



The Accord Consensus Protocol

Fast General Purpose Transactions

Elliott Smith, Benedict Miller, Alex

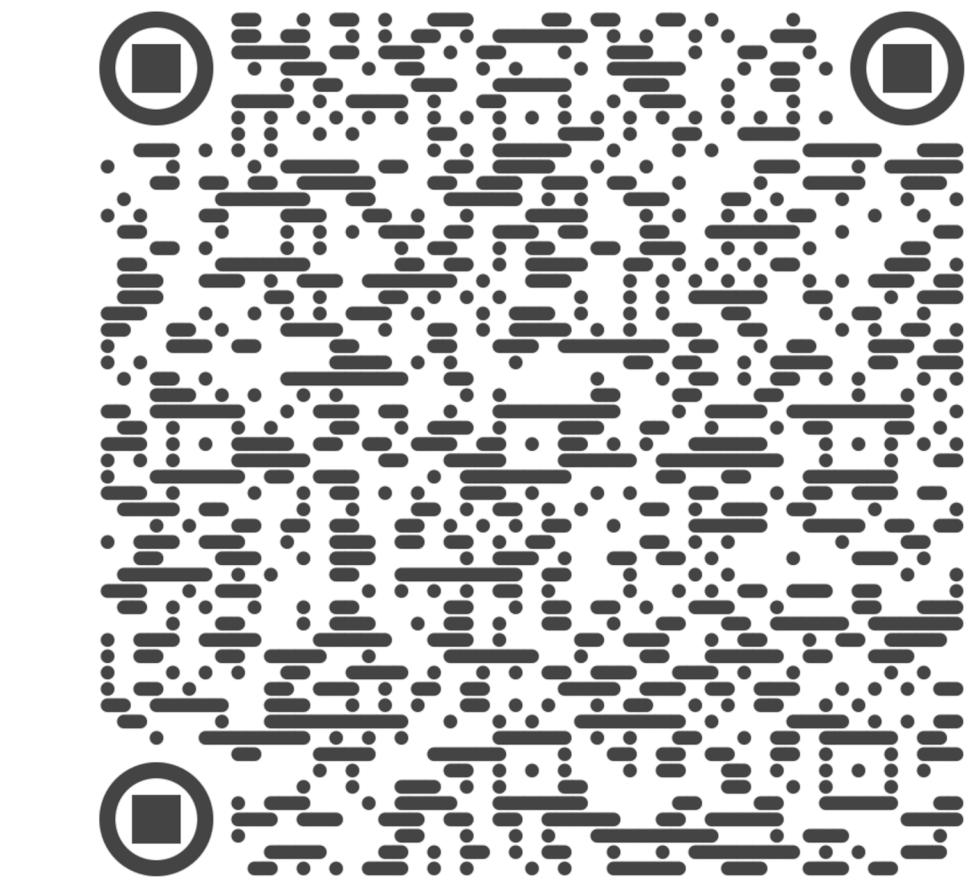
Zhang, Tony Capwell, David

Eggleson, Blake Andreas, Scott Yeschenko, Aleksey

Dr. Howard, Heidi Weisberg, Ariel

ACID Compliance
Transactions
Resilience & Linear scalability
Performance
Accessible
Novel

- Global Strict Serialisability
- Multi-partition, Multi-table
- Leaderless, Unique Recovery (No rollbacks)
- Single Network Round-trip
- Commodity Clocks
- Flexible Fast-Path Electorates
- Timestamp Reorder Buffer



