

Political Campaign Database

Phase 1

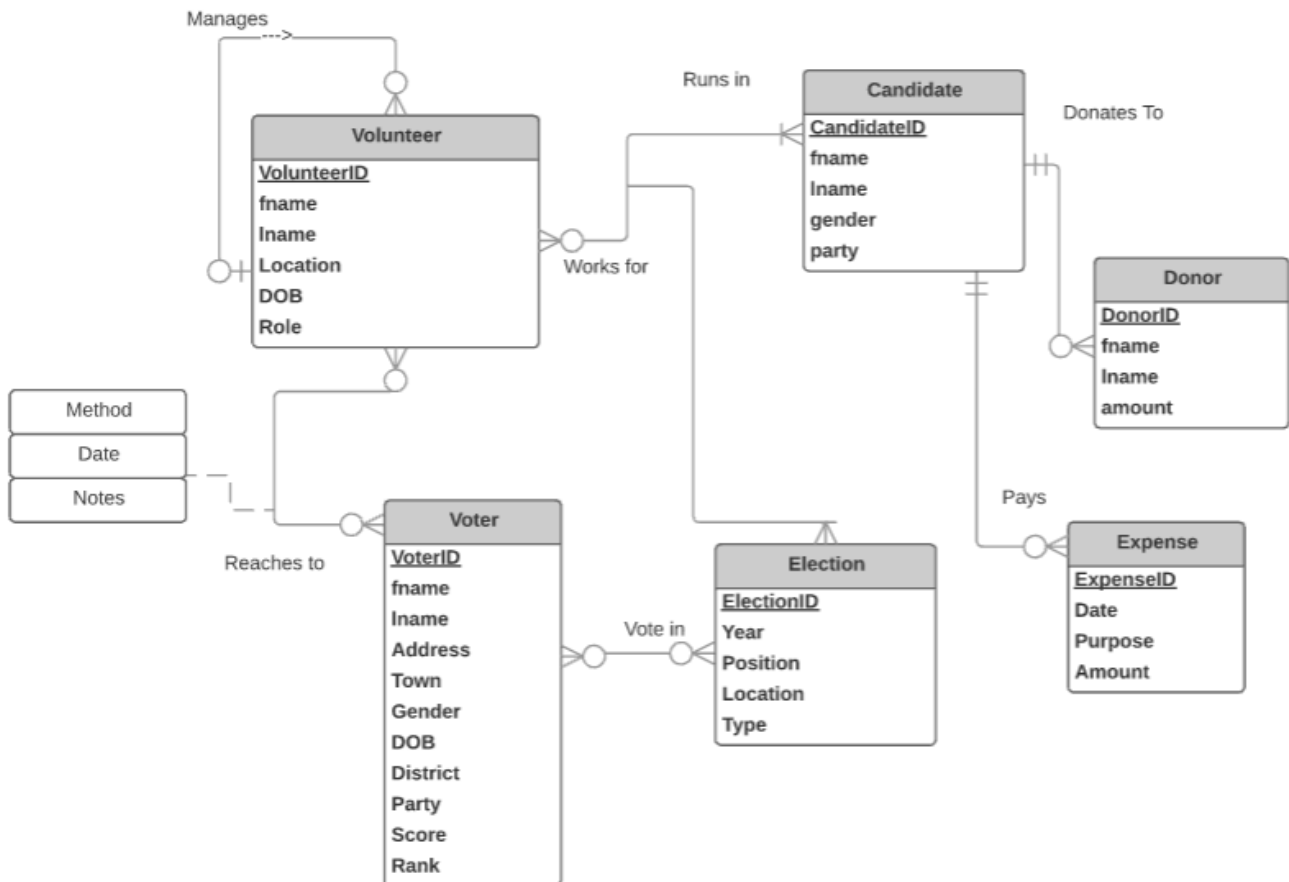
Richa Mohan, Brad Wyatt, Dominic Liu, Jeffrey Chan

