MongoDB Schema Design for Friday AI

Database: `friday_ai`

Collection: `leads`

```
{
  _id: ObjectId,
  timestamp: ISODate,
  source: String,
  status: String, // "new", "contacted", "qualified", "converted", "closed"
  name: String,
  email: String,
  company: String,
  interest: String,
  phone: String,
  job_title: String,
  budget: String,
  timeline: String,
  created_at: ISODate,
  updated_at: ISODate
}

// Indexes
db.leads.createIndex({ "email": 1 }, { unique: true })
db.leads.createIndex({ "timestamp": -1 })
db.leads.createIndex({ "status": 1 })
db.leads.createIndex({ "company": 1 })
db.leads.createIndex({ "interest": 1 })
```

Collection: `transcript_events`

Real-time streaming events (replaces transcripts.jsonl)

```
{
  _id: ObjectId,
  role: String, // "user", "assistant"
  content: String,
  timestamp: ISODate,
  source: String, // "transcription_node", "agent", etc.
  meta: Object, // optional metadata
  session_id: String, // to group events by session
  created_at: ISODate
}

// Indexes
db.transcript_events.createIndex({ "timestamp": -1 })
db.transcript_events.createIndex({ "session_id": 1, "timestamp": 1 })
db.transcript_events.createIndex({ "role": 1 })
```

Collection: `conversation_sessions`

Complete conversation sessions (replaces transcript_session_*.json)

```
{
  _id: ObjectId,
  session_id: String, // unique session identifier
  start_time: ISODate,
  end_time: ISODate,
  items: Array, // array of conversation items
  total_items: Number,
  duration_seconds: Number,
  lead_generated: Boolean,
  lead_id: ObjectId, // reference to leads collection if lead was created
  metadata: Object, // additional session metadata
  created_at: ISODate,
  updated_at: ISODate
}

// Indexes
db.conversation_sessions.createIndex({ "session_id": 1 }, { unique: true })
db.conversation_sessions.createIndex({ "start_time": -1 })
db.conversation_sessions.createIndex({ "lead_generated": 1 })
db.conversation_sessions.createIndex({ "lead_id": 1 })
```

Collection Schema Benefits:

1. **Leads**: Indexed for fast lookups by email, status, company, and interest
2. **Transcript Events**: Supports real-time logging with session grouping
3. **Conversation Sessions**: Complete session storage with lead tracking
4. **Relationships**: Lead sessions can be linked to generated leads
5. **Performance**: Proper indexing for common query patterns
6. **Scalability**: Collections can grow without filesystem limitations