Algorithm 1 Consensus Protocol

```

receive t on coordinator C from client:
1:  $t_c \leftarrow (\text{now}, 0, C)$ 
2: send PreAccept( $t, t_c$ ) to  $\forall p \in \mathbb{P}^t$ 
receive PreAccept( $t, t_c$ ) on p:
3: if  $t_c > \max(T_p) \wedge t_c > t_t$  then
4:    $t_t \leftarrow t_c$ 
5: else
6:    $t_t \leftarrow \max(T_p) \wedge t_c < t_t$ 
7: end if
8:  $t_s \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
9: end if
10:  $T_p \leftarrow t_s$ 
11:  $Accepted_p \leftarrow \text{true}$ 
12: reply PreAcceptOK( $t, t_c, deps : \{y | y \sim t \wedge t_y < t_t\}$ )
receive PreAcceptOK( $t, t_c, deps$ ) from  $\mathbb{Q}^t$ :
13:  $deps = \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
14: if  $\exists p \subseteq \mathbb{P}^t \setminus \{p \sim t\}$  then
15:   send Commit( $t, t_c, t_s, deps$ ) to  $\forall p \in \mathbb{P}^t$ 
16:   go to Execution Protocol
17: else
18:    $t = \max(T_p) \wedge p \in \mathbb{P}^t$ 
19:   send Accept( $t, t_c, t_s, t_t, deps$ ) to  $\forall p \in \mathbb{P}_t \cup \mathbb{P}^t$ 
20: end if
21:  $Accepted_p \leftarrow \text{true}$ 
22: reply PreAcceptOK( $t, t_c, deps : \{y | y \sim t \wedge t_y < t_t\}$ )
receive PreAcceptOK( $t, t_c, deps$ ) from C:
23:  $deps = \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
24: if  $\exists p \subseteq \mathbb{P}^t \setminus \{p \sim t\}$  then
25:   send Commit( $t, t_c, t_s, deps$ ) to  $\forall p \in \mathbb{P}^t$ 
26:   go to Execution Protocol
27: else
28:    $t = \max(T_p) \wedge p \in \mathbb{P}^t$ 
29:   send Accept( $t, t_c, t_s, t_t, deps$ ) to  $\forall p \in \mathbb{P}_t \cup \mathbb{P}^t$ 
30: end if
31:  $Accepted_p \leftarrow \text{true}$ 
32: reply PreAcceptOK( $t, t_c, deps$ ) to some nearby  $p \in \mathbb{P}^t$ 
33: read  $t_{read} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
34:  $deps = \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
35: send Commit( $t, t_c, t_s, deps$ ) to  $\forall p \in \mathbb{P}^t$ 
36: end if
37: reply ReadOK( $t_{read}$ )
receive Commit( $t, t_c, t_s, t_t, deps$ ) from each shard:
38:  $t_{read} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
39:  $deps_{read} \leftarrow \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
40: send result( $t, t_c, t_s, t_t, deps_{read}$ ) to  $\forall p \in \mathbb{P}^t$ 
41: send result( $t, t_c, t_s, t_t, deps_{read}$ ) to client
42:  $Accepted_p \leftarrow \text{true}$ 
43: reply ReadOK( $t_{read}$ )
receive ReadOK( $t_{read}$ ) from client:
44:  $t_{read} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
45:  $deps_{read} \leftarrow \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
46:  $Accepted_p \leftarrow \text{true}$ 
47: await Apply( $t, t_c, t_s, t_t, deps_{read}$ )
48:  $t_{apply} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
49:  $deps_{apply} \leftarrow \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
50: reply ReadOK( $t_{apply}$ )
receive ReadOK( $t_{apply}$ ) from each shard:
51:  $t_{apply} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
52:  $deps_{apply} \leftarrow \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
53: send result( $t, t_c, t_s, t_t, deps_{apply}$ ) to  $\forall p \in \mathbb{P}^t$ 
54:  $Accepted_p \leftarrow \text{true}$ 
55: await Apply( $t, t_c, t_s, t_t, deps_{apply}$ )
56:  $t_{apply} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
57: apply writes( $t, t_c, t_s, t_t, deps_{apply}$ )
58:  $Accepted_p \leftarrow \text{true}$ 
59: end if