Modern Data Management 70-455
Professor Wolfgang Gatterbauer

Executive Summary

We are focusing on developing a political campaign database for a hypothetical political candidate. This candidate or his workers would be the system user. Candidates are trying to market themselves to people, similar to how businesses need to market their products. They both need to target the right groups of people using the limited resources that they have, such as time and money. While candidates might have the best ideas, in order to win they need to efficiently run their campaign with an organized database recording all aspects of their campaign, such as donors, volunteers, and expenses. By having all the names and information organized in a data, it is much easier to contact people, making running the campaign much more efficiently. Without an organized database, the candidate may be wasting time and money on things that could have been easily addressed by a database.

ER Diagram (Part A)



Candidates have lists of voters and volunteers from several towns in the district. Candidates need to determine which houses to mail literature to, and which houses to door-knock. They aren't going to mail or door-knock everyone in the district, as there are people who don't vote, have a bad record of voting in the past, or only vote for candidates

of the opposing party. By narrowing down the list of people to mail or visit, candidates can save tremendous amounts of money. While this report is hypothetical, there is a lot of information from registered voters that are public. One of our group members has worked in a campaign before, and data has been pulled from <http://www.vlink-ma.com> if you want more information. Example: http://www.vlink-ma.com/Help/Help_Lookup.htm. A candidate can study all this information from past campaigns to identify target voters and the method to contact them.

To prepare our database structure, we decided there will be six entities consisting of Candidate, Volunteer, Donor, Expense, Voter, and Election. Candidates are the people running for a specific position in a specific election, and they have a party affiliation, and they must have ran in at least one election to be a part of this entity. Elections consist of both primary and general elections, for different positions in different years, and these elections could have had one or many candidates running in it. The candidate incurs Expenses, which records what the candidate spends on various purchases, with the dates also defining which election the expenses are for, and a candidate may have zero or multiple expenses. Donors donate money to the candidate, which the candidate uses to fund the expenses, along with their own funds, and once again a candidate may have none or many donors. Volunteers work for the candidate, but there can be volunteers that manage other volunteers. For example, there can be canvas leaders, who direct teams of volunteers to door knock in certain areas of a town. If more volunteers sign up to help the campaign, this entity would be modified. A volunteer can work for a candidate in many elections, or can work for different candidates in different elections, and this is displayed by the ternary relationship between Volunteer, Candidate and Election. Therefore, on implementing this relationship in a database, we would need a table to connect all these three. Lastly, there exists a Voter entity, which is the most extensive of all the entities with the most tuples and most attributes as well. The voter is identified by many different attributes, the most interesting of which are rank and score, which determine how likely the voter is to vote in the next election based on previous election data, and also what their party preferences and strengths are. There is also an associative entity between volunteer and Voter. This entity's attributes describe which methods were used to reach out to the voter. A voter can also vote in multiple elections, such as voting for the President and voting for the State Representative. Therefore, when we convert our database into tables, we will have a table to show which voter votes in which election,

Political elections have many business rules which candidates need to follow or else they can get disqualified. For example, candidates pay expenses which can consist of canvassing, advertising, places for hosting events, etc. But due to campaign finance laws, they need to record all of their expenses according to the date, purpose, and amount. There is also a \$500 maximum amount that can be donated to a candidate. Tracking donors can be useful for candidates because they can send thank-you letters to them, subscribe them to email lists, and keep their contact information for future records.

