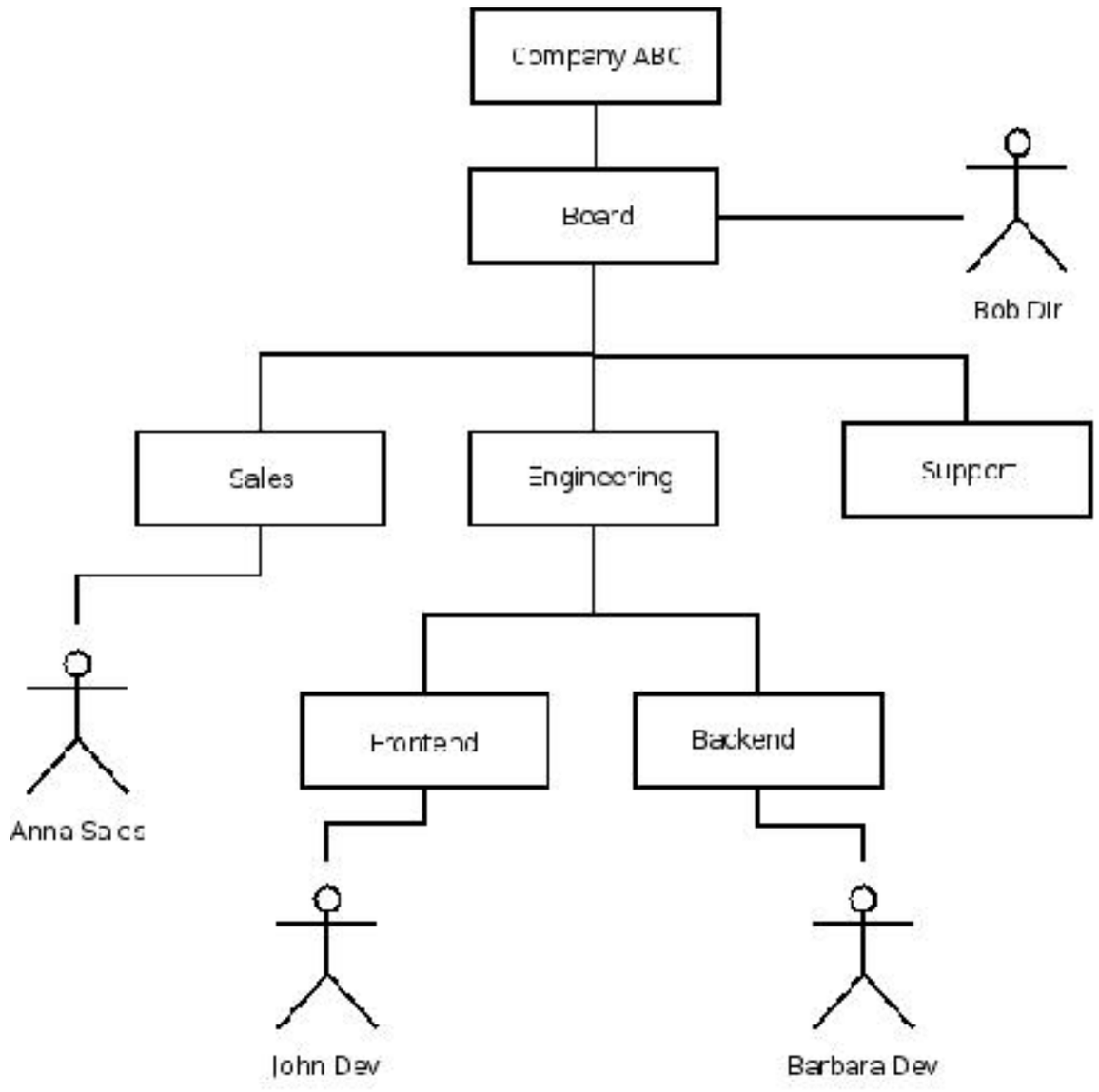
Pros: Improves performance while querying and performing updates on data

Cons: Deteriorate the performance while inserting and deleting data and also is a redundant overhead.



Access Path Dependency

The access paths for data are represented in tree-structures or networking model, but many applications fail if any of underlying model is altered.



What is proposed?

Relational Data Model

- A declarative data-access mechanism

- which remain unaffected when changes happen to physical storage representation

Relational Data Model

Given sets S1, S2 .... Sn(not necessarily distinct), A relation R is a subset of the Cartesian product S1 x S2 x S3 x S4…SN

We call S1 as the first domain of R, S2 as the second domain of R

A relation with n domains is called n-ary relation

Account

bname	account#	cname	balance
Downtown	101	Johnson	500
Lougheed	215	Smith	700
SFU	102	Hayen	400
SFU	301	Adams	1300

Property

For representation convenience, relations are presented in the form of tables, which entails following properties:

1.All rows are distinct

2.Row order does not matter

3.Column order matters

4.The identification of columns is partially labeled the domain name

“Partially”

Component(part, part, quantity)

Part	Part	quantity
P1	P2	2
P2	P3	3
P3	P4	4

To free users from remembering column order of high degree relation,

Unique role names for each domain are suggested

Operations on Relations

- Permutation
- Projection
- Join
- Composition
- Restriction

Operations on Relations(1)

- **Permutation**: The operation permutes column order

No meaning in the user view, since users identify domain by role name. Considered in the storage representation.

- **Projection**: Select certain columns from a relation, then remove duplication in rows

(*supplier*	*part*	*project*)
1	1	1
1	1	2
2	1	1
2	1	2
2	2	1

Operations on Relations(2)

R	(*supplier*	*part*)
	1	1
	2	1
	2	2

S	(*part*	*project*)
	1	1
	1	2
	2	1

R*S	(*supplier*	*part*	*project*)
	1	1	1
	1	1	2
	2	1	1
	2	1	2
	2	2	1

- Join: Image two binary relations, R and S, Join of R and S is a relation U such that π_{12}(U)=R and π_{23}(U)=S

- Composition

Operations on Relations(3)

R	(*s*	*p*	*j*)
	1	a	A
	2	a	A
	2	a	B
	2	b	A
	2	b	B

S	(*p*	*j*)
	a	A
	c	B
	b	B

R′	(*s*	*p*	*j*)
	1	a	A
	2	a	A
	2	b	B

- Restriction R by S

Redundancy

- Derivability: If a collection of operations (θ) in a certain order on relation R results in a particular relation S for all time then Relation R is θ derivable from set S.

Redundancy

Strong Redundancy: If a relation exists in a set of relations that has the projection, which can be derived from other projections of relations in the set, then the set is known as strong redundant.

For user convenience to cope with the traffic

Weak Redundancy: If a relation that has projection, which is not directly derivable form, other projections of relations but it from the projection of some join of other projections in the set it is known as weak redundancy.

For logical needs of community users.

Consistency

- The checking of inconsistency can be done on every update or insert or deletion actions performed with a penalty of some time

- every action can be logged to identify the actions led to that inconstancy or the journal should be maintained to track down all state changing transactions.

Conclusion

This paper laid the theoretical foundation for relational databases--the standard method by which information is organized in and retrieved from computers.

Q&A

Thank you