```

Algorithm 2 Execution Protocol

```

Coordinator C:
2: for  $p \in \mathbb{P}^t$  do
3:    $t_p \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
4:   send PreAccept( $t_p, t_c$ ) to  $\forall p \in \mathbb{P}^t$ 
5: end for
receive Commit( $t, t_c, t_s, t_t, deps$ ):
6:  $Committed \leftarrow \text{true}$ 
7: await Commit( $t, t_c, t_s, t_t, deps$ )
8: send Read( $t, t_c, t_s, t_t, deps$ ) to some nearby  $p \in \mathbb{P}^t$ 
9:  $t_{read} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
10: read  $t_{read} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
11:  $deps_{read} \leftarrow \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
12: reply ReadOK( $t_{read}$ )
receive ReadOK( $t_{read}$ ) from each shard:
13:  $t_{read} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
14:  $deps_{read} \leftarrow \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
15: send result( $t, t_c, t_s, t_t, deps_{read}$ ) to client
16:  $Accepted_p \leftarrow \text{true}$ 
17: reply ReadOK( $t_{read}$ )
receive ReadOK( $t_{read}$ ) from client:
18:  $t_{read} \leftarrow (\text{seq}, id) \leftarrow (t_c, seq + 1, p)$ 
19:  $deps_{read} \leftarrow \bigcup \{p.deps | p \in \mathbb{P}^t\}$ 
20: yield to competing coordinator
receive RecoverOK(*, Superseding, Wait) from  $p \in \mathbb{R}^t$ :
21: if  $\exists p \in \mathbb{R}^t \setminus \{p.Applied\}$  then
22:   send response( $t, t_c, t_s, t_t, deps$ )
23:   send result( $t, t_c, t_s, t_t, deps$ ) to  $\forall p \in \mathbb{P}^t$ 
24: else if  $\exists p \in \mathbb{R}^t \setminus \{p.Committed\}$  then
25:   send Commit( $t, t_c, t_s, t_t, deps$ ) to  $\forall p \in \mathbb{P}^t$ 
26: go to Consensus Protocol [1]
27: else if  $\exists p \in \mathbb{R}^t \setminus \{p.Accepted\}$  then
28:   select p with highest accepted ballot
29:    $t \leftarrow p.t, deps \leftarrow p.deps$ 
30:   go to Consensus Protocol [1]
31: else if  $\exists p \in \mathbb{R}^t \setminus \{p.Wait\}$  then
32:    $t \leftarrow p.t, deps \leftarrow p.deps$ 
33:   if  $\exists p \in \mathbb{R}^t \setminus \{p.p\_Wait\}$  then
34:      $t \leftarrow \max(p.t | p \in \mathbb{R}^t)$ 
35:   else
36:      $t \leftarrow \max(p.t | p \in \mathbb{R}^t)$ 
37:   else if  $\exists p \in \mathbb{R}^t \setminus \{p.Wait\}$  then
38:      $t \leftarrow \max(p.t | p \in \mathbb{R}^t)$ 
39:   else if  $\exists p \in \mathbb{R}^t \setminus \{p.Wait\}$  then
40:      $t \leftarrow \max(p.t | p \in \mathbb{R}^t)$ 
41:   end if
42: go to Consensus Protocol [1]