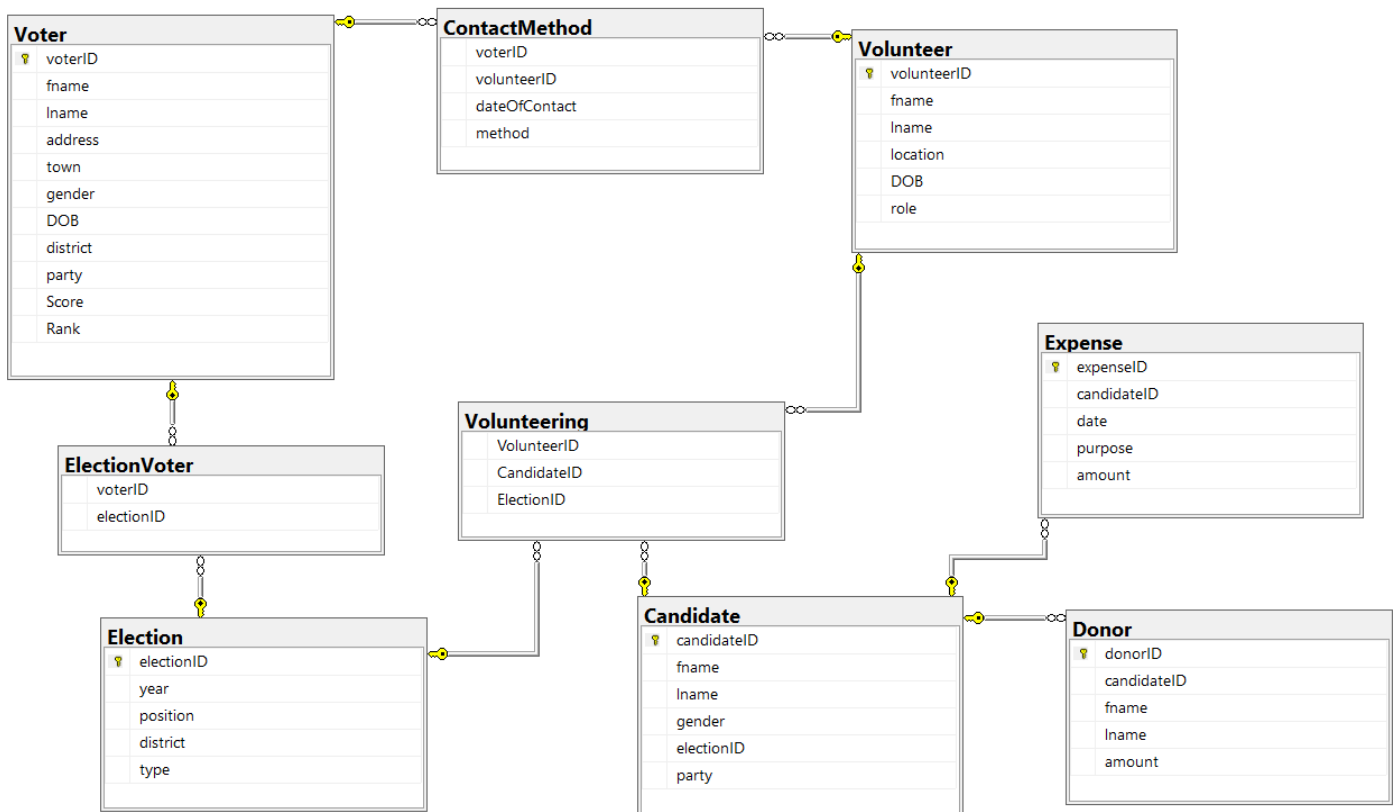
While volunteers can technically work for more than one candidate, we are assuming that volunteers will only work for one candidate during an election. Also, a candidate can have 0 or many volunteers. Volunteers also have many characteristics

which may affect their roles, such as where they are from. The characteristics are important so that they can be assigned to the right roles. For example, if a volunteer has organizational skills and is from Pittsburgh, he or she may be assigned to a managerial role.

While we had data from a previous campaign, we fabricated data for our implementation to simplify the realistic scenario. This meant that the number of tuples we had for each table were much smaller than they would be in a real table (for example, we only have 26 voters in our database). But even though we have fewer observations, we can successfully understand how to administer a database with such information, and use this data to the advantage of the candidate. Hence, after implementing the database in SQL, we run realistic queries to answer questions that would help a candidate run his campaign.

Database Schema (Part B)

After finalizing our ER diagram, we drew a relational schema for data implementation, and ensured that all our tables were in 3 normal form.



We have tables for all our original six entities (Voter, Election, Candidate,

Volunteer, Donor, Expenses), and our one associative entity (Contact Method). We also created two more tables named Election-Voter and Volunteering. Election-Voter consists of VoterIDs and ElectionIDs allowing us to identify which Voter voted in which Election. Volunteering consists of VolunteerID, ElectionID and CandidateID, allowing us to identify which Volunteer, worked for which Candidate in which election. Our associative entity, Contact method, needed VolunteerID and VoterID as foreign keys, in order to allow us to identify which Volunteer contacted which Voter. Expense and Donor also contained foreign keys to CandidateID so that we could identify the candidate who incurred the expense, or received the donation. Candidate also has electionID to identify which election he stood in.

Sample records from all our tables can be found in the appendix.

Database Implementation on SQL Server (Part C)

Example Queries Using The Database

1st important business process: How consistent registered voters in the district actually vote. Candidates already have the names of all the people in the district that are registered to vote, but they need to decide which voters should be targeted.

The first query gives the entire table of Voters along with the number of elections that each voter has ever voted in:

	voterId	lname	fname	ElectionsVotedIn
1	6	Budz	Dom	4
2	7	Cole	Brad	3
3	8	Brown	Wolfgang	3
4	9	Bombard	Clark	3
5	19	Kang	Jeffrey	3
6	1	Abemathy	Lewis	3
7	2	Ironson	Mary	2
8	3	Mascetta	Jessica	2
9	18	Lai	Tiffany	2
10	20	Smith	John	2
11	10	Haitsma	Stephanie	2
12	11	Park	John	1
13	12	1

The following query returns the voters that have voted in all elections. This is obviously important for the Candidate to spend more of their resources on connecting with these voters. In this case, Dom Budz was the only voter to vote in all four elections:

	voterId	lname	fname	ElectionsVotedIn
1	6	Budz	Dom	4

The following query returns the voters that are registered to vote, but have never voted. The candidate would likely not want to spend any time on these people (in this case it would be Paul Kim and Hilary Trump):

	voterId	lname	fname	ElectionsVotedIn
1	26	Kim	Paul	0
2	25	Trump	Hilary	0

2nd important business process: The candidates with the above average expenditure. It is common to hear that candidates that spend the most money are more likely to win elections. However there may be times, like the 2016 Republican primary election, where candidate Jeb Bush spent the most amount of money up to when he dropped out of the race, and has still lost.

In this case, the system user will be able to determine how effective having high total expenses are for winning hence, we wrote a query that allowed the user to identify the candidates who spend more than average amount on their campaign.

	lname	fname	TotalExpenses
1	Suresh	Subra	5040
2	Gatterbauer	Wolfgang	3400
3	Liu	Dominic	3200
4	Wyatt	Brad	3000

3rd important business process: Political campaigns need to effectively target voters, but they also need to target donors in order to make the most money.

Donors are likely to only donate to candidates of a certain party. Here for example, John Park donates only to Republicans, while Paul Pierce, Lucas Connors, and Andrew Bale only donate to Democrats:

	fname	lname	party
1	John	Park	R
2	Paul	Pierce	D
3	Lucas	Connors	D
4	Andrew	Bale	D
5	Sam	Smith	I
6	Christine	Nolan	I

The majority of donors are also voters. The following query find the voters who are also donors by making sure that their first and last name, and party affiliation is the same. Here, we didn't find a match for Lucas Connors, and it is possible that this was due to the fact that he didn't register the vote, or there was another Lucas Connors who registered with another party and hence didn't satisfy the match:

	fname	lname	party
1	John	Park	R
2	Paul	Pierce	D
3	Andrew	Bale	D
4	Sam	Smith	I
5	Christine	Nolan	I