```

▷ PreAccept

▷ Accept

Algorithm 3 Recovery Protocol

```

Coordinator C:
1: b ← fresh ballot
2: send Recover( $t, t_c, t_s, t_t, deps$ ) to  $\forall p \in \mathbb{P}^t$ 
receive Recover( $t, t_c, t_s, t_t, deps$ ) on p:
3: if  $p \in E_n \wedge t_c > t_t$  then
4:    $t_c \leftarrow t_s$ 
5:    $t_t \leftarrow t_s$ 
6:    $Accepted_p \leftarrow \text{true}$ 
7:    $Wait \leftarrow \{y \in Accepts | t_y < t_c \wedge t_y > t_t\}$ 
8:    $Superseding \leftarrow \{y \in Accepts | t_y > t_c\}$ 
9:    $Wait \leftarrow \{y \in Accepts | t_y > t_c\}$ 
10:  if  $\exists p \in E_n \setminus \{p.Accepted\}$  then
11:    if  $\exists p \in E_n \setminus \{p.Wait\}$  then
12:      if  $\exists p \in E_n \setminus \{p.Wait\}$  then
13:        if  $\exists p \in E_n \setminus \{p.Wait\}$  then
14:          if  $\exists p \in E_n \setminus \{p.Wait\}$  then
15:            if  $\exists p \in E_n \setminus \{p.Wait\}$  then
16:              if  $\exists p \in E_n \setminus \{p.Wait\}$  then
17:                if  $\exists p \in E_n \setminus \{p.Wait\}$  then
18:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
19:                    if  $\exists p \in E_n \setminus \{p.Wait\}$  then
20:                      if  $\exists p \in E_n \setminus \{p.Wait\}$  then
21:                        if  $\exists p \in E_n \setminus \{p.Wait\}$  then
22:                          if  $\exists p \in E_n \setminus \{p.Wait\}$  then
23:                            if  $\exists p \in E_n \setminus \{p.Wait\}$  then
24:                              if  $\exists p \in E_n \setminus \{p.Wait\}$  then
25:                                if  $\exists p \in E_n \setminus \{p.Wait\}$  then
26:                                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
27:                                    if  $\exists p \in E_n \setminus \{p.Wait\}$  then
28:                                      if  $\exists p \in E_n \setminus \{p.Wait\}$  then
29:                                        if  $\exists p \in E_n \setminus \{p.Wait\}$  then
30:                                          if  $\exists p \in E_n \setminus \{p.Wait\}$  then
31:                                            if  $\exists p \in E_n \setminus \{p.Wait\}$  then
32:                                              if  $\exists p \in E_n \setminus \{p.Wait\}$  then
33:                                                if  $\exists p \in E_n \setminus \{p.Wait\}$  then
34:                                                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
35:                                                    if  $\exists p \in E_n \setminus \{p.Wait\}$  then
36:                                                      if  $\exists p \in E_n \setminus \{p.Wait\}$  then
37:                                                        if  $\exists p \in E_n \setminus \{p.Wait\}$  then
38:                                                          if  $\exists p \in E_n \setminus \{p.Wait\}$  then
39:                                                            if  $\exists p \in E_n \setminus \{p.Wait\}$  then
40:                                                              if  $\exists p \in E_n \setminus \{p.Wait\}$  then
41:                                                                if  $\exists p \in E_n \setminus \{p.Wait\}$  then
42:                                                                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
43:                                                                    if  $\exists p \in E_n \setminus \{p.Wait\}$  then
44:                                                                      if  $\exists p \in E_n \setminus \{p.Wait\}$  then
45:                                                                        if  $\exists p \in E_n \setminus \{p.Wait\}$  then
46:                                                                          if  $\exists p \in E_n \setminus \{p.Wait\}$  then
47:                                                                            if  $\exists p \in E_n \setminus \{p.Wait\}$  then
48:                                                                              if  $\exists p \in E_n \setminus \{p.Wait\}$  then
49:                                                                                if  $\exists p \in E_n \setminus \{p.Wait\}$  then
50:                                                                                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
51:                                                                                    if  $\exists p \in E_n \setminus \{p.Wait\}$  then
52:                                                                 if  $\exists p \in E_n \setminus \{p.Wait\}$  then
53:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
54:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
55:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
56:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
57:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
58:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
59:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
60:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
61:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
62:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
63:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
64:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
65:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
66:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
67:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
68:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
69:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
70:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
71:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
72:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
73:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
74:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
75:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
76:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
77:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
78:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
79:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
80:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
81:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
82:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
83:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
84:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
85:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
86:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
87:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
88:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
89:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
90:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
91:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
92:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
93:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
94:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
95:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
96:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
97:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
98:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
99:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
100:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
101:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
102:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
103:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
104:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
105:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
106:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
107:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
108:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
109:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
110:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
111:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
112:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
113:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
114:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
115:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
116:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
117:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
118:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
119:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
120:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
121:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
122:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
123:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
124:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
125:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
126:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
127:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
128:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
129:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
130:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
131:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
132:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
133:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
134:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
135:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
136:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
137:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
138:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
139:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
140:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
141:                  if  $\exists p \in E_n \setminus \{p.Wait\}$  then
142:                  if  $\exists p \in E_n \setminus \{p.Wait\}$ 
```