It's important to identify the top donor for each candidate, because generally they have the most say in the campaigns that they donate to:

	candfname	candlname	topDonorAmount	donorfname	donorlname
1	Brad	Wyatt	300	John	Park
2	Richa	Mohan	500	Paul	Pierce
3	Jeffrey	Chan	200	Lucas	Connors
4	Dominic	Liu	150	Andrew	Bale
5	Wolfgang	Gatterbauer	500	Sam	Smith

4th important business process: We need to identify the volunteers who volunteered for multiple elections. This means they could be more likely to volunteer on our campaign.

	volunteerID	fname	lname	MultElections
1	1011	Jennifer	Huang	2

Example Database Updates

For adding new voters: Take an example of a family just moving into the district, and just registering to vote. Their scores and ranks are null because there is no prior data about them:

```
INSERT INTO "Voter" VALUES (00027, 'Arnold', 'Smith', '43 Oak Nut Street',
'Pittsburgh, PA', 'M', '01/20/1998', '4', 'D', null, null)
INSERT INTO "Voter" VALUES (00028, 'Sally', 'Smith', '43 Oak Nut Street',
'Pittsburgh, PA', 'F', '01/20/1975', '4', 'D', null, null)
INSERT INTO "Voter" VALUES (00029, 'Bob', 'Smith', '43 Oak Nut Street', 'Pittsburgh,
PA', 'M', '05/14/1974', '4', 'D', null, null)
```

For modifying a name of a Candidate: For instance, there is a candidate named 'Brad Wyatt', if we wanted to modify his name to be 'Bradford Wyatt' to be more formal, here is the update query:

```
UPDATE Candidate
SET fname = 'Bradford'
WHERE candidateID = '11';
```

Appendix

Election

	electionID	year	position	district	type
1	1	2012	Governor	4	General
2	2	2014	President	4	Primary
3	3	2014	State Representative	4	General
4	4	2016	Mayor	4	General

Candidate Table

	candidateID	fname	lname	gender	electionID	party
1	11	Brad	Wyatt	M	1	R
2	22	Jeffrey	Chan	M	1	D
3	33	Richa	Mohan	F	2	D

Expense

	expenseID	candidateID	date	purpose	amount
1	1	44	03/14/2014	Advertising	2000
2	2	33	03/15/2014	Canvassing	300
3	3	66	03/16/2014	Polling	40

Donor

	donorID	candidateID	fname	lname	amount
1	1	11	John	Park	300
2	2	33	Paul	Pierce	500
3	3	22	Lucas	Connors	200
4	4	44	Andrew	Bale	150

Volunteer

	volunteerID	fname	lname	location	DOB	role
1	1011	Jennifer	Huang	Pittsburgh, PA	04/25/1984	Campaign Manager
2	1012	Tiffany	Lai	Pittsburgh, PA	06/12/1993	Canvas Leader
3	1013	Jeffrey	Kang	Pittsburgh, PA	02/17/1989	Calling Leader

Volunteering - Connecting Volunteers, Elections and Candidates

	VolunteerID	CandidateID	ElectionID
1	1011	11	1
2	1011	77	2
3	1013	22	1

ContactMethod

	voterID	volunteerID	dateOfContact	method
1	1	1011	01/15/2015	DoorKnock
2	1	1011	05/15/2011	mail
3	1	1011	01/15/2013	Phonecall
4	2	1012	03/18/2015	DoorKnock
5	3	1013	01/15/2015	mail

Voter

	voterID	fname	lname	address	town	gender	DOB	district	party	Score	Rank
1	1	Lewis	Abernathy	20 Applepie Lane	Pittsburgh, PA	M	11/07/1993	4	R	A	1
2	2	Mary	Ironson	15 Blueberry Pie Lane	Carnegie, PA	F	12/15/1988	4	R	C	2
3	3	Jessica	Mascetta	300 Grant Street	Hazelwood, PA	F	1/19/1987	4	D	B	5

ElectionVoter - Connecting Voters with Elections

	voterID	electionID
1	1	1
2	1	2
3	1	3
4	2	1
5	2	4
6	3	1